Digital Imaging of Biological Type Specimens

A Manual of Best Practice

Results from a study of the European Network for Biodiversity Information

Editors:
Christoph L. Häuser, Axel Steiner, Joachim Holstein &
Malcolm J. Scoble

Stuttgart 2005

Editors: Christoph L. Häuser, Axel Steiner, Joachim Holstein (Stuttgart) & Malcolm J. Scoble (London)

Staatliches Museum für Naturkunde, Stuttgart
Rosenstein 1, D-70191 Stuttgart, Germany

The Natural History Museum, London
Cromwell Road, London SW7 5BD, UK

This publication should be cited as:
HÄUSER, C.L., STEINER, A., HOLSTEIN, J. & SCOBLE, M. J. (eds.) (2005): Digital Imaging of Biological Type Specimens. A Manual of Best Practice. Results from a study of the European Network for Biodiversity Information. Stuttgart. viii + 309 pp.

Contributions in this publication should be cited as:
CRICK, M. (2005): Image File Management. In: HÄUSER et al. (eds.): Digital Imaging of Biological Type Specimens. A Manual of Best Practice. Results from a study of the European Network for Biodiversity Information: 41-55. Stuttgart.

ENBI is a thematic network supported by the European Commission under the 5th Framework Programme. Contract number: EVK2-CT-2002-20020.
http://www.enbi.info

This publication can be obtained as a book or as a CD from:
Staatliches Museum für Naturkunde, Rosenstein 1, D-70191 Stuttgart, Germany, or:
ENBI Bureau, Zoological Museum Amsterdam, Faculty of Science/University of Amsterdam, PO Box 94766, NL-1090 GT Amsterdam, The Netherlands.

Printed by Oehler Offset, Fellbach

ISBN: 3-00-017240-8

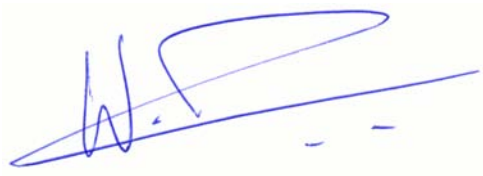
Cover graphics: JOACHIM HOLSTEIN, using photographs from this book.

Preface

by WOUTER LOS

This book addresses a number of important issues about the digital imaging of biological objects. These topics were explored in two workshops organised by the European Network for Biodiversity Information (ENBI), and the Global Biodiversity Information Facility (GBIF). With the digital imaging of a growing number of biological objects, it has become of great importance to agree on common approaches and standards. Such standardization is particularly important for natural history specimens so as to compare specimens often with only subtle differences in morphology. Emerging technologies are leading to exciting new opportunities in scientific studies and the field of biodiversity is notable among them.

I wish to thank all authors and the editors for bringing together and highlighting many new developments in this book. On behalf of the ENBI membership I also want to acknowledge the support from the European Commission for the ENBI network which made it possible to publish this book.



Wouter Los
Co-ordinator of the European Network for Biodiversity Information
(University of Amsterdam)

Contents

List of Contributors

RUUD ALTENBURG
ETI BioInformatics, University of Amsterdam, Mauritskade 61, NL-1092 AD Amsterdam, The Netherlands. E-mail:: ruud@eti.uva.nl

ARTURO H. ARIÑO
Department of Zoology and Ecology, University of Navarra, E-31080 Pamplona, Spain. E-mail: artarip@unav.es

ENRIQUE BAQUERO
Department of Zoology and Ecology, University of Navarra, E-31080 Pamplona, Spain. E-mail: ebaquero@unav.es

WALTER G. BERENDSOHN
Botanic Garden and Botanical Museum Berlin-Dahlem, Freie Universität Berlin, Königin-Luise-Str. 6-8, D-14191 Berlin, Germany. E-mail: w.berendsohn@bgbm.org

MATTHEW L. BUFFINGTON
Florida State University/CDFA, 3294 Meadowview Rd, Sacaramento, CA 95832-1448 USA. E-mail: mbuffington@cdfa.ca.gov

ROGER A. BURKS
Department of Entomology, UC Riverside, Riverside, CA 92521 USA. E-mail: rogerburks@yahoo.com

MARION CRICK
ePic Project, Royal Botanic Gardens, Kew, Kew Green, Richmond, TW9 3AE, United Kingdom. E-mail: m.crick@rbgkew.org.uk

UGO DALL'ASTA
Department of African Zoology / Entomology, Musée Royal de l'Afrique Centrale, Leuvensesteenweg 13, B-3080 Tervuren, Belgium. E-mail: ugo.dallasta@africamuseum.be

MARC DE MEYER
Department of African Zoology / Entomology, Musée Royal de l'Afrique Centrale, Leuvensesteenweg 13, B-3080 Tervuren, Belgium. E-mail: marc.de.meyer@africamuseum.be

ANDY C. DOWNTON
Department of Electronic Systems Engineering, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, United Kingdom. Email: acd@essex.ac.uk

BARBARA EINSIEDEL
Botanic Garden and Botanical Museum Berlin-Dahlem, Freie Universität Berlin, Königin-Luise-Str. 6-8, D-14191 Berlin, Germany.

SVEN ERLACHER
Zoologische Staatssammlung München, Münchhausenstr. 21, D-81247 München, Germany. E-mail: erlacher@zsm.mwn.de

DAGMAR FRITZE
DSMZ - Deutsche Sammlung von Mikroorganismen und Zellkulturen, Mascheroder Weg 1b, D-38124 Braunschweig, Germany. E-mail: dfr@dsmz.de

DAVID GALICIA
Department of Zoology and Ecology, University of Navarra, E-31080 Pamplona, Spain. E-mail: dgalpar@alumni.unav.es

FABIAN HAAS
Staatliches Museum für Naturkunde Stuttgart, Rosenstein 1, D-70191 Stuttgart, Germany. E-mail: haas.smns@naturkundemuseum-bw.de

CHRISTOPH L. HÄUSER
Staatliches Museum für Naturkunde Stuttgart, Rosenstein 1, D-70191 Stuttgart, Germany. E-mail: chaeuser@gmx.de

AXEL HAUSMANN
Zoologische Staatssammlung München, Münchhausenstr. 21, D-81247 München, Germany. E-mail: Axel.Hausmann@zsm.mwn.de

JOACHIM HOLSTEIN
Staatliches Museum für Naturkunde Stuttgart, Rosenstein 1, D-70191 Stuttgart, Germany. E-mail: holstein.smns@naturkundemuseum-bw.de

PHIL HURST
Photographic Unit, The Natural History Museum, Cromwell Road, London, SW7 5BD, United Kingdom. E-mail: p.hurst@nhm.ac.uk

RAFAEL JORDANA
Department of Zoology and Ecology, University of Navarra, E-31080 Pamplona, Spain. E-mail: rjordana@unav.es

WOUTER LOS
University of Amsterdam, Zoological Museum of Amsterdam, Institute of Biodiversity and Ecosystem Dynamics, Mauritskade 61, NL-1092 GT Amsterdam, The Netherlands. E-mail: los@science.uva.nl

PATRICIA MERGEN
Department of Zoology, Royal Museum for Central Africa, Leuvensesteenweg 13, 3080 Tervuren, Belgium. E-mail: patricia.mergen@africamuseum.be

UTA MERKEL
Botanic Garden and Botanical Museum Berlin-Dahlem, Freie Universität Berlin, Königin-Luise-Str. 6-8, D-14191 Berlin, Germany.

ROBERT A. MORRIS
Department of Computer Sciences, UMASS-Boston, 100 Morrissey Blvd, Boston, MA 02125, USA. E-mail: ram@cs.umb.edu

HEIDEMARIE NOWAK-KRAWIETZ
Botanic Garden and Botanical Museum Berlin-Dahlem, Freie Universität Berlin, Königin-Luise-Str. 6-8, D-14191 Berlin, Germany.

ALEXANDER RIEDEL
Staatliches Museum für Naturkunde Karlsruhe, Erbprinzenstr. 13, D-76133 Karlsruhe, Germany. E-mail: riedel@smnk.de

DOMINIK RÖPERT
Botanic Garden and Botanical Museum Berlin-Dahlem, Freie Universität Berlin, Königin-Luise-Str. 6-8, D-14191 Berlin, Germany.

MALCOLM J. SCOBLE
Department of Entomology, The Natural History Museum, Cromwell Road, London SW7 5BD, United Kingdom. E-mail: m.scoble@nhm.ac.uk

DARRELL SIEBERT
Department of Zoology, The Natural History Museum, Cromwell Road, London SW7 5BD, United Kingdom. E-mail: d.siebert@nhm.ac.uk

JOS SNOEKS
Department of African Zoology / Vertebrates, Musée Royal de l'Afrique Centrale, Leuvensesteenweg 13, B-3080 Tervuren, Belgium. E-mail: jos.snoeks@africamuseum.be

LARRY SPEERS
Global Biodiversity Information Facility, Universitetsparken 15, DK-2100 Copenhagen, Denmark. E-mail: lspeers@gbif.org

EDWIN VAN SPRONSEN
ETI BioInformatics, University of Amsterdam, Mauritskade 61, NL-1092 AD Amsterdam, The Netherlands. E-mail: edwin@eti.uva.nl

AXEL STEINER
Staatliches Museum für Naturkunde Stuttgart, Rosenstein 1, D-70191 Stuttgart, Germany. E-mail: a-steiner@web.de

PIET STOFFELEN
National Botanic Garden of Belgium, Domein van Bouchout, B-1860 Meise, Belgium. E-mail: stoffelen@br.fgov.be

HAROLD TAYLOR
Natural History Museum Image Resource, The Natural History Museum, Cromwell Road, London, SW7 5BD, United Kingdom. E-mail: h.taylor@nhm.ac.uk

ALAIN VANDER VELDE
National Botanic Garden of Belgium, Domein van Bouchout, B-1860 Meise, Belgium. E-mail: alainz@br.fgov.be

HUUB VELDHUIZEN VAN ZANTEN
ETI BioInformatics, University of Amsterdam, Mauritskade 61, NL-1092 AD Amsterdam, The Netherlands. E-mail: huub@eti.uva.nl

FRÉDÉRIC WAUTELET
100 rue OBECQ, B-1410 Waterloo, Belgium E-Mail : frederic.wautelet@gmail.com

I. M. WHITE
c/o Department of Entomology, The Natural History Museum, Cromwell Road, London SW7 5BD, United Kingdom. E-mail: imw@nhm.ac.uk

IRINA WILL
Botanic Garden and Botanical Museum Berlin-Dahlem, Freie Universität Berlin, Königin-Luise-Str. 6-8, D-14191 Berlin, Germany.

Introduction

CHRISTOPH L. HÄUSER[1] & MALCOLM J. SCOBLE[2]

[1] Staatliches Museum für Naturkunde Stuttgart, Rosenstein 1, D-70191 Stuttgart, Germany. E-mail: chaeuser@gmx.de
[2] Department of Entomology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK. E-mail: m.scoble@nhm.ac.uk

Images have always been of fundamental importance in taxonomy and for the documentation of natural history specimens. In recent years, however, digital photography and other imaging techniques are rapidly transforming the way in which we take and portray images in these fields. The contributions in this publication describe some of the methods and technologies currently in use for imaging biological specimens – particularly types, which are those specimens of special importance in acting as name bearers. A variety of taxa are targeted so as to illustrate some of the many problems and solutions that are emerging from the digital imaging of type specimens. We decided that this publication was best described as a "manual", because it is composed of chapters written by practitioners of various techniques and experiences. To some extent this renders the volume somewhat idiosyncratic and less well structured and planned – indeed, it developed from two workshops, a genre the very essence of which is flexibility. We hope, however, that readers will not only tolerate the approach taken, but rather that they might value it, firm in the recognition that the contributions flow from colleagues having „hands-on" experience of equipment, a knowledge of the „tricks of the trade", and an understanding of how to handle the digital results produced.

The idea for producing such a manual originated from work undertaken for a specific activity under the "European Network for Biodiversity Information (ENBI)", a thematic network project funded by the European Commission under its Fifth Framework Programme (see SCOBLE, "The Place of the Manual in the European Network of

Biodiversity Information", in this volume). The main goal of this particular component of the ENBI work programme on "Co-operation of pan European databases on biological collections and specimens" was to develop and implement technical standards or rules of best practise for type-specimen databases, particularly with regard to digital imaging. As originally planned, the standards were to be implemented via demonstrator databases for type-specimens from three exemplar taxa, butterflies, fish, and flowering plants, and the results were also to be made available in a final report (access to these demonstrator databases is available online at the BioCASE portal, http://www.biocase.org/; see also chapters by BERENDSOHN et al., HÄUSER et al., and HURST & SIEBERT in this volume).

But to encourage wider interest across Europe, two workshops were held, one in March 2004 in Stuttgart, Germany, and one in January 2005 in Chania, Crete. These events attracted considerable interest with participation extending well beyond the institutional partners directly involved in the ENBI project, and, furthermore and encouragingly, from outside Europe. The enthusiasm and interest generated at the workshops led us to expand the original work plan beyond the production of a formal report on the demonstrator databases, so as to capitalize on the multitude of experiences made available – particularly from the presentations made and the resulting discussions. The idea for a more comprehensive manual quickly took shape, with the intention of including the experiences gained from work on as many different taxa and approaches as were available. A quick survey of the equipment, procedures and standards applied by specimen imaging activities in different institutions and projects was conducted, for which a questionnaire was developed and distributed. The results of this survey have also been included in this volume under "case studies".

Some background on biosystematic research and taxonomic procedures may help the reader to appreciate better the tremendous importance and scientific value of biological type specimens, and more generally to understand why making information available from specimens in biological collections is not carried out just for academic purposes. It is not always appreciated that our basic know-

Fig. 1. Participants of the ENBI workshop, Stuttgart, March 2004. Photo: F. HAAS.

ledge of biodiversity is founded on information we have gleaned from biological specimens. Much of our understanding of natural organisms (their appearance, anatomy, occurrence in time and space, and much of their basic biology including preferred habitats, nutrition, and life cycles), is either directly or ultimately based on the specimens preserved in scientific collections and the information associated with them, which typically is recorded on labels or written notes attached directly to the specimens.

The information contained in and connected to what amounts to an estimated 3 billion biological specimens preserved in the world's scientific collections thus represent an unsurpassed treasury of information on the living world (e.g. SUAREZ & TSUTSUI 2004). Efforts to make this information more widely or even universally available have been under way for some time, and the rise of the internet and the storage and data-exchange capacities of modern information technology will greatly facilitate this huge task - a task, moreover, that the Global Biodiversity Information Facility (GBIF) has made one of its priorities. Having such a vast reservoir of specimen data readily accessible will contribute greatly to bridge what remains a considerable gap in knowledge and a significant bottleneck in successfully facing and eventually overcoming the global biodiversity crisis (HÄUSER 2004).

According to the international rules for the scientific naming of organisms (GREUTER et al. 1999, International Commission for

Zoological Nomenclature 1999), one of the fundamental principles of the system of scientific nomenclature for organisms is the "type concept", which requires that to be validly published and form part of accepted nomenclature, every scientific name must be connected to and based on some biological material or specimen to be preserved in a collection. Ideally a single specimen, the so-called holotype, should serve as the name bearer. This type specimen thus becomes the material basis for any scientific name whether for species, subspecies or higher categories. As a result, identities of taxa can be tested and verified. Such type specimens, therefore, constitute the ultimate scientific reference for any name, and constitute a lasting nomenclatural standard for taxonomy.

Although the emphasis of this manual lies with type specimens, many of the techniques and standards presented and discussed are obviously applicable to other biological specimens and materials. Our emphasis in this manual on type specimens, however, is explained by the special role they play in taxonomy and the great importance assigned to them by taxonomists. Their importance is underlined by the overall priority assigned to them when digitising specimen information including such proposals as the “e-type” initiative (e.g. BERENDSOHN & OEHLSCHLÄGER 2004, see also chapter on E-Types by SPEERS, in this volume). It is for this reason that taxonomists press for the highest possible standards in image quality and it means that they will, within reason, expend the time necessary to achieve top quality when documenting type specimens. Time and resources available, however, are unlikely to permit such standards to be extended to the vast number of non-type specimens, but they will still be useful as general guidelines.

With the entire field of information technology, particularly the area of digital photography and imaging, still undergoing rapid development, some of the technical information and equipment presented in this volume will become outdated quickly. Yet the general issues and challenges addressed in individual chapters, such as colour representation and management, effects of lighting, and techniques for handling specimens, will almost certainly persist. We hope therefore that much of the material presented in this volume will

remain useful for some time to come. But whatever its lifespan, we commend this volume to any reader with an interest in the field. And the editors and contributing authors encourage and welcome comments and additions, thus furthering the spirit of the workshops from which this manual evolved. To facilitate such iteration, and to promote the flow of information, the contents of this publication are also available in digital form from the publishers and they appear in full on the ENBI website (www.enbi.info). Efforts will be made to maintain and update this site to enrich further its contents in the future.

There is no doubt whatever that the digital age has reached and is transforming taxonomy. This manual addresses what are essentially practicalities. But we hope that the vast number of digital images of natural history specimens being created, whether of types or others, will contribute part of the infrastructure for what looks like a movement in taxonomy towards a modernised, web-based taxonomy (GODFRAY 2002; SCOBLE 2005).

Acknowledgements

Without the considerable funding for establishing the European Network for Biodiversity Information (ENBI) by the European Commission neither this volume nor much of the work it reports would have been possible. We thank the Commission for its support. The editors and the ENBI management acknowledge with gratitude also the support received from the Global Biodiversity Information Facility (GBIF), particularly towards participation of GBIF staff in the two workshops. We are grateful to all those who participated in these workshops and who helped to make them lively and productive events. We thank CHRIS JOHNSON and MELPO SKOULA and their colleagues for hosting the second workshop at the Mediterranean Agronomic Institute, Chania, Crete: both the venue and organisation were appreciated by all. ELIZABETH WATSON prepared the questionnaire for the case studies. VANESSA PIKE, the Natural History Museum, London, and MONIKA PFEFFER, Staatliches Museum für Naturkunde, Stuttgart, helped keep track of the financing of the preparation of this Manual, and we are grateful to them for their

efforts. We thank our institutions, the Staatliches Museum für Naturkunde, Stuttgart, and the Natural History Museum, London, and all those staff involved in ENBI matters and in the preparation of the workshops.

References

BERENDSOHN, W. G. & OEHLSCHLAEGER, S. (eds.) (2004): GBIF-D: German Participation in the Global Biodiversity Information Facility. Projects in the German National Programme for the Global Biodiversity Information Facility 2003-2006 - Status Report 2004. PT-DLR Environmental Research and Technology, Bonn; 156 pp.

GODFRAY, H. C. J. (2002): Challenges for taxonomy. – *Nature*, **417**: 17-19.

GREUTER, W., MCNEILL, J., BARRIE, F. R., BURDET, H.-M., DEMOULIN, V., FILGUEIRAS, T. S., NICOLSON, D. H., SILVA, P. C., SKOG, J. E., TREHANE, P., TURLAND, N. J. & HAKSWORTH, D. L. (eds.) (2000): *International Code of Botanical Nomenclature (St Louis Code). Regnum Vegetabile 138*. – Konigstein (Koeltz Scientific Books).

HÄUSER, C. L. (2004): *Global assessment of existing biodiversity – the state of and the challenges for our knowledge database.* In: BECK, E., BERENDSOHN, W., BOUTROS, M., DENICH, M., HENLE, K., JÜRGENS, N., KIRK, M. & WOLTERS, V. (eds.): Sustainable use and conservation of biological diversity – A challenge for society: 47-53. – Bonn (PT-DLR Environmental Research and Technology).

International Commission for Zoological Nomenclature (1999): *International Code of Zoological Nomenclature. Fourth Edition.* – London (International Trust for Zoological Nomenclature). xxix + 306 pp.

SCOBLE, M. J. (2005): Unitary or unified taxonomy? – *Philosophical Transactions of the Royal Society of London,* **(B) 359**: 699-710.

SUAREZ, A. V. & TSUTSUI, N. D. (2004): The Value of Museum Collections for Research and Society. – *BioScience*, **54** (1): 66-74.

The Place of the Manual in the European Network for Biodiversity Information

MALCOLM J. SCOBLE

Department of Entomology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK. E-mail: m.scoble@nhm.ac.uk

Key words: ENBI, context of manual, database cluster.

Hardly a day passes without reference in the scientific or popular media to the decline in biodiversity. Preserving biodiversity requires strong socio-political backing alongside the need for the practice of conservation based on our knowledge of biodiversity and ecology (MOONEY, CROPPER & REID 2005). But implementing the protection of biodiversity requires data and analysis. The broad goal of the European Network of Biodiversity Information (ENBI) is to coordinate European efforts on biodiversity data by improving access to the scattered biodiversity information base and the expertise on which it is founded. Although, by definition, ENBI is a European initiative, the network was founded, *inter alia*, to make an integrated contribution to the Global Biodiversity Information Facility (GBIF). GBIF, in turn, facilitates digitisation and global dissemination of primary biodiversity data.

This manual arose from a specific programme of work within ENBI involving the organization and integration of networks of European natural history databases, specifically databases to specimens and collections. The approach taken was a practical one, based on the belief that organization and integration are best implemented by means of project work, each project being valuable in its own right yet also demonstrating the value of the broader network.

The ENBI setting

Among the main aims of ENBI is the development of a strong and open network of European biodiversity centres. Biodiversity information can be made accessible on a large scale only if people co-operate to create such networks in the form of extended and distributed teams. Biodiversity information of the kind being examined in ENBI exists neither in a single electronic 'warehouse' on the internet, nor even a few such 'warehouses', although great strides have been made in Europe by connecting specimen-level and collection-level databases to the BioCASE system (see below) and to GBIF. The actual specimens from which such data are derived are housed in the extensive natural history specimen collections scattered across the globe. Potentially, European data have the capacity to occupy collectively a special place in the cyber-infrastructure because the collections of the region have the best sample of biodiversity through time. Europe houses the oldest collections and the largest number of type specimens. Coordinated linkage of this material is, therefore, of fundamental importance to ENBI. Other primary ENBI activities are the maintenance, enhancement and presentation of biodiversity databases; data integration, interoperability and analysis; and user needs – particularly in terms of products and e-services.

The preparation of this manual lies within the first of these three other clusters of activities. Within the clusters, effort is divided into parcels of work known, in projects supported by the European Commission, as 'Workpackages'. The Workpackage in question involves projects in the coordination of European databases of biological collections. One of these projects has built on a line of work that was developed through earlier EC-supported initiatives (BioCISE, Resource Identification for a Biological Collection Information Service in Europe, www.bgbm.fu-berlin.de/biocise; ENHSIN, the European Natural History Specimen Information System, www.nhm.ac.uk/science/rco/enhsin (SCOBLE, 2003); and BioCASE, a Biological Collections Access Service for Europe, http://www.biocase.org/), in which a common access system, developed in Europe, and utilising a new global data specification standard being produced by CODATA/TDWG (Committee on Data

for Science and Technology/Taxonomic Database Working Group), was adopted to link selected existing European specimen databases into a distributed European network. Within ENBI, linkage of many further databases to the system has delivered a European-focused contribution to GBIF and thus encouraged adoption of an integrated European approach to access. The greater the number of databases linked, the greater will be the usefulness and thus sustainability of the system.

Within ENBI, we have also completed the digitisation of a large and unique index-card archive housed in a European institution on the names of Lepidoptera. Each card in the archive contains basic taxonomic information: notably the name of the species or subspecies, the author who originally described it, the date of description and the place of publication, and the taxonomic status of the name (particularly whether it is a synonym or homonym). The original genus and the type locality are usually given on the card. As an archive, the information is, by definition, not up to date. However, the database into which images of the index-cards are stored is constructed so as to allow addition of new information, and the user interface was designed for appropriate input accordingly. The availability of the database at http://www.nhm.ac.uk/entomology/lepindex/ provides basic data on 10 percent of the names of all animal species online.

The focus of the programme of work covered in this manual is on type specimens. ‘Types’ act as name-bearers for species and are, therefore, reference specimens of a special kind. European institutions house a large proportion of the world’s types, representing potentially a major scientific legacy that will form, once databases are networked, a unique contribution to GBIF. Demonstrators for this project include type data and images for a group of fish and a group of butterflies (http://nx1tmp.s2you.com/platform/projects/globis), organisms that were selected to examine aspects of digitisation for differently shaped specimens.

Two other Workpackages are included in the biodiversity database cluster of ENBI. One of these is on European species bank projects, the other is on the design and promotion of common standards for

access to observational survey data. The cluster as a whole therefore covers networks of databases of observations and specimens (unit-level data), databases containing descriptive metadata about collections (examples can be seen at the BioCASE website), to nomenclature databases at the level of species names (see Species 2000 Europa at http://sp2000europa.org/ and Species 2000 at http://www.sp2000.org/).

References

MOONEY, H., CROPPER, A. & REID, W. (2005): Confronting the human dilemma: how can ecosystems provide sustainable services to benefit society? – *Nature,* **434**: 561-562.

SCOBLE, M. J. (ed.) (2003): *ENHSIN: The European Natural History Specimen Information Network.* – London (The Natural History Museum). 176 pp.

E-Types – A New Resource for Taxonomic Research

LARRY SPEERS

Global Biodiversity Information Facility, Universitetsparken 15, DK-2100 Copenhagen Ø, Denmark. E-mail: lspeers@gbif.org

Key words: Natural history collections, E-types, digital images.

Introduction

The challenge for taxonomy has always been the efficient discovery and description for the benefit of science and society of the earth's estimated 3.6 to 100+ million species. Over the last two centuries it is estimated that approximately 1.75 million species or, depending on the total estimate used, only between 2% and 49% of all living taxa that co-exist with us on this planet have been described (HAMMOND, 1995). If modern taxonomy hopes to complete the challenge of efficiently identifying and documenting the remaining 50% to 98% of life on this planet in a reasonable length of time we need to examine the bottlenecks and inefficiencies in the current taxonomic process and identify solutions that will increase the efficiency and rate of the descriptive process.

In this the "information age", one such inefficiency is taxonomic science's comparatively slow adoption of modern information management and information exchange technologies. As has been demonstrated in other branches of science, for example genomics, appropriately applying these modern technologies can greatly increase the efficiency and cost effectiveness of research. The comparatively slow adoption of new electronic technologies by the taxonomic community may be partly due to the conservative nature of this community but it is primarily a consequence of how critically dependent taxonomic science is on utilizing physical specimens and associated historical descriptive literature as its basis.

Traditionally, taxonomic research entailed collecting new specimens through field work, sorting and identifying this material, depositing these physical vouchers in a natural history collection and, through comparisons of this new material with previously deposited material, and often specimens borrowed from other institutions, describing new taxa where appropriate. Over the last 200 years, this process has resulted in a proliferation and expansion of the world's natural history collections so that now their combined holdings provide us with our most complete record of life on this planet. These collections contain the specimens that are the *sine qua non* for taxonomic research and provide nearly three hundred years of historical information vouchering the historical existence of organisms at particular times and places.

These natural history collections also contain the "type" specimens selected by earlier workers as a standard of reference for the correct assignment of a published biological name and as an exemplar of the circumscription of the taxon. These nomenclatural types are recognized by the scientific community as essential to the stability and validity of biological names (DASTON, 2004) and examining this material is a crucial step in defining the circumscription of existing named taxa. The reliance on these critical specimens to provide nomenclatural stability makes the science of taxonomy different from other sciences. Other sciences like chemistry and physics can to a large extent rely solely on a widely duplicated and in general broadly accessible literature base to document the progressive development of their scientific knowledge. Taxonomy, however, relies for historic stability not only on its literature base but also on single physical specimens or series of specimens that have been and are being deposited as types by specialists in the world's natural history collections.

In the vast majority of cases, however, these collections have not been developed through any long-term strategic plan and as a result it is generally very difficult to predict the taxonomic, temporal and geographic coverage of any one collection's holdings. Taxonomic research and the resulting historic expansion of individual natural history collections have been relatively opportunistic with the growth of any individual collection and its complement of type holdings being

dependent on the particular taxonomic interests of its curatorial and research staff over time; the changing emphasis of funding agencies toward funding collecting activities; the historic opportunities for staff to join various collecting expeditions to different geographic areas and the personal networks of various researchers that have facilitated the exchange of specimens. A result of this pattern of expansion is that the critical type material that is the basis of our nomenclatural system and which must be examined by current workers is widely distributed across thousands of collections; some types are even in private individuals' collections. The traditional solution for accessing this material has been for the taxonomist to physically visit the institutions holding critical material or to request these institutions to exchange this material on 'loan' for examination. In the 'age of the internet' this reliance on moving people and material rather than the more efficient approach of moving information is one of the major challenges for taxonomy.

An additional consequence of the historical pattern of taxonomic research and its resulting impact on collections development is that traditionally most taxonomic research has been conducted by researchers from developed countries and the material they collected while conducting this research has been deposited in their home institutions. As a result, the vast majority of type material is held in collections in developed countries. If the taxonomic community is to meet the challenge of describing the biota of these mega-diverse countries it is essential to facilitate access to this historic type material so it can be studied by taxonomic researchers working at the local level. The high cost of the traditional approach to taxonomic research based on moving people and material is a significant impediment to taxonomists working in these mega-diverse countries. It is clear that through the appropriate application of modern information and communication technology we can increase the efficiency and cost effectiveness of taxonomic research while at the same time making critical data and information available to users in mega-diverse countries.

E-Types

Recent advances in digital imaging technology, broadband internet connectivity, computer software for manipulating digital images and image compatible database software now make it practical and efficient for natural history collections to make digital images of their type specimens globally accessible. The term “E-Type” originates from the 2002, the All Species Foundation “E-Type Initiative” <http://www.all-species.org/summer/initiatives.html> (accessed on 29/VII/2005) and refers to a digital image of a type specimen. Many institutions are already taking advantage of these technological developments and the number of images of types accessible over the internet is rapidly increasing. The following are just a few examples of web sites that are currently serving images of type specimens.

- The New York Botanical Garden’s - Vascular Plant Types Catalog <http://sciweb.nybg.org/science2/hcol/vasc/index.asp> (29/VII/2005)
- Digital specimen images at the Herbarium Berolinense < http://ww2.bgbm.org/Herbarium/default.cfm> (29/VII/2005)
- Type Specimens In The University Of Florida Herbarium <http://www.flmnh.ufl.edu/herbarium/types/digitalimaging.asp> (29/VII/2005)
- Linnean types in the Swedish Museum of Natural History < http://linnaeus.nrm.se/botany/fbo/types.html.en> (29/VII/2005)
- Collections Database Nationaal Herbarium Nederland < http://www.nationaalherbarium.nl/virtual/> (29/VII/2005)
- Harvard University Herbaria < http://www.huh.harvard.edu/databases/cms/rock-types.html> (29/VII/2005)
- Smithsonian Institution Botanical Type Specimen Register <http://ravenel.si.edu/botany/types/index.cfm> (29/VII/2005)
- MCZ Type Database @ Harvard Entomology < http://insects.oeb.harvard.edu/mcz/> (29/VII/2005)
- Type Specimens of Japanese Ants at MCZ, Harvard University <http://ant.edb.miyakyo-u.ac.jp/MCZ/index.html> (29/VII/2005)
- Images of NZ Lepidoptera type specimens http://www.landcareresearch.co.nz/research/biodiversity/invertebratesprog/lepidoptera/index.asp (29/VII/2005)

- California Academy of Sciences Department of Ichthyology Primary Types Imagebase <http://www.calacademy.org/research/ichthyology/Types/about_types.html> (29/VII/2005)

As the number of these sites rapidly expands it is clear that, to allow the taxonomic community to take full advantage of this exciting development, there is an urgent need for metadata standards to document these images, for imaging protocols to ensure data quality, and for specific recommendations from taxonomic specialists as to what views and details of the type specimens for their particular taxonomic groups would be must useful. A number of these issues are discussed in other papers in this volume. In addition, there is a need for some form of Globally Unique Identifiers (GUIDs) that will allow the unique identification of each individual image. This will facilitate referencing the image when it is utilized. The most critical need, however, is for the development of a registry of these data resources that will support the indexing of each resource and thus support rapid global searching for type images. The Global Biodiversity Information Facility (GBIF) (www.gbif.org) (29/VII/2005) already has through its prototype data portal much of the infrastructure in place to support these searches and is working with the Taxonomic Database Working Group (TDWG) (www.tdwg.org) (29/VII/2005) to refine and enhance these capabilities.

Access to e-type images will allow taxonomists world wide to examine type material without the need for accessing the physical specimen. This will greatly increase the efficiency of the taxonomic process by allowing taxonomists to easily identify exactly what type material needs to be examined in more detail and thus greatly reduce the need for travel and greatly reduce requests for the loan of type material. Nevertheless, it should be recognized that, in many cases, having access to digital images of types will not be a substitute for researchers physically examining critical specimens. Having access to these images will significantly reduce the need for travel and the loan of specimens but these aspects of taxonomic research cannot be completely eliminated as many taxonomically important characters, such as molecular data can not be derived from images.

References

DASTON, L. (2004): Type specimens and scientific memory. – *Critical Inquiry,* **31**: 153-182.

HAMMOND, P. (1995). The current magnitude of biodiversity. In: HEYWOOD, V. H. and WATSON, R. T. (eds.): *Global Biodiversity Assessment:* 113-138. – Cambridge, U.K (Cambridge University Press).

Colour management

ROBERT A. MORRIS

Department of Computer Sciences, UMASS-Boston, 100 Morrissey Blvd, Boston, MA 02125, USA. E-mail: ram@cs.umb.edu

Key words: Color, sRGB, CMYK, ICC Color Profile, CIE XYZ Color Model, Chromaticity, gamut.

Several issues make the management of colour a difficult problem as anyone knows who has struggled with making a printed image look the same as an image on the screen. The main problem is that colour appearance, although heavily influenced by the spectral energy reaching the eye, is first and foremost determined by computations made in the viewer's brain. For this reason, colour models that are based on human perception have been heavily used in the graphics arts for many years. The so-called CIE XYZ colour model uses a coordinate system which is, at its heart two sets of light spectra and an intensity axis, based on colour matching experiments using a small collection of human observers. These were determined by averaging the outcomes of a number of psychophysical experiments in France in 1931. In practice, to normalize for intensity this model is often normalized to X+Y+Z=1 to produce an intensity independent two-dimensional model with x=X/(X+Y+Z), y= Y/(X+Y+Z), This results in a representation of the visible spectrum as shown in figure 1. With this kind of representation it is common to display the set of colours a particular device or colour model is capable of representing, independent of the intensity of the colours, called the *gamut* of the device or colour model. The numbers x and y are the *chromaticity* of the colour and range between 0 and 1. For highly technical colourimetry, the colour of an object is often described by its *reflectance spectrum*. This is a function $r(\lambda)$ describing the proportion of light of wavelength λ reflected by the object. Typically λ encompasses the visible spectrum from 450 to

850 nm, but ultraviolet reflectance is of biological interest in part because some plants use ultraviolet colours as pollinator attractors. Typically such measurements are made at 1 or 5 nm intervals, although the corresponding functions for physical illuminants can sometimes be computed from the underlying physics of the illumination. The CIE XYZ colour model actually has three such functions, also called X, Y, and Z, and among their properties are that every reflectance function limited to the visible spectrum can be represented as a linear combination of the X,Y,Z basis functions. If the functions are named $X(\lambda),Y(\lambda),Z(\lambda)$ this means that we can always write

$$r(\lambda) = X_r * X(\lambda) + Y_r * Y(\lambda) + Z_r * Z(\lambda)$$

The numbers (X_r, Y_r , Z_r) are the XYZ coordinates of the colour represented by $r(\lambda)$. One advantage of this colour model over others is that colours close in this space will be regarded as close by human observers having normal colour vision.

From figure 1 one can see the problem of gamut mismatches which is often addressed by image processing software such as Adobe Photoshop. Here we see that the popular sRGB colour space often used to describe digital images presented on a monitor can represent some colours that a printer can not. Worse, each printer make and model – and sometimes even individual printer – will have a different gamut. Typically, image processing software gamut adjustment consists of mapping the colours that are out of, say, the printer gamut, to those which are in it. Often this is semi-automatic and the software's mapping choice may or may not be biologically appropriate if it renders a colour that is of some particular significance, e.g. a character state like "flower colour". In addition, mapping an RGB colour to a different RGB colour does not necessarily survive a change of printer as being an appropriate choice. There are in fact two slightly different RGB colour spaces in common use, the colour space and associated colour profile sRGB issued as a standard by the Worldwide Web consortium, and a related one defined by Adobe. Figure 1 shows two gamuts: sRGB is shown in the larger triangle and a simplified typical printer gamut in the smaller one. The data for the latter are derived from real measurements of printed images for a

particular set of inks reported in Table 2 of VIGGIANO & WANG (1991). Several things are apparent from this picture (and a few are not). The most important is that monitor gamuts are usually much larger

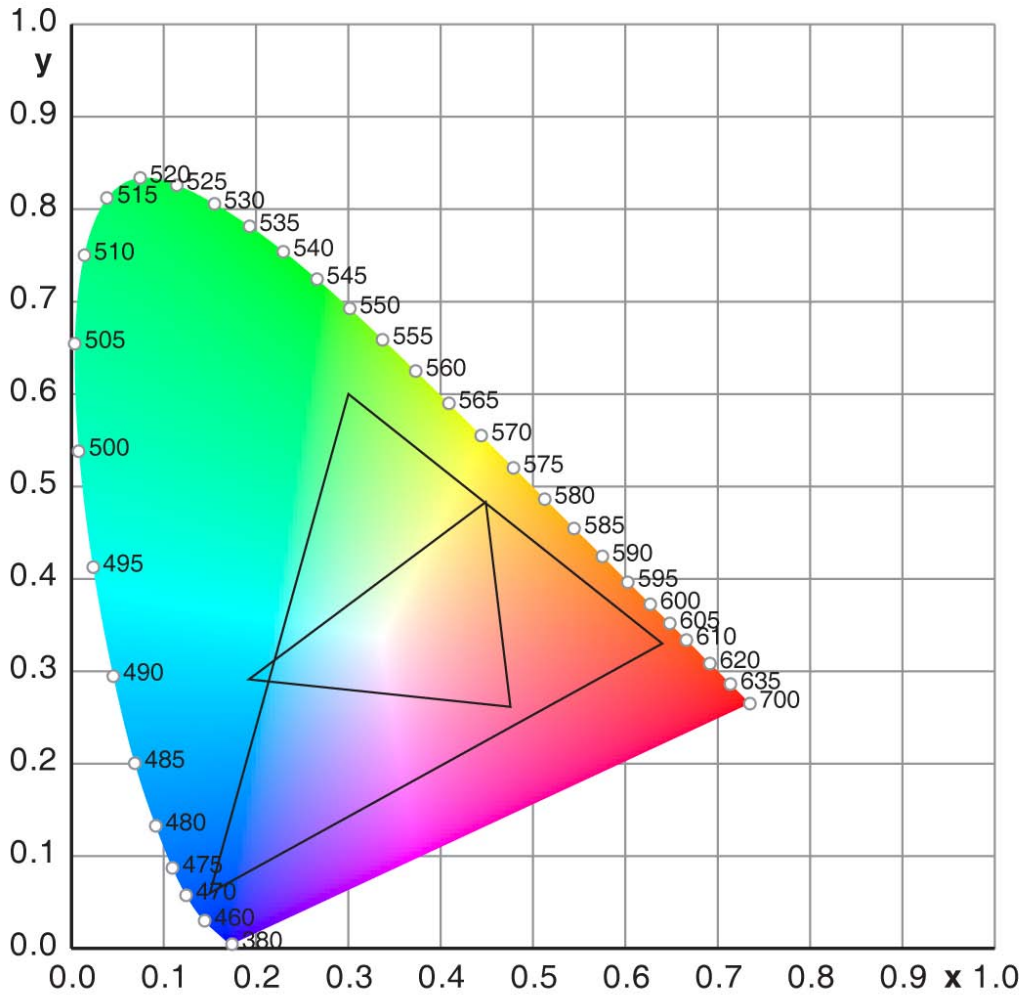


Fig. 1. Chromaticity diagram showing sRGB gamut and the smaller simplified printer gamut computed based on measurements of actual printer inks (see text). The diagram was computed with a minor modification of the PostScript program of Gernot Hoffman. Substantially more sophisticated treatment of colour may be found on his web site http://www.fhoemden.de/~hoffmann/ciegraph17052004.pdf. By the very principles illustrated, readers should not expect that a printed version and an online version of this figure will look the same. Border colours are pure spectral colours, given by the wavelength on the border. Interior colours are mixtures. Further discussion is at http://wiki.cs.umb.edu/twiki/bin/preview/BDEI/ColorManagementIllustrations.

than printer gamuts, as is the range of colours capturable by digital cameras and scanners. For this reason, high-end image processing tools like Photoshop provide, in a more or less easy way, for users to have their images displayed on screen with only colours that are representable by the printer. One implication of this difference is that colour manipulations made to a monitor gamut will often be unsatisfactory upon printing. If you mean to support both a screen and a printer version of your images, you need to either be satisfied with the printer mapping of your screen image, or consider managing two different copies of the image, one for each device. The surest way to be most satisfied with both is to edit to both gamuts. However, the automatic gamut mapping provided by most image processing software is often deemed satisfactory for many purposes. A fuller treatment of colour models and gamuts may be found in BERNS (2000). For an appreciation of what is involved in producing a printed book of high colour fidelity, see BERNS & REIMAN (2002).

In the biological sciences it is rare to represent colours in the CIE XYZ space or the corresponding xy chromaticity space. However, these spaces are at the heart of another kind of colour description in wide use in the colour imaging community. These are the ICC (International Color Consortium) Colour Profiles. Such profiles are files that characterize, among other things, the colour aspects, including the gamut, of some particular device (or, sometimes, of a hypothetical abstract device based on a colour model. For example, there is an ICC Profile for sRGB which is what many CRT manufacturers design to. Colour profiles are standardized, named data sets with a standard file format that image colour management software can read. It is common for printer manufacturers to provide an ICC profile for their devices, though sometimes it is available only on request. Using ICC profiles, image processing software like Photoshop can generally do reasonable gamut mapping between the acquisition and output devices, *if* both of them actually correspond to their published profiles. In recent years, modestly priced monitor and scanner calibration packages have come to market with both software and colourimetry hardware suitable for building ICC profiles for your equipment in case it has diverged from the original. See http://www.adobe.com/support/techdocs/321382.html for some

details and a brief tutorial on ICC profiles, http://www.color.org/profile.html for a more complete description, and http://www.color.org/faqs.pdf for the ICC FAQ. The impending generation of LCD displays will be backlit with Light Emitting Diodes ("LED"s) instead of the current use of fluorescent back lighting which limits the gamut to something smaller than most CRT displays. LED backlit displays are brighter and with larger gamuts than CRTs. For this reason, the colour mismatch between printers and screens is likely to increase in the coming years, and rigorous use of these ICC profiles is likely to be even more important.

BEST PRACTICE: Attempt to acquire the ICC Colour Profile for your imaging devices and store it as part of your image library, with a reference to it in the image Metadata

Scene Illuminant. In a typical use of an image, a file is presented with each pixel having three coordinates in some colour space. In general, this may not be enough to reconstruct the colour reflectance spectrum at the point represented by the pixel, and without this it is difficult to know that a rendering from those values will appear the same as it might have to an observer at the original scene. Even if you have a good characterization of the reflectance spectrum of the object imaged – or you are prepared to accept some representation of each colour as given in some colour model – it is necessary to know the spectral distribution of the illuminant for a perfect reconstruction of the image. The emission spectra of some typical illuminant sources is often expressed as "colour temperature" in degrees Kelvin, being the spectrum of an ideal black-body radiator of that temperature. An illuminant designated D65 meets an international standard for approximating mean noon daylight and traditional "daylight" films are balanced for this. However, most digital camera flashes and auxiliary flash are xenon flash tubes, which have a very uniform spectrum close (in the visible spectrum) to natural daylight or a Black Body Radiator at 6500°K. Because of this, most flash photography can be done without sophisticated characterization of the illuminant.

BEST PRACTICE: Image your specimens with xenon flash (unless special requirements apply; see chapters in "Current Approaches" section).

Colour cards. There are colour target cards widely available that comply with the IT8 standard for scanner calibration. Inexpensive ones are available from http://www.targets.coloraid.de/. Including such a target with your scanned images and photographed images can help make colour calibration, hence comparison, easier.

BEST PRACTICE: Include an IT8 colour target in your images where possible.

References

BERNS, R. S. (2000): BILLMEYER and SALTZMAN's Principles of Color Technology. 3rd Edition. – John Wiley & Sons.

VIGGIANO, J. A. S. & WANG, C. J. (1991): A Suggested Method for Computing Colorimetric Densities. – *Proceedings*, *Annual Technical Conference of the Technical Association of the Graphic Arts,* Henrietta, NY, 5-8 May, 1991: 196-215.

BERNS, R. S. & REIMAN, D. M. (2002): Color managing the third edition of BILLMEYER and SALTZMAN's Principles of Color Technology. – *Color Research & Application,* **27** (5): 360-373.

Image Metadata Standards and Practices

ROBERT A. MORRIS

Department of Computer Sciences, UMASS-Boston, 100 Morrissey Blvd, Boston, MA 02125, USA. E-mail: ram@cs.umb.edu

Key words: Image metadata, DIG35, Z39.87, EXIF, IPTC, SDD, ICC Color Profile.

Image metadata may be thought of comprising three somewhat distinguishable parts: technical, content, and curatorial metadata. Technical metadata consists of information about the image acquisition system, including enough information to assist in colour and other image manipulations. Some technical metadata is typically determined by the acquisition system itself and available either embedded in the image file or available to external applications. Content metadata describes the image itself and is not usually produced by the image acquisition system, with the exception of date, time, and for GPS enabled cameras, the location of the acquisition. Curatorial metadata refers to information that assists in the location of the image, intellectual property rights, and similar such issues of concern to the storage and distribution of the image. There are a number of image metadata standards presently accepted by one or another standards body. In the next section we will present recommendations for minimal metadata for images, and then examine the prevailing standards with respect to those data. In a later section we will discuss each standard in greater detail.

Summary recommendation for minimal metadata representation:

Use the information represented either in DIG35 or Z39.87, including encoding of metadata captured by the imaging device.

Exchange this metadata in XML form using the corresponding schema.

Minimal metadata:

Technical metadata

Original digitization resolution
Original digitization resolution units
Original digitization size
Original digitization units
Image acquisition system manufacturer and model

Colour data

Colour model or ICC Profile. See MORRIS on "Colour Management", in this book.

Curatorial metadata

Date of data acquisition
Time of data acquisition
Intellectual Property Information

Content metadata

Biota of interest

Representation of metadata:

In this section we focus on a number of existing image metadata standards briefly.

EXIF. The EXIF standard EXIF.org is a popular metadata representation often used by digital camera manufacturers to encode camera technical data and, typically, date and time of the exposure. The cameras usually embed the data in the header of jpeg files and TIFF files. Many image processing applications can retrieve and edit it, and there is programming support for embedded EXIF in most widely used programming environments, include Java, .NET, PHP, Perl, and Python. An example of EXIF data is shown below. As illustrated after that example, most of the data encoded in EXIF can

also be represented in the DIG35 and Z39.87 standards, so we do not dwell on it here.

DIG35. The “Digital Image Group” was an industry association now absorbed by the International Imaging Industry Association (I3A, International Imaging Industry Association). It promulgated an extensible image metadata standard called “DIG35” which has emerged as DIG35 v 1.1. Many DIG35 elements are optional, so do not present a burden for systems that cannot represent its rich collection of metadata objects. In addition the standard provides for the embedding of DIG35 <METADATA> elements in arbitrary XML documents.

A DIG35 document consists of a root <METADATA> and a sequence of six types of optional subelements, the last of which are 0 or more <METADATA> elements which provide for a recursively defined tree of metadata.

The second level DIG35 elements:

BASIC_IMAGE_PARAM. This includes optional metadata about the image file format, a Unique Identifier, image size, preferred rendering size, and colour model.

IMAGE_CREATION. This includes optional technical data about the image acquisition system, but also about the human and organization responsible for the creation of the image. The technical data can be quite extensive, including data about the optics and colour models in play and other information for sophisticated characterization of the imaging device, as well as details of the imaging event itself, such as exposure data, the name and organization of the creator, etc.

CONTENT_DESCRIPTION. Content description element contains detailed optional information about the location of the scene and the date, time and season of its acquisition. Location information can be encoded as a place name (including postal address) or as detailed GPS information, including map datum. Of most potential interest are several subobjects which describe parts of the picture in greater detail. There are a number of potential overlaps with the SDD

schema, which is beyond the scope of this work. These subobjects can be 0 or more PERSON items, 0 or more THING items, and 0 or more ORGANIZATION items. Each can have its position and extent within the picture given and also it geolocation specified in the same manner as the location of the original scene. The THING objects are recursively defined, so that it is possible to describe a complex decomposition of an object in the image. Of particular interest for field photography is an element EVENT, which can describe events depicted in an image, including the participants, (as PERSONs, THINGs, etc.) and to a limited degree, the relations of the events to one another. Finally, there is support for property lists - more or less characters/state pairs - which can also be mapped into Descriptions in the SDD schema. These lists have a limited form of name scoping by the use of a subobject called a DICTIONARY that essentially is a namespace for the PROPERTY, which is provided with a reference to a DICTIONARY object

HISTORY. This element provides for the image manipulation history of the image. It allows the recording of the fact that a small number of particular manipulations have taken place, and for human-targeted free text describing what that manipulation was.

IPR. Besides the rather detailed, but expected, specification of IPR ownership and usage requirements, this element has a HISTORY subelement which is a sequence of IPR elements, providing for a tree of IPR objects.

METADATA. At the end a METADATA element there is an optional list of METADATA objects, providing for a recursively defined tree of METADATA objects.

Z39.87 (NISO 2002) is a draft standard of the U.S. National Information Standards Organization (NISO). Its concern is almost entirely with technical metadata of the acquisition system and it is extremely sophisticated about issues such as colour management and image quality analysis. In addition, it is somewhat stronger about technical metadata for scanners than is DIG35. Finally, it has an expression, MIX, (Library of Congress and NISO 2004), in the Metadata Exchange and Transmission Standard (Library of

Congress 2005), which is widely used in the Digital Library community. The standard has been receiving attention in cultural heritage image collection management communities where faithful rendition of non-digital art is of great importance. Z39.87 provides for the identification and location in the image of one or more standardized targets (e.g. colour or grayscale cards; see MORRIS on "Colour Management", in this book). Projects with scientific imaging for which faithful colour is of high importance should examine Z39.87. Adobe's eXtensible Metadata Platform (XMP), a metadata standard implemented across many of the company's products is based in part on Z39.87. It is likely that Z39.87 will be a supported metadata type for JPEG2000. Because Z39.87 does not provide for content metadata, its use would need to be augmented by elements from another standard.

Most of the element names of DIG35 are global. MIX has a single top level global element, ***mix,*** but its XML complexType definitions are all global, this makes it slightly easier to embed DIG35 elements in multiple schemas, since they will all have the same name and not require any mapping between element names, which, however, is not difficult.

The second level Z39.87 elements:

BasicImageParameters. Documents the file structure with sufficient parameters to insure a reasonable image can be rendered for viewing.

ImageCreation. Describes how, when and by whom the image was created.

ImagingPerformanceAssessment. Contains the metadata pertaining to the image quality, such as colour management, digitization resolution

ChangeHistory. Supports tracking of image processing that has been applied to the image.

Other significant metadata standards

IPTC. The International Press and Telecommunications Council has promulgated a standard for the representation of technical and content metadata. The content metadata is aimed at describing topics of interest to news media. One interesting aspect of the IPTCv4 standard is that it has representation of limited ontology information by providing enumerated subjects and categories of content (i.e., a "kind-of"), such as "Forecast", "History", "Summary", etc. Attempting to extend this for biological data does not seem profitable since other ontology standards have evolved for the Semantic Web and are likely to be more widely adopted.

TDWG UBIF and SDD. The Taxonomic Data Working Group (TDWG, http://www.tdwg.org) has a number of standards presently in the review process which may serve the interests of those wishing to provide image curation and content metadata expressed in XML. The most general of these is the Unified Biological Information Schema (UBIF) proposed for the general metadata concerns of biological data exchanged on the internet. UBIF is discussed at http://wiki.cs.umb.edu/twiki/bin/view/UBIF/WebHome. A second standard, SDD (the Structure of Descriptive Data) is meant to address a large class of problems about describing objects. SDD (which uses UBIF for some of its concerns) supports the representation of community controlled vocabularies that can be shared among users concerned with a similar group of taxa, sufficiently similar that common terminology can be used for descriptive characters and states and the relations between them. SDD is discussed at http://wiki.cs.umb.edu/twiki/bin/view/SDD/WebHome.

JPEG2000. A new ISO standard, JPEG2000 is a complex multipart standard using a compression scheme that supports the ability to decode parts of an image without decoding all of it. More important, JPEG2000 supports more structured embedded metadata than classical JPEG. Parts of the standard are rapidly being adopted by camera and graphics arts vendors such as Adobe. Although tools for the manipulation and decoding of JPEG2000 are not yet widely available, it is likely that this will change over the next few years. The

groups managing Z39.87, DIG35 and other existing standards are, or soon will be, in the process of registering to be an optional embedded metadata type for JPEG2000.

TABLE 1. Correspondence of some common metadata elements in Z39.87 (MIX) and DIG35. Upper level XML elements containing these are omitted.

Metadata Item	**Z39.87**	**DIG35**
Creation date/time	DateTimeCreated	GENERAL_CREATION_INFO/CREATION_TIME[a)]
Image pixel width	ImageWidth	IMAGE_SIZE/WIDTH
Image pixel length	ImageLength	IMAGE_SIZE/LENGTH
Source width	Source_Xdimension	N/A[b)]
Source width units	Source_XdimensionUnit	N/A[b)]
Image length	Source_Ydimension	N/A[b)]
Source length units	Source_YdimensionUnit	N/A[b)]
MIMEType	MIMEType	FILE_FORMAT/MIME_TYPE
Color Space[c)]	ColorSpace	COLOR_INFO/COLOR_SPACE
ICC Profile[c)]	ProfileName	COLOR_INFO/COLOR_SPACE
Image Producer	ImageProducer	
Imaging Device	DeviceSource	
Scanner parameters		
Manufacturer	ScannerManufacturer	SCANNER_CAPTURE/SCANNER_INFO/MANUFACTURER
Model	ScannerModelName	SCANNER_CAPTURE/SCANNER_INFO/MODEL
	ScannerModelNumber	SCANNER_CAPTURE/SCANNER_INFO/MODEL
SerialNumber	ScannerModelSerialNumber	SCANNER_CAPTURE/SCANNER_INFO/SERIAL
Software	ScanningSoftware	SCANNER_CAPTURE/SOFTWARE_INFO
Pixel Size	PixelSize	SCANNER_SETTINGS/PHYSICAL_SCAN_RES/SCAN_RES
Physical Resolution	XphysScanResolution	SCANNER_SETTINGS/PHYSICAL_SCAN_RES/WIDTH
	YphysScanResolution	SCANNER_SETTINGS/PHYSICAL_SCAN_RES/HEIGHT
Digital Camera Parameters		
Manufacturer	DigitalCameraManufacturer	CAMERA_CAPTURE/CAMERA_INFO/MANUFACTURER
Model	DigitalCameraModel	CAMERA_CAPTURE/CAMERA_INFO/MODEL
Serial Number		CAMERA_CAPTURE/CAMERA_INFO

TABLE 1. Correspondence of some common metadata elements in Z39.87 (MIX) and DIG35. Upper level XML elements containing these are omitted.

Metadata Item	**Z39.87**	**DIG35**
		/SERIAL
Exposure Data		
F Number	FNumber	CAMERA_SETTINGS/F_NUMBER
Exposure Time	ExposureTime	CAMERA_SETTINGS/EXP_TIME|R_EXP_TIME[d)]
Subject distance	SubjectDistance	CAMERA_SETTINGS/SUBJECT_DISTANCE
Illuminant	SceneIlluminant	CAMERA_SETTINGS/SCENE_ILLUMINANT
Lens Focal Length	FocalLength	CAMERA_SETTINGS/FOCAL_LENGTH
Flash type	Flash	N/A[e)]

Notes:

a) Also supported in attribute TIMESTAMP on element CAMERA _SETTINGS

b) DIG35 provides for recording the subject distance but no way to record the size of the field of view for cameras. For scanners, this can be computed from the scanning resolution, but for cameras it is good practice to include a ruler in the scene.

c) DIG35 permits only the name (or URL) of an ICC Colour Profile, whereas Z39.87 provides for detailed specification of the colour space in terms of the CIE model (see MORRIS, "Colour Management", in this book) as well as an ICC Profile. This is not a major difference because most users will prefer to acquire an ICC Profile for their device and use that, since many image processing tools support it directly.

d) May be either a double (EXP_TIME) or a rational number (R_EXP_TIME)

e) Z39.87, but not DIG35 provides for recording the flash model, but for scientific photography, the scene illuminant is what is important. Most modern flashes are pulsed Xenon lights and correspond closely to the D65 standard daylight illuminant. See MORRIS, "Colour Management", in this book, for further information.

The XML at the right represents the portion of the EXIF from the above image which is extracted by the widely used *jhead* program (Wandel 2005), enhanced with jhead-xml(LA POUTRÉ 2004). Below we exhibit metadata for the same image, including further content metadata, represented in both Z39.87 and DIG35. Original image by Jennifer Forman Orth, available at http://gallery.cs.umb.edu/gallery/NFSPlants/IrisVersicolor3 under a Creative Commons Attribution-ShareAlike 2.5 license.

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<ImageList>
  <ImageInfo fname="IrisVersicolor3.jpg">
        <FileName>IrisVersicolor3.jpg</FileName>
        <FileSize units="bytes">1003189</FileSize>
        <FileDateTime isodate="20050516T002848Z">
          Mon, 16 May 2005 00:28:48 %z
       </FileDateTime>
        <CameraMake>NIKON</CameraMake>
        <CameraModel>E995</CameraModel>
        <DateTime isodate="20040602T094021">
          Wed, 02 Jun 2004 09:40:21
       </DateTime>
        <Resolution>
              <Width>2048</Width>
              <Height>1536</Height>
              </Resolution>
        <IsColor>true</IsColor>
        <FlashUsed>false</FlashUsed>
        <FocalLength units="mm">19.3</FocalLength>
        <ExposureTime units="s" equiv="1/189">
          0.005
      </ExposureTime>
        <ApertureFNumber>f/4.4</ApertureFNumber>
        <ISOequivalent>100</ISOequivalent>
        <Whitebalance>cloudy</Whitebalance>
        <MeteringMode>matrix</MeteringMode>
        <ExposureProgram>program (auto)
      </ExposureProgram>
        <JpegProcess>Baseline</JpegProcess>
  </ImageInfo>
</ImageList>
```

Below is shown the representation of this in both DIG35 and Z39.87 (MIX) form. The latter supports no content metadata, but for specimens one might consider augmenting it by use of DarwinCore Species Analyst or ABCD TDWG metadata. Both standards support extensive description of Intellectual Property Rights, not shown here.

DIG35 Representation

```
<?xml version="1.0" encoding="UTF-8"?><METADATA
xmlns="http://www.digitalimaging.org/dig35/1.1/xml"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xsi:schemaLocation="http://www.digitalimaging.org/dig35/1.1/xml
DIG35v1.1RAM.xsd" xmlns:dig35="DIG35v1.1RAM.xsd">
      <BASIC_IMAGE_PARAM>
            <BASIC_IMAGE_INFO>
```

```
                <FILE_FORMAT>
                    <FILE_NAME>IrisVersicolor3.jpg</FILE_NAME>
                    <MIME_TYPE>image/jpeg</MIME_TYPE>
                </FILE_FORMAT>
                <IMAGE_SIZE>
                    <WIDTH>2048</WIDTH> <HEIGHT>1536</HEIGHT>
                </IMAGE_SIZE>
            </BASIC_IMAGE_INFO>
            </BASIC_IMAGE_PARAM>
        <IMAGE_CREATION>
            <GENERAL_CREATION_INFO>
                <CREATION_TIME>2004-06-02T09:40:21.0Z</CREATION_TIME>
                <IMAGE_CREATOR>
                    . . .
                </IMAGE_CREATOR>
            </GENERAL_CREATION_INFO>
            <CAMERA_CAPTURE>
                <CAMERA_INFO>
                    <MANUFACTURER>
                        <ORG_NAME>Nikon</ORG_NAME>
                    </MANUFACTURER>
                    <MODEL>E995</MODEL>
                </CAMERA_INFO>
                <CAMERA_SETTINGS>
                    <EXP_TIME>0.005</EXP_TIME>
                    <F_NUMBER>4.4</F_NUMBER>
                    <METERING_MODE>matrix</METERING_MODE>
                    <SCENE_ILLUMINANT>Daylight</SCENE_ILLUMINANT>
                    <FOCAL_LENGTH>0193</FOCAL_LENGTH>
                    <FLASH>false</FLASH>
                    <EXPOSURE_INDEX>100</EXPOSURE_INDEX>
                    <CAMERA_LOCATION>
                        <ADDRESS>
                            <ADDR_NAME>UMASS Field Station</ADDR_NAME>
                            <ADDR_COMP>Nantucket</ADDR_COMP>
                            <ZIPCODE>02564</ZIPCODE>
                            <COUNTRY>US</COUNTRY>
                        </ADDRESS>
                        <COMMENT>Virtual Nature Trail Stop 8. See
http://efg.cs.umb.edu/nantucket</COMMENT>
                    </CAMERA_LOCATION>
                </CAMERA_SETTINGS>
            </CAMERA_CAPTURE>
        </IMAGE_CREATION>
        <CONTENT_DESCRIPTION>
            <THING>
                <PROPERTY>
                    <NAME>Taxon</NAME>
                    <VALUE>Iris versicolor</VALUE>
                </PROPERTY>
                <PROPERTY>
                    <NAME>Vernacular</NAME>
```

```
                    <VALUE>blue flag</VALUE>
                </PROPERTY>
                <PROPERTY>
                    <NAME>Rank</NAME>
                    <VALUE>Species</VALUE>
                </PROPERTY>
                <PROPERTY>
                    <NAME>Family</NAME>
                    <VALUE>Iridaceae</VALUE>
                </PROPERTY>
            </THING>
        </CONTENT_DESCRIPTION>
</METADATA>
```

Z39.87 Representation

```
<?xml version="1.0" encoding="UTF-8"?>
<mix xmlns="http://www.loc.gov/mix/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xsi:schemaLocation="http://www.loc.gov/mix/.xsd">
    <BasicImageParameters>
        <Format>
            <MIMEType>image/jpeg</MIMEType>
            <PhotometricInterpretation>
                <ICCProfile>
                    <ProfileName>sRGB</ProfileName>
                </ICCProfile>
            </PhotometricInterpretation>
        </Format>
        <File>
            <FileSize use="bytes">1003189</FileSize>
        </File>
    </BasicImageParameters>
    <ImageCreation>
        <ImageProducer>Jennifer Forman, University of Massachusetts-
Boston</ImageProducer>
        <DigitalCameraCapture>
            <DigitalCameraManufacturer>Nikon</DigitalCameraManufacturer>
            <DigitalCameraModel>E995</DigitalCameraModel>
        </DigitalCameraCapture>
        <CameraCaptureSettings>
            <FNumber>4.4</FNumber>
            <ExposureTime>0.005</ExposureTime>
            <MeteringMode>Pattern</MeteringMode>
            <SceneIlluminant>Daylight</SceneIlluminant>
            <FocalLength>0.0193</FocalLength>
            <Flash>No</Flash>
            <ExposureIndex>100</ExposureIndex>
        </CameraCaptureSettings>
        <DateTimeCreated>2004-06-02T09:40:21.0Z</DateTimeCreated>
        <Methodology/>
    </ImageCreation>
    <ImagingPerformanceAssessment>
```

```
            <SpatialMetrics>
                <ImageWidth>2048</ImageWidth>
                <ImageLength>1536</ImageLength>
            </SpatialMetrics>
        </ImagingPerformanceAssessment>
        <ChangeHistory>
            <ImageProcessing>
                <SourceData>IrisVersicolor3.jpg</SourceData>
            </ImageProcessing>
        </ChangeHistory>
</mix>
```

Managing image metadata

File names. There is a single best practice for managing media resource metadata that is most important but widely ignored: don't encode metadata in names of files or directories. While one often sees file or directory names that abbreviate taxon names, pixel size, and the like, such names frequently do not survive moves to other operating systems (For example, Windows file names are case insensitive, but Unix and Macintosh file names are case sensitive), or to some kinds of archival storage media. CDROM and DVD file system requirements now or in the future may impose yet different restrictions on the character sets, name and directory path lengths which may be incompatible with your current names. Finally, when used as keys to queries made across the internet, log names may interfere with some of the mechanisms that transport those queries. Although file systems are often thought of as though they were databases, it is preferable to use almost any kind of database to store your image metadata. From an information management point of view, images are just another kind of specimen, and a database similar in spirit and technology to your specimen management database should be your first consideration.

Globally Unique Ids (GUIDs). As with specimens, your media resources should have a globally unique identifier that can be used as a primary key into your image management database. An excellent introduction to requirements for GUIDs by DAVID THAU is at http://www.tdwg.org/2004meet/EV/TDWG_2004_Papers_Thau_1.zip. It has a brief description of the Life Sciences ID (LSID) system for GUIDs, and if you are adopting this for your specimens, your support

for that will serve you well for GUIDS for your images. Whether your specimen GUID scheme (e.g. institution code + accession number) is sufficient for adopting as an image GUID may depend on whether your present scheme is satisfactory in an internet environment. It should meet at least these requirements:

- Global across the Internet.
- Persistent for all time. That is, if the image moves to another location or institution, the GUID should not change
- Resolve to a unique media resource.
- Provide a key to metadata as well as media data.

Databases A small set of data may be adequately served by a flat file database such as a spreadsheet or a low-end database system, but such systems are not generally robust enough for large collections. They are difficult to interface to internet services, and are subject to the kinds of data entry errors that enterprise-level systems protect against. For example, if the location of the specimen acquisition is a named place that you wish to have in your metadata, and your images represent many specimens from the same location it is better to have a separate database table with the attributes of those locations and use database relations to insure that simple data entry errors, e.g. different spellings of the same place name, do not pollute the data. In addition, if you have such tables in your specimen database, you can integrate the specimen records and image metadata and insure further integrity. Commercial enterprise-level relational databases including Oracle, Microsoft SQL-Server, and IBM DB2 are often available to academic and cultural institutions at very reduced pricing. Two popular open-source relational databases, MySQL and PostgreSQL also can serve as well as the commercial products in many cases, and have large communities supporting them. No database solution will be without support costs, and it is important to budget for them.

When considering a database system for storage of your metadata, it is important to keep in mind whether you intend to provide some or all of your metadata as a web service. This is the contemporary way of sharing data of any kind on the internet, and the recommended way to do this entails offering your metadata in XML form using an

agreed-upon XML controlled vocabulary conforming to a published XML Schema. This is discussed elsewhere, but here it should be noted that you should plan for this service at the time you plan your metadata database and make technical and budgetary plans for such a service. All of the above-mentioned enterprise-level databases have, or can be connected to, mechanisms for providing XML, and you can share expertise (and perhaps expense) if you collaborate with others in your institution who also have need to offer web services, even if unconnected to media resources or specimens. Commercial relational databases typically have high-performance XML support, but there are also a number of open source "native" XML databases, including eXist (http://exist.sourceforge.net/), BerkeleyDB XML (http://www.sleepycat.com/products/xml.shtml), and Xindice (http://xml.apache.org/xindice/). These could be used if you wish to store your XML metadata in native format. Such a solution requires more sophisticated programming at the moment, because these systems typically are not complete solutions, but rather something to be built into an integrated data management system custom built.

Representing time and date. Use UTC-referenced times for all your time-based metadata. It is otherwise difficult or impossible to know the time and date at which the image was acquired except to within 24 hours. This may entail setting the date and time on the imaging equipment or providing for the conversion in the software implementing the metadata storage. The Worldwide Web Consortium recommends storing dates and times in the ISO 8601 standard, and this is common in interchange formats, so should be used where possible in image metadata. ISO 8601 dates contain self describing timezones, e.g. 2005-01-16T19:20+01:00 to indicate January 16, 2005 at 19:20 in the time zone UTC+1. 2005-01-16T18:20Z designates the same time UTC. Thus, the ISO 8601 standard permits you to use local time, while at the same time insuring that the actual time can be known independently of metadata about the location of the media resource acquisition. See http://www.w3.org/TR/NOTE-datetime for details.

Storing the metadata in the images. TIFF and JPEG support embedded EXIF metadata and this is used by most digital camera manufacturers to record camera technical data and exposure data in the image file. This has the advantage that the metadata can travel with the media data but brings the risk that embedded metadata will be inconsistent with an external image metadata base in cases where they hold the same metadata such as that imposed by the camera manufacturer including date and time of the image acquisition. To avoid such inconsistency one should extract such metadata from the image as it is intended to store externally. Several simple EXIF extraction programs can assist with this, including JHEAD (http://www.sentex.net/~mwandel/jhead/), METACAM (http://www.cheeseplant.org/~daniel/pages/metacam.html), and the exiflib package for C programmers (http://sourceforge.net/projects/libexif). It is worth experimenting with these and such others as may become available, because none of them extract all the information from every camera model. Some digital scanners also embed EXIF data in the images they provide.

The recent ISO standard JPEG2000 has an extraordinary story about embedded metadata. JPEG2000 is rapidly being implemented by the digital media industry, but easy to use integrated applications are a few years away.

References

EXIF.org (2004): http://www.exif.org/ [accessed 11.XII.2004].

International Imaging Industry Association (I3A-HOME) (2004): http://www.i3a.org/ [accessed 12.XII.2004].

LA POUTRÉ, J. (2004): Jhead for xml: http://www.xs4all.nl/~jlpoutre/BoT/jhead-xml/ [accessed 15.V.2005].

Library of Congress (2005): Metadata Encoding & Transmission Standard (METS): http://www.loc.gov/standards/mets/ [accessed 15.V.2005].

Library of Congress and NISO (2004): MIX: NISO Metadata for Images in XML Schema: http://www.loc.gov/standards/mix/ [accessed 15.V.2005].

NISO (2002): Z39.87-2002: http://www.niso.org/standards/standard_detail.cfm?std_id=731 [accessed 15.V.2005].

Species Analyst (Darwin Core v2) (2002): http://tsadev.speciesanalyst.net/documentation/ow.asp?DarwinCoreV2 [accessed 2002].

TDWG (CODATA/TDWG Working Group on Access to Biological Collection Data (ABCD)): http://www.bgbm.org/TDWG/CODATA/Schema/default.htm [accessed 2003].

WANDEL, M. (2005): Exif jpeg header parser and thumbnail remover, version 2.3: http://www.sentex.net/~Emwandel/jhead/ [accessed 15.V.2005].

Image File Management

MARION CRICK

ePic Project, Royal Botanic Gardens, Kew, Kew Green, Richmond, TW9 3AE, United Kingdom. E-mail: m.crick@rbgkew.org.uk

Key words: Digital imaging, image file management.

Image file management is an aspect of specimen imaging which runs right through the digitisation process.

There are a number of issues to be taken into account in image file management

- Image file standards
- Image storage
- Image organisation
- Software

Image file standards

One if the key issues in image file management is ensuring the longevity of the images produced. This is done by producing benchmarks for your image file standards and ensuring that they are adhered to. Benchmarks will change according to the way in which the image will be used.

When imaging you will create archival files, master files and delivery files. Different benchmarks will apply to the different types of file.

File formats

Image file formats can be divided broadly into lossless and lossy formats and into open and propriatary formats.

Lossless file formats are ones in which the original image data is stored intact or compressed in such a way that the data is unchanged when the image is later decompressed. As a result, the image files are more faithful to the original. They are also very large. Examples of lossless file formats are TIFF (uncompressed in its plainest form) and PNG (uses lossless compression). Lossy file formats use compression within the image file. The compression works by removing image data. As a result the image file is very small when compared to the lossless format. The most well known lossy format is JPEG. One of the more recent formats available is JPEG200 which comes in both lossy and lossless versions.

Open formats are those which are the source code for the format is available to software developers. It means that the file can be read in a variety of formats and that the format is likely to be sustainable in the long term. Examples of open formats are TIFF (this is not fully open but has been so well documented that it is widely supported), PNG, JPEG and JPEG2000. Proprietary formats are those which can only be read in a limited number of applications. Their maintainance is dependent on the company which developed them. Examples of proprietary formats are MrSID and PSD.

File compression

Images stored in lossless formats tend to be large and require large amounts of storage. For this reason, many organisations prefer to compress images to reduce the storage requirement. There are two ways in which files can be compressed.

***Internal compression*:** These are compression algorithms which are specific to the image file format and may be lossy or lossless. PNG files use lossless compression within the format. Both lossy (jpeg) and lossless (LZW) compression is available for colour TIFF files.

***External compression*:** Images can also be compressed using an external data compression tool to created a “zipped” file. When using external compression, it is advisable to use one which is open source (ie the source code for the algorithm is available to you so that it can be maintained). There are two open source compression tools which perform well – gzip and bzip2. Of the two, bzip2

produces a slightly higher rate of compression and halves the storage requirement for images.

If any compression is used, it is essential to note within the image metadata which compression type has been used and to maintain information about how to decompress the image within your organisation.

***Archival files*:** These are files which are as close to the original captured image as possible. Where possible they should be stored in the original capture format. These files are very rarely accessed. If possible you should store the archival file in addition to the master file.

***Master files*:** These are files from which all delivery images are created. For this reason the files should be high quality, high resolution files. Master files should be stored in a lossless format such as TIFF, PNG or the lossless form of JPEG2000. If compression is used then the form of compression should be noted.

***Delivery files*:** These are files which are created for fast delivery to users. For this reason size is the most important issue. These files should be stored in a compressed or tiled (eg MrSID) format. Using open source formats will allow for easier access for users. These files can be processed from the master file at any time – they do not need to be created at the capture stage.

Image organisation

When you build up a store of digital images it will become necessary to organise and document the images in such a way that they are easily accessible. There are three elements to this.

Naming the image – make sure the image name will be unique - do not duplicate an image name. Always use the appropriate 3 letter file extension - TIF for a TIFF file, JPG for a JPEG file.

Organisation of images – If you have more than a very few images it is advisable to create a series of folders for the images. Placing too many images in a single folder can affect the performance of your

computer. No metadata should be recorded in the folder names, unless it is also recorded in the image metadata.

Metadata – All images should be accompanied by metadata. All image file metadata should be contained in an image management system, specimen metadata can be contained in the same database or else in a specimen database linked to the image database.

Image storage

There are a number of image storage methods which can be considered. It is essential that, when you select a method for image storage, you consult closely with the IT department within your organisation. They can advise best on current costs and on solutions that fit within your organisation's existing infrastructure.

CD/DVD:

Specimen imaging produces large image files. This makes storage on CD impractical as CDs hold up to 700mb of data. DVD-R stores up to 4.7GB of data, which makes it more viable for a digitisation project. If storing images on CD or DVD is it essential to follow the following guidelines.

1. Use archival quality discs
2. Do not burn the disc at top speed
3. If you intend to use the images on a regular basis, create a second "working" copy of each CD.
4. Check each disc after it has been written.
5. Store the discs in suitable conditions – a cool, dark, dry place.
6. Re-check the discs on a regular basis and replace the discs periodically (approximately every 5 years).
7. Create a filing system for the discs and ensure that the location of each image is recorded in its metadata.

Local hard drive:

Images can be stored on a PC hard drive. This method of storage means that images can be quickly accessed. However, it does mean

that access to the images is limited and the storage capacity is limited by the capacity of the hard drive. There is an additional problem in that the images are vulnerable as they could be lost if the PC breaks. Images stored on local hard drives must be backed up to another format on a regular basis.

Networked storage:

There are a number of ways of approaching networked storage.

1. Storage on a file server.
2. Network addressed storage.
3. Content addressed storage.
4. Storage area network.

The most appropriate of these types of storage will depend on your organisation's infrastructure. Selection of networked storage in particular should be discussed with your own IT department.

The advantages of these types of storage are that you can store far larger numbers of images as there is capacity for extending the storage, the images can be shared easily within the organisation, the images are held in a central location and the images can be backed up easily.

Image management and serving software

Although it is possible to manage a small number of images without additional software, if you are building a large image collection, it is advisable to use some sort of image management software to do so.

Before approaching any image management software provider, it is essential that you determine your image management requirements and consult with your organisational IT department about your technical infrastructure.

TASI[1] identifies four different categories of image management application.

Folder views – these do not really manage the images in any way but provide an access route to the images either through browsing or a very simple search. Some allow very basic metadata to be captured – but this metadata is contained within the image file rather than held in a separate database. These systems are very cheap, often available as shareware.

Simple image management systems – these systems are often used by image users with small collections rather than by image librarians. They create basic flat file databases for the image metadata. They are generally suitable for small to medium collections of images but sometimes have server editions which allow larger databases to be built. Because the underlying database is generally in a proprietary format they are often not interoperable with other databases. Development of the database is often limited but some do offer scripting solutions. Some applications allow basic image processing – e.g. creation of thumbnail and screen resolution images, cropping and rotation of images. These systems can be created fairly cheaply. Always find out whether free systems are proprietary or open-source. In the former case, you might find that what is free now might not be free a few years later. In general, users of proprietary software (whether free or not) need to be concerned about being locked in by the use of proprietary exchange standards. Thus, if a management system has a good export story for metadata (e.g. plain text, XML, spread sheets, etc.) then users have some insulation against loss of their metadata should they have to migrate to another system.

Full image management systems – such image management systems have are similar to simple image management systems in overall appearance but very different in underlying structure. Some applications allow basic image processing – e.g. creation

[1] http://www.tasi.ac.uk/advice/delivering/comparison-ims.html

of thumbnail and preview images, cropping and rotation of images. The databases offer greater extensibility and interoperability and the development allows for a more complex database structure. The disadvantage is that these systems are somewhat more expensive than a simple content management system.

Bespoke image management systems – many organisations have created their own image management system in the absence of commercial software applications which fulfilled all their needs. If outsourced this is an extremely expensive approach to take, but does allow for a precise specification to be met. As Kew has an applications department the expense is not such a great factor but the specification would need to be carefully planned.

Within these categories, the boundaries are becoming blurred – as simple image management systems become more advanced the increased functionality and ability to user server based additions means that the difference between the simple and advanced image management systems has become much smaller.

Collection/Digital asset management systems

In addition to the systems described by TASI, a number of applications are broadening their scope to handle a wider range of digital files. There seems to be a trend towards complete asset management rather than purely handling image management.

Image serving software

At the simplest level, an image serving application is one which makes images available for other applications, specifically one which allows images to be viewed through a web browser. This function is supplied by some image management applications through web interfaces.

There are also specific image serving solutions which can be used in conjunction with full/bespoke image management systems to improve delivery of images.

***Image authoring*:** Some image applications offer image authoring tools which allow images to be saved in proprietary formats so that they can be fully utilised in the image serving application.

***File formats*:** Many image serving applications will support a variety of image file formats and make them available to web browsers without having to store an additional set of JPEG or PNG surrogate files.

***Scalable image delivery*:** Some applications will use “tiled” image formats such as SID or Flashpix to deliver scalable images through web browsers. It is only necessary to store a single copy of the image and the software automatically generates images on the fly from the original. This means that the software reads only as much of the image as is required for the specific viewing request and delivers only the portion of the image required. This may be the entire image at different resolutions or parts of the images. This feature means that network traffic is greatly reduced and that higher quality image files can be delivered across limited Internet bandwidth.

***Zoomable images*:** Applications often have features which allow the user to zoom into an image to see details at higher resolution.

***Navigation within images*:** Applications often allow the user to navigate within the image – a small portion of the image is shown inside a frame and the user can use tools to view adjacent portions of the image.

***Choice of ways to access the images*:** The images served through the applications may be viewed either through a free browser plug-in, which has to be downloaded by the user, Java applets or even through plain html pages.

***Image security*:** Some software applications offer the facility to give the appearance of a watermark on the image file when the images are served. Some actually watermark the image.

Examples of image management applications

Image authoring and editing software

Product: Zoomifyer
Manufacturer: Zoomify
Website: http://www.zoomify.com/zoomifyer/competitive.asp
Features: Zoomifyer is an authoring tool rather than a server application. Images can be viewed by end users using plug-in (Netscape and Mac users), ActiveX controls (IE users), Java applet, clientless Javascript option all of which are generated by the authoring tool. Generates images in proprietary PFF format. The next release of Enterprise should directly support JPEG2000. Compatible with any web server. Visible watermark can be placed in the viewing window. Zoomifyer Enterprise can work as a desktop application or as an automated application on a network server (Mac or Windows, but not Unix).

Product: ImageZoom
Manufacturer: Scalado
Website: http://www.scalado.com/index2.html
Features: Allows you to create “zoomable” images – where sections of the image can be enlarged. The images cannot be panned across. Creates multiple image files – a large image and several higher resolution zoomed sections of image - rather than a single image file. This works as a desktop application which runs on either Windows or Mac platforms.

Product: GeoExpress
Manufacturer: Lizardtech
Website: http://www.lizardtech.com/
Features: Writes images to the SID format, which is a multi-resolutional, tiled format, allowing portions of very high resolution images to be viewed selectively. Compresses the image file. Supports image file sizes of over a gigabyte.

Product: ImageMagick
Manufacturer: ImageMagick
Website: www.imagemagick.com
Features: Open source desktop or server based imaging application. The software can be run as a script or used through a variety of interfaces, including Java. The application supports the majority of file formats, including JPEG2000 and also supports some vector graphic formats. It has a number of different utilities which perform different functions. Convert and mogrify do one time image conversion, display offers on the fly conversion of image files, montage automatically builds thumbnail galleries.

Image serving software

Product: eyespy
Manufacturer: AXS technologies
Website: http://www.axs-tech.com/html/products/eyespy/index.html
Features: Creates tiled images files for web delivery – images can be zoomed into and panned. Supports 2 major file formats and plans to support JPEG2000. Allows different interfaces for internal and external users.

Product: MediaRich.
Manufacturer: Equilibrium
Website:
http://www.equilibrium.com/Internet/Equil/Products/MediaRich/Product+Tour/index.htm
Features: Image creation and processing tools, can be easily integrated with open standard databases, supports wide range of file formats, delivers an appropriately sized image for the viewing device, allows image files to be repurposed for many uses from a single source image, creates images which can be zoomed into and panned across. MediaScript creates the javascript code which allows the images to be viewed through a web interface. Requires Microsoft or Solaris operating systems and IIS or Apache web server.

Product: Zoom Image Server
Manufacturer: iSeeMedia
Website: http://www.iseemedia.com/products/zoom/index.html
Features: Scalable image delivery from a single image, zooming and navigation within images. Images can be viewed with a plug-in

or a Java applet, supports Flashpix and IVUE file formats. Can be used on Windows, Linux and Solaris web servers. Enterprise edition allows images to be watermarked.

Product: Web Image Server
Manufacturer: Vimas Technologies
Website: http://www.vimas.com/ve_image_server.htm
Features: Supports TIFF, JPEG, PNG, TGA file formats. Application will only run on a Windows NT or Windows 2000 server. Allows images to be edited through a web interface and allows the user to resize the displayed image.

Product: Express Server
Manufacturer: Lizardtech
Website: http://www.lizardtech.com/
Features: ExpressServer allows SID images, as well as high resolution JPEGs, TIFFs and other image formats to be delivered as scaled JPEG files through a web browser, either with our without a web plug in.

Product: Image Web Server
Manufacturer: Earth Resource Mapping Pty Ltd
Website: http://www.earthetc.com/
Features: Images available to web browsers and to other applications, mainly GIS applications. Saves images in compressed ECW format. Images can be viewed through a web browser using an automatically downloaded plug-in. The plug-in allows images the user to define the extent of the image file that they wish to download in each instance (resolution, portion of image). The application is only suitable for Windows servers and works as an extension to Microsoft IIS.

Image management applications

Product: iBase Manager
Manufacturer: iBase
Website: http://www.ibase.com/site/software/index.shtml
Features: Allows images to be catalogued and batch processed to create smaller image files – the format of the batch-processed files is not specified. Additional modules are available to create a touch-screen browser and a web interface. The application has cross

platform client capability but will only run on a Microsoft Windows Server.

Product: Fotostation/ Index Manager
Manufacturer: Fotoware
Website: www.fotoware.com
Features: Software developed mainly for use in publishing, specifically newspaper, industries. Two applications are used in conjunction – Fotostation allows image creation and manipulation workflow to be streamlined and provides the thumbnail interface, Index Manager created proprietary format databases in "open industry standard" and has a search engine. Index uses IPTC standard fields and allows 20 custom fields. Windows NT based.

Product: IDA II (Professional edition)
Manufacturer: Aequitas
Website: www.aequitas.co.uk
Features: Image management application that handles imaging workflow and supports relational databases held in SQL server and Oracle. The database is customisable and the application supports JPEG, BMP, TIFF and PCD images. Both the server and client software only run on a Windows platform.

Product: Picassa
Manufacturer: Google
Website: http://picasa.google.com/
Features: Propietary free system for Windows. Designed for home use, but suitable for many sizes of collections. Application has a very flexible album structure, captions and other metadata said to be stored in IPTC standard (see MORRIS, in this book, pp. ##-## on Metadata Standards). There is an active discussion group about Picassa on Google Groups (http://groups.google.com/group/picasa/) and some complaint that the IPTC data is not fully portable to other IPTC aware systems, though the consensus is that exchange with many important image annotation systems, including Adobe Photoshop and Menalto Gallery (see below), is generally satisfactory. A documented CSS template-based export

Collections/Digital asset management systems

Product: The Museum System (also TMS light and embARK)
Manufacturer: Gallery Systems

Website: http://www.gallerysystems.com/
Features: This is a museum administration database rather than an image management system. Image files are handled in the "media" module but can be catalogued in the collections module. This is a complex system which has a number of sophisticated data entry options and allows the user to link to an integrated thesaurus. The database can be held in SQL server or Oracle. A web interface is provided through the e-museum application which needs to be purchased separately. The applications are heavily reliant on Microsoft and optimum performance is only gained with a Microsoft database server.

Product: Index+
Manufacturer: System Simulation
Website: www.ssl.co.uk
Features: This is a full digital asset management system which can handle resources other than images. This is a bespoke system which is tailored to each organisation's needs. The metadata is held in an object-oriented database which is fully interoperable. System Simulation also builds interfaces which allow Index+ to interact with existing organisational databases.

Product: Portfolio
Manufacturer: Extensis
Website: www.extensis.com/portfolio
Features: A simple image management system. It allows you to build a customised set of database fields, easy import and export of data as a delimited text file and can be used to generate preview files for images. It also automatically captures technical image metadata. The data is held in a proprietary file format but can be exported to a delimited text file. Images and metadata can also be exported to web based galleries. There are opportunities to develop the software interface further through scripting in Visual Basic or Applescript. The software also allows for records to be created for hard copy images as well as digital images. This application is being as an interim image management system until the image server has been created. Cross Platform capability. Runs as standalone or on Windows or Mac servers.

Product: Canto Cumulus
Manufacturer: Canto

Website: www.canto.com
Features: This application is described as a digital asset management system rather than an image management system, as it handles a large range of digital files. The underlying database is held in a proprietary format which is ODBC compatible and the field structure can be customised. Previously the data could only be exported in a proprietary data format but the latest release allows data to be exported to XML. There is a 4GB limit on the database size. There are a number of additional applications which Canto provide to enhance the software. Of chief interest are the database connector application, which allows you to connect the application interface to a variety of database types, and the web publishing application which serves the images and database through a web browser. Neither are integral to Cumulus but are designed to work alongside it. Canto provide a software development kit to customise the client application using visual basic or applescript. Both PC and Mac clients. The workgroup and enterprise editions are compatible with a Unix server.

Hybrid applications

In addition to the applications which offer elements of the image server, there are some applications which have all of these features – offering some image management, manipulation and serving functions.

Product: KE Emu
Manufacturer: KE Software
Website: http://www.kesoftware.com/emu/
Features: Collections management system with an images module, which allows images to be connected to other databases within the system. It is server based and runs on Windows PCs, connected to either Windows or Unix servers. Data follows established museum standards. Uses KE Texpress database, which is an open systems database. Web serving software is available.

Product: Toadview and ToadHMS server editions
Manufacturer: Oxford ArchDigital
Website: http://www.oxarchdigital.com/toadview/
Features: This is a server based image database based on open source applications (MySQL, imagemagick). The application has a web browser interface so there is no need to install client software

and both PC and Mac use is supported. The application supports several image file formats and creates thumbnail and JPEG preview images when the image files are loaded into the system. Allows basic image manipulation. Also acts as an image server – resizes and watermarks original images for web delivery. Metadata can be stored in a variety of database formats including MySQL. Adaptable metadata framework. Has advanced image security and access features.

Product: Insight
Manufacturer: Luna Imaging
Website: http://www.lunaimaging.com/insight/index.html
Features: Java based application. The application is used through a web browser. This application combines an image management application with an image serving application. The Insight image server incorporates LizardTech's MrSID Photo code. The server side of the application allows scalable image delivery from a single file, zooming and navigation within the image files. No plug-in is required to view images. The management application can be based around a variety of databases including Oracle, MS SQL and Sybase. Has a number of thesauri loaded. Has workflow management features.

Product: Menalto Gallery
Manufacturer: Menalto.com (open source)
Website: http://gallery.menalto.com/
Features: Mature open source, web-based image server software that manages albums, automates thumbnailing, and provides for limited collaborative image annotation and web search against that annotation. A companion program, GalleryRemote provides for uploading from Windows and Macintosh machines directly to the server. The system supports mirroring, whereby copies of images can reside on multiple servers with the web client able to select which mirror to use for best service. It can be embedded in a number of content management systems, including the widely used open source Mambo system (http://mamboforge.net/) Gallery2 as completed two release candidates and is nearing its final release at this writing. While version 1 uses its own database, Gallery2 is SQL based and includes support for MySQL, Oracle, and Postgres databases.

Online Acquisition of Scientific Archive Documents – A Survey and Manual

ANDY C. DOWNTON

Department of Electronic Systems Engineering, University of Essex, Colchester CO4 3SQ, UK. E-mail: acd@essex.ac.uk

Key words: Digital imaging, archive documents.

Background and Introduction

Digital archive construction from historic paper archives is a major document image analysis application of research interest both for cultural and scientific applications. It is also a challenging computer science research problem. For example, 17 papers out of 53 presented at the most recent Document Analysis Symposium (held in Florence, September 2004) were concerned with analysis of historical documents (MARINAI & DENGEL 2004), and a special issue of the *International Journal of Document Analysis and Recognition* is currently in preparation dedicated to the topic of Analysis of Historical Documents (ANTONOCOPOULOS & DOWNTON in press).

Scientific archives differ from more conventional textual data such as books, letters and reports because they are often stored in well-structured taxonomies (e.g. libraries, scientific specimen indexes, tables and censuses) where the structure extends across the index as well as within the layout of each document. Construction of the digital archive must therefore preserve the inter-document as well as the intra-document structure. Ancient scientific archives are however often recorded using text which challenges off-the-shelf OCR (SPITZ 2004), not only because of poor document quality and/ or decayed typescript or handwriting, but also because standard commercial OCR systems cannot infer the data structure inherent within the records without human guidance. Conversely, commercial form-processing OCR systems are seldom suitable for pro-

cessing archives, because they are designed to process 'cooperative' fixed-format pre-defined forms designed for OCR using background drop-out colours and/or tabular guidelines to maximise performance, rather than arbitrary pre-existing document archives.

To address archive applications, a user-configurable archive document processing system is required. This must integrate document image analysis and text post-processing tools with a configurable OCR package (which may be available as commercial off-the-shelf software, or as a programmer's development kit (PDK)), to generate text content that can be fed direct into a target online database or digital library. User configurability is essential to allow the same system to be re-targeted to process different archives, and also variable layouts within the same archive. Unfortunately, although bespoke systems of this general type have been developed for high volume commercial applications such as postal address recognition (IEEE Trans) and cheque processing (ICDAR), such applications benefit from large continuous (rather than smaller batch-mode) data processing requirements, fixed and restricted context, and business data source quality, none of which normally apply to historic document archives.

This paper surveys the available technologies, systems and solutions for converting paper to online archives, and identifies both promising developments and continuing problems. We illustrate the current technical state-of-the-art by describing in detail a particular solution, for scanning and recognising biological archive card indexes, that has been developed as a collaborative multi-disciplinary project between the UK Natural History Museum (NHM), and the Department of Electronic Systems Engineering, University of Essex[1]. The evaluation data now available for this system gives a realistic indication of currently achievable performance for scientific archive acquisition, and also indicates how this performance

[1] VIADOCS (Versatile Interactive Archive Document Conversion System), Research contract 84/BIO11933, awarded under the joint EPSRC/BBSRC Bioinformatics research programme, 2000-2003.

depends critically on exploiting available layout and contextual cues to maximise OCR performance.

Scanning Document Sources

A fundamental issue in acquiring an online version of a paper-based archive is to identify an efficient and effective archive scanning mechanism. Archives normally consist of a large number of records, whether recorded on paper, cards, microfilm, in ledgers, or books. A particularly difficult class of textual archive data is that associated with specimen labels, because of the complex manual handling required to scan each label. The major concern is not simply to identify an *effective* scanning technology for a particular type of document source (in terms of resulting image quality), but to develop an *efficient* multi-document scanning system that is capable of acquiring thousands or even millions of images within a limited time-frame.

Imaging Technologies

Most document imaging is nowadays carried out using CCD sensors. One-dimensional (linear) CCDs are used in mechanically scanned systems such as flatbed scanners and two-dimensional CCDs are found in cameras.

Flatbed scanners

High-resolution (1200-2400 dpi) colour flatbed scanners are now widely available as PC peripherals at negligible cost, and even the cheapest scanners are capable of excellent quality document imaging. Resolution in one axis is defined by the size of the linear CCD, while in the orthogonal axis it depends on the stepping resolution of the transport motor that drives the CCD across the document. Document scanning using standard PC scanners is slow (typically a page every few seconds maximum), limited by the mechanical scanning rate across the page.

PC flatbed scanners are designed for occasional use under full manual control, and are therefore unsatisfactory for scanning even

moderate size archives, because they require virtually continuous manual supervision, and have very limited throughput rates. High volume professional flatbed scanners with built-in document feeders are available for professional digital archiving and document processing applications, and these may also be more suitable for scientific archive acquisition, albeit at a cost 10-100 times more than basic PC scanners. However, professional flatbed scanners are designed to process standard consistent modern document formats (such as reams of A4 paper), and are often unable to handle archive documents efficiently. By comparison with modern paper, archive documents use a much wider variety of paper and card sizes, and exhibit inconsistency and varying quality and paper texture, as well as fragility, which make efficient paper handling a difficult problem to overcome.

Mail and cheque scanners

Instead of transporting a linear CCD across the document to produce a scanned image, the document can be transported past a

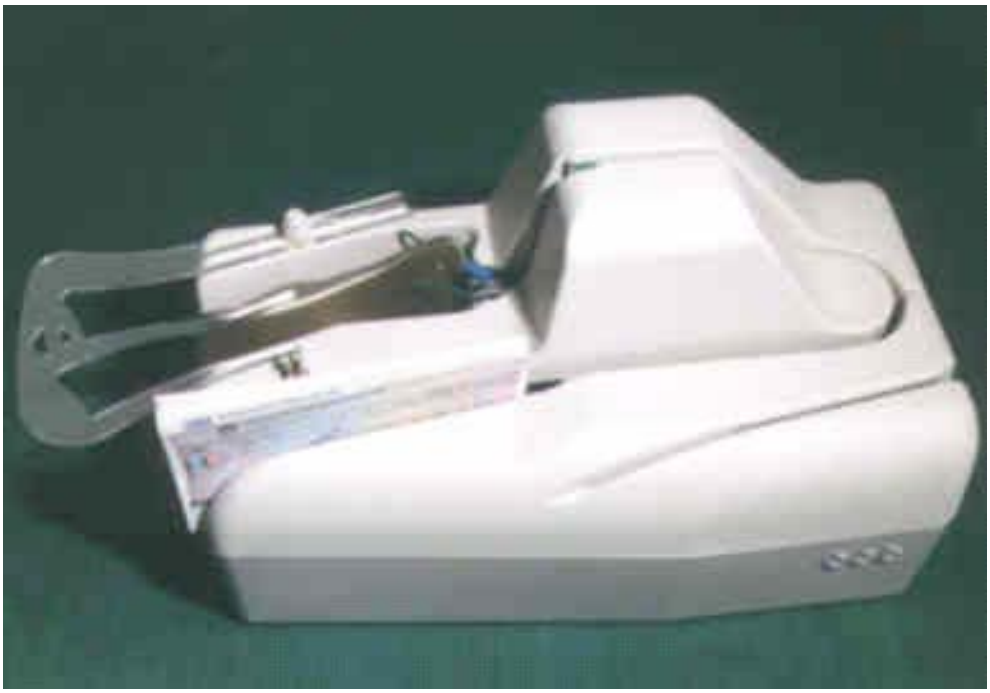


Fig. 1. Cheque scanner (courtesy SEAC-BANCHE). Up to 50 cheques (or index cards) are loaded into the font hopper, and are then fed one-by-one past twin vertically mounted CCDs (so both sides of the document are scanned simultaneously). A user-specified text string can also be printed on the document during scanning.

fixed linear CCD. This is the principle used in mail and cheque scanners, where stacks of mail or cheques are fed one-by-one past the CCD, held by moving belts either side of the document. Large scale systems used for postal sorting typically scan and process 10 letters per second from hoppers filled with thousands of mailpieces, while small cheque scanners (fig. 1) of the type found at bank tills can scan about 1 cheque per second from a hopper holding about 50 cheques. The large high-speed systems are built to order, and thus are relatively expensive, but small cheque scanners are available at more modest cost (around £2000), and are quite well-suited to scanning archive card indexes of the type typically found in many museums and libraries (see Section 5).

Hand-held scanners

Another possible scanner configuration is the hand-held scanner. Typical commercial implementations are modelled on pens, with varying width scanning heads. They are generally designed for selective scanning of document components, and often intended to scan one or a few lines of text at a time, though some models have the capability to stitch together a group of overlapping raster scans into a single document page. Some models are designed for operation in conjunction with a host computer and record raster images internally until downloaded to the computer via an associated pen-holder; others are intended as independent peripherals, and include some OCR capability within the pen. Due to their limited input data rate and processing capabilities, such devices are of limited value in archive document acquisition, though they could be used for acquiring small images from irregular document fragments such as specimen labels.

Book scanners

The technology described so far is optimised for scanning single document pages rather than bound documents such as books or ledgers. Perhaps the most suitable device for scanning archive books is the Xerox PARC bookscanner (fig. 2). This scanner has been designed specifically to handle rare and fragile books without damaging them, by using a two-paned glass wedge platen to

position each pair of opposing pages without opening the book to more than 90 degrees. This minimises geometric distortion and warping of the page images (which otherwise occurs due to inability to fully flatten the book spine). Between scanning each pair of pages, the platen is raised and lowered and the page turned manually: not surprisingly the scanning rate is thus slow (2 pages/minute colour; 8 pages/minute grey-scale).

Fig. 2. The Xerox PARC rare book scanner (from http://www2.parc.com/eml/members/ready/parc_bookscanner.htm)

Camera technology

Digital camera technology, based on two-dimensional CCD arrays, has developed rapidly in the last decade, with the result that cameras now offer a viable alternative to flatbed scanners in many document image acquisition applications. A 6 Mpixel camera can in principle image an A4 page with a resolution of nearly 300 dpi, and even higher resolution CCDs are now readily available at modest cost. Professional quality cameras are available with 20-40 MPixel CCDs. Cameras are flexible, portable and fast compared with scanners, and for some abnormal document sources (Papyrus, Palm leaves, wooden and stone tablets etc) may be the only viable imaging source. Significant research has also been carried out on techniques for mosaicing large document images (e.g. up to the size of a complete desktop) from a series of adjacent camera image frames[2].

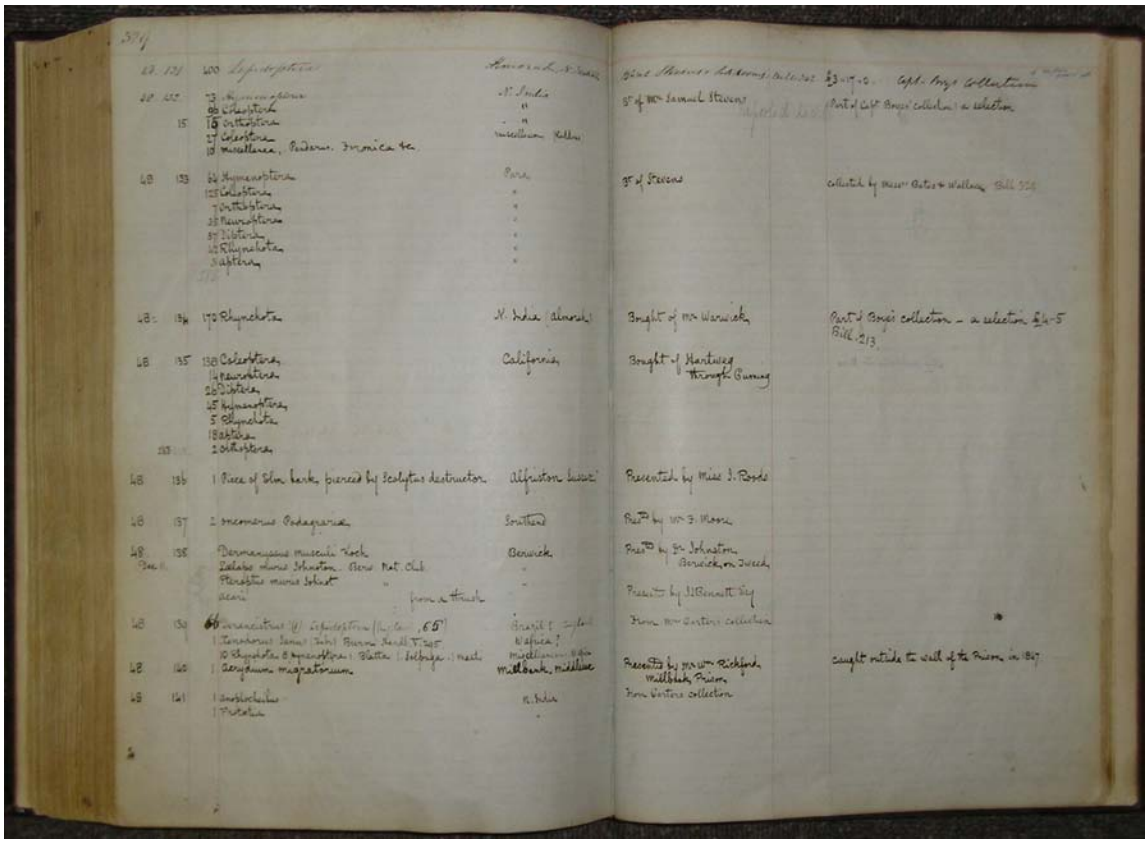


Fig. 3. Ledger page images from the Natural History Museum, showing effects of affine and page-curl warping, and illumination intensity variation due to point-source lighting.

[2] Desktop document mosaicing.

The disadvantage of using cameras to acquire document images is that it is more difficult to accurately control geometry, lighting, focus and image noise for camera-based pictures, than for flatbed or linear scanners. Geometric distortions arise from non-orthogonal alignment of the camera with the scanned page and include perspective, affine and non-linear warping of the document image[3], sometimes requiring sophisticated 3D shape models to unwrap curled pages (fig. 3). Lighting for camera-based document image acquisition is also more difficult to control, and typically suffers from highlights and significant intensity variation due to linear or point-source illumination of the document. In view of the increasing prevalence of digital cameras, considerable research is now in progress on techniques for correcting geometry, lighting and noise defects in camera-based document images[4] .

Document Sources

Scientific archives originate in many different types of documents, but can be classified into two major classes, *primary sources* and *index records*. Primary sources are archives where the document itself is the scientific source, for example, log books and descriptions of their research written by famous scientists[5], which have value both as scientific sources and historical records. Such sources are idiosyncratically organised by the writer, and will not in general have any semantic structure beyond the source itself which can be used to help infer document content.

In contrast, index records are essentially lists of pointers to other scientific artefacts (which may themselves be documents), and hence will inherently have an implied index superstructure. They may be based upon the use of a common physical recording medium (e.g. index cards), but this doesn't guarantee a consistent index structure or layout, since large scientific indexes have usually been accumulated over many years by significant groups of individuals, each of whom may have used their own idiosyncratic recording style.

[3] Unwrapping page curl reference.
[4] Compensating illumination variation
[5] Famous researchers log books online

Nevertheless, a consistent physical storage medium is the first pre-requisite for efficient online acquisition, since it introduces the possibility of a 'production-line' mechanical scanning process.

Not all scientific archives can be rigidly classified as sources or records: some historic documents (e.g. Jepson's *Flora of California*[6] published in several volumes spanning decades, fig. 4) fulfil both roles, and Herbaria sheets typically include both specimens and documentation on the same sheet.

2. **L. punctata** Goodding. LILAC SUNBONNET. (Fig. 392.) Low flat-topped plants 1 to 2 inches high, 2 to 7 inches across, seeming as if prostrate; herbage minutely tomentulose or rarely glabrate; leaves with deltoid 3-toothed or 3-lobed apex, sometimes with a pair of teeth or lobes below the 3 terminal teeth, all the teeth bristle-tipped and the petiolar or cuneate base with simple or 2 or 3-forked bristles; flowers subsessile; calyx-lobes ⅘ to as long as corolla-tube; corolla lilac, 7 to 10 lines long, subregular, its lobes about 2 lines wide, purple-dotted, each with 2 very shallow longitudinal channels from above the middle towards the base and ending below in a lunate yellow ridge; capsule narrowly oblong, acute, 3-sided, the cells 3 to 9-seeded.

Gravelly hills and mesas, 900 to 4500 feet: Inyo Co.; central and eastern Mohave Desert. East to Nevada and Arizona. Apr.-June.

Locs.—Inyo Co.: Silver Cañon, White Mts., *Heller* 8308; Bishop, *Almeda Nordyke;* Argus Range (n. end), *C. N. Smith* 139; Hanaupah Cañon, Panamint Range, *Jepson* 7010. Mohave Desert: betw. Kelso and Baker, *Jepson* 20,582; Amboy, *Newlon;* Ludlow, *Jepson* 5494; Lavic (5 miles s.), *J. Grinnell;* Newberry (lava fields sw.), *Newlon* 498; Ord Mt., *Jepson* 15,496; Barstow, *Jepson* 5361.

The anthers, borne on very short filaments inserted unequally in the throat, approximate about the style at mouth of throat, leaving pin-hole entrances to the tube beween the filaments.

Refs.—LANGLOISIA PUNCTATA Goodding, Bot. Gaz. 37:58 (Jan., 1904); Hel., Muhl. 1:57 (Feb., 1904); Jepson, Man. 808, fig. 780 (1925). *Gilia setosissima* var. *punctata* Cov., Proc. Biol. Soc. Wash. 7:72 (1892), type loc. Surprise Cañon, Panamint Range, *Coville* 716. *Navarretia setosissima* var. *punctata* Cov., Contrib. U. S. Nat. Herb. 4:154 (1893). *Gilia punctata* Munz, Man. 402 (1935). *L. lanata* Brand; Engler, Pflzr. 4[250]:169 (1907), type loc. Candelaria, Nev., *Jones* 3965.

Fig. 392. LANGLOISIA PUNCTATA Goodding. *a*, habit, × 1; *b*, leaf, × 1; *c*, long sect. corolla, × 1½; *d*, calyx and pistil, × 1½.

Fig. 4. Example of Jepson's *Flora of California* (from http://ucjeps.berkeley.edu/jepson-project3.html)

[6] http://elib.cs.berkeley.edu/docs/rare.html and http://ucjeps.berkeley.edu/jepson-project3.html

Document archives are also stored in a variety of different physical formats, not all of which are easily imaged (see below). Therefore the primary issues to be resolved when planning to convert a scientific archive into an online database are:

1. What imaging techniques are practicable and efficient for the archive?
2. What imaging quality standards need to be met to ensure sufficient quality both for currently planned and possible future uses of the online archive?
3. What document image analysis problems does the archive present, and to what extent do these exceed the capabilities of current off-the-shelf OCR systems?

Card Archives

Card archives were widely used throughout the 20th Century for organising scientific and literary sources, a role now fulfilled by computer databases. The development of storage boxes and trays to allow card indexes to be easily searched on their primary key, together with the fact that card rather than paper was used for each record, means that most indexes are in good physical condition (albeit sometimes faded) and can be scanned without damaging the original source.

However, different card indexes may be subject to very different recording conventions. Over the last century, information has usually been type-written, but a minority of cards (and sometimes complete indexes) are entirely hand-written. Hand-written annotations on typed cards are also common where the card taxonomy has been updated. Some archives have index information on both sides of the card, or index information on one side, and source data (such as a specimen drawing) on the other.

Card archives can be scanned using a bank cheque scanner, modified by the addition of specially developed software which allows configuration of the scanning process, to build a large image archive from a series of batch scans which may take place over days or weeks. Individual cards are scanned at a rate of about 1 card/second, but the overhead of transferring small blocks of cards backwards and

forwards between the scanner and their archive trays, and subsequent checking the quality of the images produced, means that a typical scanning rate is no more than 10.000 cards/day (based on experience of scanning several archives totalling over half a million cards at the UK Natural History Museum[7]). This is a very high image acquisition rate compared with other sources however, and the high payoff per unit cost in terms of safeguarding copies of unique source indexes has encouraged the UK Natural History Museum to scan significant parts of their archives, even when funding does not permit further indexing of the images immediately.

Books and Ledgers

The most efficient way to scan books and ledgers is to cut off the spine and convert the book into a ream of individual paper sheets, which can then be scanned using a document-fed single or double-sided flatbed scanner. This destructive method is reasonable where the book has no inherent historic value (especially if multiple copies are available), but is hardly acceptable for historic artefacts.

Where non-destructive scanning is essential, the process is likely to be slow, expensive and labour-intensive, requiring equipment such as the PARC book scanner described above.

Manuscripts

Ancient and rare manuscripts present a variety of document handling and scanning problems. They may be constructed from non-standard materials, physically fragile, of abnormal size and with poor text/background contrast due to decay of the paper and ink fading. As a result it may be physically impossible to scan them using flatbed scanner technology, in which case a digital camera may be the only feasible image acquisition mechanism.

Microfilm

Many archives were converted to microfilm during the last century,

[7] This rate assumes that no other manual indexing is being carried out while scanning, otherwise the scanning rate will be very much slower.

but often at a significant cost in quality, as judged by the standards considered desirable today. Recognising textual content of microfilmed documents is a research area that has received some attention in the document image analysis community[8], but is a challenging restoration process, which mainly serves to highlight the importance of ensuring that the original scanning processes generate images of high enough quality for all foreseeable future as well as current applications (see figs. 5 and 6). Self-evidently, this is not achieved by imaging processes which compromise the colour rendition and/or spatial resolution of the original document.

Specimen labels

Scanning of specimen labels is a unique requirement of scientific archives which has so far received little attention within the document image analysis (DIA) community. In some simple cases, text region identification techniques pioneered in other natural scenes (e.g. recognising signs in street scenes)[9] may be of some relevance, but in other cases, where specimen labels are partially obscured by the specimen itself, or by other labels (fig. 7), the time required to make the label visible for imaging might be better spent directly labelling the specimen's image within the online database.

Scanning costs

Document scanning for online digital libraries is now a major web activity[10], and as a result there is increasing accumulated experience of the costs that scanning incurs. A recent symposium on the US National Initiative for a Networked Cultural Heritage (LESK 1997)

---

[8] Historic editions of the New York Daily Times are online at: http://www.nyt.ulib.org/

[9] Text recognition in natural scenes

[10] JSTOR digital library (University of Michigan and Princeton) includes an archive of imaged copies of 29 Botanical and Ecological Journals dating back to 1867, at: http://www.jstor.org

Making of America is a digital library of imaged primary sources in American social history, containing about 8,500 books and 50,000 journal articles, at: http://www.hti.umich.edu/m/moagrp/

The Price of Digitization: New Cost Models for Cultural and Educational Institutions, NINCH Symposium: April 8, 2003, New York City, http://www.ninch.org/forum/price.report.html

The New-York Times.

VOL. XIV.....NO. 4235. NEW-YORK, MONDAY, APRIL 10, 1865. PRICE FOUR CENTS

HANG OUT YOUR BANNERS

UNION VICTORY!

PEACE!

Surrender of General Lee and His Whole Army.

THE WORK OF PALM SUNDAY.

Final Triumph of the Army of the Potomac.

The Strategy and Diplomacy of Lieut.-Gen. Grant.

Terms and Conditions of the Surrender.

The Rebel Arms, Artillery, and Public Property Surrendered.

Rebel Officers Retain Their Side Arms and Private Property.

Officers and Men Paroled and Allowed to Return to Their Homes.

The Correspondence Between Grant and Lee.

OFFICIAL.

THE PRELIMINARY CORRESPONDENCE.

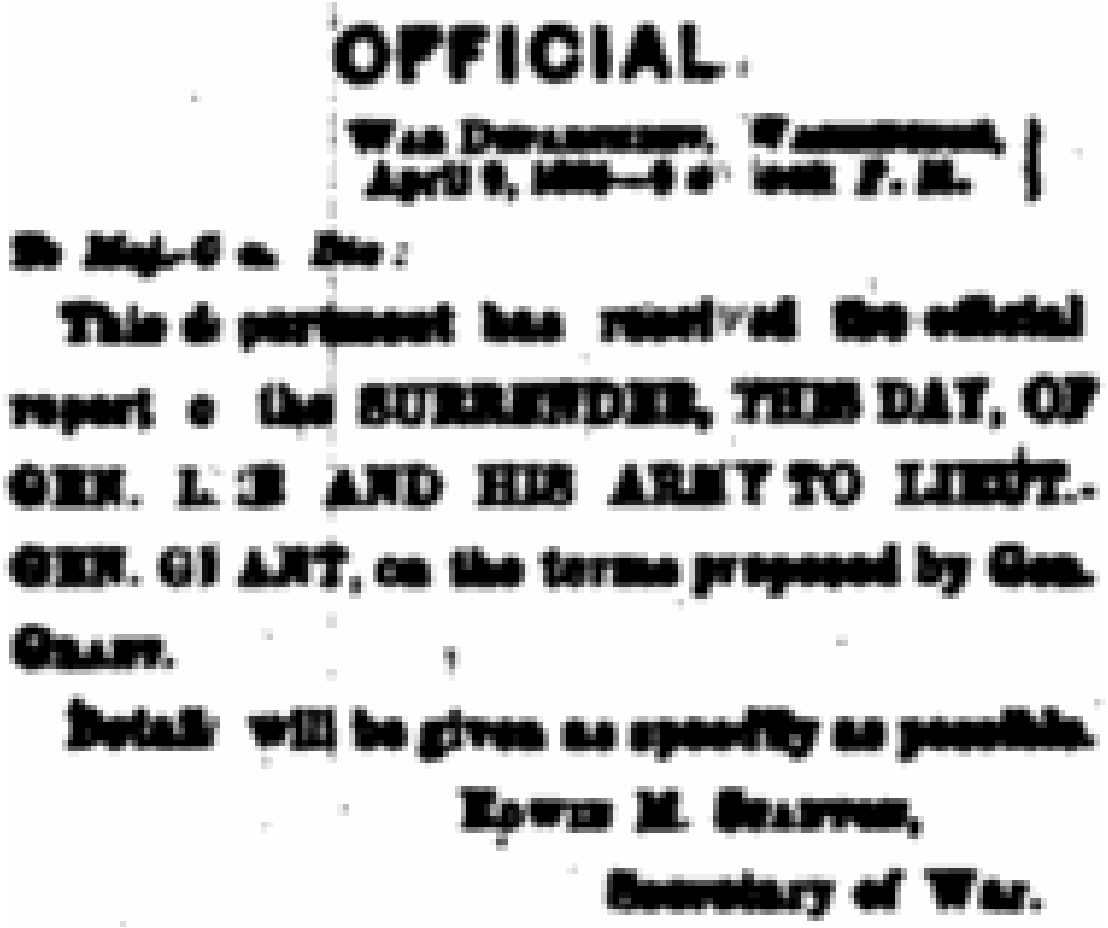
THE VICTORY.

Thanks to God, the Giver of Victory.

Honors to Gen. Grant and the Gallant Army

A NATIONAL SALUTE ORDERED.

FROM THE PACIFIC COAST.



OFFICIAL.

Fig. 6 Blow-up of a column from the same New York Times showing virtually unreadable text.

Fig. 7. Obscuration of label text by other labels and the specimen itself (note also the extremely small script: the moth wingspan is about 4cm).

estimated that the cost of digitizing a book could vary between $ 4-$ 1000, depending on the quality control that is applied to the process. Cheap scanning is based on low quality, fast, binary imaging; high quality implies individual handling of document sheets scanned in high-resolution colour. Inevitably, scanning of scientific archives will be towards the expensive end of this spectrum. Surveys also indicate that only about 1/3 of the total cost of capturing document image content is associated with the scanning process, OCR, correction and markup. The remaining costs are split equally between the costs of cataloguing, description and indexing; and quality control, file maintenance and administration. (It is often forgotten that once an archive has been scanned – often using short-term 'soft' project funding – significant costs will continue to be incurred in backing up the archive, and periodically migrating it to new physical media and new database standards. Failure to do so will rapidly render the archive unreadable.)

Scanning standards are subjective: quality control usually involves checks that all pages are physically scanned, but much higher costs are incurred if it is required to check each page for human and/or machine intelligibility. Only rarely are documents scanned to quality levels that guarantee their suitability for future uses, yet paradoxically once an archive has been scanned, there is little prospect of funding being available for re-scanning for many years (BECCALONI et al. 2005). Hence a philosophy of 'right first time' is essential in scanning primary scientific archive sources.

Image Quality and Archive Image Formats

Binary documents

Rapid progress in computer technology over the last few decades has had the unfortunate side-effect of rendering many early attempts to computerise archive documents obsolete. Until the last decade, most documents were scanned only in binary (black and white), constrained by computer processing speed, scanner capabilities and storage capacity. Binary images are readily compressed using facsimile coding standards such as JBIG and JBIG2, and most off-

the-shelf business OCR software also uses character recognition algorithms that operate on binary images. Since initial applications of document scanning and OCR were invariably high quality business documents, which were also typically black and white originating from monochrome printers, these technologies were perfectly satisfactory and sufficient for most business purposes.

Archive future-proofing

In scanning archive and historic documents however, it is generally essential to reproduce the original document as exactly as possible, since the objective is not only to recognise and index its content, but also to preserve it as an artefact. Making the archive available in electronic form requires both that the original source document can be viewed as an image, and also that it is fully indexed within an online database to allow searching and location of the required data (and image). Unless index data are to be added manually, document analysis (to identify relevant database fields) and OCR (to recognise their content) are needed[11]. A major difficulty with the indexing process is to be able to predict what information will be important to future users of the archive. For example, current OCR technology may be able to index the typed content of an index card, but is unlikely to recognise handwritten annotations, although these may represent an update or correction of the original typed information. Hence compatibility with future needs and technological developments demands that an accurate copy of the original image is included in the archive (and not just recognised textual content), so that improved analysis tools can be re-applied to the archive image in the future.

Grey-scale and colour image coding

Scanning documents in grey-level typically requires each pixel to be represented by a byte of data (8 bits) rather than a single bit,

[11] An alternative approach, popular in digital libraries, is to recognise and index all textual content of the document, so that any word can be a search term. This is more suited to continuous text documents than archive indexes, where OCR is often limited to specific localised well-formed fields for which full contextual support (e.g. a dictionary) can be deployed.

while scanning in full colour requires 3 bytes (24 bits), a byte each for the three primary colours Red, Green and Blue. Hence grey-level image storage requirements are typically 8 times as large as binary, and colour images 24 times as large, in uncompressed formats. At a resolution of 300 dpi, a single uncompressed A4 page requires around 1 Mbyte of storage in binary, 8 Mbytes in grey-scale and 24 Mbytes in full colour. Compression can reduce these figures by a factor of 10-100, depending on whether lossless or lossy coding is acceptable.

Whereas only a few (facsimile-based) image coding standards exist for binary images, there are many alternatives for grey-scale and colour images, reflecting the much wider range of applications for colour imagery. Some image formats (e.g. BMP, TIFF, PNM) are based upon a pure raster (pixel-based) image representation: such formats are often used internally in applications for manipulating the image, but are not efficient for storage. Other formats (e.g. JPEG for natural imagery and GIF for graphical imagery) are widespread because of their adoption as web standards, but neither is optimised for document representation. Finally, there are some image coding standards that have been developed specifically for document compression and coding.

One hybrid algorithm is DjVu, which has become a de facto commercial standard (BOTTOU et al. 1998), and is based upon segmenting the document image into three layers, foreground, mask and background. The mask layer, coded using facsimile binary coding techniques separates foreground textual detail from the background, which is coded using the JPEG2000 natural image wavelet coding method. Other document image coding techniques utilise the Region-of-Interest mechanism within JPEG2000 to identify rectangular areas of textual detail which can then be preferentially coded (YIN et al. 2004). Still other formats are explicitly designed for mixed image-and-text representation (e.g. PDF, HTML, RTF and PostScript) or for pure text-only applications (e.g. ASCII/Unicode).

Therefore it would be premature to assume that there is currently any single ‘standard’ to which scanned archive document images should adhere, and which will guarantee to avoid obsolescence. The

reality is that creating online archives is not a once and for all operation, but rather an evolutionary process, where archive content and format will continue to develop over many years as computer technology offers progressively more sophisticated document analysis and recognition tools.

The conclusion from this review is that, while future archive document image processing and storage formats are likely to continue to evolve, the major bottleneck at present is in acquiring source data through document scanning. Hence, where scanning is undertaken, it is essential that quality standards are met that will not compromise future document image analysis, OCR and/or hand-writing recognition. A minimum specification is to scan archives in full colour at 300 dpi or greater, since US surveys of commercial OCR systems have shown that performance is compromised below 300 dpi resolution, but is not significantly improved at higher resolutions (RICE et al. 1996). Useful guidance on best commercial practice in document scanning can be found at the website of the Association of Information and Image Management (AIIM)[12], for example one current initiative is to develop an international standard to define the use of PDF for archiving and preserving documents[13].

Document Image Analysis and Recognition

Scientific archive documents are often highly structured, with the structure extending over the complete archive rather than being confined to each individual document. The language they contain may be arcane and subject-specific (e.g. Latin species names, proper names or place names) and image quality may be poor due both to poor formation of the original text (e.g. using early typewriters) and subsequent ageing and deterioration.

To maximise computer recognition of the content of such documents requires a four stage processing system (described below). At present, only the third of these stages (OCR) is readily available as

[12] Association for Information and Image Management International, 1100 Wayne Ave., Suite 1100, Silver Spring, Maryland, 20910, http://www.aiim.org
[13] AIIM PDF-Archive committee, http://www.aiim.org/standards.asp?ID=25013

an off-the-shelf commercial software package, with the result that effective online acquisition of scientific archive data remains, in general, an unresolved research and development problem.

- The first stage, **image preprocessing** is responsible for pixel-level operations applied to the scanned image to separate foreground textual (and possibly line-drawing) content from the document background. This reduces the subsequent processing load and maximises the performance of the following document image analysis and OCR stages. Stand-alone OCR packages usually include internal thresholding/binarisation, but these may be sub-optimal for archive documents, since they are more usually configured to maximise performance on business documents.
- **Document image analysis (DIA)** is used to help identify different semantic components of the archive, which will typically map to different fields of the corresponding online database. Identification of these components serves two purposes: first it identifies which database field a particular text string should be mapped to; and second, it allows OCR of that field to be optimised by the application of field-specific dictionaries.
- **Optical Character Recognition (OCR)** is applied to semantically-labelled sub-images of the complete document. Its performance is maximised by optimal image thresholding to compensate for variability in the archive document background, and by the application of field-specific dictionaries to improve word recognition rate. OCR may be based upon the use of configurable off-the-shelf technology, but current off-the-shelf OCR packages are unable to achieve good system performance unless augmented by the other system components described here.
- The OCR system produces a raw text string as its output, but this string requires further **textual post-processing** to configure it for insertion into the online database. First, regular expression matching may be required to extract the required sub-string from the raw text string corresponding to an identified semantic field (for example a year date and/or a name may need to be extracted from a string providing a full literary reference). Second,

further application of field-specific dictionaries as part of the text post-processing can be used to correct some OCR errors. Finally, post-processing can be used to configure the output text strings for direct insertion into the target online database, using a standard database input format such as CSV, suitable embedded SQL commands, or an XML representation.

An Example System – VIADOCS

System Concept

VIADOCS (Versatile Interactive Archive Document Conversion System) was a research project to develop a user-configurable archive document recognition system, funded by the UK Biological and Botanical Sciences and Engineering and Physical Sciences Research Councils (BBSRC and EPSRC) under their joint Bioinformatics research programme. It integrates image analysis and text post-processing tools with a configurable commercial OCR package, to generate text content that can be fed direct into a target online database, and includes an efficient archive card scanner for initial document image acquisition based on a commercial cheque scanner.

User configurability is essential to allow the system to be re-targeted to process different archives, and also variable layouts within the same archive. The pattern recognition aspects of the system (ranging from colour segmentation, to document structure classification, to artefact identification and removal) are uniformly implemented using a fuzzy classification scheme which is parameterised within the user interface. OCR performance is optimised using configurable user dictionaries linked to semantically labelled text fields identified using the document image analysis sub-system. The raw output of the OCR system corresponding to each labelled sub-field of the document image is then post-processed using a regular expression engine to convert it into the required database format (e.g. LepIndex; BECCALONI et al. 2005). The system was developed in conjunction with the UK Natural History Museum (NHM), and hence has largely been evaluated on their archive data, although it is also intended to be more widely applicable to other structured document archives.

NHM card archives

Archives at the UK Natural History Museum (NHM) are recorded in card indexes, which contain bibliographical data and other information for one scientific name on each card, laid out in a standardised format (fig. 8) for each archive. However, different archives may be subject to very different recording conventions. Information is usually type-written, but a minority of cards are entirely hand-written and hand-written annotations are common.

Card archives are scanned using a bank cheque scanner (fig. 1), modified by the addition of specially developed software which allows configuration of the scanning process, to build a large image archive from a series of batch scans which may take place over days or weeks. Using this system, over 0.5 million Lepidoptera (Moths and Butterflies) and Coleoptera (Beetles) cards have so far been scaned

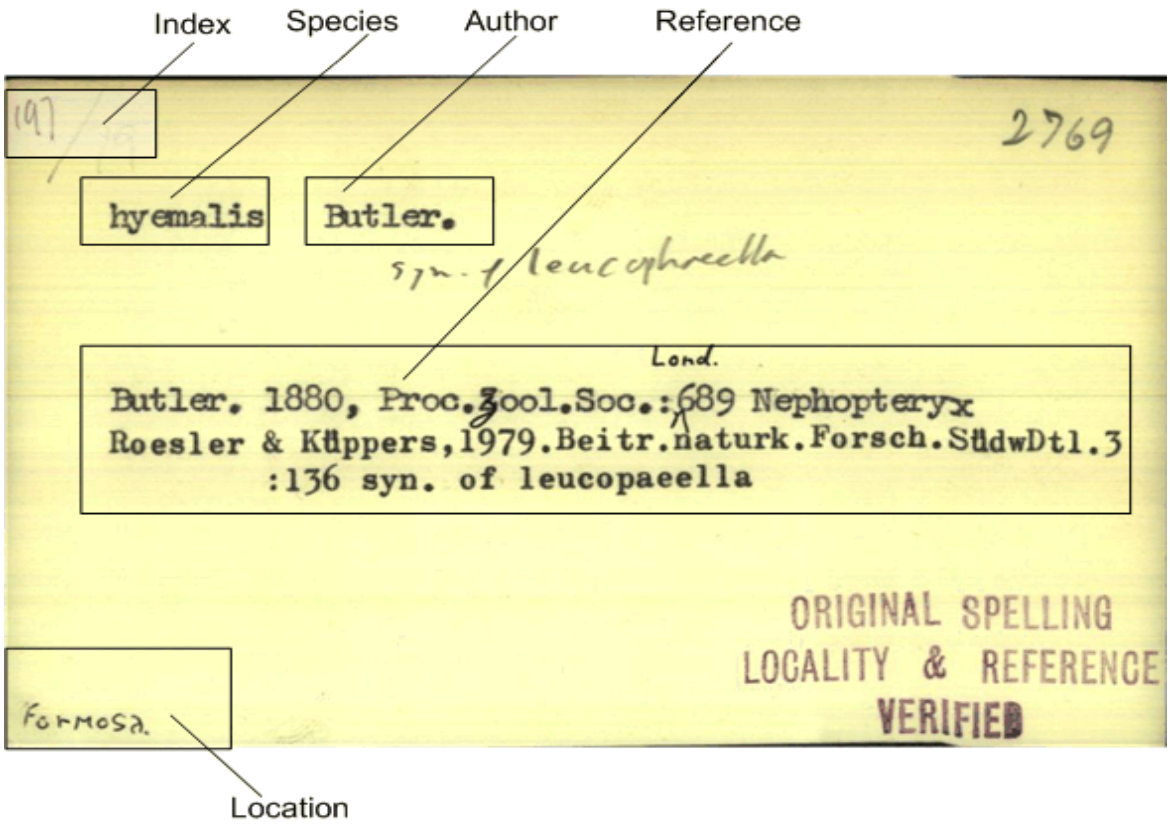


Fig. 8. An index card with multiple hand print and handwriting annotations showing components to be extracted.

at the NHM. The scanner has the capability to scan both sides of a card simultaneously in colour and/or monochrome at 200 pixels/inch resolution at a rate of about 1 card/second, and stores the resulting images in JPEG format. It is also able to print a reference file number on the back of each card for cross-checking against the electronic file archive.

Because different card archives record varying semantic information with different layouts, user configuration of the system is required before commencing analysis of an archive, to define suitable card template layout(s), identify dictionaries to be used with the OCR system, and specify any text post-processing required to interface the system output to a target database. The system is configured using fuzzy templates specified using a graphical user interface, which are described in detail elsewhere[14].

System Overview

The overall system (fig. 9) consists of four main components, pre-processing, document analysis, OCR (using a commercial OCR engine) and post-processing. Pre-processing reads the original JPEG document images, and converts them from either colour or grey level into binary for semantic labelling purposes. Document analysis then segments and semantically labels important text fields. The output normally consists of labelled text field colour or grey-level sub-images in the system's internal image format (PNM). However, pre-processing can also be recalled after Document Analysis so that the labelled text field sub-images can be converted into binary under system control rather than relying on the OCR's internal binarization algorithm. The OCR system recognises labelled images text-field by text-field and converts them into raw text, which is further processed by regular expression post-processing to meet final database input requirements. The components are integrated into a complete archive batch processing system with the user interface shown in fig. 10.

[14] Cheque processing – ICDAR

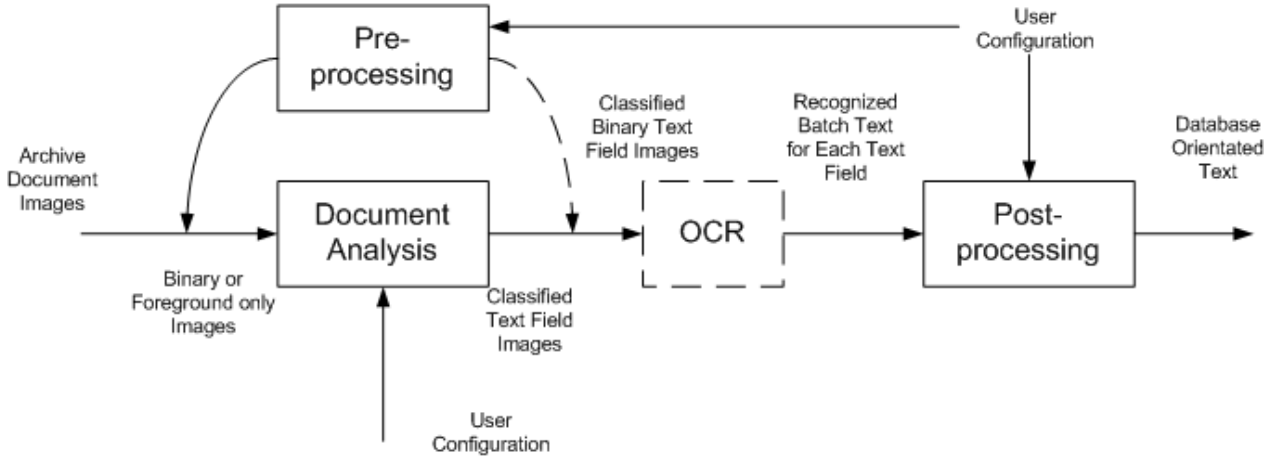


Fig. 9. Overall VIADOCS System Diagram.

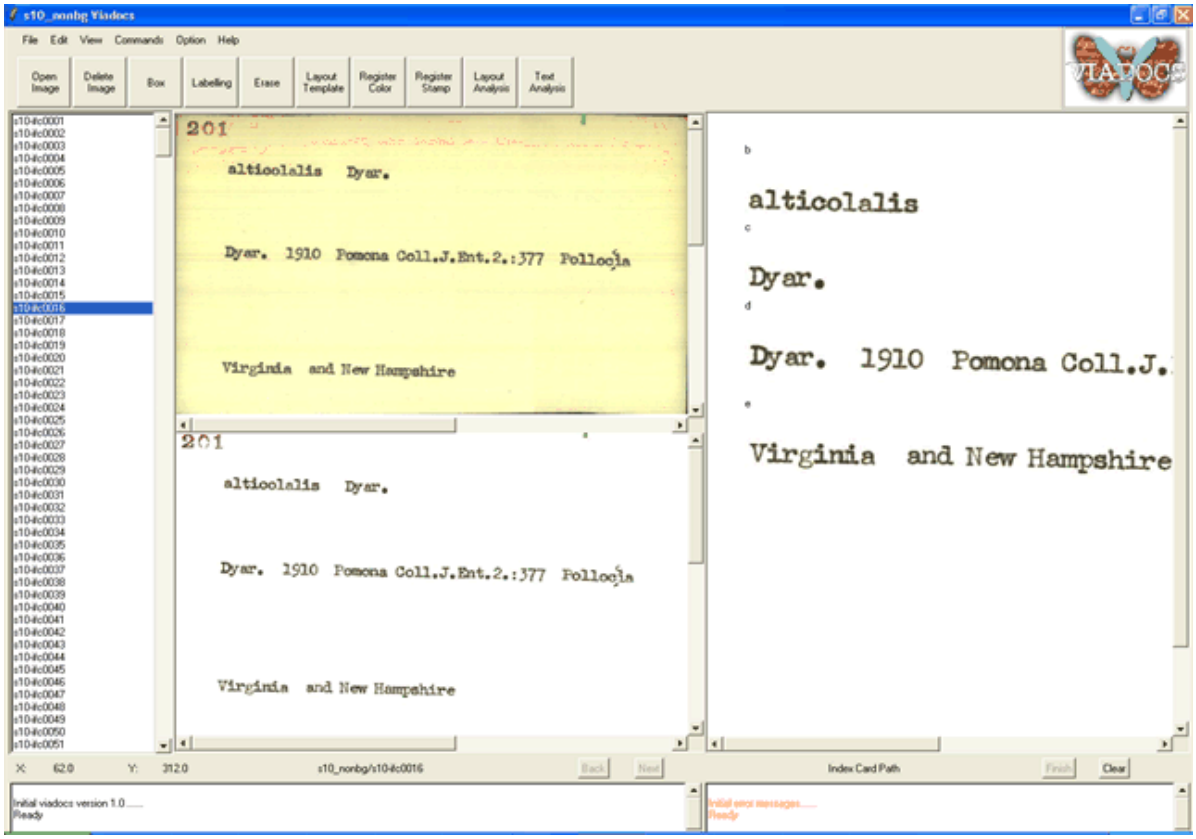


Fig. 10. VIADOCS Graphical User Interface

Document Image Pre-Processing, Analysis and OCR

The format of archive index cards consists of several independent blocks of text, and each block contains one or more logically related text fields. Blocks retain a fairly consistent mutual layout over a complete archive, but the layout of text fields within each block is

not strictly fixed. Pre-processing provides a number of tools, including five alternative binarization and color segmentation algorithms to separate text from background, any of which can be applied to input images. A top-down hierarchical text region segmentation algorithm then extracts and stores the contents of each index card into a hierarchical tree structure (the so-called X-Y tree), consisting of text blocks, lines and words.

In addition to segmentation, document image analysis labels each segmented region of each card (as shown in figure 8), in this case based upon the best-match template layout pre-registered during system configuration for the batch of cards being processed. Labelled image fields allow the OCR system to be configured with field-specific dictionaries, and raw text output from the OCR to be fed to the correct database field.

The OCR used in the proposed system is a commercial product, ABBYY FineReader 6.0. Since it is designed for stand-alone use, it includes its own internal image processing (e.g. binarization) integrated with OCR, but can also accept pre-processed images in a variety of different image formats including binary, allowing VIADOCS users to substitute alternative binarization algorithms. Abbyy is used as middleware working in combination with the other components of the VIADOCS system. For example, figure 11 shows texts on the card that have been extracted and labelled into 5 classes of images: Index, Species, Author, Reference and Location. To recognise the class Author, a specific name dictionary (provided by the NHM) is added to the default OCR English dictionary. A different field-specific dictionary is used for each semantic field. The OCR output of this processing is the raw text which is saved into a separate text file for each text image.

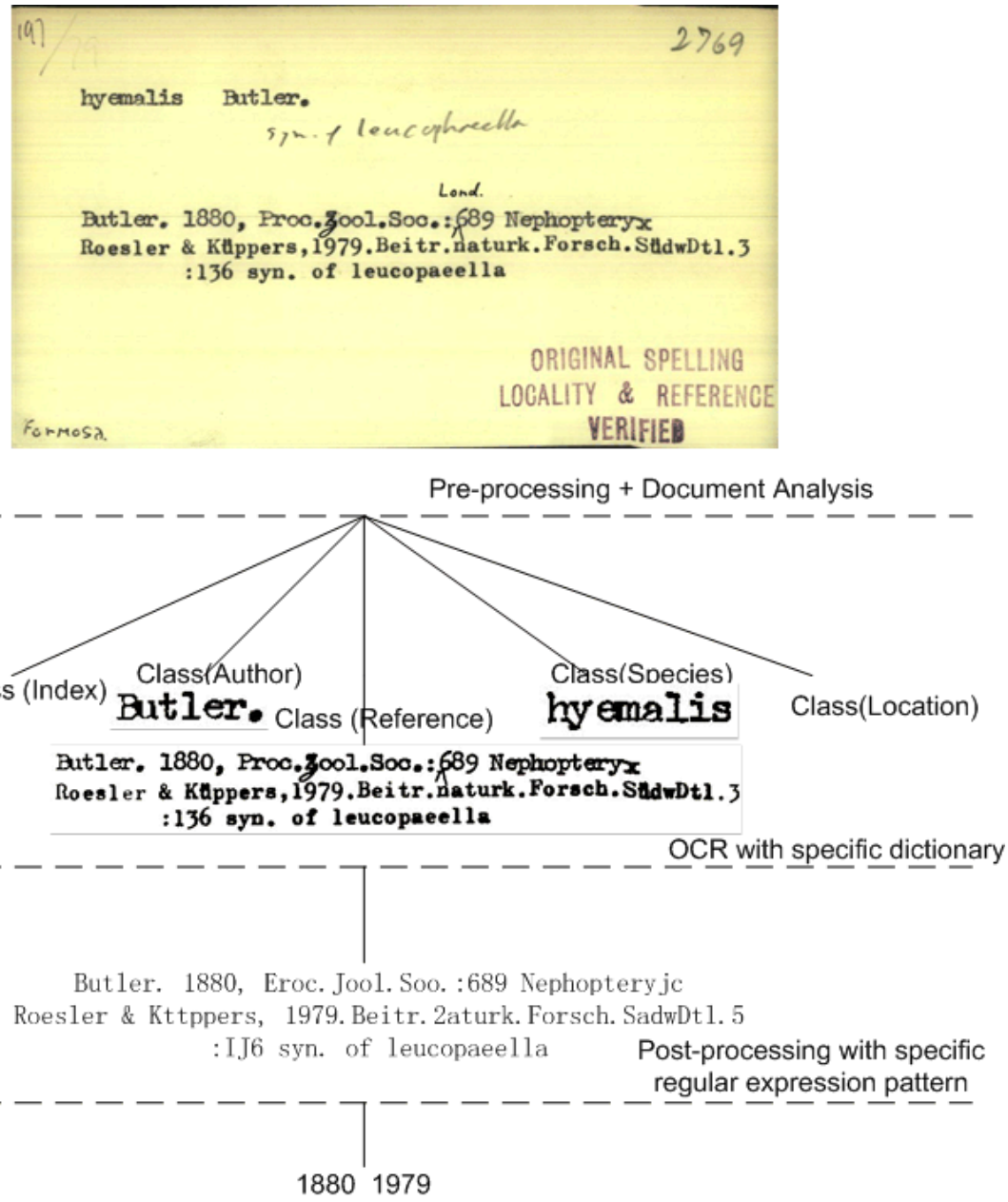


Fig. 11. Example of Card Processing

Post-processing

The purpose of text post-processing is to generate database-oriented text strings for input to the online database. The texts obtained from OCR are regarded as “raw” in comparison with the requirements. For example, the NHM online database only needs the published year in the recognized reference to be stored as a search key; hence other contents can be ignored. Another example is the author name, where the database requires complete author

names, but abbreviated author names (terminated with a full stop) are frequently found in the original images, e.g. the author “Warren” may be abbreviated to “Warr.”. The corresponding complete author names need to be retrieved and substituted for each abbreviation in the online database.

Tcl regular expressions specified within another part of the system interface (fig. 12) are the main tool to manipulate the raw OCR text output. For example the regular expression to parse the published year from a reference is:

regexp {([1][7-9][0-9][0-9])} $reference year

where *regexp* is the regular expression command, *{([1][7-9][0-9][0-9])}* is the parsing pattern (which searches for a year between 1700 and 1999), *$reference* represents the reference raw text

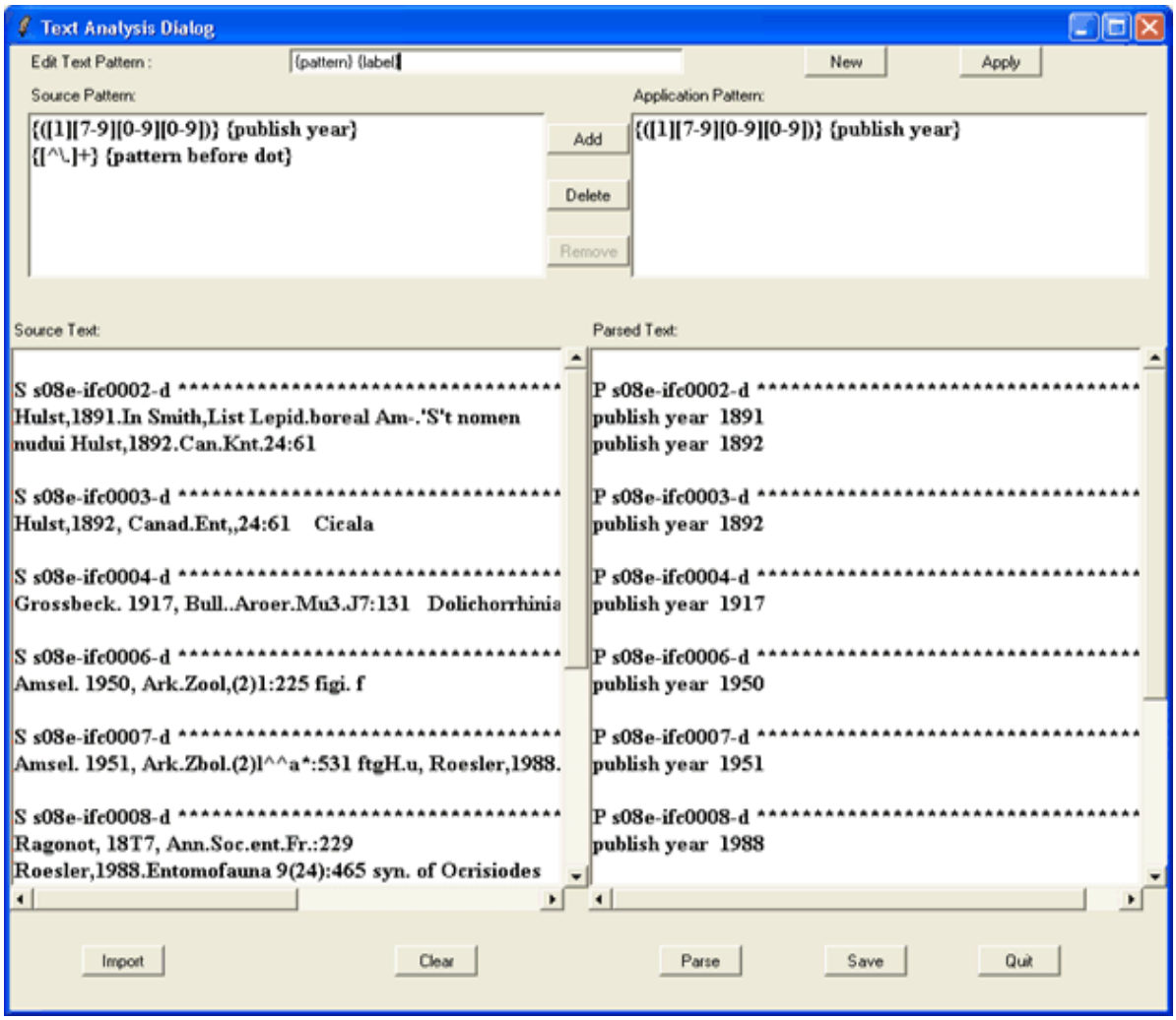


Fig. 12. VIADOCS user interface for post-processing.

generated by OCR, and *year* contains the 4 matched digits output to the 'year' database field. A similar technique is used to select the author name from the raw text, and can even be used, in conjunction with a dictionary of author names, to correct limited OCR errors or expand a contracted author name to its full equivalent.

System Evaluation

The system has recently been evaluated on two sets of sample cards from the NHM archives. One set of 4,435 cards was randomly chosen from the Pyraloidea dataset of 27,578 archive cards, for which full truth data was independently available. A second set of 10,000 cards was processed from the Curculionidae subset of the Coleoptera archive. The Curculionidae testset used different card layouts and dictionaries from the Pyraloidea testset and therefore provided an independent dataset for validating system performance, and also for estimating whether sufficient user (re)configurability had been allowed in the system design. For both datasets, the text fields extracted from each archive card for evaluation were: genus/species name, author name and the date sub-field within the reference, since these fields are currently indexed in the museum's online archive (BECCALONI et al. 2005).

Analysis of errors in the first dataset (table 1) shows that 15% of overall errors occurred when document image analysis wrongly extracted or labelled text fields, and 13% resulted from truthing errors (e.g. see fig. 13) including abbreviations. The remaining 72% of errors were generated by the OCR system, often caused by touching typewritten characters. 16.4% of these errors were subsequently corrected by text post-processing.

Since author recognition was carried out with an incomplete Author dictionary, the word recognition rate for this field is lower than for Species/Genus, where a full dictionary was available. The poorer result for Year was mainly caused by the OCR, which was less accurate in recognising digits than characters (nearly 89% of total errors for Year were caused by OCR errors compared with the

average of 72%). Another cause of poor performance is that quite a few years are handwritten.

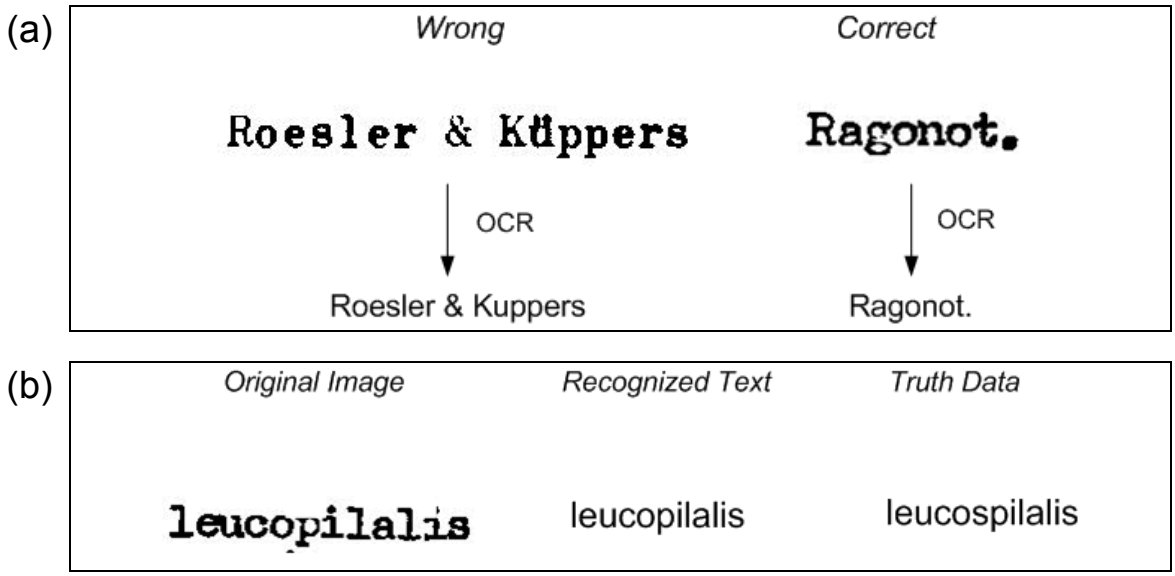


Fig. 13. Example of text field recognition errors: (a) showing an OCR error where the German ü was not available in the OCR character set; (b) showing a species name error, although in fact the OCR has correctly transcribed the typescript.

Table 1. Error Analysis for Pyraloidea Dataset of 4435 card images

Text Field	Species/ Genus	Author	Year
Document Analysis errors	149/4435 -3.4%	166/4435 -3.7%	140/4435 -3.1%
OCR errors	460/4435 -10.4%	711/4435 -16.0%	1080/4435 -24.4%
Truthing Data errors	50/4435 -1.1%	365/4435 -8.2%	0
Post-processing corrections	165/4435 +3.7%	346/4435 +7.8%	0
Correct Text. fields	3941/4435 88.9%	3539/4435 79.8%	3215/4435 72.5%

Truth data was only available for 881 cards from the Curculionidae dataset of 10,000 archive cards. The text fields extracted from each archive card for evaluation in the second evaluation

dataset were genus name, species name and author name. On the image (fig. 14), the top block is Genus and the second (reference) block contains both Species and Author data. Most of the time, Species is the first the word in the block. Author, in most cases, is located in the middle of the block, and terminated with a comma. Its initial letter is always capitalized. Suitable regular expressions were used to search for these fields embedded within the 'raw' OCR output for the reference sub-image.

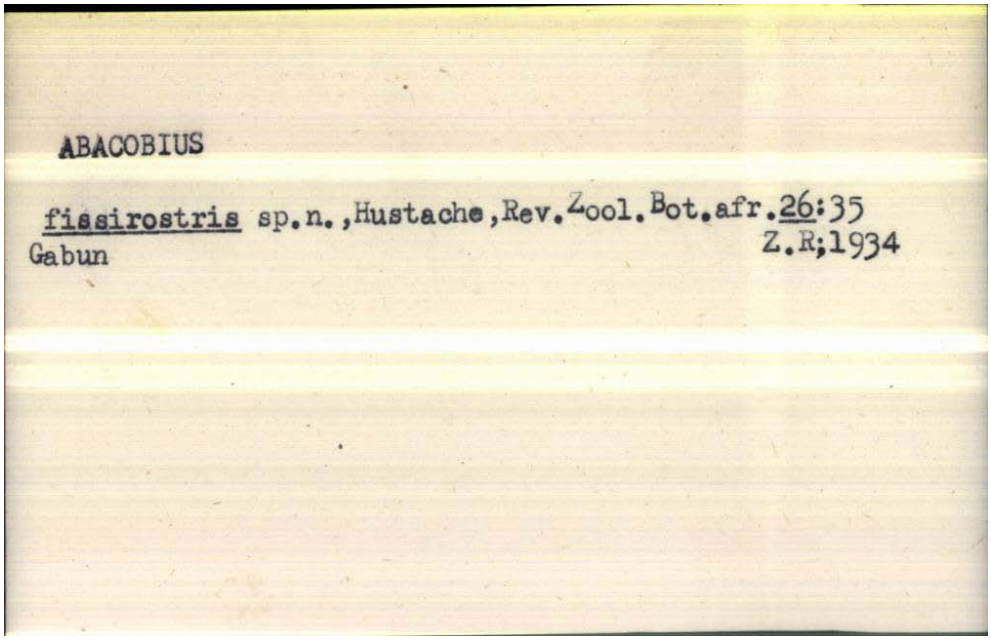


Fig. 14. Different Index Card layout found in the Curculionidae Dataset

Table 2 summarises the evaluation results. Analysis of errors showed that 8.1% of overall errors occurred when document image analysis wrongly extracted or labelled text fields, and the remaining 91.9% of errors were generated by the OCR system, often caused by touching typewritten characters and complex surroundings. 12.6% of errors were corrected by the text post-processing stage. As species recognition was carried out with an incomplete Species dictionary, the word recognition rate for this field is poorer than the other two. Both species and author fields have more complex adjacent text surrounding them than the genus field, which is a separate sub-image. This results in a lower recognition rate for both species and author in comparison with genus.

Table 2. Evaluation Results for Curculionidae Dataset

Text Field	Genus	Species	Author
Text fields	881	753	783
Document Analysis errors	28/881 -3.2%	17/753 -2.3%	17/783 -2.2%
OCR errors	170/881 -19.3%	286/753 -38.0%	222/783 -28.4%
Post-processing corrections	36/881 +4.1%	10/753 +1.3%	47/783 +6.0%
Correct Text.fields	719/881 81.6%	460/753 61.1%	591/783 75.5%

Conclusions

Efficient methods for online acquisition of archive scientific documents are essential to avoid a bottleneck in populating taxonomic databases to meet future research requirements. Manual data input of legacy data is now gradually being superseded by semi-automated methods based around document image acquisition using scanners or digital cameras. In some cases (e.g. herbaria specimen sheets), both specimen image and textual description can be acquired simultaneously, while in others (e.g. archive index cards), the specimen description is held independently, and hence a separate document image acquisition system is required. A fast and efficient image scanning process is required to deliver high-quality document source images (at least 8 bit colour at a minimum of 300 dpi resolution), regardless of how subsequent document content recognition is achieved.

Until recently, the specialised structure and vocabulary of scientific archives, together with typically low quality typescript or handwritten textual annotation has challenged the capabilities of OCR. However, recent research has shown that useful recognition performance is now becoming achievable using augmented off-the-shelf OCR and handwriting recognition systems, albeit that further correction and validation by either curators or users of the online database will continue to be necessary. The potential of current systems is

illustrated by the VIADOCS archive index card scanning and conversion system develop by the University of Essex and the UK Natural History Museum. This system has already been used to scan around 0.5 million index cards, and with further optimisation of the user configuration process, can potentially save considerable time in generating online archives semi-automatically from card index sources, compared with the current alternative of manual entry of all database fields.

References

ANTONOCOPOULOS, A. & DOWNTON, A. C. (eds.) (2005, in press): International Journal on Document Analysis and Recognition, special issue on Analysis of Historical Documents.

BECCALONI, G., SCOBLE, M. J., ROBINSON G. & PITKIN, B. (2005): The Global Lepidoptera Names Index. http://www.nhm.ac.uk/entomology/lepindex [accessed 2005].

BOTTOU, L., HAFFNER, P., HOWARD, P. G., SIMARD, P., BENGIO, Y. & LECUN, Y. (1998): High quality document image compression with DjVu. – *Journal of Electronic Imaging*, **7** (3):410-425.

LESK, M. (1997): *Practical Digital Libraries: books, bytes and bucks*. San Francisco, Ca (Morgan-Kaufmann).

MARINAI, S. & DENGEL, A. (eds.) (2004): *Document Analysis Systems VI*, Proceedings of 6th International Workshop, DAS2004, Florence, Italy, September 2004, Springer Verlag, LNCS 3163.

RICE, S. V., JENKINS, F. R. & NARTKER, T. A. (1996): The 5th annual test of OCR accuracy. – Technical report, ISRI, University of Nevada at Las Vegas, 1996, ISRI TR-96-01. http://www.isri.unlv.edu/pub/ISRI/OCRtk/AT-1996.pdf [accessed 2005].

SPITZ, A. L. (2004): Tilting at windmills: Adventures in attempting to reconstruct *Don Quixote*. – In: MARINAI, S. & DENGEL, A. (eds.): Proceedings of 6th International Workshop, DAS2004, Florence, Italy, September 2004: 51-62. Springer Verlag, LNCS 3163.

YIN, X.-W., DOWNTON, A. C., FLEURY, M. & HE, J. (2004): A Region-of-Interest Method for Texturally-Rich Document Image Coding. – *IEEE Signal Processing Letters*, **11** (11): 912-917.

Taxonomic-Grade Images

ARTURO H. ARIÑO & DAVID GALICIA

Department of Zoology and Ecology, University of Navarra, E-31080 Pamplona, Spain. E-mail: artarip@unav.es

Key words: Digital imaging, morphometrics, precision.

To many lay people, a small red woollen cap identifies a diver in a photograph of marine theme. JACQUES-YVES COUSTEAU and his crew almost invariably wore them in his movies. Were red caps trademarks, insignias of divers? Well, no, for black caps were the unspoken norm among divers. But at some time, or so legend has it, COUSTEAU settled for red caps for his crew because they looked crisper on his documentary films. The not-so-legendary *'red shirt' school of photography*, as it was known among National Geographic Society professional photographers of the '50s, often asked their subjects to wear red garments when shot on a landscape (MOLDVAY 1981) because red rendered nicely on the relatively new Kodachrome film, yielding brighter, more compelling pictures. So, red caps were not "depicting reality". Rather, they were an artistic license, a clever eye-catching device.

This chapter does *not* deal with the "artistic" aspects of the digital imaging of specimens. 'Black caps' must populate our pictures: we want our images to be accurate rather than beautiful. We will *not* explain how to take a "nice" picture; this is beyond our field of expertise and on this, we cannot offer advice. Neither shall we attempt to set standards for "correct" images. Rather, our goal is to highlight some issues that we have encountered while imaging specimens for scientific purposes. These issues might eventually influence the use of the images as a research or documentary device.

Our experience with type series is limited to microfauna, but many of the issues have been experienced by us while working with non-

type research specimens outside the microfauna. Morphometrics, a research field where images are increasingly used to measure features for taxonomic purposes, is most affected by many of these issues. But the breadth and overlap of these matters warrant a general discussion and, moreover, one that may be applied to type and non-type material.

Uses of digital images of type specimens

The Internet now teems with images – Google lists 880 million at the time of writing. Of these, about 9000 have been labelled as “holotype”, “paratype”, “lectotype” or “neotype”. A preliminary survey shows that, indeed, at least 80% actually correspond to images of type series of species.

Types have been imaged from the very beginning of taxonomy and systematics. Though an image cannot replace a formal description, its ability to convey recognition is unmatched. A good drawing was often the only way to complement any description, or to dispel imaginative interpretations of these descriptions.

Photography brought a quantitative jump, but not always an entirely positive one. A “real copy” of what the eye could see replaced the interpretation of the draughtsman. Fidelity could be achieved, and the imaging process took a small fraction of time compared with drawing. But there were several limitations. Small specimens, in particular, suffered from lack of definition because of depth of field problems, whereas a drawing allowed such problems to be overcome. So, photography did not, and in many cases, could not, replace drawing.

Nowadays, digital imaging systems are overcoming many of the problems associated with conventional photography. Photographs can be taken accurately and rapidly, and the much lower cost per image enables the capture of many pictures from the same specimen, from different angles and under different observation conditions. Large sets of images that can be kept in digital storage and recalled at will promise to be as good for observation as the object itself. Digital processing of the images allows particular features to

be highlighted, regions and structures marked, and even mixed and pasted into high quality composites.

We are close to the point at which the quality of some digital images are such that they might even the original specimen for scientific study. In some cases in taxonomy, it would seem that no new knowledge can be extracted from the type specimen than could be captured from a complete set of images of the specimen.

Effectively, digital images offer a real possibility of allowing type specimens to be stored out of harm's way and used for reference when the images fall short of what is required. Many types have been lost in the mail, broken under the microscope, burned with the laboratory, or have simply vanished. Digital images can be copied and replaced *ad infinitum* and, in principle, without any degradation or loss of quality.

There are, of course, other uses for images. They can be posted and disseminated not only for use by scientists, but also for technicians, managers, students, or the general or lay public, allowing everyone the opportunity to see a particular species. Among scientists there are many possible uses. Some will need the images as references to confirm the identification of a specimen; others, to learn the morphology of an unfamiliar group. Others will need them for research purposes.

Each use has its own requirements. The need to confirm identification for managerial purposes in a case of pest control may be less demanding than when the known range of a given subspecies is being revised. Fidelity, minute detail, or accuracy are increasingly necessary as one centres the research on the taxon itself. Requirements for pictures that are to be used in systematics are the most demanding: any taxonomic feature, which is tantamount to saying any observable aspect of the specimen, has to be captured in the utmost detail and veracity.

Can digital images of type specimens be used instead of the original? We postulate that, theoretically, yes, as long as images are of *taxonomic grade*. By which we mean that any feature of actual or

potential relevance has been captured on the image(s) and is as observable on it as it would be on the specimen itself.

Even within taxonomic-grade images, there may be different requirements. A taxonomically relevant feature could be colour. Nudibranchia, for instance, include colour pattern among their taxonomic features. Thus an accurate capture of colour pattern is important for these organisms, while their dimensions may be less important as their soft bodies are essentially deformable. In other taxa, however, dimensions may matter while colour is irrelevant. Nematodes are a classical example. In such cases, images would be able to represent accurately scale, position, and relative placement, akin to what photogrammetry can do. We could call these *morphometric-grade* images, and they are the basic data provider for the field of morphometrics. This discipline studies shape comparisons quantitatively (ZELDITCH 2004), and is used for ontology, evolution, environmental studies, or taxonomic purposes.

While we cannot offer a comprehensive account of taxonomic grade images, we will discuss some of the problems encountered when trying to approach this aim, and how these problems affect the grade of the images. Many other aspects or specific details will be found in other chapters in this manual. Examples will be presented from our current or recent work to illustrate the discussion. We conclude that all these problems, as well as those not identified here, must be overcome if images are ever to substitute for the original specimens.

The Need for Spatial Resolution

Perhaps the single most important factor to account for the taxonomic usefulness of an image is the spatial resolution achieved, which is somewhat akin to the discrimination power of a microscope. Both concepts share the separation of distinct features, but spatial resolution also requires metric accuracy and geometrical fidelity. A well-resolved image should accurately capture the morphology and morphometry of the specimen and of its taxonomic features.

Taxonomy by Measure

The need for accurate measurements stems from current taxonomic practice in many taxa groups. Be it for lack of other conspicuous features, or for subtle, consistent differences in conspicuous ones among known separate taxa, measurements from body parts or taxonomic characters are often found in the descriptions of types. In such cases, a representative and accurate sample of measurements may serve to identify taxa or to define species. Therefore, types must be well characterised metrically.

Some Examples

In Nematoda, the size of certain organs such as odontostyle or vulva, plays a major role in their identification. LOOF (1985) provides an example of the heavy use of measures in this group. Bone dimension in vertebrates is another well-known case. Rodent crania have been used to characterise groups by diastema length and other metric features at several taxonomic levels, while measures of fish structures may even yield better specific clues than genomics. Images of scales and bones have been used, for instance, to identify pure specimens in hybrid populations of Owens tui chub, *Gila bicolor,* through measurements (MADOZ et al. 2003; see fig. 1).

Comparative morphometrics

Morphometrics deals with spatial positions as well as measurements. Angles between defined base lines, both planar and solid, together with distances, characterise features. Distances and angles can be inferred from images by identifying *landmarks*, standardised points corresponding to the object whose coordinates are recorded. Images are in nature planar representations of 3D objects, and landmarks placed on them are therefore projections of these points on 2D space (fig. 2). Sets of images taken from different angles can represent projections of the same 3D object on different planes, and can be used to reconstruct the complete 3D set of coordinates (see VELDHUIZEN & SPRONSEN, "3D Images of Birds", in this book, on 3D imaging.

The ability to detect both morphological consistencies among specimens of one group, and consistent differences between different groups, is the basis of the taxonomical use of morphometrics. For this, it is of paramount importance to be able to specify landmark positions as accurately as possible. Images have distinct advantages over physical measurements (i.e. callipers, rulers), such as repeatability and reduced risk of damage to the specimen, which is particularly important for type specimens. They also have considerable drawbacks, some of which will be dealt with later in this chap-

Fig. 1. Scale of the highly hybridised Owens tui chub, *Gila bicolor*. Measurements from this set of pictures led to the discovery of the pure population in Death Valley, USA (MADOZ et al. 2003). Cropped from the original 2272 pixels wide image. Pixel represents 5.47 µm. Canon Powershot G2 on Leica MZ6 stereomicroscope.

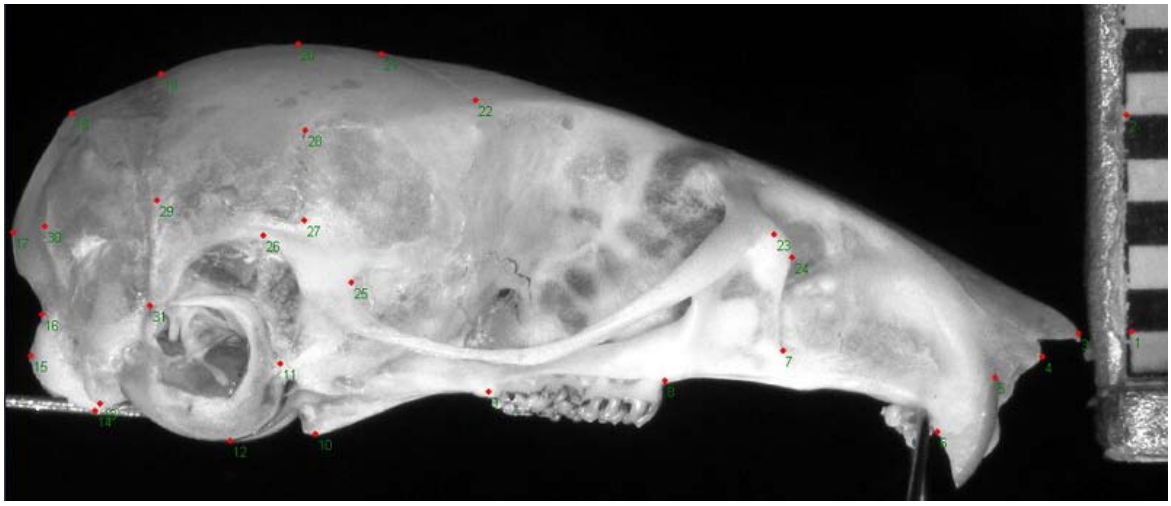


Fig. 2. A field mouse skull in lateral view imaged for morphometric analysis. Landmarks have been set on the image. At right, 10-mm scale placed on the sagittal plane of the cranium. Cropped from the original 2272 pixels wide image. Pixel represents 11.92 µm. 8-bit greyscale from green channel. Canon Powershot S45 on Leica MZ6 stereomicroscope.

ter. A balance must be struck, almost on a case-by-case basis, in order to use images for morphometry, but it is becoming increasingly apparent that the disadvantages of using images are diminishing if one compares the state of the art today with that of the 1990s (compare with descriptions by GARCÍA-VALDECASAS 1996 and FINK 1990).

Precision and Accuracy

RICHARD AVEDON once said, 'There is no such thing as inaccuracy in a photograph. All photographs are accurate. None of them is the truth.'[1] This may be only partly true with respect to the digital imaging of types: truth is always approached and never reached, but inaccuracy exists in its own sense.

In standard biometry, precision and accuracy have well-defined meanings (KREBS 1999). A measure that correctly estimates a true value is accurate. A measure that can be expressed to enough significant figures is precise. When taking repeated measures, a plot of the frequencies of the measures against the assumed true value of the measured magnitude synthesises both concepts (fig. 3).

An ideal digital image of a type that is to be used for morphometry is one that renders most accurate and precise measurements. In the spatial domain, there are a number of factors affecting measurements, stemming from the fact that the digital image is a discrete sample of a virtual continuum: the image being formed on the sensor. Accuracy and precision can then both be an expression of fidelity to the real object. A landmark for measurement can be placed the more correctly the better the position can be ascertained in the sampled image. A well-focused, well-contrasted image renders better accuracy: a landmark can be situated with less uncertainty. Precision is, in turn, a function of the number of pixels into which an image is divided: the more pixels for a given spatial magnitude (size), the higher precision, for the coordinates of the landmark can then represent a narrower region of real space. Figure 4 illustrates these concepts for digital images.

[1] Quoted by DANIEL OKRENT in: No Picture Tells the Truth. The Best Do Better Than That. Article in The Public Editor section, *The New York Times*, Jan 9th, 2005.

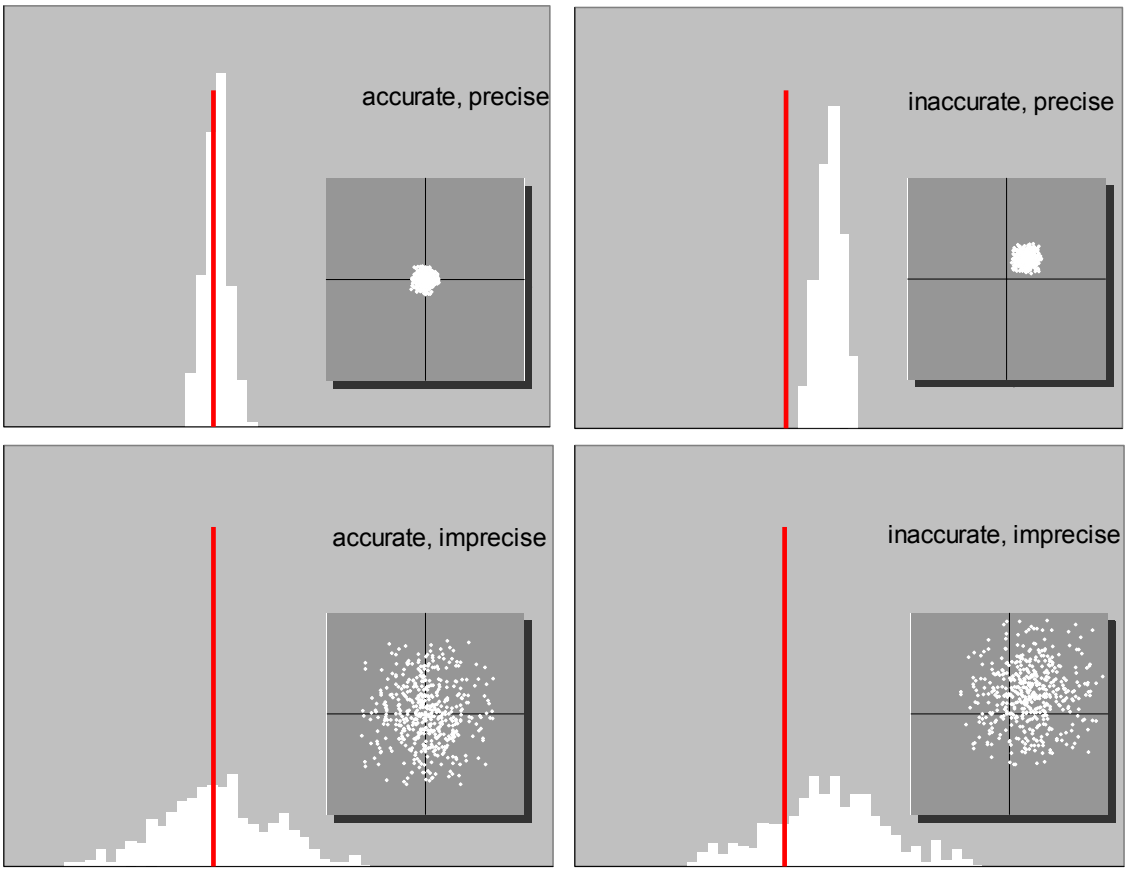


Fig. 3. Histogram of x-coordinates of 500 hypothetical landmarks placed at the theoretically true coordinate position indicated by the red line. The shape and placement of the histogram indicates the achieved accuracy and precision of the measures. Inserts represent the measured coordinates in x,y space relative to the true position (crosshairs).

Sensitivity

Sensitivity encompasses the level of response of the sensor elements to a given amount of incident energy (light), the form of this response (whether it is linear or not), and the accuracy and repeatability of this response. Most digital cameras can have their working sensitivity (i.e. the precise form of this response to a given light conditions) selectable, the range of which will largely depend on the purported quality of the camera. A Nikon D100, for instance, can be set between 200 and 1600 ISO and will accept a large variety of lighting conditions.

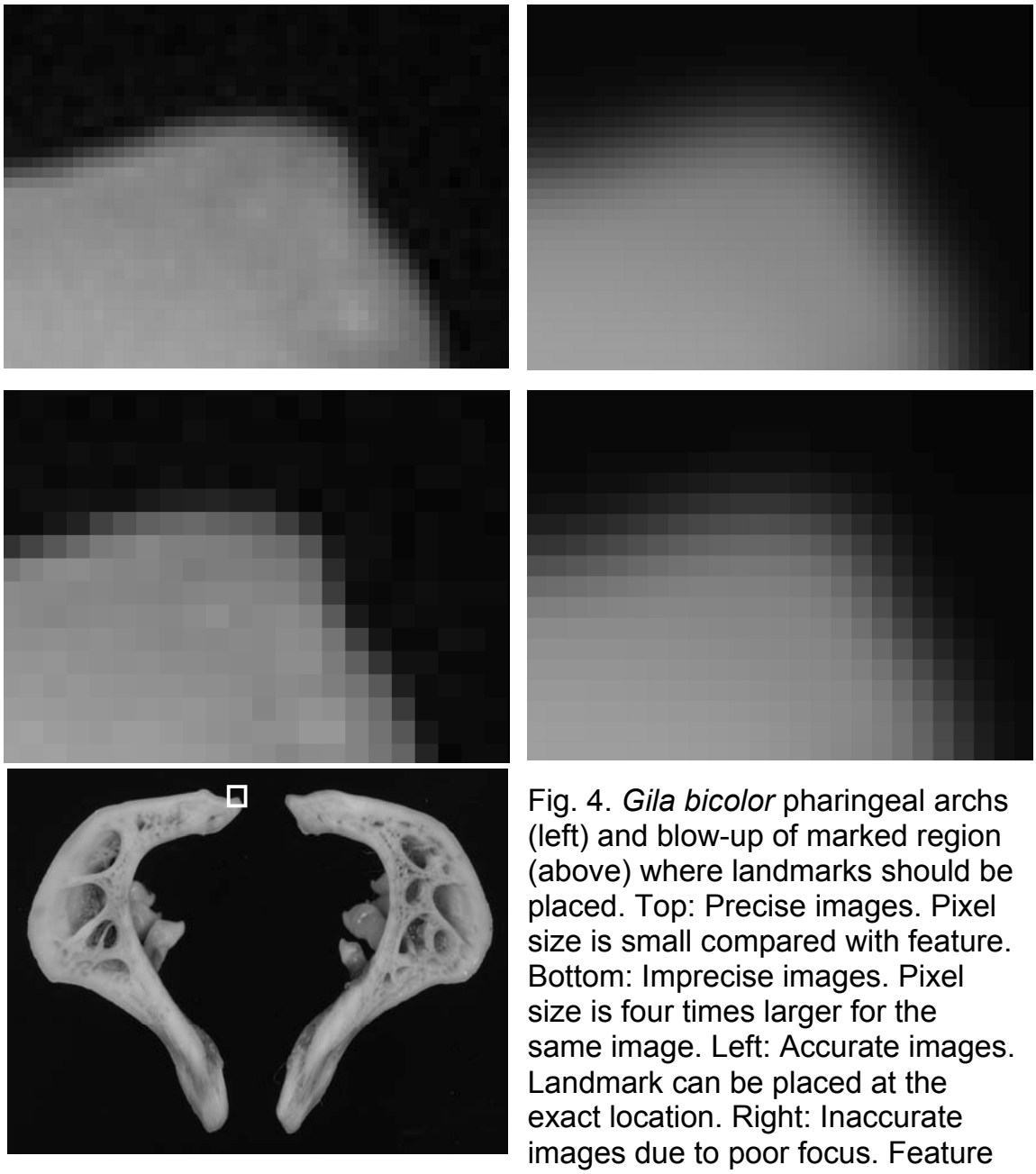


Fig. 4. *Gila bicolor* pharingeal archs (left) and blow-up of marked region (above) where landmarks should be placed. Top: Precise images. Pixel size is small compared with feature. Bottom: Imprecise images. Pixel size is four times larger for the same image. Left: Accurate images. Landmark can be placed at the exact location. Right: Inaccurate images due to poor focus. Feature location uncertain. General image (left): central centimetre cropped from original 3-cm 3008x2000 image. Nikon D100, 180-mm Micro Nikkor (Photo J. Madoz).

Working with still specimens in a laboratory environment means that the exposure time does not need to be short for images to be stable. Even when lighting conditions are deliberately poor, long exposure times, up to one to two second, can be used without visible image degradation. However, digital images from CCDs can suffer

from noise, saturated pixels, and possibly blooming when very long times are used. Thus, high sensitivity coupled with accuracy and a large sensor area is generally to be preferred.

Three-sensor cameras have a higher implicit sensitivity and can operate under lower light conditions than single-sensor cameras. Most available light is directed to one of the three chips by a prism according to its wavelength and very little is discarded. In standard single-chip cameras, at least two thirds of the light entering the lens is not used. Each pixel on the CCD is covered by one of three colour dyes (red, green, and blue) that filters out light of the remaining wavelengths. Thus, aggregated sensor surfaces being equal, the sensitivity to light of a three-sensor camera trebles that of a single-chip one; and for equal sensitivities, a three-sensor camera will need three times less aggregated sensor surface (or produce a threefold resolution picture) than single-chip cameras.

Light conditions, sensitivity, and the amount of energy hitting the sensor (in turn, a function of aperture and exposure) are thus coupled. The target is to optimise the dynamic range of the image, i.e. to minimise the number of saturated and zero pixels, while extending the histogram of pixel response to these limits. This ensures the largest probability of discriminating elements in the image based in changes of colour or shadow on the specimen.

Noise

Noise in the image is most often a function of exposure time. Underexposed stills can be made to look brighter but at the cost of making the transistor response more random. Excess exposure time, in turn, may saturate pixels and reduce contrast, as well as have saturated points appearing on the picture as salt grains. This effect occurs most often in exposures longer than 1-2 seconds.

Lighting

Adequate lighting is necessary to ensure a good, contrasted, equilibrated and faithful image where all features of interest are visible. Both intensity and spectrum must be taken into account, as explained earlier. Artistic lighting (shadow enhancement, colourising,

etc.) is generally not of concern for taxonomic-grade images. Certain shadows do contribute to the volume perception, but in general it seems more suitable to obtain an even, homogeneous illumination that shows all features without obscuring them. Diffuse lights help to avoid unwanted shadows.

In microscopy, Koehler illumination supplies field of homogeneous intensity. For macroscopic specimens, large illuminating surfaces help fading unwanted shadows on the specimen better than spot-lights. Circular arrangements (circular strobe lights, prismatic cubes, light cubes) are generally used.

Figure 5 shows a mobile light cube that we call *"photochariot"* used for decimetric specimens, with a blue chroma deployed. Large-surface fluorescent lamps are used. A very convenient light-box with a circular fluorescent lamp inside a reflective box is used at Stuttgart's Naturkundemuseum for butterflies (see HÄUSER et al., "Digital Imaging of Butterflies", in this book)

Backgrounds and contrast

The background of digital images of specimens can be changed at will and selected so as to maximise the contrast with the subject. Light and colour can be used: a background can be much lighter or much darker than the relative brilliance of the specimen, or could be of a colour that is not present on the specimen. In both cases, any feature, especially in the contour of the specimen, is well separated from the background in the final image.

Chemical photography can accomplish similar results by selecting specific backgrounds, but once the image is taken, the background is fixed. In digital form, the background can be isolated ("masked") at post-processing under certain conditions, and this mask digitally changed to the colour or shade that best enhances contrast with the specimen. Or, it may be made transparent (in some file formats) and superimposed on any background.

Masking the background is greatly simplified by selecting a homogeneous backdrop ("chroma") at the time of taking of a colour not

generally present in the specimen. Blue is generally a good choice for many subjects.

Fig. 5. 'Photochariot' used at UNAV for imaging museum specimens directly on site. Three large-format fluorescent lamps form a light cube. 36 W Philips TL-D 965 full spectrum tubes are used. Native colour temperature is 6500°K; CRI=98. Polycarbonate diffusers. A blue chroma provides background. Two auxiliary small-format fluorescent, 5500°K lamps (not shown) are added to both sides of the camera. White balance is set to direct sunlight+1 to match colour temperature.

When chroma is not used, white, neutral grey, or black seem preferable, according to the specimen's main colours. Bones are generally shot against a jet-black background, and colourful subjects such as butterflies show well against white. Black and white, however, are more tricky if they are to be masked out at post-processing, for the

selection algorithms can also capture very bright or very dark features on the specimen itself.

Chroma isolation also has a few drawbacks. Pixels surrounding isolated features against the background (e.g. hair or feathers in skins) tend to wash out into the background, and then they may not persist in full once the background is subtracted or changed. Figure 6 shows a typical specimen shot against a blue chroma, masked at post-processing for contrast-enhancing background selection.

Colour fidelity

Colour assessment

Colour can constitute a taxonomic character. Although differences in tone are often too subtle for the human eye to perceive, a fact from which high-compression image formats benefit (see MORRIS, in this book), they do exist and should preferably be captured in master images. Human perception of colour, however, often needs a reference to "detect" a particular hue or tint and can be very easily fooled, as many experiments show. When shooting a specimen, lighting conditions and camera settings will produce a picture that may look "normal" to the eye, but that may or may not match the "real" colours of the object. In turn, "real" colours can only be defined in terms of a given standard. White light is often standardized as a colour temperature, i.e. the spectrum of a black body radiating at a given temperature (typically 5500°K for sunlight). Perceived colours are then the parts of the spectrum that are reflected from the specimen (i.e., not absorbed) or created through interference processes.

Spectrum

When illuminated by artificial light, the spectrum of the lighting will determine, along with the surface properties of the specimen, which colours will reach the sensor. Any light that lacks a part of the spectrum will be unable to provide these colours to the sensor even if the specimen would have reflected them. Thus, it is of the utmost importance to illuminate the specimen with full-spectrum lamps, so as not to miss any possible colour on the picture. However, full spectrum does not mean that all components are equally balanced: some may

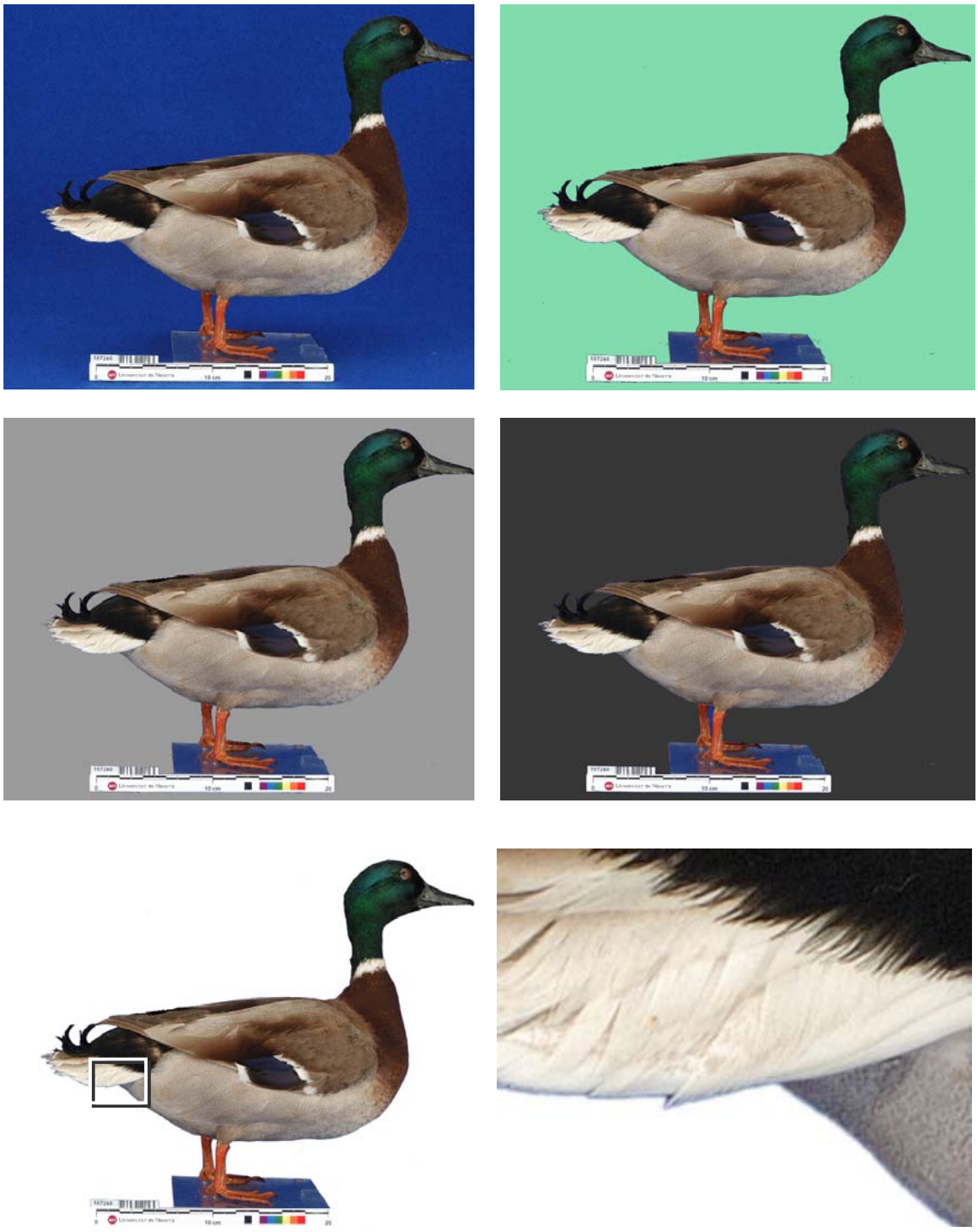


Fig. 6. Museum specimen shot against blue chroma on 'photochariot' (top left) and after automatic masking and substitution of background for light green (top right), neutral and dark grey (center) and white (bottom left). At bottom right, blow-up of a 285-pixel wide region of the white background image showing selection artefacts in undefined contours such as feathers (blue halo). Original image is 3008 x 2000 pixels, each representing 0.17 mm on the plane of the scale.

predominate. Actually, even "standard" white light (that defined by black body radiation, as it is found on tungsten or tungsten-halogen lamps) does not have a uniformly distributed spectrum; Wien's law shows that the dominant wavelength will depend on black body temperature in a non-linear manner.

Low-temperature lamps

Lights whose physical principles are not based in black-body temperature (i.e. virtually any lamp not based on filament heating) will have a non-continuous spectrum with components not obeying Wien's law, and the camera or post-processing should equilibrate the spectrum to "fill in" the troughs in the spectrum. This is not always achieved.

A distinct advantage of non-Wien based lights is that their spectrum usually lacks much infrared. This translates into lower heating of the specimen, and reduction of potential damage.

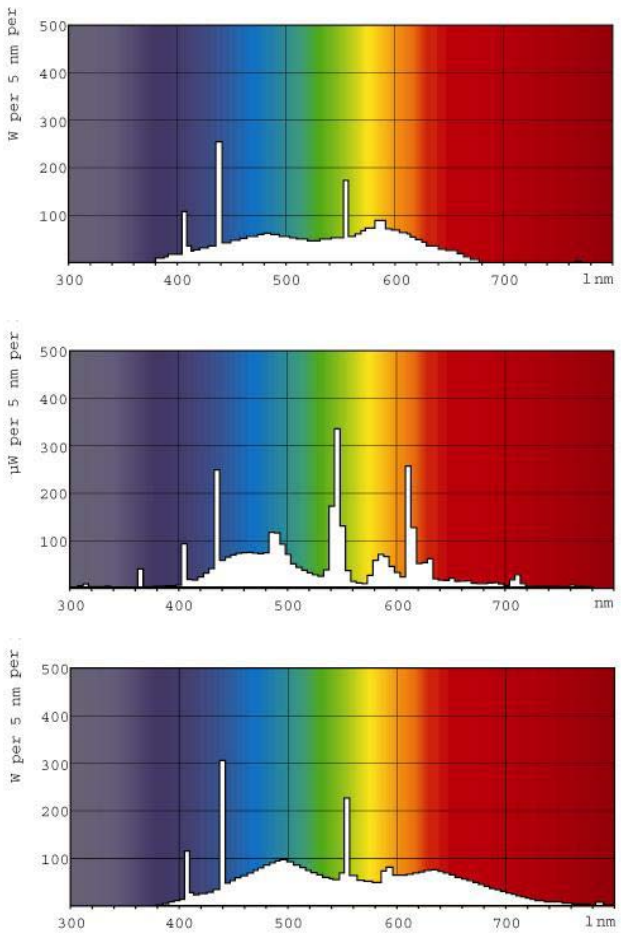


Fig. 7. Declared radiance spectra of a cool-white fluorescent tube, Philips TLD 765 with low CRI (top) and full-spectrum tubes, Philips TLD 965 with high CRI (middle and bottom; note wider range). www.prismaecat.lighting.philips.com.

Colour rendering

Current technologies have made available a variety of low-temperature lights. Most are based upon fluorescent or compact fluorescent tubes. These can be classified according to several parameters:

- **Colour temperature.** The general tone of the light is rated against the average tone of indirect sunlight (5500°K). Tubes of

2700°K render soft, yellow-red tones, and tubes of 6500°K cooler, bluer tones. Tubes with sunlight rating are best for colour fidelity, although some others can be compensated at post-processing or take times.

- **Number of phosphors.** Traditional tubes would have only one or two-band phosphor, yielding a very concentrated spectrum around the fluorescence lines of the medium. Modern tubes may have three or more different phosphors, covering a wider spectrum. For colour-critical lighting of type specimens, only certain multi-phosphor lamps can carry all wavelengths that may potentially be reflected from the specimen. Not all multi-phosphor lamps are adequate for imaging. The choice of phosphors is designed to match specific needs. For instance, tubes used in greenhouses or culture chambers have been optimised for the parts of the spectrum most used by chlorophyll and other photosynthetic pigments, and aquarium multi-phosphor tubes are optimised for enhancing certain interference colours from fish scales and favour photosynthesis by plant chlorophyll more than by algae pigments.
- **CRI, colour-rendering index**. This index (0-100) compares the spectral response of standard test cards when illuminated with a given lamp to the same response when illuminated with sunlight. The higher the index, the more daylight-like the spectrum of the lamp. Common household tubes have CRIs around 70-85, whereas three-phosphor lights to be labelled as 'sunlight' must have CRIs above 90. Figure 7 compares the spectra of a common cool white and full-spectrum white fluorescents.

High CRI and colour temperature alone do not guarantee exact rendering. Diffusers do alter the spectrum, and the white balance must be set either to the actual spectrum, obtained from a photographic white test card, or selected to match a preset by experimenting (see below).

White Balance

As they depict the exact output of the sensor, RAW files are themselves unaffected by the white balance setting of the camera, which is usually included as an in-file tag, and can be corrected at post-processing. However, TIFF and JPEG formats include white balance correction at the time of generation, and cannot therefore be com-

pletely corrected afterwards: adequate white balance should be best achieved at take time.

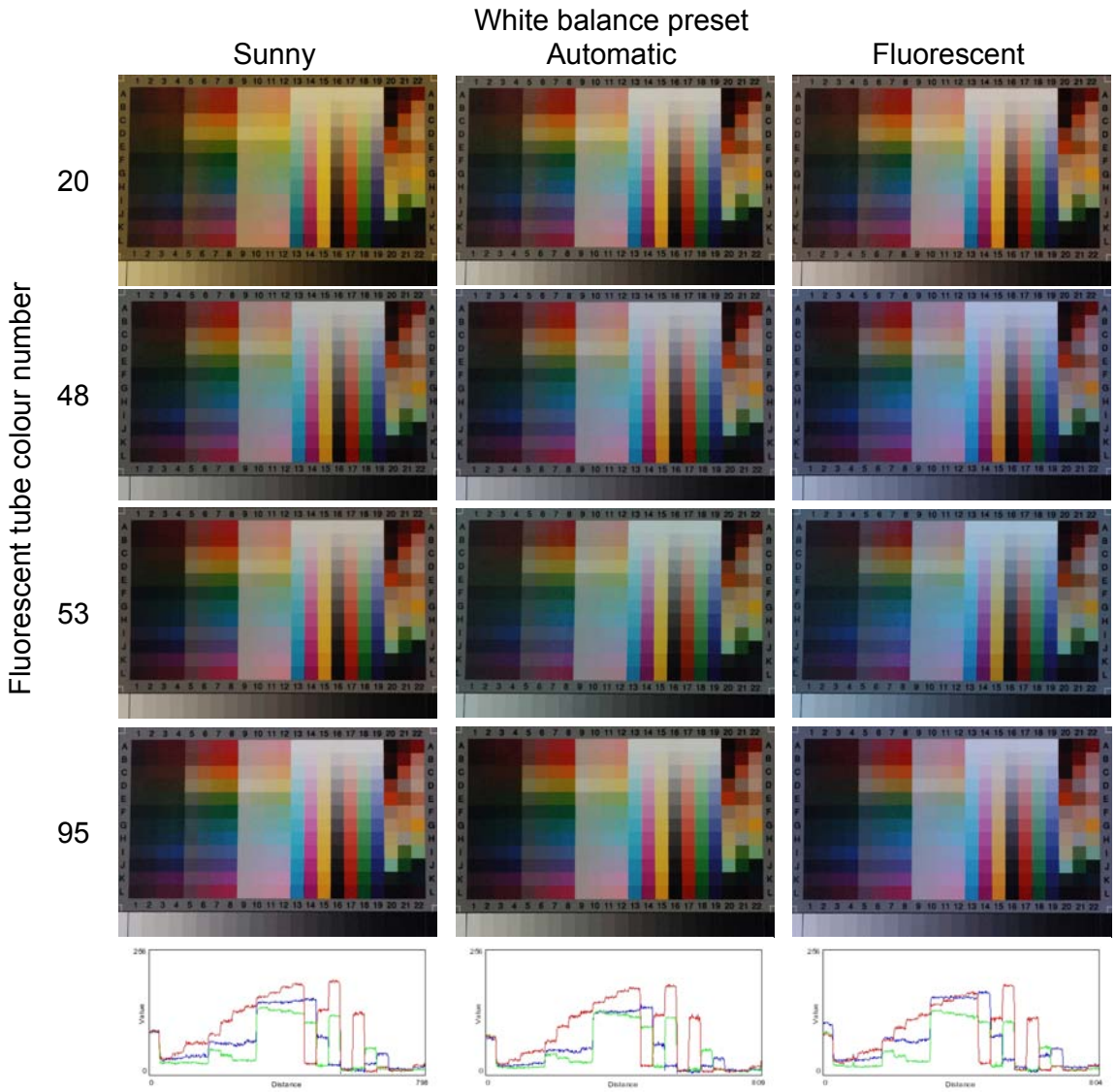
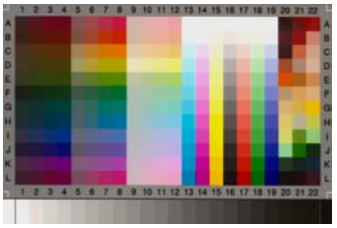


Fig. 8. Some samples of digital photographs of a colour test card (right, scanned) taken under certain lighting conditions and various white balance presets with a Nikon D100 SLR. (The full set contains 117 images). Fluorescent tube colour number corresponds to different spectra; 95 and 48 are full spectrum tubes. The plots show the RGB components of the bottom row of colour blocks of the bottom row of images, used to determine the colour similarity to the master.



Auto white balance is normally not very useful for picturing specimens. This algorithm tries to figure out the white balance from a col-

our-equilibrated subject, but quite often specimens have dominant colours and/or are shot against a homogeneous chroma backdrop that may not be white or grey. Thus, the best option is a custom white balance, obtained generally by shooting a studio-quality neutral grey or white test card. This is not always possible, i.e. when shooting microscope slides. Most cameras can also be set to specific light conditions.

It must be borne in mind that white balance, illuminating spectrum, and colour are bound together. Experiments will best determine the adequate white balance setting for a particular studio.

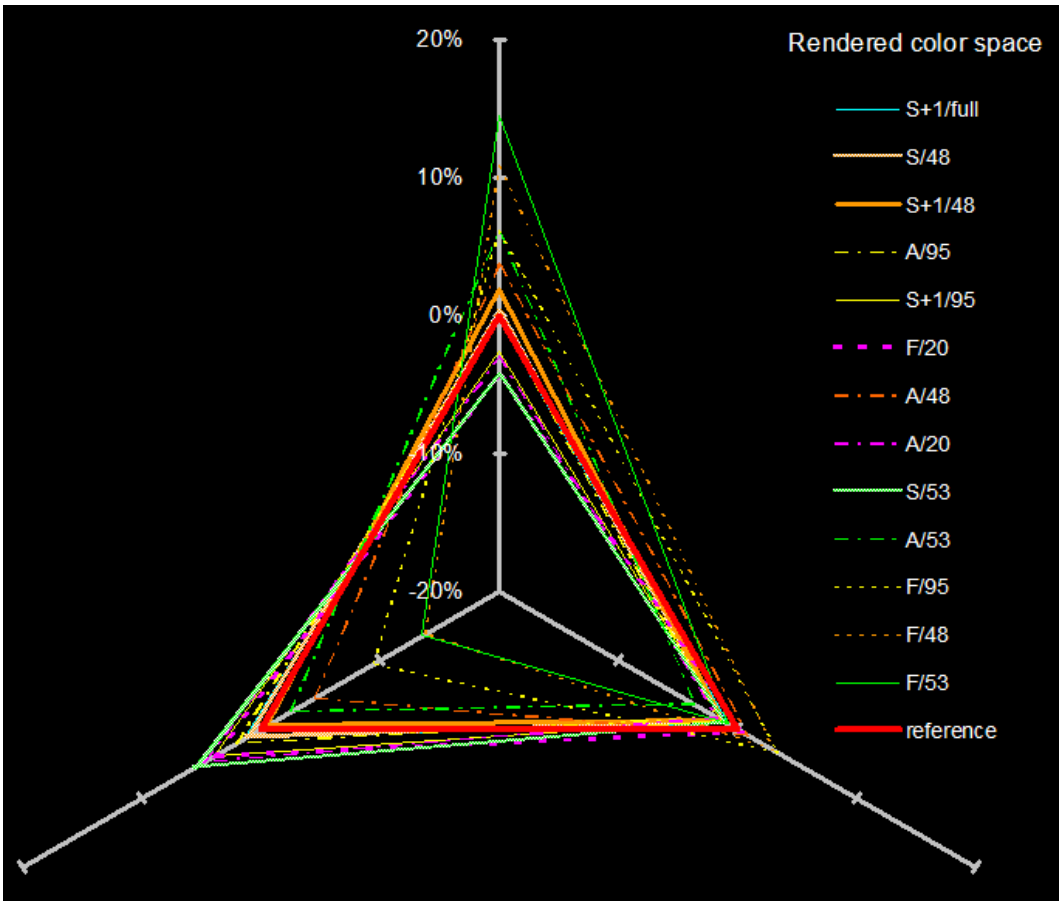


Fig. 9. RGB components of the white test block of some digital photographs of a colour test card. Conditions and codes as in fig. 8. Each triangle corresponds to one image. Vertices indicate each component's percentage of deviation from the reference (scanned card's values). Best is as close as possible to reference (zero) at the three axes.

Figures 8 and 9 show some results of such a test performed on the 'photochariot' described above in order to select the best lighting and white balance conditions for a large set of images of museum specimens. A standard colour test card was scanned with a high-resolution scanner to obtain a colour baseline. Then, the same card was shot under different lighting and white-balance adjustments, and colour profiles were obtained from two sections of the card (white sample and bottom colour row), by extracting and averaging the individual RGB components from most pixels of each colour block. These averages were compared with the numerical values from the baseline, and represented in a triangular diagram. The best white balance and lighting conditions were selected by picking up the plot that most closely approached the baseline plot, i.e. the one that yielded the lowest sum of squared differences between the case and the baseline for each RGB component. In this case, automatic white balance was not the best option, nor any white-corrected preset. Colour was rendered most accurately when lighting with full-spectrum 965-series (6500°K) tubes of 93 CRI, with white balance set at hard sunny preset on a Nikon D100 DSLR.

Sand in the cogs

The above discussion dealt with some of the main goals that a taxonomic-grade image must meet. Now we shall explore some specific problems encountered while imaging both type and non-type specimens. Our list is far from complete; other researchers have found other issues as will be seen elsewhere in this Manual.

Image size and DPI

All digital images have an explicit size measured in pixels. This number of pixels ultimately limits detail if there is no other factor: no details in the image can be represented that are smaller than the pixel size. The more pixels, the more detail, as long as diffraction, focus, aberration or blur allow smaller details to be formed on the image plane. Conversely, high-pixel images are useless if these other factors concur to blur detail above pixel size. Figure 10 shows an example of this limit being reached. ARIÑO et al. ("Imaging Soil

Mesofauna", in this book) discuss this interplay in more detail in the context of microscopy images.

Relevant image size is, therefore, relatively independent of actual, printed size on medium, for it is measured in pixels. The number of dots per inch (DPI) in a printed photograph depends on the line screen used for halftone and the enlargement. Digital images can be scaled down so that the limiting resolution comes from the printing medium, or enlarged so that each pixel expands across several printed dots and becomes the resolution limit.

In practice, images of specimens down to a centimetre can encounter the ultimate limitation in pixel size. A 6 Mpx DSLR camera produces an image that is 3000 pixels across, little more than half the definition of quality 35-mm film. When viewed on a 21-inch 1600-dot computer screen, two image pixels are merged into one single screen pixel. If printed in standard 133-lpi offset paper and using the typical 2x halftone factor, this image would be 28 cm across (almost one landscaped A4-page) before individual pixels would appear.

Fig. 10. A research specimen. Original image taken with Nikon D100, 3008 pixels wide, and 4.5x and 55x blow-ups. The level of image detail in the last blow-up shows that contrast is commensurate to pixel size.

Enlarging details of the image, however, would encounter the pixel limitation. To be able to yield detailed enlargements, the image resolution should be finer. For on-screen presentation (typically at 72 dpi) a 6 Mpx image can be enlarged to almost twice its breadth on a 21" screen or 2.3 times on a 17" flat panel to match pixel to pixel.

Larger enlargements would need more megapixels, scan cameras, or high-definition scanners. It should be noted, however, that in scanners objects are acquired directly at the set resolution and are limited by this resolution. At 8400 true (i.e. *not* interpolated) DPI, a high definition scanner produces an image of a 9 mm object that is 3000 pixels across, equal to a 6 Mpx DSLR image. This allows for insects to be imaged with great detail, as has been impressively demonstrated by JOSEPH SCHEER's artwork. Image size is not comparable, though. A conventional 180 mm macro lens (equivalent to 270 mm for 35-mm film) can capture about the same detail as a flatbed scanner at 2400 DPI. However, with these optics (not using any duplicator or additional lens) the smallest object that can fill the DSLR field is about 3 cm across, whereas the scanner can image a surface that is more than one hundred times larger, as long as it is essentially flat. Cameras fitted to microscopy systems can help capture specimens much smaller. For larger objects, a scanner with such definition rating may not be suitable both because of depth of field problems (for 3D objects) and the size of the resulting image. Figure 11 compares objects captured with a flatbed and a DSLR fitted with a macro lens. The additional resolution theoretically attained with a 9200-DPI scan is offset (for a small field, and mainly due to the much smaller scanner optics) by the higher speed, better depth of field and sharper image of the DSLR lens.

Plant sheets are usually scanned at somewhere around 1000 DPI (600 DPI being now generally considered the absolute minimum requirement), which renders images in the hundred-megabyte range. At this resolution, an 8-cm object is 3000 pixels across and the smallest captured detail is theoretically 27 micron. Throughput is also of interest, as a full-plate image scanned at high resolution can take many minutes and may end up with scanning artefacts.

Scan-cameras that use wobbling-sensor technology may produce large images on a par with a 600-DPI scanner without touching the specimen, allowing for a deeper field if adequate optics and illumination are used (see above section on sensitivity). They also have the relative advantage of producing separate images for each chromatic channel and luminance (greyscale). Although they are very expensive at present, when used in conjunction with apochromatic lenses they may constitute the best available solution for very accurate images especially if colour is not critical (see section below on chromatic aberration).

Scale errors

Many measures from specimens make sense when they relate to each other. In these cases, the absolute magnitudes need not be known; just the proportions. For instance, elongation in a wing, the span to chord ratio, may have a similar value in a herring gull and a kittiwake although the actual span is quite different. If elongation is of taxonomic importance, then geometric accuracy in the image is important while scale is not. In some cases, the absolute measures must be known. In such a situation, scale is to be inferred from the image. A scale line can be present on the image, i.e. can be imaged along the specimen (see fig. 10) or added by the imaging system, or can be 'implicit' (see below).

Implicit scaling

While a scale for an image may not be physically present, it may still exist ('implicit'). The breadth of digital image can correspond to a given breadth at the focal plane. (It is assumed that the geometry is square; see later).

The Gauss formula shows that the ratio between object distance (z_o) and image distance (z_i) is a function of focal length (f). Geometrical optics, in turn, defines magnification as the ratio between object height (h_o) and image height (h_i). Since by definition sharp focus is achieved when z_i is equal to the actual lens plane-sensor plane dis tance, z_i is known and z_o can be calculated (through f). It follows that magnification is known from the image height (in pixels), and since

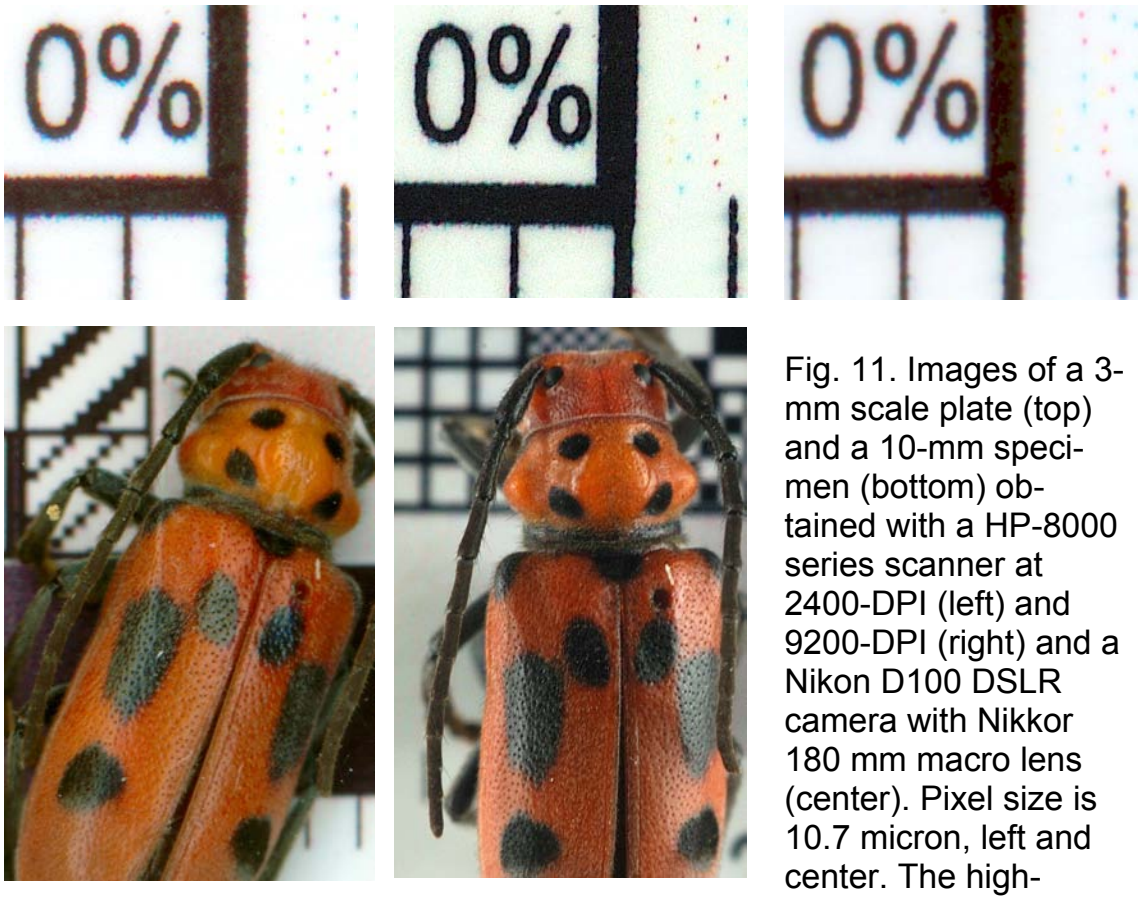


Fig. 11. Images of a 3-mm scale plate (top) and a 10-mm specimen (bottom) obtained with a HP-8000 series scanner at 2400-DPI (left) and 9200-DPI (right) and a Nikon D100 DSLR camera with Nikkor 180 mm macro lens (center). Pixel size is 10.7 micron, left and center. The high-resolution scan (top right) provides pixels 16 times smaller but does not add to the overall sharpness, which seems higher for the camera.

$h_o/h_i = z_o/z_i$ the exact height of the object can be calculated. It should be noted, however, that z_i is subject to great inaccuracy: sharp focus is difficult to ascertain due to field depth. The ratio tends to be very high due to short image distances, and the imprecision grows with object distance.

Therefore, in practice a scale or ruler is imaged at exactly the same distance (and/or magnification) as the specimen. All images under identical conditions should represent the same real world measure per unit length on the image. Since digital images are a square mosaic of pixels, the image breadth (in real world) divided by the number of pixels across defines the size of the pixel, i.e. the real world size of the part of the image represented by one pixel. This can

then be tabulated against take conditions for reuse. Images can be resized for paper printing or screen visualization and as long as they are not resampled, a pixel will always represent the same distance.

Several problems do appear with implicit scale. These can be grouped into two categories:

- *Resampling problems*. An image, say 2048x1536 pixels, might be resized to 1024x768 pixels for web posting. If objects appear to occupy the same proportion of space in the resulting image, this has been actually resampled. The same object is represented in fewer pixels. If this resampling has not been documented, the scale changes automatically by a factor of two: one pixel represents twice the original measure. The same situation pertains to increased sampling (which will never compensate low resolution: more pixels would be used to represent the same blur). On the other hand, cropping does not change the scale, just that portion of the image that is visible.
- *Changed take parameters.* As explained above, magnification (which is 'undone' for finding scale) is a function of object distance and focal length. Changes in any of these two parameters (i.e. by displacement of the object, displacement of the focal plane, zooming in or out, or changing lenses) automatically changes magnification. If these are not documented, the pre-tabulated scale is rendered invalid, although (overcoming the problems explained above) could be recalculated using Gauss (or Newtonian) formula.

Precision of scales

In a morphometric-grade image, it is important to know the inherent precision of the scale. For implicit-scale images, the precision is 1/image breadth (in pixels) so this can readily be translated into metric units. For explicit scales, it is 1/total length of the scale.

A practical consequence is that any explicit (i.e. on-image) scale that is shorter than the full breadth of the image has proportionally less precision. Measurements in a digital image are effected using the scale by placing landmarks at the extremes of the scale of known length (for visible scales) or by using the breadth (for implicit) as vir-

tual landmarks. A scale that spans 300 pixels in a 3000-pixel image has a precision of one decimal point less than the full scale. Landmark placements can still be accurately placed, but the allowable maximum measurement can be divided into 300 units, not 3000 (KREBS 1999). A one-pixel error in the placement of the landmarks on the scale would translate into a ten-pixel error when measuring a feature spanning the whole image. If the landmarks had been placed at the extremes of a 600-pixel scale, the maximum measurement error of a one-pixel error on the scale would have been five pixels.

Therefore, visible scales should ideally be as long as possible within the image for maximum precision. It should be noted, however, that one part in 300 is an acceptable degree of achieved precision for most studies (SOKAL in KREBS 1999). Figure 12 shows an example of a visible scale that is too short.

Fig. 12. A specimen with a too short explicit scale. A five times larger scale could have been fitted vertically for better precision.

Defocus and image imprecision

An out-of-focus image obviously cannot produce a precise image (see above section on accuracy and precision): the edges of the features that are needed to place a landmark cannot be precisely located. The focus must be as sharp as possible, but artificial focusing at post-processing can also introduce artefacts that affect a measurement.

Out of focus images can result from a number of causes. Among them are:

- *Improper focusing.* In microscopic images where autofocus is out of question, the focus plane of the digitiser may not be congruent with the focus plane of the oculars. This may force the photographer to focus through the imager. A small screen, such as that found on compact cameras, does not offer enough resolution to see the proper focus point. A computer screen or auxiliary monitor fed by the sensor data stream affords a much better alternative. However, SLR cameras do not offer generally this option: the image is only formed at shoot time. Continuous focusing must be done through the camera's optical system.

- *Mechanical vibration.* Any small movement transmitted to the camera at shoot time results in motion blur if the image moves across more than one pixel. Precision is lowered by a factor of two for each new pixel that an edge in the image crosses. Images should be taken after ensuring no mechanical vibration. This means almost invariably either remote trigger or delay-activated take (as occurs in self-photographs), especially if the camera is mounted atop some stack such as a phototube. SLR cameras may also benefit from a secondary 1-sec delay between mirror-up and shutter activation that removes vibration due to the mirror movement. The effect is much less important for short exposures. However, under laboratory lighting conditions, often low to avoid excess damage to types (or under complicated microscopic lighting such as is used for differential interference contrast), exposure times tend to approach the 1-sec mark. Naturally, the imager must be solidly fixed to a stable mount (fig. 13).

- *Optical characteristics of the microscope*. In microscopic images, focus is limited by the optical system (see ARIÑO et al., "Imaging Soil Mesofauna", in this book).

Fig. 13. A typical macrophoto stand. 180 mm lens, 5500ºK cold full-spectrum lamps.

- *Automontage techniques*. When moving from one focal plane to another, the relative distance from the object to the sensor plane actually changes. This introduces a small parallax problem (see below): objects seem to “displace” on plane space, marring the accuracy of the resulting measures. This is particularly acute for very narrow depth of field situations, such as microscopic images, as discussed below. When trying to mount such images, portions of the final image that have been assigned to a different Z-plane may appear displaced with respect to the continuation of such feature on a different Z-plane. The overall effect is a loss of definition and metric accuracy.

- *Depth of field and aberrations*. See below.

Depth of field

Trade-off f-stop, sharpness

Diffraction occurs whenever light beams pass through a photographic diaphragm, used to regulate the amount of light reaching the sensor. Light waves hitting the diaphragm’s edges bend around them slightly, diverting from their straight paths. The result is a set of concentric rings of alternating light/shadows forming in the image on the sensor. Their effect on the picture is that of reducing the overall sharpness or definition. This effect is noticeable only at very small aperture sizes (very high f-stop). Large aperture sizes essentially do

sharpness or definition. This effect is noticeable only at very small aperture sizes (very high f-stop). Large aperture sizes essentially do away with it. However, it comes at a price: depth of field reduces as aperture increases. The trade-off between sharpness and depth of field size is one best found by trying several adjustments when imaging a particular object.

Figure 14 shows a specimen imaged with macro lens, showing the degradation that occurs at both too low and too high apertures. As a rule, flat objects such as herbaria specimens will always benefit from

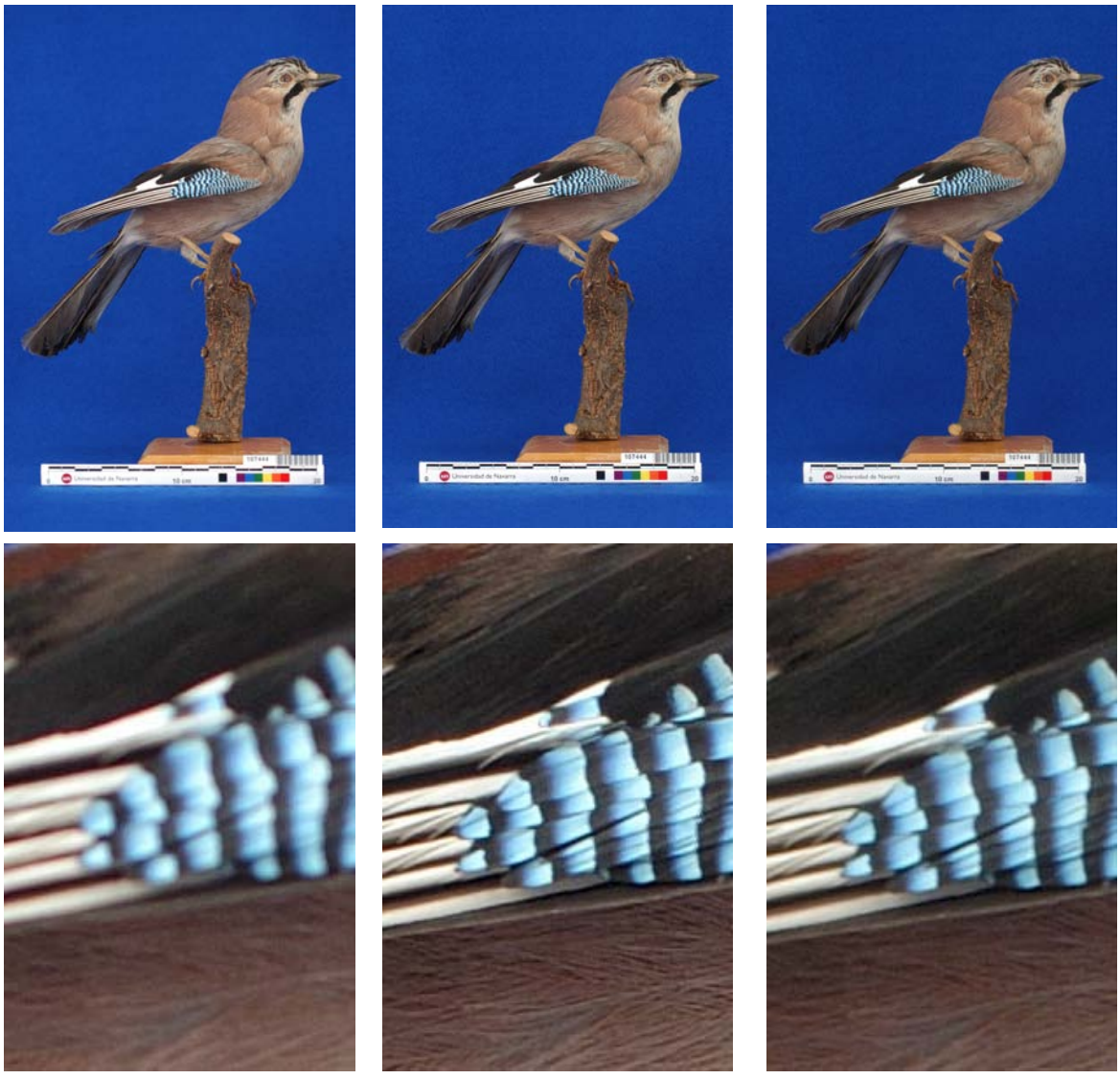


Fig. 14. An Eurasian jay (*Garrulus glandarius*) specimen photographed with a 180 mm macro lens at *f*=4.5 (left), 11 (centre) and 32 (right). Bottom, blow-ups of a right wing section. Note the lack of field depth for the widest aperture, and the diffraction-induced loss of definition for the smallest diaphragm.

large apertures, whereas in specimens having volume (e.g., a skull) it is best to use the widest aperture that still allows the whole object to be in focus. It should be noted that the depth of field is also a function of the focal length (see above); therefore, short focal lengths should be preferable for deep objects as long as distortion does not appear or can be controlled.

Automontage in microscopy

Particular depth of field problems exist with certain microscopic images. Whole specimens that are observed by transparency cannot normally be focused simultaneously and specific focal planes are selected at a time ("optical slicing"). It may be possible to use automontage techniques or algorithms to have a stack of images come to a certain focus, but it also comes at a cost as above:

- Objects spanning across different optical slices (e.g. a seta, hair, leg, etc which is not perfectly horizontal on the slide) may change relative position due to the changed working distance. In the resulting automounted image, the object may lack continuity. See figure 15 for an example.

- Objects located at opposite sides of the specimen (e.g. dorsal and ventral features) may be fused and presented as being on the same plane. This specific problem is further discussed by ARIÑO et al., "Imaging Soil Mesofauna", in this book.

Although these problems might prevent the routine use of automontage for morphometric grade images, there remains the possibility of using stacks of images. A CAD program can be used to reconstruct the 3D model of the landmarks if the interval between Z-planes is known; BYTHEL et al. (2001) give an example.

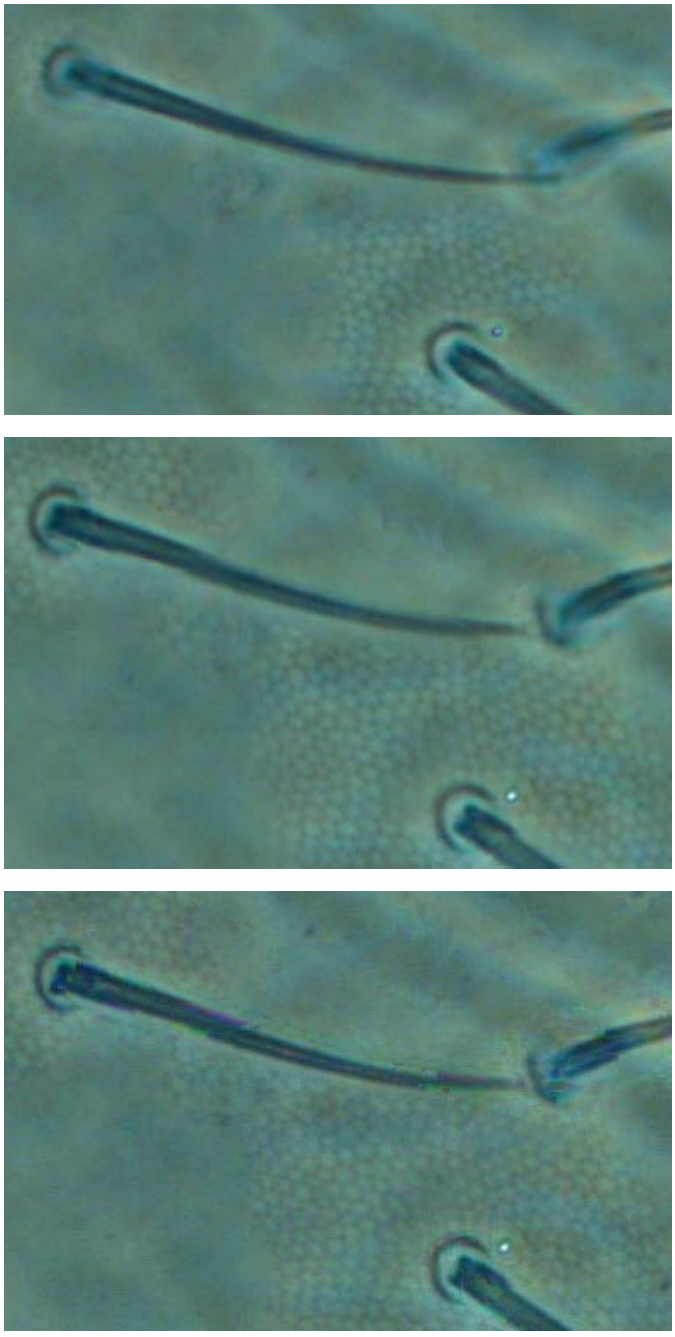


Fig. 15. Detail of the leg of the holotype of a collembolan. Field width, 0.018 mm; pixel size, 46 nm. Nikon E995 on Olympus BH50, 100x phase contrast objective. Top: one of the frames from a five-image stack, unprocessed. Centre: focused stack (Syncroscopy's Auto-Montage). Plane transitions are smoothed to give an illusion of continuity. Bottom: focused stack (ImageJ). Plane transitions kept sharp to reveal displacement among image pieces.

Perspective and Parallax

Virtually any digital image of a specimen is a 2D projection of a 3D object. Except for flat objects, such as herbaria sheets, or optical slices seen through a microscope, where this effect can be neglected, all images suffer from parallax. Two main problems can be recognised for morphometric-grade images as originating from parallax:

- Hidden features due to inadequate angle of view, and
- Scale foreshortening and inaccuracy.

Figure 16 illustrates both effects. A 1-cm long section of a 3-cm diameter tube has been imaged on a graph paper. Features on the outer wall of the tube cannot be seen. The angle formed between the wall and the ray trace to any point in the outer wall prevents viewing, as the wall itself intercepts the line of sight. Moreover, if the upper lip of the tube should be measured by reference to the paper scale, it would yield a 6.7% larger diameter. Being closer to the lens, it appears larger. This difference is already important for morphometrics, for it means that the maximum achievable precision for any measure in an object of this size would be one part in fifteen, one half the minimum required precision according to Sokal's criterion (KREBS 1999).

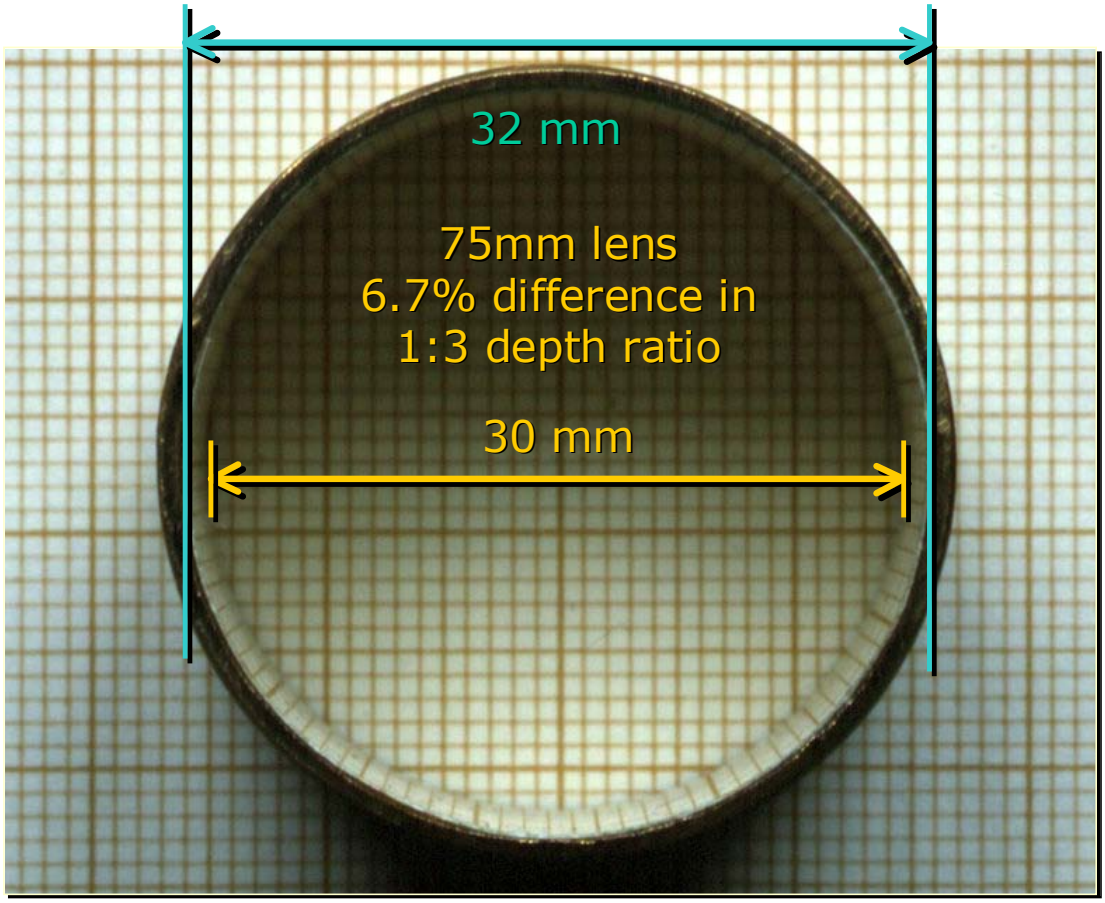


Fig. 16. A tube showing a parallax-induced measurement error (see text).

Parallax is more apparent with close objects, short focal lengths, or near the borders of the image. The relationship between factors, however, is not completely linear. Distortion due to lens configuration conspires with the other factors to make it difficult to predict parallax for a given configuration. It is best analysed by experiments, such as imaging 3D objects of known dimensions in order to tabulate particular combinations of imager, optical system, and Z distance between the feature and the scale plane. An alternate solution would be to use sets of images from different angles and do a 3D reconstruction with a CAD or morphometry program (see BYTHEL et al. 2001).

3D direct modelling

Yet another possibility to correct parallax lies in the use of 3D direct modelling. This can be done classically by means of stereo pairs, or with contact 3D scanners, reflex microscope, or laser 3D scanners.

A contact scanner consists of a stylus that follows closely the surface of the specimen, noting its position in space. Hard, large objects such as shells or skulls could be acquired this way. Once acquired, any projection could be obtained in 2D space. But it cannot be used on small or delicate specimens.

Microscopic specimens can be imaged and the position of certain points of interest recorded in Z-space with a reflex microscope. A reference point (light from a laser or a diode) is focused precisely on the feature being imaged, the system recording the exact Z value of the focal plane. Images obtained as such are planar, but include the Z reference for selected points.

Laser 3D scanners are a natural evolution of contact scanners, where a laser beam that determines the distance between the probe and the surface replaces the feeler. The model is reconstructed in much the same way as in a contact scanner. The image can be taken rapidly and Z-marked for all pixels, but heir resolution is not very good at present. The spatial resolution of 300,000 pixels of the Minolta laser scanner image, with a maximum accuracy of 0.05 mm, may be typical.

Angle of take and orthogonality

A taxonomic-grade image of a 3D specimen must generally be accompanied by complementary images. Features may have to be visible from specific angles. The angle of view with respect to some reference plane or axis in the specimen (e.g. the sagittal plane) should be specified in order to help calculations, especially if exact placement location has to work in 3D space, which is the general case.

Mutually orthogonal images, a particular case of angle, are very useful at both helping the accurate placement of landmarks, and in the correction of parallax. These images can be obtained by rotating the specimen, by rotating the imager, or by creating virtual images.

Rotating the specimen is usually a delicate operation, and should best be achieved by movable stages. The specimen should be fixed to the stage to enable rotations, although these manipulations increase the risk of damage to the specimen. Moving the imager, in turn, will generally require heavier, more complicated equipment such as motorised, robotic arms, but that arrangement has given excellent results for 3D reconstruction (see VELDHUIZEN & SPRONTEN, "3D Images of Birds", in this book.

An alternative to rotating systems is the use of mirrors. Figure 17 shows a stage used for centimetre-sized specimens. We have extensively used it for imaging the delicate crania of voles and mice. The specimen is placed on a stand in the form of a wire frame, with mirrors that allow taking images of the sides and the underside without moving, rotating or removing the specimen by pointing either to the subject or to its reflections. Mirrors are set at a 45° angle to the vertical plane. Figure 18 shows three such images.

Mirrors used in any imaging procedure must be of good quality. In order to avoid ghost images we use first-surface mirrors. Distortion is avoided if the mirrors are polished to a high degree. We use astronomy-grade flat mirrors for general image, which are a good price-quality compromise. They are made of glass, coated with aluminium, and polished to a few wavelengths.

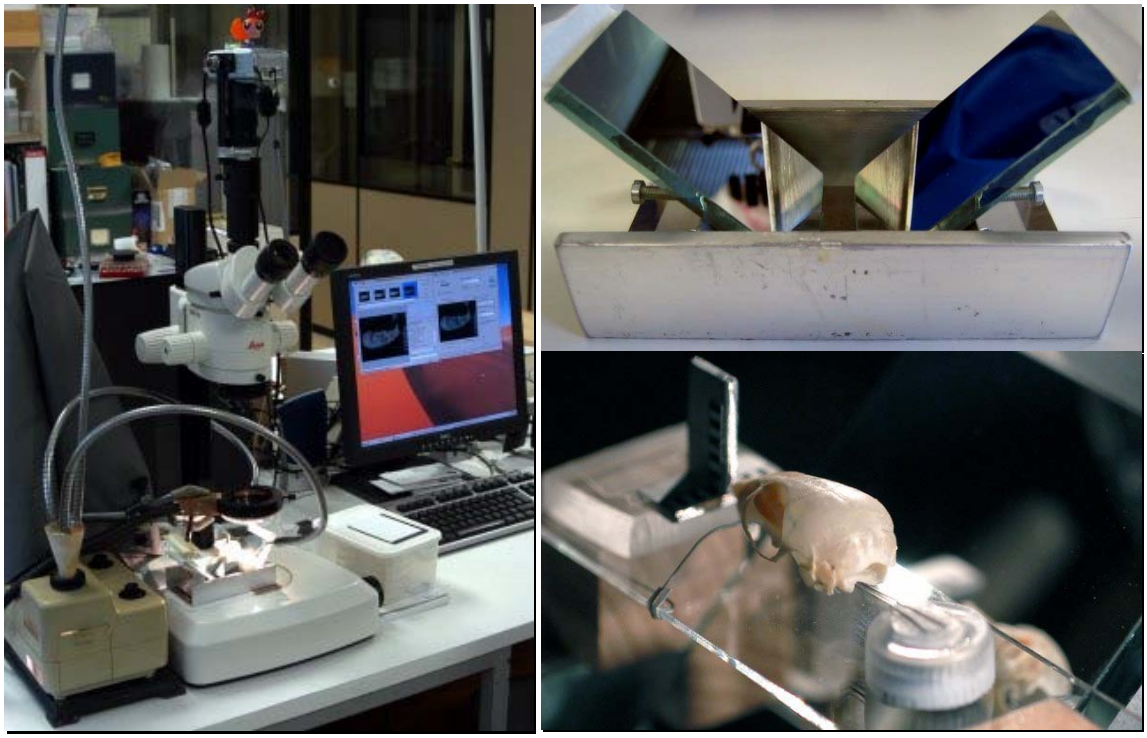


Fig. 17. Mirrored stage used for orthogonal images of centimetre-sized, delicate specimens without tilting. The stage is displaced under the stereomicroscope to obtain either direct images, single-reflection images (sides) or double-reflection images (underside).

When using stereomicroscopes for small specimens, additional care should be exercised in positioning the specimen and the stand, including mirrored stands. Typically, the optic path for the imager follows only one of the twin visual paths. This configuration may result in an aberrant image of the specimen if it is centred with respect to the microscope's axis, which lies between both optical paths; in fact, the slightly slanted image is actually one frame of a stereo pair. An axial carrier that can laterally move the specimen allows for unslanted images, by ensuring that the camera path axis (as opposed to the microscope's axis) is directly over the subject.

For scaling, we use an explicit T-shaped scale whose insertion point lies along the central axis of the device, which is where the specimen is placed for imaging. All images therefore include an image of the scale, which is landmarked.

Distortion

Planar (aspherical) lenses

Most photographic lenses are constructed of several single lens elements having spherical surfaces. Large aperture lenses made of spherical elements tend to suffer from spherical aberration, which may be noticeable as distortion near the edges of the image. Aspherical lenses correct spherical aberration because their free-curved surfaces are not spherical and can be designed so as to provide a basically planar, distortion-free field of view. Their production being much more difficult than standard spherical lenses, aspherical lenses are present in high-end optics.

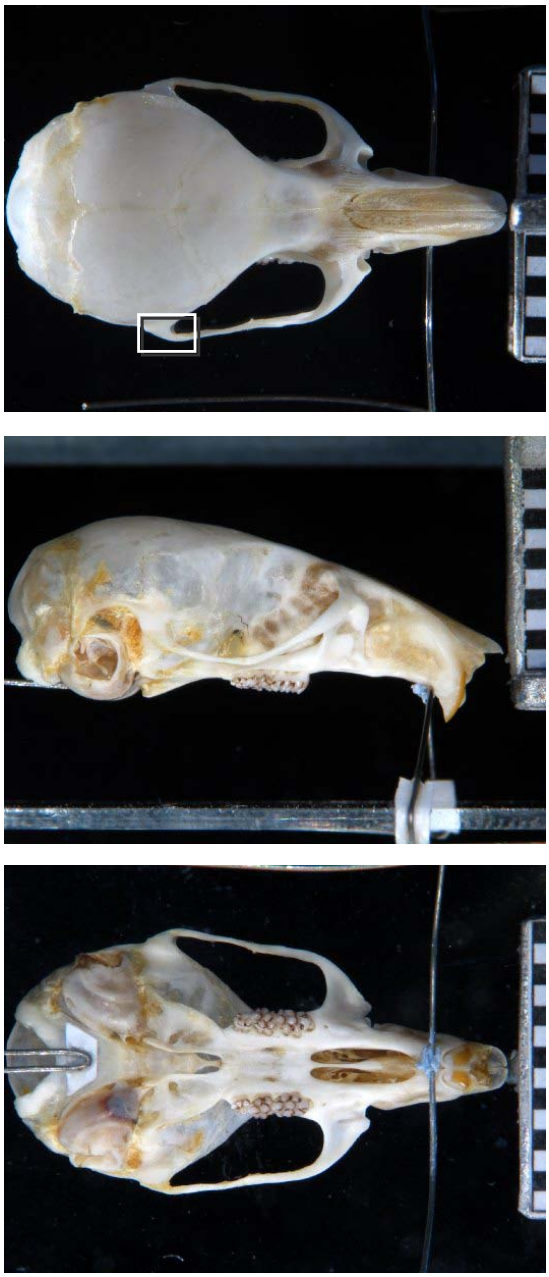


Fig. 18. Upper side, right side and underside of the skull of a wood mouse (*Apodemus sylvaticus*) specimen photographed on a mirrored stage. Scale appears at right. The square on first image marks the area for figure 20.

Chromatic aberration

Chromatic aberration appears because the refraction power of a medium is a function, among other things, of wavelength. Longer wavelengths (reddish) are refracted more than shorter wavelengths (bluish). Light entering the lens will be focused at slightly different focal planes (axial chromatic aberration) and/or at different positions on the sensor (transverse chromatic aberration), appearing as differential magnification according to wavelength. The overall effects are a reduction in sharpness and the appearance of

cyan and red fringes contouring the objects. Figure 19 shows an example of lateral chromatic aberration in a specimen.

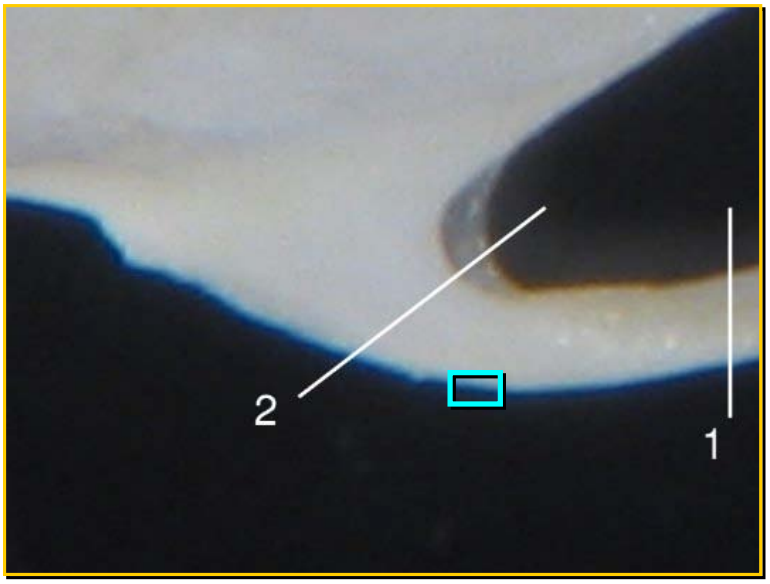


Fig. 19. Blow-up of the first frame of figure 18, showing chromatic aberration around the edge of the bone. Lines are experimental transects. Accurate measurements are needed to characterise this particular morphotype.

The typically small diameter of standard, consumer digital camera lenses, which translates into wide path angles, greatly contributes to this problem. Also, a microlens array on top of the sensor, which is often used to convey the light to the photosensitive element of each pixel, as well as the blooming effect (charge smearing to pixels in the vicinity of saturated ones), can be a source of chromatic aberration (BOCKAERT 2004).

The more a lens "bends" the beam paths, i.e. near the border or on a high-power lens, the greater the separation, or prism effect, of the light components. Longer focal lengths also increase chromatic aberration.

SLR digital cameras having large-diameter lenses for macroscopic specimens, or CCD cameras with no lens at all such as those used in some microscopes, can help reducing this effect in these cases. Also, stepping down the lens can also reduce the axial component of chromatic aberration.

Achromatic and apochromatic lenses

The chromatic aberration can be greatly corrected by using a set of two single lenses of different materials (i.e. of different refractive

power, such as optical glass and fluorite) found in achromatic lenses, so that their respective aberrations for two given wavelengths cancel each other. Three elements are used in apochromatic lenses, which basically correct all visible wavelengths to the maximum extent. Most top quality macro lenses and microscope objectives are apochromatic. However, almost any component that lies in the optical path and is not apochromatic itself (i.e., the objective lens of a compact camera being used on top of an apochromatic microscope objective lens, or an aspherical eyepiece lens) could eventually reintroduce chromatic aberration in the image seen by the sensor.

Pixel imprecision

Perhaps the single most important problem of chromatic aberration for morphometric-grade images is the way borders are lost. Contrasting borders are essential to adequately place landmarks. When chromatic aberration appears, borders become diffuse. Figure 20 shows the imprecision area of the border across the region marked in fig. 19 due to blue aberration. This means that the landmark will be placed in any pixel from a range, thus reducing accuracy. The effect is akin to defocusing.

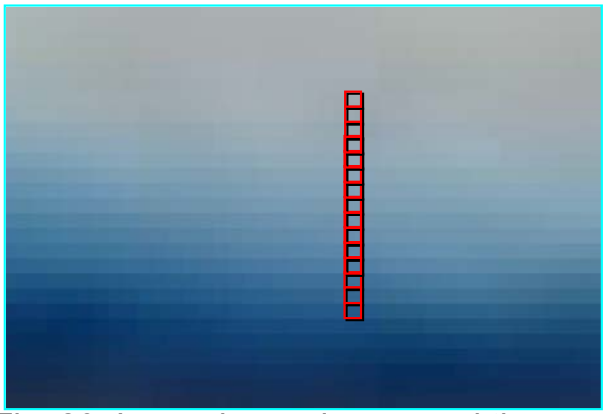


Fig. 20. Imprecise region around the edge marked in figure 19. The red plot marks the pixels any of which could represent the true edge.

Discarding some wavelengths can mitigate the problem. In our case, the analysis shows (as expected) that the green channel of the RGB image is more or less centred between blue and red channels. By selecting only the green component of the image, a greytone image is produced that is better defined in terms of contrast than its RGB counterpart. This may also be achieved in analog VC by capturing only the green signal, and is routinely done in microscopy. In digital cameras, using the green channel has a secondary advantage. Most CCDs have a colour pixel distribution in which for each blue or red pixel there are two green

pixels. Thus, green channel has twice the definition of either blue or green channel.

Figure 21 shows the RGB components of an image where colour is unimportant.

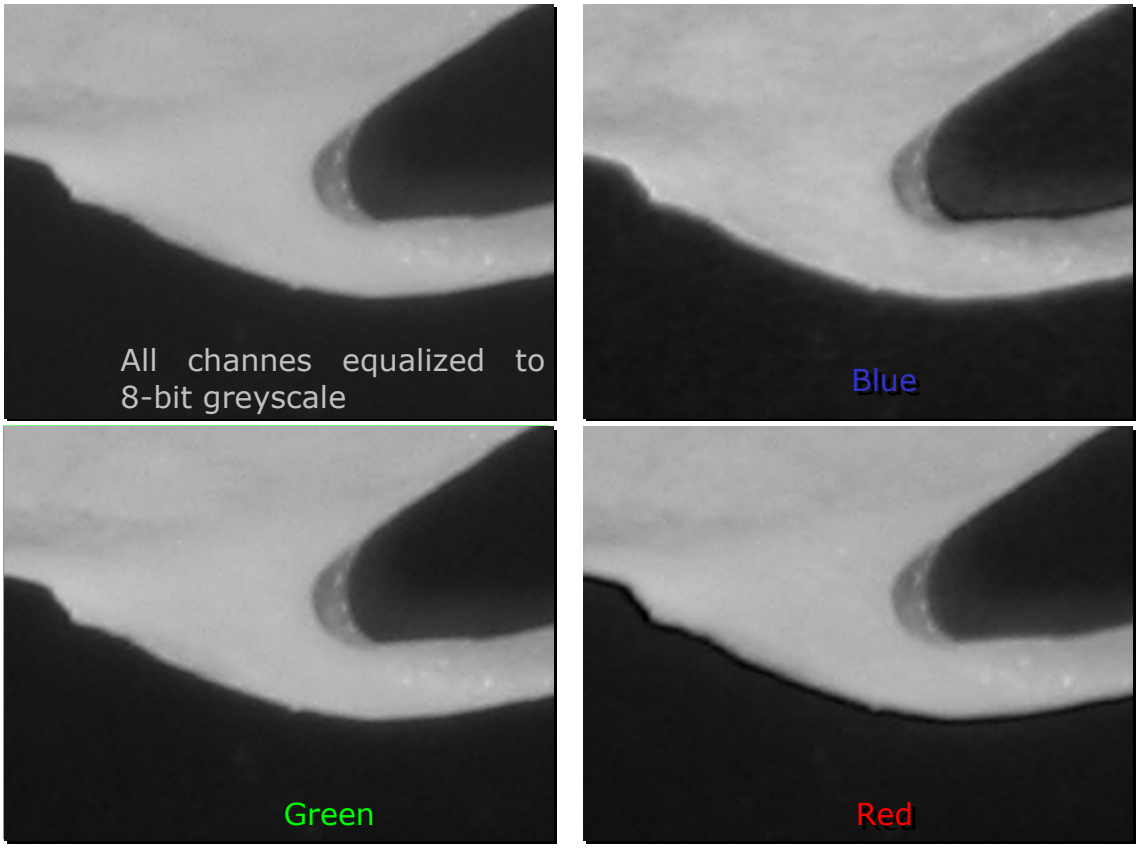


Fig. 21. Channel separation of the region of interest in the image of figure 16. The green channel is selected as the best defined.

Acknowledgements

We wish to thank JAVIER MADOZ, ANA AMÉZCUA MARTÍNEZ and NATALIA JÁUREGUI (UNAV) for kindly providing some of the pictures used in this chapter, and Dr. MALCOLM SCOBLE (NHM) for very valuable comments and suggestions.

References

BYTHELL, J. C., PAN, P. & LEE, J. (2001): Three-dimensional morphometric measurements of reef corals using underwater photogrammetry techniques. – *Coral Reefs,* **20**: 193-199.

FINK, W. L. (1990): Data acquisition for Morphometric Analysis in Systematic Biology. – In: ROHLF F.J. & BOOKSTEIN F.L. (eds.) (1990): *Proceedings of the Michigan Morphometrics Workshop. Special Publication No. 2.* – Ann Arbor, Michigan (The University of Michigan Museum of Zoology).

GARCÍA-VALDECASAS, A. (1996): Two-dimensional imaging. – In: MARCUS, L. F., CORTI, M., LOY, A., NAYLOR, G. J. P. & SLICE, D. E. (eds.) (1996): *Advances in Morphometrics, NATO ASI Series A: Life Sciences,* **284**: 71-81.

KREBS, C. J. (1999): *Ecological Methodology,* 2nd edition. – Benjamin Cummings.

LOOF, P. A. A. (1985): Taxonomic studies on the genus *Prodorylaimus* ANDRÀSSY, 1959 (Nematoda: Dorylaimina). – *Revue Nématol*., **8** (3): 193-213.

MADOZ, J., PARMENTER, S., LEUNDA, P. M., ARIÑO, A. H. & MIRANDA, R. (2003): Morphometric analysis of scales of the Owens River Basin *Gila bicolor* populations. – In: HENDRICKSON D. A. & FINDLEY L. T. (eds.) (2004): *Proceedings of the Desert Fishes Council*, **35**: 47-48.

MOLDVAY, A. (1981): *National Geographic Photographer's Field Guide*. – Washington (National Geographic Society).

RUSS, J. C. (1994): *The Image Processing Handbook*. – CRC Press.

RUSS, J. C. (2004): The Image Processing and Measurement Cookbook. http://www.reindeergraphics.com/tutorial/index.shtml [accessed 9.I.2005].

ZELDITCH, M. L., SWIDERSKI, D. L., SHEETS, H. D., & FINK, W. L. (2004): *Geometric morphometrics for biologists: a primer*. – Elsevier.

A Photographer's Viewpoint

HAROLD TAYLOR

Natural History Museum Image Resource, The Natural History Museum, Cromwell Road, London, SW7 5BD, UK. E-mail: h.taylor@nhm.ac.uk

Key words: Digital imaging, photographic technique, botany, entomology, zoology.

This chapter summarises techniques and equipment from a photographer's viewpoint. Equipment and image formats are reviewed briefly, and some practical entomological, zoological and botanical examples are then illustrated.

The mention or recommendation of commercial products in this chapter reflects solely the personal views of the author.

Technique

Mounting the camera on a copystand is the best arrangement for many types of subject. Using the camera vertically also allows the easiest means to 'level' the camera with a small spirit level (Maplins #BP61R), this goes someway to avoid image distortion and aids overall sharpness by keeping the camera 'film plane' parallel with the subject.

Fig. 1.
Spirit level.

A scale can be placed beside the object if required (Filofax # 173609). A good quality ruler, preferably white plastic (non shiny to avoid reflections) is a good choice.

Including a Kodak grey card at the start of the session (Kodak # 847-8174) will help with image editing later, this is only of use if the lighting is kept unchanged throughout the session.

Recognised as an industry standard reference for exposure evaluation and grading, the card is comprised of a large 18% neutral grey area bordered by 3% and 90% black and white patches to provide further reference for post capture grading (establishing a neutral colour balance, highlight and shadow within an image using Photoshop, especially important if the subject being photographed lacks these). The surface is also specially treated to minimize glare.

The Kodak separation guide, colour patch (Adorama # 1527654) does a similar job to the grey card but has a series of control colour patches included.

Fig. 2. Kodak grey scale and colour patch.

Fig. 3. Kodak grey card.

Tools of the Trade: Digital Cameras

A selection of SLRs available today:

Canon 300d (Rebel) 18 MB RGB, USB1.1
Canon 10d 18 MB RGB, USB1.1 (discontinued)
Canon 20d, 24 MB RGB, USB2
Canon 1Ds, 31 MB RGB, Firewire
Canon 1DS mkII, 47 MB RGB, Firewire & USB2
Nikon D1, 7.5 MB RGB, Firewire (discontinued)
Nikon D1X, 18 MB RGB, Firewire (discontinued)
Nikon D100, 18 MB RGB, USB
Nikon D70, 18 MB RGB, USB

Professional digital cameras now come in a wide variety of guises but generally they follow two main camps; Nikon or Canon lens fittings. Both manufacturers make superb quality camera bodies and suitable optics. Fuji also produces a fine body (S2 pro*)*, which takes Nikon fitting lenses.

The single-lens-reflex is the type I would recommend for scientific photography, though perfectly good results can be obtained with the lesser expensive viewfinder types such as the Olympus, Nikon Coolpix or Fuji Fine pix range.

The **SLR** (single-lens-reflex) allows viewing the image directly through the camera's taking lens, so 'what you see is what you get' (generally speaking). Digital cameras now use either a **CCD** (charge-coupled-device) or a **CMOS** (complimentary-metal-oxide-semiconductor) sensor to capture the image.

CMOS is the newer emerging technology, it is cheaper to manufacture and uses less power (if battery life is particularly important)

Figs. 4-6. Nikon D1 Canon 1Ds Canon 300d

For the last two years the author has used (and been perfectly happy) with a Nikon D1 digital camera. This has a CCD sensor and provides a 7.5 MB RGB file (JPG, TIF or RAW format), a file size perfectly useable for many applications as images tend to be used as part of a final larger plate for publication. It is only when a full size single image is required to be printed around A4 that the D1 sensor's shortcomings become evident (A4 size requires around 27 MB for reproduction at 300 dpi) though for inkjet report printing at 170-200 dpi, 7.5 MB is still sufficient.

I have now upgraded to a Canon 1Ds CMOS body, this delivers a file of around 31 MB RGB that is easily sufficient for A4 reproduction. An 8 MB file is also available.

All of the Pro range of SLR digital cameras are suitable for this kind of photography, though the Nikon and Canon range are my preferred recommendation: both have a superb lens range (especially the macro types); and both have software available for using the camera 'tethered' to a host computer workstation. This is my preferred workflow for scientific photography; the camera uses a tethered data cable transferring the captured image directly to a dedicated workstation, where the image appears seconds after the actual capture, ready for inspection.

Remote capture software allows a computer to remotely fire a digital camera connected to it. Two key benefits are that images can be stored directly onto the computer's hard disk and that images can be immediately previewed on the computer monitor instead of on the small LCD of the camera.

Nikon uses a 'firewire' interface (IEEE 1394) to connect the camera to the host workstation (which must also have a firewire card). Canon also use 'firewire' and some models also use USB2 which has a potentially faster data transfer rate (the latest 1Ds mkII has both). The cheaper Canon 300d uses an older (slower) USB1.1 interface. Most computer workstations now have USB interfaces embedded within their motherboards, few also have firewire. This can easily be added using an add-on card, Adaptec provide a good range.

Camera Supports

Holding the camera vertically, pointing downwards towards the subject is a very practical method of working for most scientific photography. Using a copy stand or studio style camera stand is the best means to accomplish this (figs. 7-9). The studio stand in particular is very flexible (assuming you have the space available) since the camera can be easily 'swung' out of the way allowing un-hindered access to the specimen work area.

Fig. 7. Cambo Studio stand.

Figs. 8-9. Kaiser Copy Stands.

Fig. 10. Nikon Multiphot.

Fig. 11. Zeiss Multiphot II.

Specialist Macro Set-ups

Nikon, Zeiss and Leitz have all manufactured specialised dedicated macro camera systems in years gone by (figs. 10-11). Unfortunately these beautifully crafted pieces of kit are no longer available and quite difficult to find second-hand. All had a set of macro lenses with dedicated illumination systems specifically designed for the job in hand. These set-ups are ideal for higher magnification work from around X1 upwards.

Macro Lenses

Macro lenses are specifically designed for working at close distances, although some can be used for normal photography, they really only give their best performance when shooting macro.

Nikon and Canon both produce excellent macro lenses, and Tamron and Sigma also produce top quality optics. The 50 mm focal length is a good all round choice, the longer 100/105 mm versions allow a greater working distance between the camera lens and subject. The Tamron 90 mm is also particularly good.

These lenses allow continuous focusing down to 1:1 (lifesize, X1) although tubes can be added to allow higher magnifications.

Nikon 60 mm | Nikon 55 mm | Sigma 50 mm | Tamron 90 mm | Canon 65 mm MPE

Fig. 12. Lenses.

Specialist Macro Lenses

Specialist macro lenses, which are designed for higher magnifications, can also be obtained. The Canon 65 mm MPE lens in

particular is exceptional (at a price) but can only be used within the X1-X5 magnification range (or higher with tubes)

If you can find them, Zeiss produced a superb set of macro lenses for their Ultraphot microscope system, these 'Luminars' still produce great images, each focal length was matched to a specific magnification range. To use these lenses you would need to fabricate an adaptor, although the four smaller focal lengths use the RMS microscope thread, adaptors are available for these.

Nikon also produced a superb macro set for their Multiphot system; 19, 35, 65, 120 mm focal lengths.

Fig. 13. The Zeiss Luminar range.

Lighting for Macrophotography

Coming from a photographic background I would not hesitate to recommend using flash (strobe) lighting rather than tungsten or fluorescent systems. Flash systems give out lots of daylight balanced consistent light in a fraction of a second (typically around 1/500th). Tungsten bulbs get very hot and exposures can be long, adding the possibility of some camera movement (camera shake). Fluorescent systems, although bright and while emitting less heat, have an incomplete spectrum. This results in some subject colours reproducing rather oddly.

Small Flash Units

The smaller low powered units from Metz, Sunpak, Sigma etc. can give perfectly good results, indeed for small-dedicated set-ups they may be adequate. However not having an in-built modelling light (focusing light) they can be fiddly to set-up.

Ringflash provides almost shadowless frontal lighting, which can be a little too flat in some situations.

Fig. 14.
Canon ringflash.

Studio Flash Units

The professional studio systems are very versatile and powerful, offering rugged dependability (at a price). Broncolor, Elinchrom, Bowens, Balcar all produce very capable flash systems.

Fig. 15. Broncolor studio flash.

Photography

Again, using the camera mounted vertically is perhaps the best option. A rack and pinion mounting for the camera is also recommended since this allows the whole camera and lens assembly to be racked up and down finely to achieve focus (rather than use the lens focusing ring which would change the magnification each time it is focused) Most copystands (certainly the Kaiser range) have this feature included.

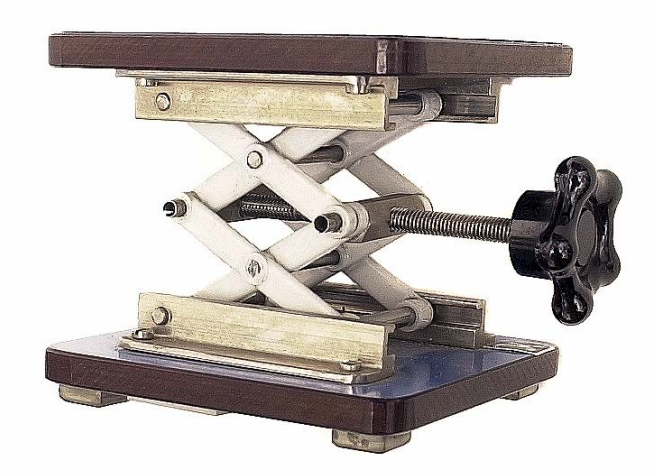


Fig. 16.
Laboratory jack. Useful for focusing.

Fig. 17. Kaiser X-Y stage. This allows focusing plus an East-West movement of the camera.

When photographing pinned specimens (in fact anything small) you have a choice when focusing; either move the camera and lens, or move the specimen to focus, this also allows you to maintain a fixed magnification on the camera lens.

The small laboratory jacks as supplied by the large scientific suppliers (Fisher Scientific) are useful for 'jacking' the specimen up or down to obtain focus. Something similar can be made from an old photographic enlarger (which can often be found very cheaply) though some work will be needed to turn the focusing mechanism into a workable unit.

Fig. 18. Old photographic enlarger

Fig. 19.
An old focusing mount from a photographic enlarger can be turned into a useful macro support stage for specimen focusing. Indeed the enlarger baseboard and column from the same unit can be equally used as a copystand with a little work.

Image Editing Software

Adobe Photoshop is the *'de facto'* standard in image editing software, used by design companies, photographers and repro houses around the world. It offers many tools and features indispensable when working with digital files or scans. Photoshop CS (or version 8) is the latest incarnation; you will need this if you intend working with the raw format.

The files we produce with a digital camera are 'raster' files or 'bitmap' files (not to be confused with Windows .bmp files). They are composed of pixels (many thousands/millions of them) which make up the final digital image.

Digital Camera File Formats

When saving files from the camera you can choose in which format to save the files.

JPEG or JPG (Joint Photographic Experts Group)

This general purpose format is useful when space or bandwidth restrictions apply. JPG is the Windows version, JPEG the Macintosh.

(JPG files are a compressed 'Lossy' format, this means the file is compressed at the expense of throwing away some image data. However the data that is removed is usually from parts of the image that will not be missed.) JPG can be saved as RGB, CMYK or GREYSCALE. The format is favoured for any hi-res images used in web pages. The degree of compression can usually be varied, with trade-offs between image quality and file size. Highly compressed images exhibit a strange 'blocky' texture within the image area when viewed in detail.

TIF or TIFF (Tagged Image File Format)
This format is most flexible format to supply a repro-house for printing. TIF is the Windows version, TIFF the Macintosh version. TIF files are cross-platform compatible, that is they can be read by all computer platforms and operating systems (Windows, Unix, Macintosh). TIF can be saved as RGB, CMYK or GREYSCALE. TIF is an uncompressed format; no data are lost so the file consumes a lot of space when saved to disk. The format is also widely supported by all image editing and DTP applications.

RAW (from an extract by VINCENT BOCKAERT, DPR: see http://www.123di.com/about_auth.htm)
The RAW data option when saving files with a digital camera offers many advantages (and some disadvantages). A RAW file provides all the capture data with no in-camera processing. This work flow gives the photographer greater control of the final image. RAW data is has 12 bits (compared with 8 bits for a JPEG) of data, so any adjustments made have more data available resulting in a final image with possibly smoother tonal values. Similarly highlight and shadow detail tend to be better preserved following and adjustments.

Drawbacks with this procedure are:

Speed, because of the extra data involved and required adjustments, the time taken to produce a finished file may be longer that would otherwise be the case.

Storage, again because of the extra data involved, the files are larger and take longer to open within the software (CS1 and

CS2, the newer Photoshop programmes both handle RAW data).

Dedicated software is required to view the RAW files. They cannot simply be viewed in Windows Explorer as can JPEGS.

RAW files can be thought of as digital negatives and kept as the original backup copies of the subject.

Printing Terms, Reproducing the Final Plate in a Publication

dpi (dots-per-inch)

DPI (dots-per-inch) is a printing term used to describe an output device resolution (inkjet printer, imagesetter) though the term can be used alongside PPI (pixels-per-inch) which is the on-screen pixel dimension since both describe the smallest viewing or printing element being used.

A digital file's size and dimensions often cause some confusion. It really just depends on how the figures are presented. If we take an A4 size scan and open it in Photoshop, checking the image size reveals

30 cm x 21 cm at 300 dpi resolution (25.1 MB)

We can change the resolution to 150 dpi (just acceptable for inkjet printing) and the dimensions jump to

60 cm x 42 cm at 150 dpi resolution (still 25.1 MB)

Or we can increase the resolution to 600 dpi (overkill for most purposes) and watch the dimensions shrink to;

15 cm x 10.5 cm at 600 dpi resolution (still 25.1 MB)

So we can take an image file and either use it stretched with bigger dimensions at a lower resolution, or shrunk with smaller dimensions at a higher resolution. The file size remains the same, we are merely changing the dimensions and the output resolution.

lpi (lines-per-inch) lpcm (lines-per-cm)

These are the units used for the printing screen ruling or screen frequency or halftone screen (three names for the same printing screen).

A screen frequency of 133 lpi or 150 lpi is generally regarded as a common screen used for good quality reproduction, though the Repro-House will use whatever screen that fits the job depending on what type of publication is being printed and the quality of paper stock they will be using.

As a general guide, digital files should be supplied at twice this screen. However, a lot less is acceptable depending upon image content. So a publication using a 150 lpi screen will require scans at 300 dpi, if the Repro house use a 133 lpi screen then the scans are rendered a little bigger (It is best to have them larger than smaller).

Botany

Herbarium sheets are the standard form in which most botanical type specimens are kept and studied. Two forms exist: bound sheets stitched into a large book or single sheets, commonly called 'Botany sheets'.

Type specimens are likely to be of vintage quality and invariably become brittle with age, handling can be troublesome and of course the utmost care must be taken to ensure any loose parts including labels are kept together.

Naturally ensure you have plenty of desk space available to arrange the sheets before and after photography. Use storage boxes and avoid stacking the sheets too high before and after use.

Bound volumes create special problems as care must be taken not to stress the binding too much when opening and preparing the pages, invariably some volumes can never be fully opened flat, all that can be done is ensure the page being photographed is as horizontal as possible. Book cradles commonly used in libraries are useful for support. You could also 'swivel' the mounted camera to

match the angle of the page being copied if that helps, this will slow down the whole procedure though.

Lighting herbarium sheets

I have found the best form of lighting for this material is overall diffuse using a 1 m-size softbox over your flash head, or bounce the flash from the ceiling (assuming your ceiling is 'whitish' in colour).

Fig. 20. Herbarium sheet samples.

This form of illumination provides near shadowless lighting, ideal for specimens that are stuck down upon a page mount. The camera is best mounted vertically using a copy stand or studio stand (which has the added advantage of being able to 'swing' the camera out of the way during page preparation).

Smaller specimens: Lichens

Lichen specimens and bryophytes are stored in small paper packets along with identifying labels and other data. Type material should be kept dry but other specimens may be sprayed with water to bring them back to life (in appearance at least). The change is quite dramatic once the excess water has been absorbed.

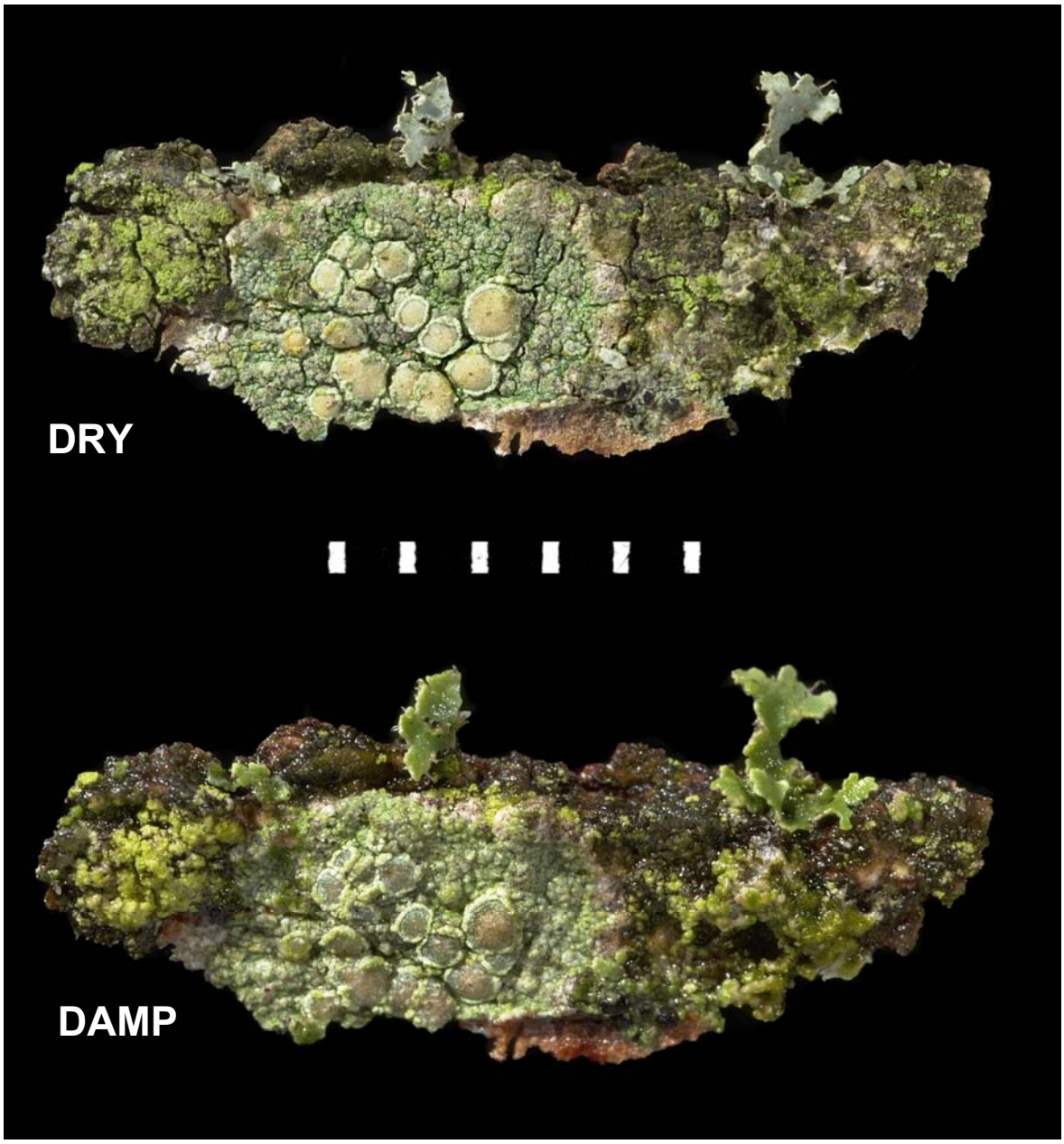


Fig. 21. Lichen sample. Above: dry. Below: damp. Again, black velvet has been used as a background for simplicity, lighting is a simple top-left flood (reflector head) which gives harder lighting than a softbox.

Entomology

Insects, butterflies and moths are almost universally pinned after collecting, except for the smaller insects where small card supports may be used, glued to the specimen. Because of the age and brittle nature of dried biological material, a steady hand and plenty of patience is required.

Fig. 22. Pinned butterflies ready for photography, with their labels removed and placed beside. Photographing the whole tray like this is useful as a reference print later. Canon 10d, Sigma 50 mm.

Wherever possible it is recommended that the curator handles the material, placing the specimen for photography and removing afterwards. Curators have the accumulated experience and deftness with forceps required. The tiny pinned labels are often removed for photography.

Fig. 23. A neutral grey background is a useful compromise when photographing many specimens which have a wide colour variety (very pale and very dark).

Pinned butterfly photography

The images below show the set-up used for a project to photograph approximately 500 pinned butterfly specimens for use in a publication/database. A Nikon D1 camera was fitted with a 60 mm macro Nikkor lens. The camera was tethered to a host computer running Nikon Capture to record and save the images. Specimens (both sides had to be photographed) where placed on a clear Perspex support. A Kodak grey card was used as the background. A scale was placed at the edge of the frame.

Figs. 24-25. Set-up for butterfly photography.

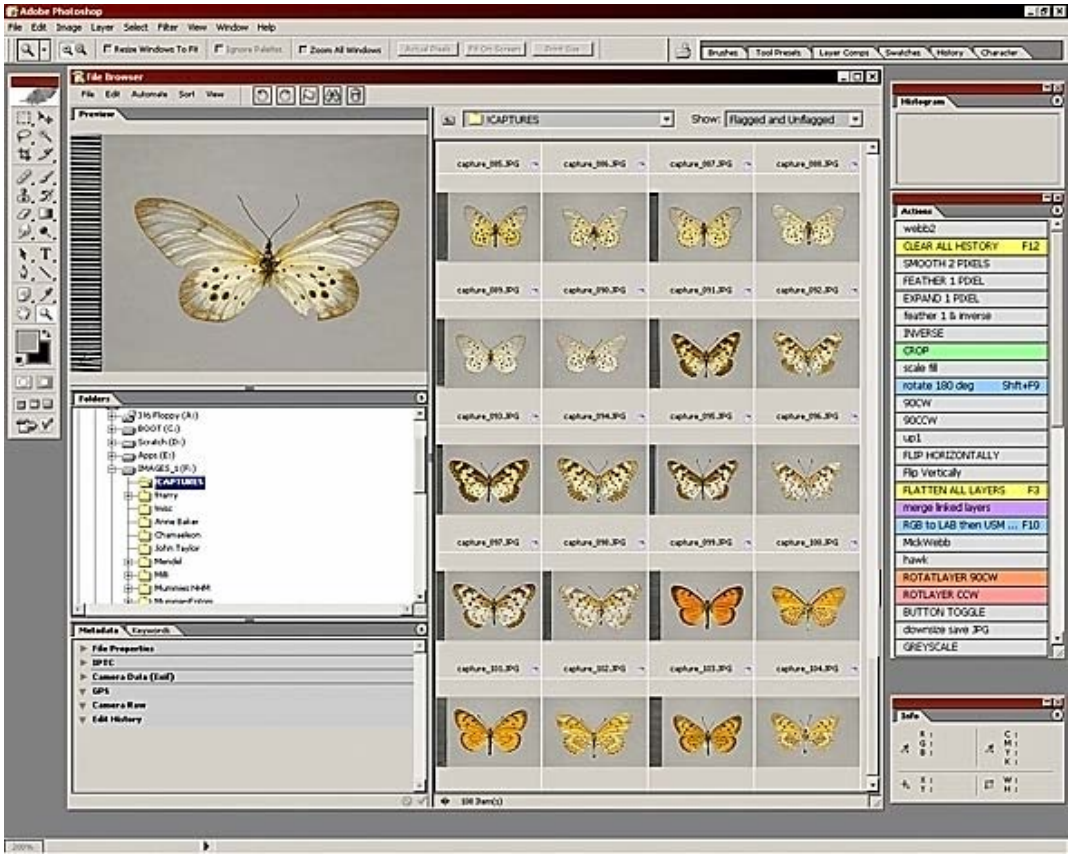


Fig. 26. Screendump from Photoshop file browser.

Zoology

Perhaps the most diverse collections are zoological, with dried bird and animal skins, stuffed trophy heads, mummified remains, fish and marine specimens in spirit, dry shells in cotton wool. These can all present their own particular handling problems.

Skins

It has been common practise in the past to impregnate arsenic dust (a toxin) in skins to repel insect pests. It is necessary always to wash one's hands thoroughly and avoid touching the face and mouth while handling such material.

Fig. 27. A single birdskin has been photographed from both lateral sides and combined together as a single plate. Separate views of the labels have also been added. Data included on labels frequently needs to be included for research papers. Simple soft top-left lighting, Canon 10d, Sigma 50 mm

Bird skins are commonly mounted on a wooden support (rather like a lollypop). The skin is supported internally using suitable stuffing to fill it out and help retain the original shape and character. In this form the skins are easily studied and handled.

Fig. 28.

An ancient type specimen of a young brown bear.

The specimen took approximately 30 mins to dry the surface fur (all that was needed). Lighting was a simple large softbox.

Canon 10d, Sigma 50 mm.

Spirit specimens

Many types are stored in Isopropyl alcohol (spirit). It offers a good compromise as a preservative for biological material. An 80% concentration is a common alcohol strength. The biggest problem perhaps is the risk of fire and it goes without saying that fire precautions must be taken when photographing such material.

Once again the curator should really handle the specimens, as they will need some form of drying out before they can be photographed.

Although you can attempt to photograph such material submerged in the alcohol, the idea of large quantities of spirit lying around exposed is not to be recommended.

Snails, Marine Specimens

Small specimens can be easily photographed under spirit (the curator may request this if the specimen could be damaged by drying out). A small glass or Perspex box is ideal to hold the spirit, ideally one with flat sides to avoid lighting distortions. A small sheet of glass is best kept to cover the full box when not being used to help stop evaporation.

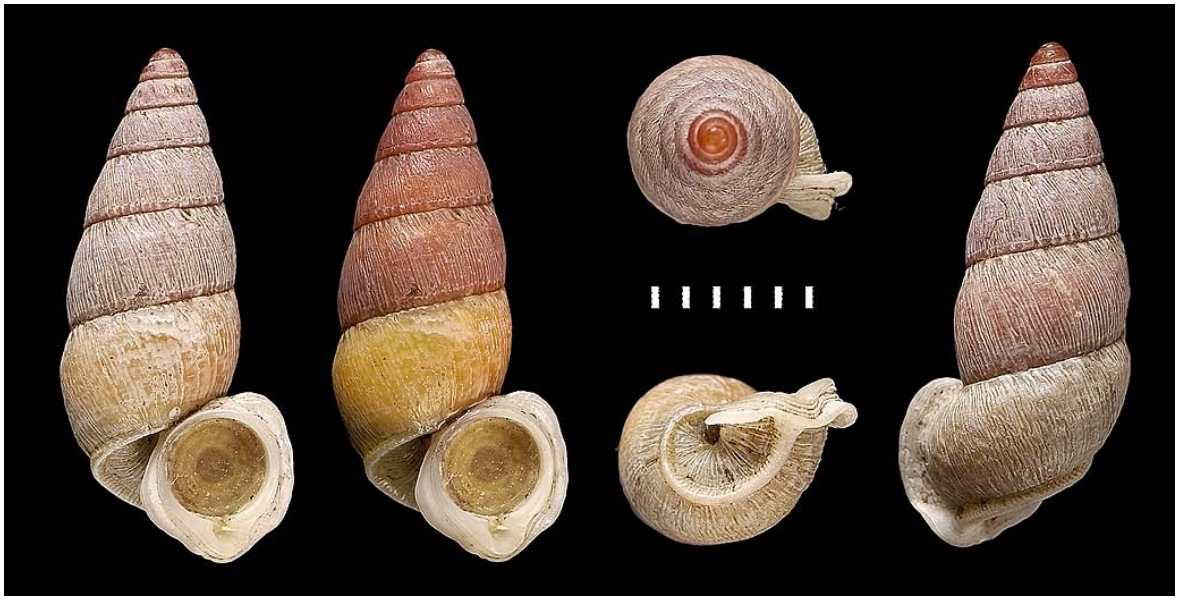


Fig. 29. Sri Lankan Snail shells, dried and photographed, a composite plate of six images. Simple soft top-left lighting. The specimens where mounted on pins, a black velvet background was used. Nikon D1, 60 mm Nikkor.

When photographing small specimens such as these snail shells, mounting them on a stalk (metal nail or pin) using a piece of 'sticky material' such as plasticene or blue-tak, works very well. If one has access to engineering facilities it is worthwhile having a 'support set' made such as a series of different sized metal pins (3 and 5 mm diameter) with a small heavy base with holes drilled to matching sizes.

There are several 'sticky' mountants available; Plasticene, BlueTak, Proptac and Stickywax (the last two are photographic prop sundries available from Calumet Photo).

Fig. 30.
StickyWax by Condor Foto.

Fig. 31. Proptac by Calumet.

This method gives an image 'floating in space' when combined with a black velvet background.

Velvet, such as used for this image, is perfect for obtaining a 'black void'. The short pile absorbs light very effectively.

Figs. 32-33. Snails on black velvet.

Bones and Skulls

Skulls can be difficult to support because of their irregular shape. The use of a small sand tray can be very effective for supporting such objects. Here again, black velvet is a perfect background effectively hiding any background shadows. The very fine white sand used for children's playground sandpits provides the best support. Just empty some into a shallow tray and cover it with the velvet. The specimen can be pressed gently into the sand, which will then provide the necessary support.

Fig. 34. Quite often a very simple arrangement will suffice (and provide a pleasing composition). Here the bones where simply placed on a sheet of white laminate as a background. Canon 10d, Sigma 50 mm.

Fig. 35. Another simple arrangement, a mounted chameleon already fixed to a branch. Moving the white background far enough behind the specimen ensures that no shadows are cast. Canon 1Ds, Sigma 50 mm.

Digital Imaging of Prokaryotes for Taxonomic Purposes

DAGMAR FRITZE

DSMZ - Deutsche Sammlung von Mikroorganismen und Zellkulturen, Mascheroder Weg 1b, 38124 Braunschweig, Germany. E-mail: dfr@dsmz.de

Key words: Digital imaging, prokaryotes, micro- and macrophotography.

Introduction

Two technical aspects should be mentioned prior to addressing particular details. First: the size of the objects in question usually lies **between 0.2 and 1 µm**, whereas some of the larger prokaryotes can reach up to 3 µm in diameter. Second: in bacterial and archaeal taxonomy we are working in most cases with the **living organism**, represented by individual cells (visualized microscopically) or accumulations of cells, colonies, which develop on the surface of growth media and which are viewed under low magnification with a stereomicroscope.

Both these observations underline the need to develop and use high quality techniques and devices without which research in this area would not be possible.

Identification and characterization of bacterial and archaeal species is based today on polyphasic approaches using various genetic and phenotypic methodologies, including micro- and macroscopically captured morphological information. Being one of the 'traditional' techniques, a large body of literature has been published over the centuries on the visualization of microscopic biological entities. Outputs are found particularly in textbooks for university classes or through documentation from producers of microscopes themselves (see e.g. KAPITZA 1997).

The focus of this chapter will be on light microscope techniques as

applied in bacteriology with emphasis on phase contrast. This is because for taxonomical purposes features like length, width and shape of **living** cells, intracellular bodies, extracellularly deposited material or mobility of cells is important. Such features are revealed most effectively through phase contrast microscopy. Fixation of specimen and staining of cells or cell components, which is a typical procedure in medical applications is usually not recommended as it may alter cell shape and size, and uneven staining of cytoplasma may lead to false interpretations. One counter example is the classical technique of Gram-staining which is widely applied and uses indeed fixing, tanning and staining, but this method relates to the visualization of the retention of the staining agent and not to the shape of the cell. Other contrasting techniques in transmitted light and reflected light, in fluorescence microscopy, and in electronic techniques like SEM and TEM, have their applications in various research areas and less in taxonomy.

While the hand drawings and paintings of the early days of bacteriology have long been replaced by less laborious microphotography, more regularly employed microphotography was made possible since the era of e.g. ‘autodeveloping’ polaroid pictures, which were further time saving. However, full routine pictorial documentation was made feasible only with the introduction of digital processing of microphotographs. This development exactly reflects how the various strains and cultures held at DSMZ are, to a quite differing level, optically documented.

Taxa Group: Bacteria/Archaea (prokaryotes)

Institution: German Collection of Microorganisms and Cell Cultures (Deutsche Sammlung von Mikroorganismen und Zellkulturen), Mascheroder Weg 1b, 38124 Braunschweig, Germany.

Person Responsible: no specific individual charged; all curators plus some technicians, each responsible for a specific group of organisms.

Number of Person Hours per year devoted to imaging: 500 - 1000 h/per year.

Microorganisms: some details on the smallest of living organisms (adapted from WFCC 1996)

Without microbial activity, life on earth would not be possible. The diversity of microorganisms plays a major role in maintaining the biosphere and provides a vast resource for humankind.

Microorganisms are to be found in every ecological niche, performing recycling roles and interacting with other living forms in ways that we are only now beginning to understand. Their total numbers are unknown and their study in-situ is difficult.

They cannot be accurately enumerated; estimation of a 'base line' for inventorying purposes is not possible.

They may be transferred across borders by wind, water, the movement of animals or humans.

They are unlikely to be depleted by sampling; however, the loss of hosts could lead to the loss of dependent microbial species; examples are known for fungi.

Within a species, isolates/strains may show slight genetic variation, also e.g. depending on sampling time; thus individual isolates are of considerable significance in terms of genetic expression.

Microorganisms that are isolated from natural or manmade environments are typically conserved in culture collections, be it public service collections or specialized research collections.

For the purposes of describing new taxa and to act as essential reference standards for future study and use, the conservation and accessibility of type strains (those on which the taxonomic description is based) and other representative isolates, is fundamental.

Furthermore, the ex-situ conservation of all isolated micro-organisms that have been studied and reported in the scientific literature, would equally be important if science is to progress. The uncertainties associated with re-isolation of strains underline the need for deposit in a culture collection, where conservation skills and ready accessibility to specimen tagged with a conserved and unique reference number are provided.

Without this, scientists would constantly need to conduct the skilled and expensive processes of isolation, characterization and identification at the start of each new study.

Careful documentation of all conserved material is essential and the availability of such data is encouraged and supported by the WFCC.

Number of Images captured and stored each year: ~ 1500 captured; of these about 50% stored; includes also non-type material

Purpose of Image:
Target Audience

(1) Morphology is one out of many features that needs to be recorded for the description of bacterial/archaeal cultures.
(2) Together with other phenotypic or genotypic details it forms a mosaic of information useful at all levels of taxonomy and for a variety of diagnostic purposes and serves to characterize each specimen.
(3) The main purpose at present is for in-house documentation, and, to a smaller extent, for publications. In contrast to photographs for in-house documentation, photographs for scientific publications generally need to be of a higher quality, with all taxonomically or otherwise important details readily identifiable and properly arranged.
(4) Customers of microbial collections (research institutions, universities, schools, quality control laboratories, industries, etc.) would benefit from the availability of such information, which is normally only found in the original publications, if at all.

Form (e.g. Website, GBIF portal, paper publications)

Older pictures such as negative film and/or positive paper photographs; more recent pictures stored as files on hard disk and/or CD, or as printouts for documentation; no data base as yet.

Wish-list:

It would be useful to be able to
(1) scan digitally old hard copy microphotographs,
(2) produce images of all type material (type strains) macro- and microscopically (one to several from each sample, colonial details, cell shape, cell details, etc.),
(3) produce them (representatives) in sufficient quality to suit eventual publication purposes (print, on-line),
(4) produce an on-line searchable data base connected to the DSMZ internet catalogue of strains.

Capturing:
Handling Specimens for microphotography:

Before taking pictures, a bacterial culture needs to be cultivated to an optimal physiological state. This culture then needs to be processed in a way that allows the preparation of a specimen in a form suitable for taking high quality pictures (e.g. suspension of appropriate density, etc.). It also needs to be taken into consideration that some sensitive bacteria obligately depend on certain culture conditions (e.g. pH, salt concentration, gas atmosphere, temperature) and may require the respective conditions also for mounting procedures. Spore forming organisms often develop spores only as a response to certain culture conditions. Even then, locating a suitable section of the specimen for optimal photographic documentation may prove time consuming.

Mounting Specimens / Specimen Preparation
For light microscopy:
Standard Routine Methods

Usually non-fixed, non-stained cell suspensions on slides for use under the microscope; for higher quality, agar-coated slides are used (see text box for description of method).

Some typical examples for microscopic appearance of bacteria are given in figs. 1-4 in terms of shape and size of cells and quality of capture; halos around structures are artefacts. Fig. 1 shows the unsatisfactory result of a simple aqueous mount. Usually, the thin water layer between the slide and the cover glass still allows free movement of the cells (resp. too many cells do not lie in focus) which makes observation largely impossible. Figs. 2-4 show the results obtained using the agar-coated slide technique. Fig. 2 represents the final image as it should be used for the data base. The information necessary for scientific interpretation, such as accession number of organism, magnification, scale, culture conditions and date of capture, is given within the picture. Details on culture conditions like growth temperature, length of incubation and composition of growth medium need to be documented because bacterial cell morphology is culture dependent.

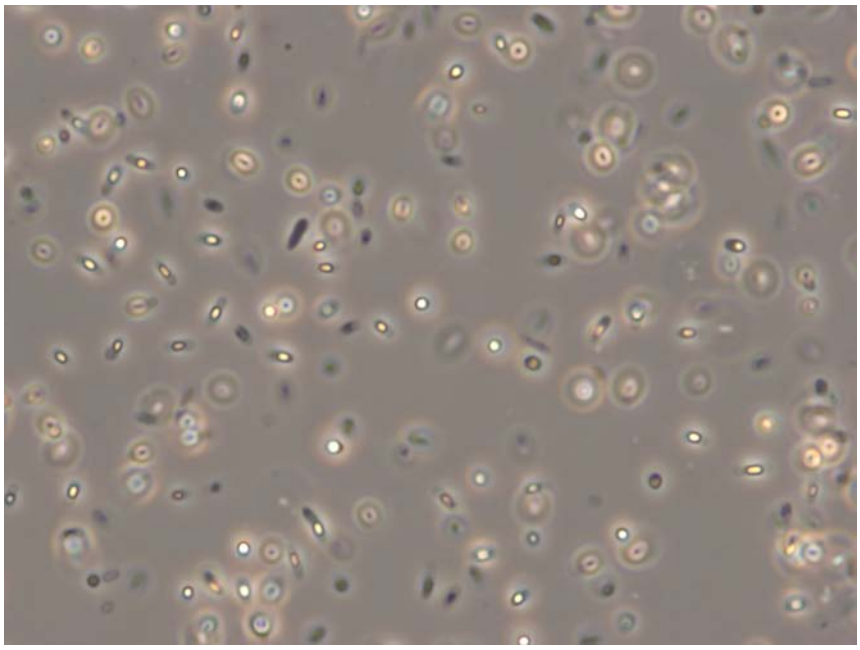


Fig. 1. Microscopic view of a simple aqueous mount of an endo-spore forming bacterial culture: individual vegetative cells, spore bearing cells and free spores; phase contrast oil immersion objective (x100); the thin water film between microscopic slide and cover glass still allows free movement and thus most cells are not in focus.

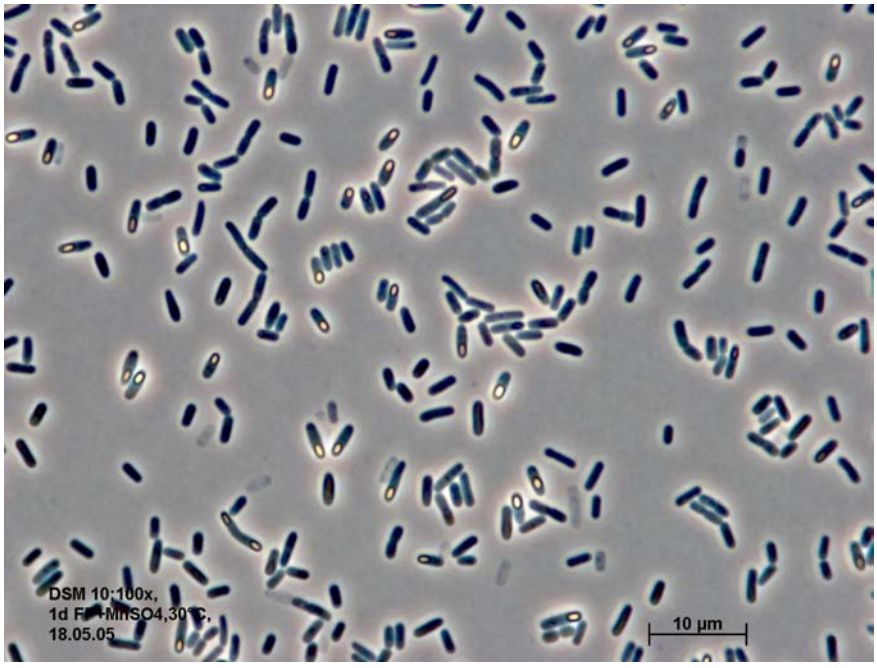


Fig. 2. Microscopic view of *Bacillus subtilis* subsp. *subtilis* DSM 10^T (type strain of the species) using the agar-coated slide technique; individual vegetative cells, spore bearing cells and some empty cells are seen using phase contrast oil immersion objective Zeiss Neofluar (x100); cells are mostly in focus as residual water is taken up by agar and cells are immobilized between agar and cover glass.

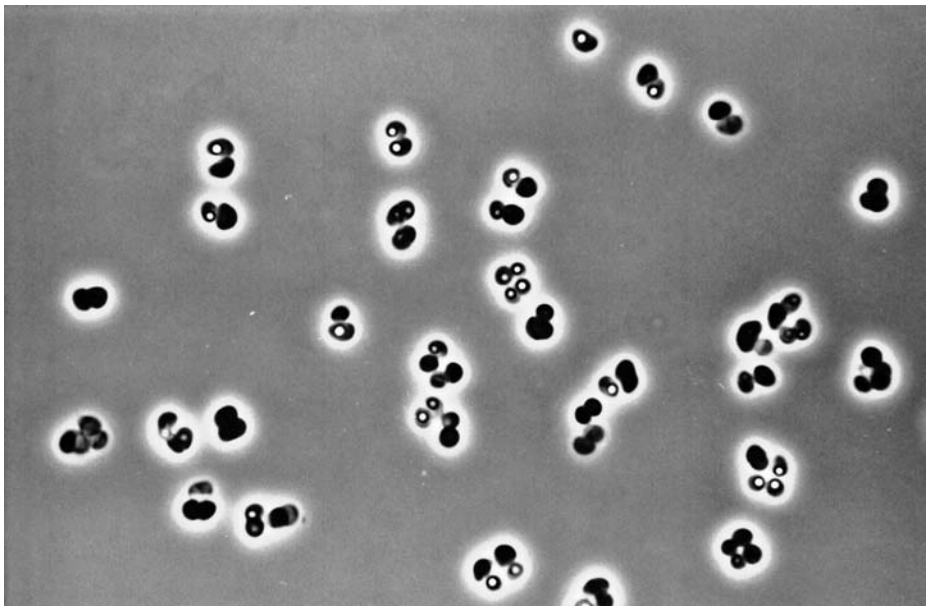


Fig. 3. *Halobacillus* (ex *Sporosarcina*) *halophilus* DSM 2266^T , a tetrads (sarcinae) forming coccus; type strain of the species. Vegetative cells and spore bearing cells, oil immersion objective (x100).

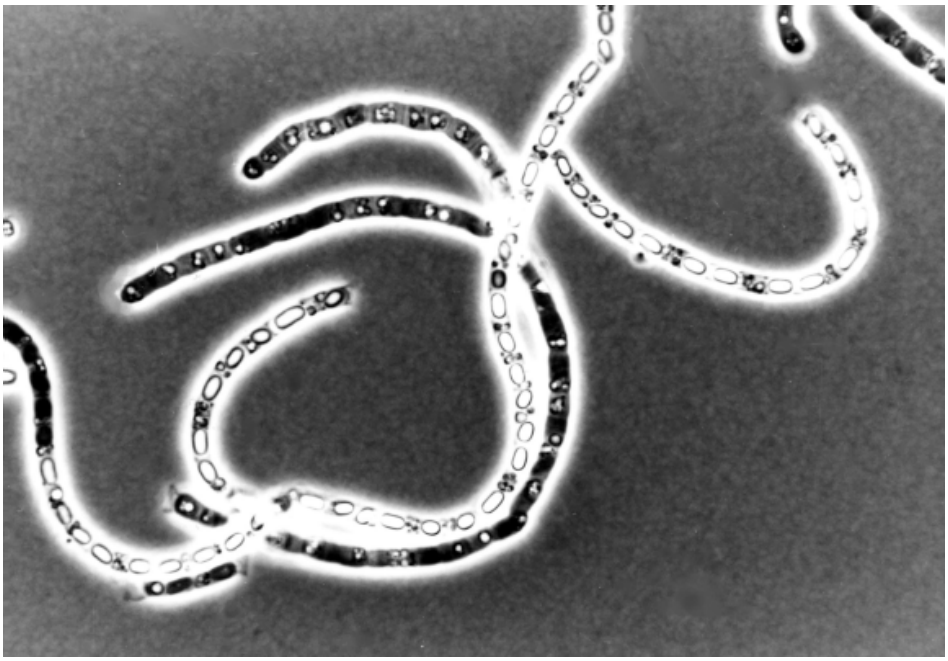


Fig. 4. *Bacillus funiculus* DSM 15141^T**,** filaments consisting of vegetative and sporulated cells containing additional deposits; type strain of the species; oil immersion objective (x100).

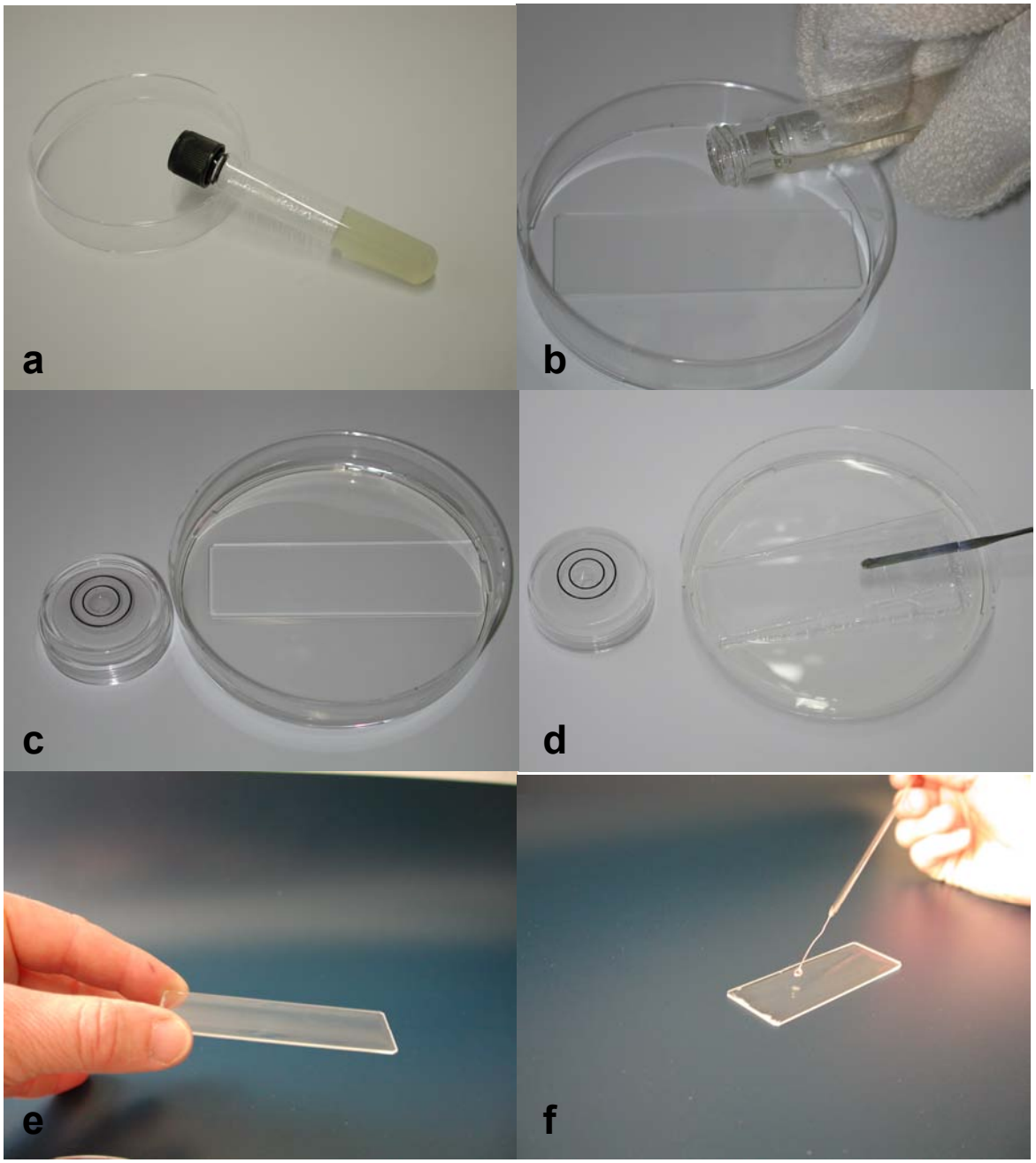


Fig. 5. Agar-Coated Microscopic Slide Technique.

Microscopical Examination and Microphotography of Bacteria

(according to CLAUS & BERKELEY, 1986); see also home page of the World Federation for Culture Collections under http://www.wfcc.info/tis_menu.html

The microscopic appearance of vegetative or spore-bearing cells is best observed with freshly prepared cell suspensions under a phase contrast microscope. Staining and fixing of cells should be avoided as this often has a strong influence on shape and general appearance of the cells. Vegetative cells will appear darkly contrasted, while mature spores will appear as bright white bodies (round, ellipsoid, cylindrical, banana shaped). Other cell enclosures (crystals, parasporal bodies, fat globules etc) appear in varying shades and forms. Only in cases the nature of the inclusions is not absolutely clear, staining may be used in addition. Objective magnifications x40 and x100 (oil immersion) should be used.

It should be borne in mind to mount organisms in a suspension fluid of similar ionic strength as the culture medium. Considerable differences in ionic strength may have an effect on cell shape or may even lead to rapid lysis of cells. This is especially important for halophilic organisms.

For high quality microscopic examination, slides freshly coated with a thin (0.5 mm) layer of 2 % water agar (purified agar) should be used (fig. 5 a-f). To coat the slides, place a slide in a 90 mm Petri dish, add 5 ml of hot molten agar and gently but quickly rotate the plate on the bench so that the whole of the bottom of the dish and the slide are covered. Close the lid and leave the plate in an absolutely horizontal place to settle. Wait for a few hours until use to enable evaporation of excessive humidity. Then cut out the coated slide and clean the lower side. Add a small drop of an evenly turbid suspension of the bacterium to be examined onto the surface of the agar, place a cover glass on the preparation and examine.

The advantage of this method lies in the complete immobilization of cells as all excessive mounting liquid is absorbed by the agar. Additionally, all cells are brought into focus through this effect.

Other, not routinely applied, but useful techniques for microscopic examination:

Gelatine coated slides

Gelatine-agar coated slides (using agar medium supplemented with 14-30% gelatine) provide a medium of high refractivity which helps to visualize internal details of bacterial cells (fig. 6). The higher concentrations are suitable for Gram-positive bacteria, whereas Gram-negative bacteria may need lower concentrations. The problem of halos obscuring the boundaries of especially the larger bacterial cells can be reduced and – dependant on the concentration – even a kind of negative contrast may be obtained (fig. 7). Best results need to be determined by experiment for each kind of organism and the detail to be observed. A less precise method to reduce the halos is to use slides covered with a thin film of 15-20 % gelatine and dried. These are then ready to be used for wet mounts (ROBINOW 1975). Polyvinylpyrrolidone, used instead of gelatine, provides similar results and makes work independent from the low melting point of gelatine (SCHAECHTER et al. 1962).

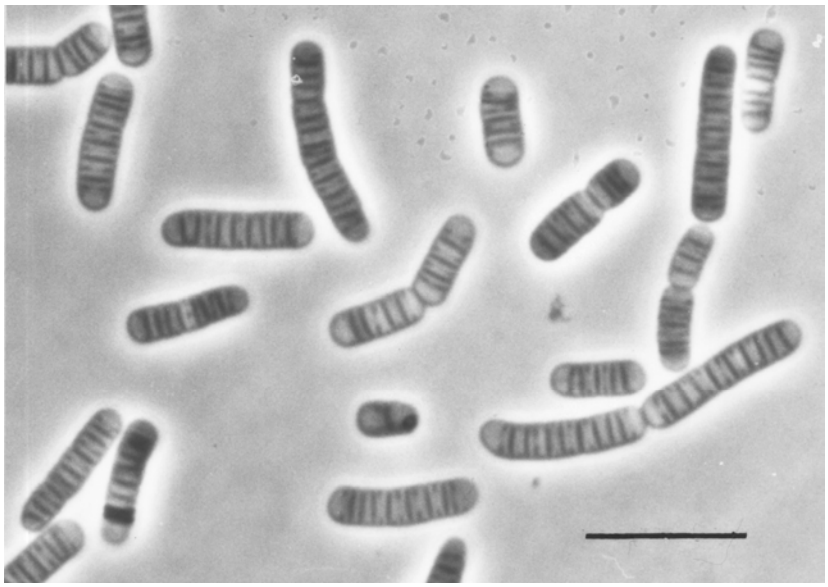


Fig. 6. Trichomes of *Caryophanon latum*, strain Lenglern. Phase contrast, gelatine coated slide (5% w/v, dried), bar 10µm. Note internal structure of trichomes showing cross walls between cells and growing septa at various stages of closure within cells.

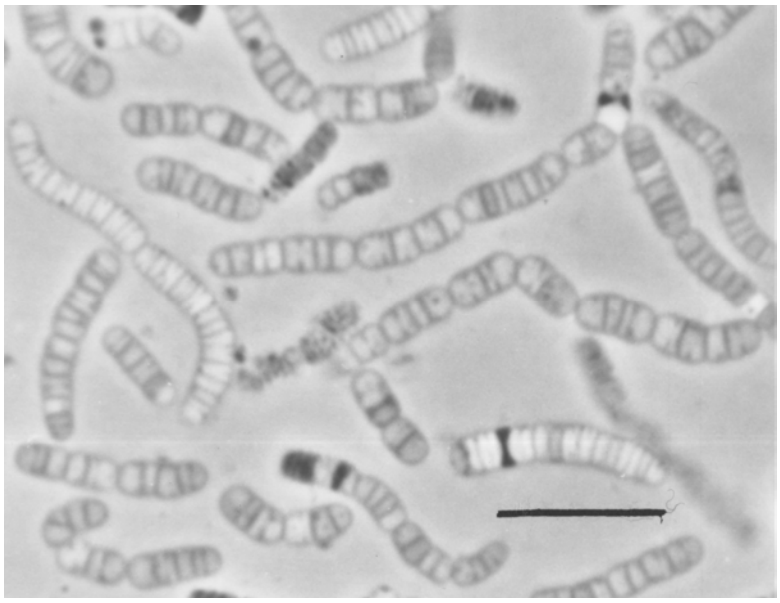


Fig. 7. Trichomes of *Caryophanon latum*, strain Lenglern. Phase contrast, gelatine coated slide (20% w/v), bar 10 µm. Note 'negative contrast' and absence of halos.

Microcultures

In special cases, it might be desirable or necessary to follow the growth behaviour of single bacterial cells over longer times in order to obtain insights in specific morphological changes during their growth cycle. Agar- or gelatine-coated slides inoculated with microbial cells and incubated under suitable conditions can provide adequate recordable information during longer observations (HEINZEN & ENSIGN 1975). Note that the latter method is not applicable with gelatine degrading organisms. Maintaining favourable growth conditions (e.g. pH, gas atmosphere, water content) is crucial.

Electron microscopy (SEM/TEM): the DSMZ uses external services (example fig. 8); for methodology extensive published literature is available (e.g. SOMMERVILLE & SCHEER 1987 or HAJIBAGHERI 1999).

Capture Device (Camera, Scanner, Lenses, etc.)
Photomicroscope (Zeiss Axioscope), digital camera (Axiocam MRc Rev.2, Jenoptik ProgResC10+) (fig. 9).
Stereomicroscope (Zeiss Stemi 2000C), digital camera (axiocam MRc Rev.2) (fig. 9).

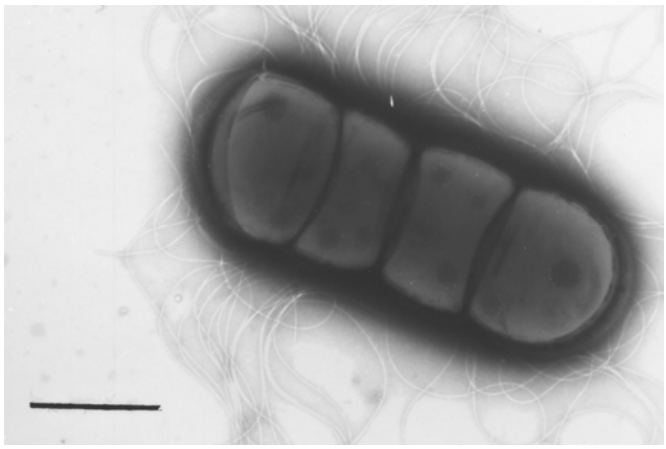


Fig. 8. Trichome of *Caryophanon tenue*, strain Uelzen with flagella. TEM, negative staining, uranylacetate, bar 2 µm. (Courtesy of Dr. M .MADCOUR, Univ. Göttingen)

Fig. 9. Microscope Zeiss Axioscop, x40 objective (Plan-NEOFLUAR, Ph2) and x100 oil-immersion objective (Plan-NEOFLUAR, Ph3), bright field, phase contrast; digital camera Axiocam MRc using software Axiovision. Stereomicroscope Zeiss Stemi 2000C, x0,8 - > x2 magnification, reflected and transmitted light; same camera as for microscope.

Viewing Cultures under Low Magnification: Stereomicroscopy

To judge bacterial cultures for purity, they are usually inoculated onto solid media, on the surfaces of which they develop into accumulations of cells, called colonies. These can be differentiated according to structure of rim, interior and surface, and according to shape, colour and opaqueness. Some examples are given in figs. 10-14. For interpretation of colours it is recommended to use e.g. RAL reference cards for colours. Similar like the microscopical appearance, morphological characteristics of colonies depend from culture conditions, which therefore need to be recorded in all detail (e.g. composition of medium, temperature of incubation, age). When judging and comparing colony morphology, it should be borne in mind that the size of colonies often depends from the volume of growth medium on which they develop and from the number of colonies that develop (SHAPIRO 1992).
Though today, characterization and identification of bacterial cultures relies strongly on genetic methods, the feature of colonial appearance is a useful recognition tool for daily laboratory work, especially when working with large numbers of cultures.

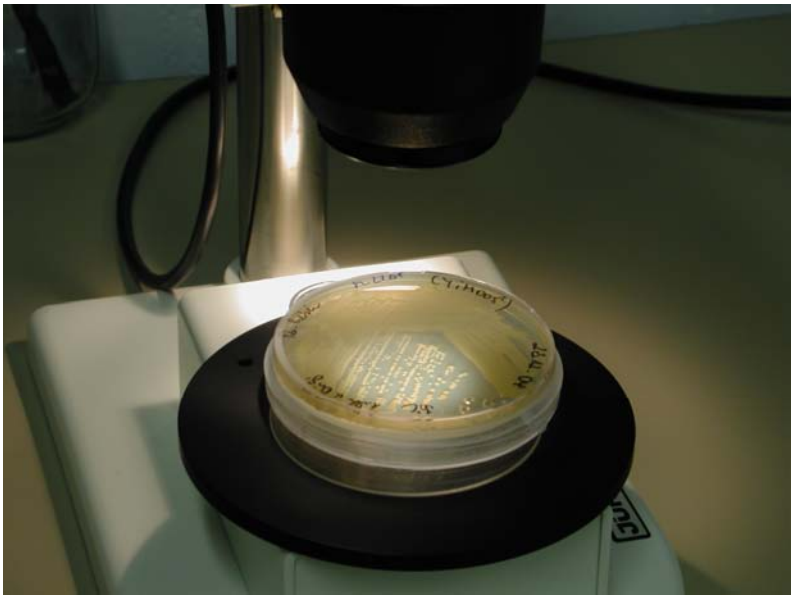


Fig. 10. Stereomicroscope: individual colonies and confluent bacterial growth in Petri-dish (ø 9 cm) containing nutrient medium, viewed from bottom side of the agar dish, transmitted light.

Fig. 11. Bacterial colonies and confluent growth on nutrient medium in Petri dish (ø 9 cm); viewed through Stereomicroscope (Petri Dish upside down), reflected light. Details of culture conditions are written directly onto culture vessel during actual work with the culture.

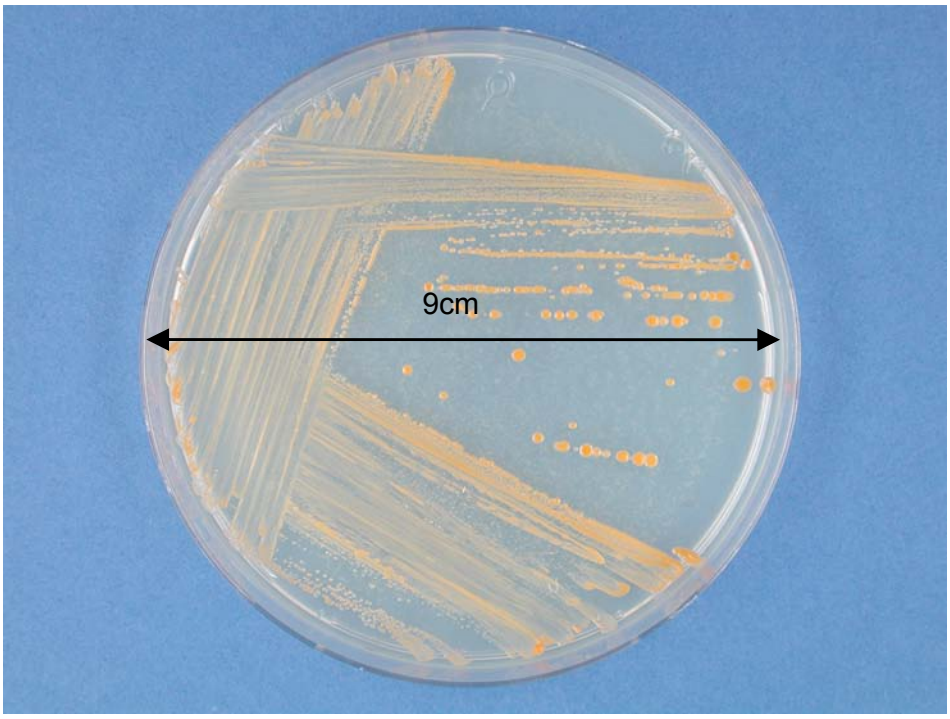
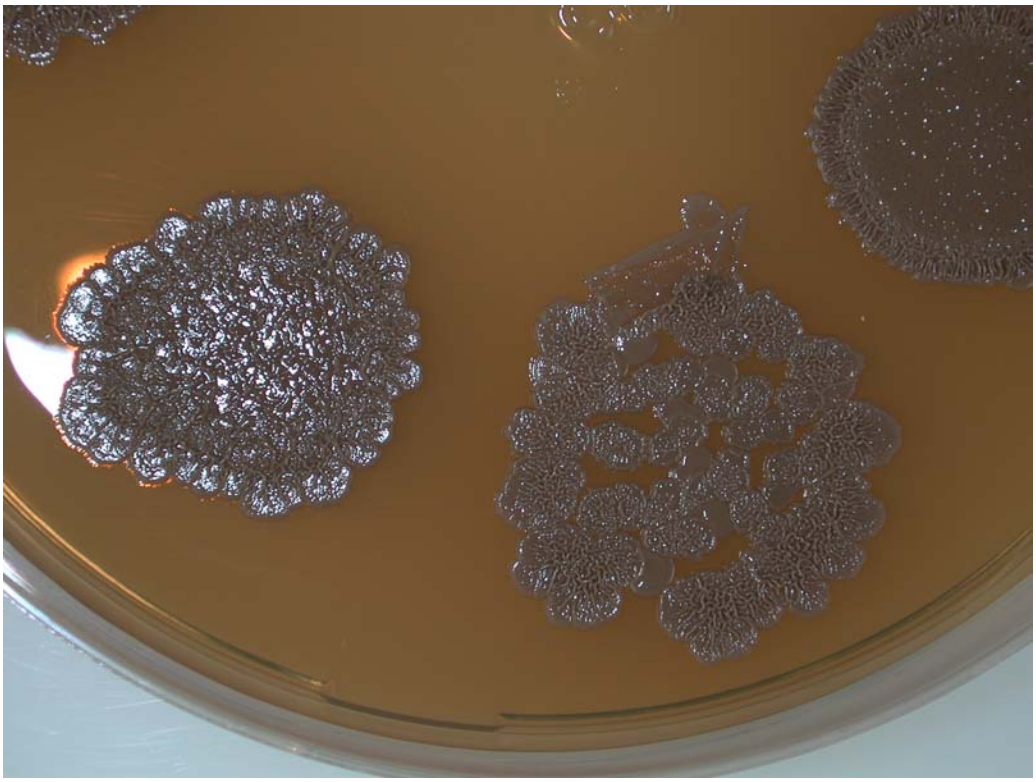
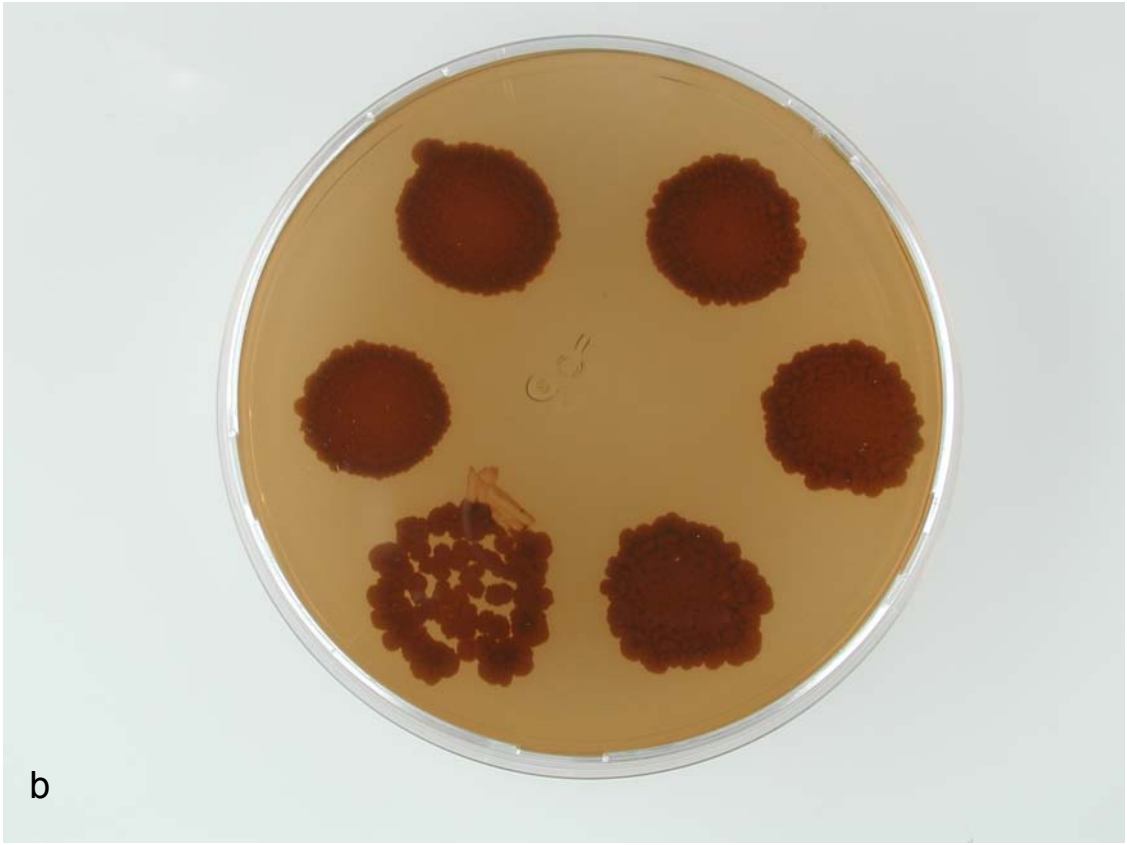


Fig. 12. Orange coloured colonies and confluent growth of *Halobacillus halophilus* DSM 2266^T (type strain of the species). Medium: Difco Marine Agar + 10mg/l $MnSO_4$, pH 7.6, incubation temp. 30°C, incub. time 3 days. Viewed from top, reflected light. Using the light blue background, colour fidelity is most closely met in reproduction.



a



b

Fig. 13. Bacterial culture. (a) Colonies appear blackish-glistening under reflected light, viewed from top. (b) Colonies appear brownish-reddish using transmitted light, viewed from bottom.

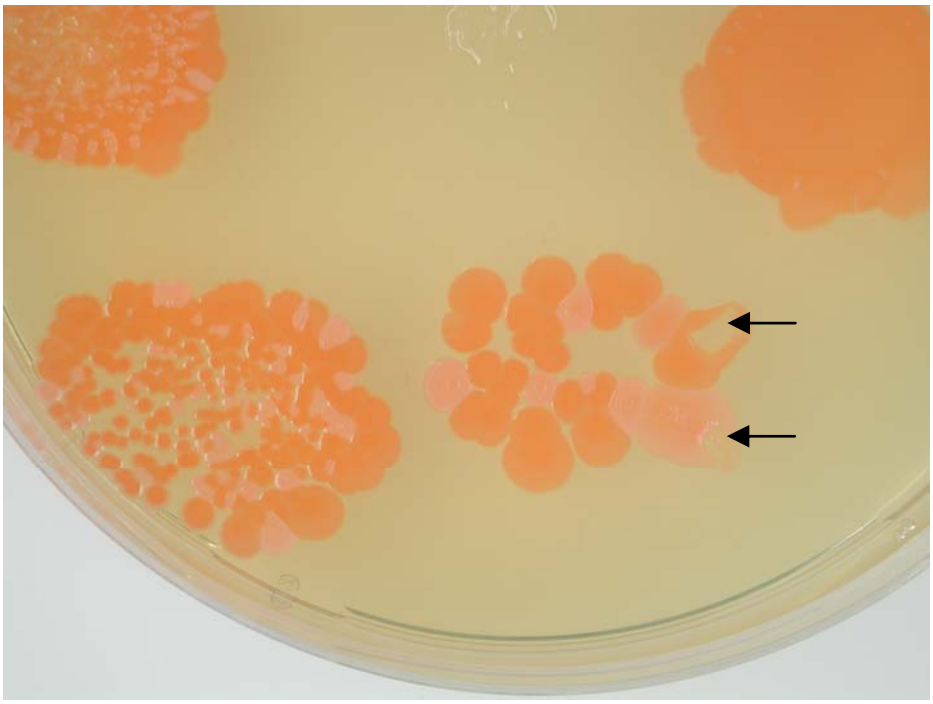


Fig. 14. Bacterial colonies of different colour as seen through the stereomicroscope; reflected light. Arrows indicate where some material has been taken off from colonies for microscopical examination.

Lighting (including light spectrum)
Light microscopy: usually phase contrast, more rarely bright field, dark field, UV light
Resolution (dpi), **pixel**: 1300 x 1030, or 1040 x 770 or 2080 x 1542

Software: ImagePro, Axiovision

Scale: Objects are usually 0.3 - 3 micrometer in diameter, therefore microscope objectives with a magnification of x40 or x100 (oil objective) are necessary. For photography, quality of objectives should be comparable to e.g. Zeiss Plan-Neofluar; scale bar is stored on picture (fig. 2); for (sub)cellular structures electron microscopy is used.

Data items captured
always: DSMZ-accession number of specimen and magnification
optionally: name
Other data such as e.g. more detailed information on handling procedures and culture conditions are recorded on paper and stored in hard copy file.

Wish-list:
To accelerate image production, easier access to devices should be possible. Presently only two microscopes equipped with digital cameras are available.
(1) At least one additional photomicroscope stand would be necessary (purchase costs for appropriately equipped Zeiss Axioscope about 17,000 €).
(2) Additionally, for preparing high quality images of the collection of type specimen to suit the purpose of web publication for GBIF and other goals, a person trained in the photomicrographic techniques used for prokaryotes would need to be employed.

Working with Images

Storage Formats: tif, jpg.

Software: Photoshop, Corel Photopaint.

Linking to/within Databases: Not yet.

Enhancing Images: (Cropping, Colour adjustment/Contrast, Image size, Filters)
Anything possible; but with light microscopy not much done as usually not necessary/not recommended.

Quality Control: more important during capturing, see there.

Naming Images: Images are named by accession number (collection acronym plus numerical) of specimen; through this, unambiguous linking to existing data bases and other related data is possible; it is essential to use the accession number system for microbial specimen. This became especially obvious when microbial databases were linked on an international level.

IPR policy for Images: Not yet.

Wish-list:
Long term goal would be to build a data base which is linked to existing strain data bases, thus being able to offer the combined, searchable information to the public via internet; see wishes above.

Maintaining databases
Back up: Presently no data base; only files on hard disk and/or CD.
Storage capacity: High.
Maintaining links: Not yet.
Linking to other systems: Not yet.

Wish-list: (unlimited budget and unlimited staff)
Final product should be a fully internet-compatible data base, searchable and linkable.

Acknowledgements

Ms. C. WAHRENBURG (DSMZ) is thanked for assistance with photographs; Dr. S. SPRING (DSMZ) is thanked for reading the text. Dr. D. CLAUS is especially thanked for thorough review of and discussion on the manuscript.

References

CLAUS, D. & BERKELEY, R. C. W. (1986): Genus *Bacillus* Cohn 1872. In: SNEATH, P. H. A., MAIR, N. S., SHARPE, M. E., & HOLT, J. G. (eds.), *Bergey's Manual of Systematic Bacteriology*, vol. 2: 1105-1139. – Baltimore (Bergey's Manual Trust, Williams & Wilkins).

HAJIBAGHERI, M. A. N. (ed.) (1999): *Electron Microscopy. Methods and Protocols.* Methods in Molecular Biology, vol. 17. – Totowa, New Jersey (Humana Press).

HEINZEN, R. J. & ENSIGN, J. C. (1975): Effect of growth substrates on morphology of *Nocardia corallina*. – *Arch. Microbiol.,* **103**: 209-217.

KAPITZA, H. G. (1997): *Microscopy from the very beginning*. 2nd revised edition. – Jena (Carl Zeiss Jena GmbH).

ROBINOW, C. F. (1975): The preparation of yeasts for light microscopy. – *Methods Cell Biol.,* **XI**: 1-22.

SCHAECHTER, M., WILLIAMSON, J.R. & KOCH, A.L. (1962): Growth, cell and nuclear divisions in some bacteria. – *J. Gen. Microbiol.,* **29**: 421-434.

SHAPIRO, J. A. (1992): Photographing bacterial colonies. Section 26: 1-9. In: MILLER, J. F. (ed.), *A Short Course in Bacterial Genetics.* (Handbook, Section 26) – Cold Spring Harbor (Cold Spring Harbor Laboratory Press).

SOMMERVILLE, J. & SCHEER, U. (eds.) (1987): *Electron Microscopy in Molecular Biology: A Practical Approach.* – Oxford-Washington (IRL Press).

WFCC (1996): *Access to ex-situ Microbial Genetic Resources Within the Framework to the Convention on Biological Diversity.* Background document to the UNEP/CBD/COP/3/Inf.19 information document. Ed. B. KIRSOP for World Federation for Culture Collections.

Digital imaging at the Herbarium Berolinense

WALTER G. BERENDSOHN, B. EINSIEDEL, U. MERKEL, H. NOWAK-KRAWIETZ, D. RÖPERT & I. WILL

Botanic Garden and Botanical Museum Berlin-Dahlem, Freie Universität Berlin, Königin-Luise-Str. 6-8, 14191 Berlin. E-mail: w.berendsohn@bgbm.org

Key words: Digital imaging, herbaria.

Introduction

In June 2000, after a prolonged period of testing, the BGBM started to provide access to high resolution images of its herbarium holdings on the World Wide Web. As of May 15, 2005, 36,170 images of specimens collected in 198 different countries can be inspected online through the BGBM image server (BGBM 2005).

This article documents the decision-making process which led to the procedures adopted. To be of assistance to others it also covers some of the dead ends we ran into and some of the unsolved problems ahead. We remain convinced, however, that for herbarium specimens of most groups the provision of high resolution images on the Internet represents a major step ahead in our current struggle towards a more efficient taxonomic work process.

Prerequisites set and problems solved before starting the process

Photographic imaging of specimens, especially of those which we were not sending out on loan, and of type specimens actually sent out, was an ongoing practice at B for decades. Some important collections were published on MicroFiche (e.g. the WILLDENOW Herbarium, HIEPKO 1972). Considerations of replacing the photographic process by digital imaging started in the 1990ies. Discus-

sions resulted in setting several pre-requisites, which somewhat delayed the actual start of the process but have been very useful in streamlining the process of imaging and data provision afterwards.

Full sheet vs. details: We decided to always take images of entire herbarium sheets, and we required an image resolution sufficiently high to show details of the plants and label texts. The threshold was put to be roughly equivalent to viewing the original sheet with the naked eye. In addition, the equipment had to be affordable as long as no major external funding came our way (which indeed did not happen until recently). Scanners were considered inadequate because they would require upside-down placement of the specimen, with the resulting damage and loss of material. Moreover, the format of the pteridophyte herbarium, a particularly valuable part of our collection comprising more than 300,000 specimens, was too large for the common A3+-size scanners. A camera was needed, but by the end of the 1990ies, one-shot cameras providing the required degree of optical resolution were still priced beyond our possibilities. We therefore opted for a scan camera, which offers high-resolution at a reasonable price.

Bandwidth: Another problem to be solved was how to make high resolution images available over the network - even compressed our files are several Megabytes in size. The solution came in sight when a consortium of several big companies announced the FlashPix image format in 1996 (DONOVAN 1998). Put simply, a FlashPix file contains the image in several resolutions. A FlashPix image server can thus adapt the transmission of data to the image and resolution actually needed at the receiving end, to display on the user's computer screen. This typically brings down the quantity of data transmitted to an amount that can be viewed within seconds even using slow modem connections (see table 1 for details).

At that time, the only available FlashPix image server for the WWW (MGI Zoom Server) was prohibitively expensive, being geared at commercial catalogues with a per-image licensing fees. However, the company understood our licensing requirements (many-images / low volume usage) and offered us a very reasonable special deal in

1999. This allowed us to install the implementation of the image server, which transferred only the amount of data necessary for the specific display over the network.

Table 1. Data transmitted over the network using a FlashPix image server.

Image file size	*Viewer window size (pixel x pixel)*	*Data transmitted (KB)*	
		Initial	*Full zoom*
~ 10 MB	545 x 388	11 KB	138 KB
~ 10 MB	1000 x 750	53 KB	500 KB
~ 95 MB	545 x 388	20 KB	172 KB
~ 95 MB	1000 x 750	97 KB	500 KB

Textual data and verification: Much of our discussion centered upon the extent and quality of the textual information accompanying the image. Clearly, the ideal would be to provide a full transcription of the original label's text as well as any annotations found on the sheet itself, accompanied by an up-to-date interpretation of, e.g., the collection locality. The identification should be made by a specialist and the nomenclature adapted to the latest usage. However, lacking specific funding, the amount of personnel resources we are able to dedicate to this process is very limited. We had to acknowledge that very few of our specimens would make it into the network if we'd adhere to these conditions. As a default, we thus decided to largely publish the information as found in the herbarium, and confine textual data input to the very minimum. Currently, this comprises (i) the barcode of the specimen (which mirrors earlier numbering schemes, where these were applied, see BGBM 2005), (ii) the scientific name under which the specimen is stored in the herbarium, (iii) the modern country of origin (where it can be easily deduced), and (iv) the original designation of the collection site. Optionally, (v) collector and (vi) collection number are also registered.

We had many doubts about publishing unverified information on the WWW. Our decision was drawn towards publishing because the image supported verification of the information, and because we hope to put remote annotation mechanisms in place which will allow us to improve the quality of identifications in the future, as well as to add high quality data on the collection locality etc. of the specimen.

Withholding the specimen from the public would make that impossible. However, image data are served with a clearly spelled out *caveat*.

Initiation of the digitization process

In 1998, we started to look at available cameras and found that scan cameras would suit our purposes and be available at an affordable price. As compared to one-shot cameras, there is a certain performance penalty in the form of the longer time it took to actually take the picture. However, this was offset by other bottlenecks in the information transfer and storage system, so the differences were considered tolerable. After extensive testing (see technical implementation below) the system was set up and but for technical improvements and changes of equipment it has basically remained unchanged since.

Priority setting: The severe cost cutting environment under which this system evolved actually spared us one decision: that of priorities for digitization. There was no way we would be able to support imaging per-se, it had to be either connected to soft-money funded projects or be integrated with internal procedures, preferably with a cost reduction off-setting the costs caused by the imaging itself. In short, we had to be very opportunistic to start the process.

Digital specimen loans: The internal procedure that provided a chance to attain a certain cost advantage through the imaging process was that of loans to other institutions. In many cases, most of the received material is investigated only superficially by the specialist, confirming the determination of a well-know taxon, or looking at the distribution or other textual data on the sheet; or trying to confirm the type status of a specimen through features like hand-written annotations etc. In these cases, an image fully serves the purpose. So we decided to replace loans as far as possible by the provision of digital images. Apart from saving mailing and packaging costs, the procedure has additional advantages. Sending plant material by mail is often slow and unreliable, and may result in damage or sometimes loss. We therefore never sent all the material available. In contrast, a loan request will now typically offer images of

our entire holdings of the taxon requested by the investigator. The process is also normally much speedier than sending the material by mail. Where necessary, the specialist can request individual specimens for direct inspection, their selection based on the images received. We offer these images on our image server, but to serve loan requests compressed images are also be provided for download. If this poses difficulties we also offer sending the files by email (compressed to a size of 500 KB and 2 MB each), or - as a last resort - sending electronic copies on disk or paper printouts by mail.

Production: Up to now, digitization effected by our core staff as a result of loan requests has produced the majority of our images served online. A small but notable addition is provided by in-house curatorial and specialist interest (e.g. for types of pteridophytes, supported by the Association of Friends of the BGBM, and for types of Compositae, supported by the BIOLOG program of the Federal Ministry of Education and Research). A major increase in the production of images has been achieved through the participation of our herbarium in the “Types of African Plants” project funded by the Andrew W. Mellon Foundation to support the Aluka initiative.

The African Plants Initiative (API) project at B: Aluka's mission is to build and support a sustainable, online database of scholarly resources from the developing world, beginning in Africa, with content that is important for research and teaching both in the countries of the region and in the worldwide scholarly community. One of the first content clusters Aluka focuses is „African Plants“, which will aggregate data beginning with a comprehensive database of high-resolution images of African type specimens from leading herbaria and botanic gardens around the world (Anonymous 2005). The BGBM signed up to this project and agreed to provide high resolution digital images of in-between 8,000 and 10,000 specimens of vascular plant types and original material from the herbarium B. Two technical staff were responsible for the actual imaging and handling of specimens, while a botanist acted as the project botanist and carried out coordinative tasks, material selection, the additional data input, and quality control. Acquisition of equipment and initial supervision of imaging was the responsibility of our herbarium curators, the technical and programming side was covered by two

further members of our core staff. The project posed several challenges as compared to the previous' two years of digitization. First of all the camera equipment had to be changed to obtain the higher resolution required by Aluka. As detailed below, this brought about an extended period of testing and adjustments, apparently we had hit the cutting edge. A second challenge was the actual location of types in the herbarium – some types are not marked as such, that material could not be included. Finally, this was apparently the first time that a geographic selection of material of this scale had been undertaken in the herbarium, and the material was not always separated and marked according to geographic regions as it had been originally planned. A major obstacle was that we had to work without any guidance as to the occurrence of taxa in Africa (except for North Africa, where the Med Checklist was of some help, for the Gramineae and for the Loranthaceae, where the herbarium revision of GEBAUER (1992) was particularly helpful). A lesson taken from this experience was that in a large herbarium selection of a geographical subset for digitization should start with the preparation of at least a list of generic names under which relevant material may be stored. Imaging activities under the project were now concluded, with the result of adding 11,448 images to the BGBM's image server.

Technical implementation

Camera and lighting: In January 2000 we started image production with a Pentacon Scan2000 camera, made by the German company PRAKTICA (see table 2 for technical details). This camera has been in use since then and has produced more than 22,000 images up to now. The major initial difficulty concerned the lighting equipment, the original lamps were of the wrong type and caused strong color deviations, furthermore they were generating too much heat which caused the CCD chip to change its properties. The equipment was replaced accordingly and we now use for all cameras two Kaiservision lamps each (see fig. 1) with 2 or 4 36 W light bulbs. Selecting the appropriate lighting equipment with respect to spectrum and effect on ambient temperature is an essential prereqisite for high-quality images. In May 2001 we added a second

Fig. 1. Setup of the imaging station.

camera, this time of the Pentacon Scan3000 type, with a Schneider-Kreutznach lens. Adjustment of focus, aperture and depth of field turned out to be much more complicated and time-consuming as compared to the Scan2000. The camera was therefore mostly used for special tasks, for example in the digitization of the SCHWEINFURTH collection (fig. 2). The lesson we learned here was that the lens has a profound effect on the quality of the digital image, which seems to be much more pronounced than in traditional photography.

About 6,000 images were taken with this camera until, in the course of preparations for the TAP project, it was changed for a newer model (Pentacon Scan5000) providing higher resolution, with a second one of that type bought in addition. As mentioned above, we experienced rather annoying difficulties with this equipment. Once more, lenses had to be changed to obtain reasonable depth of field and properly focused images (especially in marginal areas of

Fig. 2. A sheet in the Schweinfurth collection, a set of historical documents damaged by water which is not accessible to the public due to its fragile state of conservation. "*Dendrosicyos socotranus* / Cucurbitaceae / 18' [Fuß] hoch" shows the only arborescent Cucurbit of the world. It was drawn by SCHWEINFURTH during his 1881 expedition to the island Sokotra, the only place in the world where it can be found.

Table 2: Camera equipment.

Camera (Pentacon)	*Max. resolution (pixel)*	*Lens*	*Software*	*Interface*	*Flashpix file size*
Scan2000	3.640 x 4.624	Nikon	Silverfast V4 / Adobe Photoshop7.0	SCSI	~ 20 MB
Scan3000S	5.363 x 5.363	Schneider Kreuznach	Silverfast V4 / Adobe®Photoshop 5.5	Fire-Wire	~ 40 MB
Scan5000	8.192 x 12.000	Schneider Kreuznach APO COM-PONON 4,0/60	Silverfast V6/ Adobe Photoshop7.0	Fire-Wire	~ 95 MB

The Scan3000 is no longer in use at the BGBM.

the image). For one of the cameras, stronger lights were installed to be able to use a narrower aperture for increased depth of field (thick objects). Another problem consisted in variations in the quality of the pictures between cameras as well as within the images taken by the same camera. Tests were run to explain this in terms of ambient factors, hardware used or the software setup, to no avail. Several upgrades of the software had to be installed before the system was functioning properly for any prolonged period of time. Lack of appropriate software documentation led to a delay in finding the appropriate parameters (e.g. grid factor) to adjust the software and adopt it to the chosen resolution. All this did cost substantial time for testing and implementation, but finally we obtained a working system constantly producing images of high quality (although some variation can still be observed). 11,350 images were produced using the Scan5000 cameras up to now (1 May 2005). The resolution of phanerogam specimens images is equivalent to slightly more than 600 dpi, that of fern specimens (with larger sheet size) is slightly lower.

Data processing and storage: our imaging workstations are normal high-end PCs, currently 2,6 GHz P4 processor, 1 GB RAM, 128 MB graphics and a 100 Mbit network connection, the current operating system is Windows 2000 SP4. The Pentacon Scan5000 cameras are connected via an IEEE1394 “FireWire” interface and

are controlled by Silverfast Software. Image processing is effected with Adobe Photoshop. The TIFF files produced are converted by a batch process to FlashPix (compression factor 1) and to two JPEG-format images, one with 2-4 MB for downloading and a thumbnail version of ca. 20 KB for the selection list (fig. 3). Since October 2004 the TIFF files are compressed using the lossless ZIP format. The original TIFF files are backed up to the University's central backup facility. All other files are stored locally, in the file system, using directory setup reflecting the barcode numbering scheme. Files are secured by RAID systems and internal backups. Storage solutions were constantly upgraded and extended in capacity since 1999, always using standard equipment and taking advantage of the considerable improvements of computer data storage that took place during that period. Currently (May 2005) 3 servers are in use with RAID storage systems of a total capacity of 2,8 TB. Our lesson learned is that the cost of storage is not the limiting factor for imaging in an institution like ours.

Serving the data: From the beginning, we used web-based technologies for serving the data both outside and inside the BGBM. As mentioned above, until August 2003 we used the MGI ZoomServer, but subsequently changed the solution to the TrueSpectra Scene7 server software. Our main reason was the functionality of the client, which for the MGI server was restricted to HTML and thus offered very limited functionality. The Java viewer offered did not work very well. Moreover, only FlashPix format images could be displayed. In contrast, the TrueSpectra server showed higher performance, accepts TIFF, JPEG, PSD and other formats and a comfortable standard viewer for the client (based on Macromedia Flash). This offers zooming, turning, and adjustment of color density, contrast, and brightness of the image. The software is installed as a separate web server under Linux. To further increase performance, we also installed the ZoomCache Server (Neptune-laps), which caches once asked-for zoomed image details and thus increases display speed by a factor of about 4. The installation is completed with the Neptunelaps FSI-Viewer, which adds the possibility of direct length measurers on the zoomed images (fig. 3). All images prepared become available on the network after 24 h. This

includes other digital images (for example taken from specimens on loan), which, however, are not served to the public. The publicly available images can be accessed directly on the BGBM's home-page, or via user interfaces using the GBIF compatible BioCASe protocol (DÖRING et al. 2005), such as the GBIF Data Portal, the BioCASE portal, or the website for the German GBIF node for Botany.

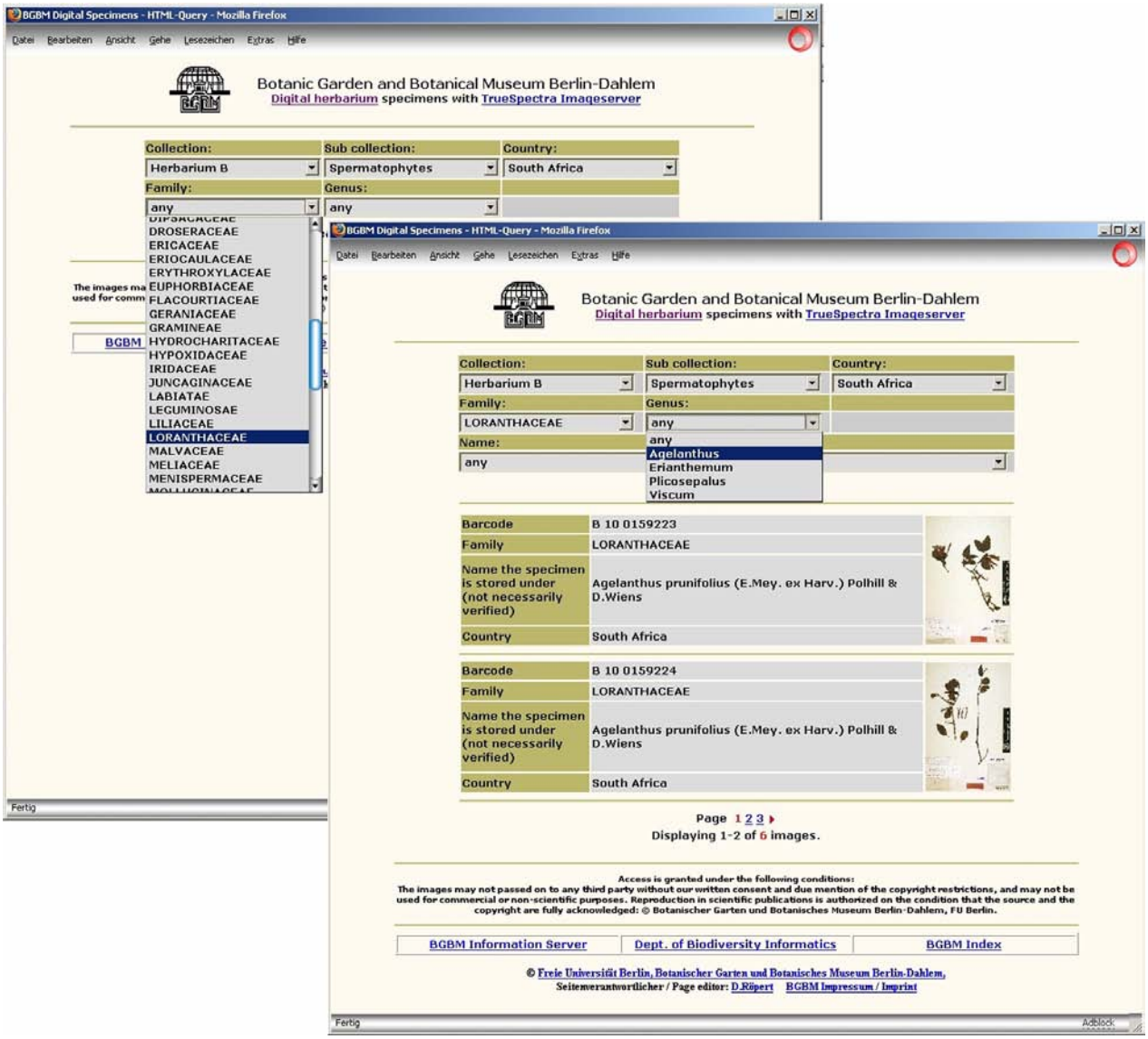


Fig. 3. Selecting a specimen image on the BGBM portal. Selection of collection, country of origin, and family or (part of) name is used to restrict the query. Thumbnails are shown and can be paged through once only a limited number remains. Clicking on the thumbnail image (or on a name in the alternative list view) opens a window offering the full image server functionality (fig. 4).

The imaging practice

Basic setup and process: Images are taken with a constant distance and standard parameters for a given sheet size. A pre-scan is used to control focus, depth of field and color adjustments, followed by the main scan of the specimen into Photoshop. Scanning the specimen's barcode with a barcode reader serves to name the image file and connect it to the BGBM's database. If the basic text data are not present, they are entered now, using an MS Access front-end application developed for this purpose. The image is than saved on a BGBM network server as an uncompressed tiff file.

Setup of camera and lighting: A number of preparatory steps are necessary to obtain high image quality. This starts with ensuring that the camera is in a horizontal position (use a spirit level). Lighting should be maximized by adjusting distance and angle of lamps, but excessive flare must be avoided– areas with tonal values >250 in the RGB-histogram should be rare. A uniform illumination should be achieved, testing is done by scanning a white sheet and checking the white-level with a reflection densitometer, the values should vary only moderately. It is recommended to use a luxmeter to measure the illumination and to optimize accordingly. All this may be done with different lamps and arrangements of lamps to achieve an optimal setup. Since a scanning process is taking place, obviously any commotion or vibration of the object or camera must be avoided. In setting up the workplace the properties of the building must thus be considered, too. Even in a relatively solid place like our museum building, event such as slamming the door result in multicolor fault lines on the image. Places with heavy traffic or other causes of vibration may be completely unsuited for setting up the imaging equipment. Reflective objects (in our case the ruler with the mm

Fig. 4. Online manipulation of images. The image server only transmits the resolution needed for the actual display. Top left: view of entire sheet, as displayed after clicking on thumbnail (fig. 3); right: zooming in and taking a measure. Bottom: turning the image; bottom right shows the controls for contrast, colour and brightness.

scale and the BGBM logo) should be positioned as close to the center as possible to avoid flare. In our case positioning the ruler towards the middle of the long margin of the sheet was sufficient, since lengh to width ratio of the herbarium sheets exceeds that of the image taken by the camera. Ambient conditions may also have an effect on the paper of the sheet or the labels. During the winter month we observed from time to time that the paper warped upwards, probably because of low relative humidity. We are presently doing without it, but probably an airconditioned room would

be of advantage to avoid such effects on the objects as well as on the technical equipment.

Focusing: Finding the optimal focus (first manually and subsequently by software) should commence with maximum aperture using a test sheet with lines. To test for depth-of-field, a 45° wedge with a millimeter scale on the slope has proven most advantageous. With thicker objects (branches etc.), the focus should be fixed at 50% of the object's thickness, which of course should not exceed total depth of field. Using ad-hoc optimization is possible by focusing on different levels until optimal focus for the paper and the highest point in the object is reached, but this is a time-consuming process. Using a different aperture changes depth-of-field and makes re-adjustment necessary. From time to time, a check should be made if the entire area of the scan is in focus. A sheet with small print covering the entire area is useful for that purpose. It is scanned for that purpose, and checked especially in the marginal areas. Faulty areas may be the result of the camera being unleveled, of heterogeneous illumination, or of dirt on the lens.

Imaging: Adjustments made to calibrate the scanning software and the image editing program vary according to the software programs used and we are therefore not going into detail here. In general, where both programs offer adjustments, these should be set to the same values (which is sometimes rather difficult to figure out, if terminology varies). Different cameras also offer widely varying options for adjustments. Automatic settings for unsharp masking and adjusting color balance should be avoided, since our objects are not the average photographer's choice. Using a histogram to analyse the picture has proven essential, because most image corrections can be diagnosed by looking at the histogram. For example, if the histogram is too narrow, changing the aperture and/or exposure time (scanning speed) can be used to achieve a much improved result. Correction of color saturation is achieved by adjusting the color channels in the histogram. Adjusting the white balance is often necessary because ambient light conditions influence the color quality. Color cast correction is achieved by choosing a constant white area (e.g. the white in the color scale, or part of the barcode label) and adjusting the red-green-blue values to read close to 250 in

the densitometer. The scanned picture can be further optimized by changing brightness and contrast, preferably using a gradation curve. Many tutorials for improvements of digital images with different programs are available on the Internet, we recommend to set aside some time for training and studying for the personnel involved in the digitization process.

Conclusion

The technology is available to produce high resolution images and to make them available in a usable form, although technical improvements to speed-up the work process are needed. We are convinced that the availability of digital images of herbarium specimens (as well as other images of organisms and parts of organisms) will soon form an indispensable part of primary data provision for basic research, and that it will strongly influence the work practice of the individual taxonomist. Taxonomy today is being squeezed between growing demand and urgency for classification and identification of organisms on the one hand, and diminishing personnel resources (as measured in time available for research) on the other hand. Bringing the data to the specialist is thus a top priority for infrastructural improvements. While the physical specimen cannot be replaced as a permanent source of primary data (see FUNK et al. 2005), the features inherent in the specimen's superficial physical appearance can be mobilized and thus shared very efficiently by using digital images on the Internet. Digital imaging of specimens is thus a major contribution to the new suprainstitutional collaborative infrastructure taxonomic research needs to tackle the challenges ahead.

References

Anonymous (2005): Aluka, Mission and Audience. http://www.ithaka.org/aluka/index.htm [accessed 16.VI.2005].

BGBM (2005): Digital specimen images at the Herbarium Berolinense. ww2.bgbm.org/herbarium/default.cfm [accessed 16. VI.2005].

DÖRING, M., DE LA TORRE, J., GÜNTSCH, A. et al. (2005): The BioCASE unit-level access protocol v. 1.3. http://www.biocase.org/dev/protocol/index.shtml [accessed 16.VI.2005].

DONOVAN, K. (1998): The Promise of the FlashPix Image File Format. RLG DigiNews 2(2). http://www.rlg.org/preserv/diginews/diginews22.html#FlashPix [accessed 16.VI.2005].

FUNK, V. A.; HOCH, P. C.; PRATHER, A. L. & WAGNER, W. L. (2005): The importance of vouchers. – *Taxon*, **54**: 127-129.

GEBAUER, R. (1993): The type specimens of African *Loranthus* taxa (Loranthaceae) ... kept in the general herbarium at Berlin-Dahlem (B). – *Willdenowia,* **23**: 257-289, 291-304.

HIEPKO, P. (ed.) (1972): Herbarium WILLDENOW. Alphabetical Index. Herbarium C. L. WILLDENOW (1765-1812) – Microfiche-Ausgabe. Zug (IDC).

Protected names of products used in this article: Adobe®, FireWire®, FlashPix®, MGI®, Neptunelaps FSI-Viewer®, Photoshop®, Silverfast®, TrueSpectra®.

Imaging Soil Mesofauna

The Land in Between

ARTURO H. ARIÑO, ENRIQUE BAQUERO & RAFAEL JORDANA

Department of Zoology and Ecology, University of Navarra, E-31080 Pamplona, Spain; artarip@unav.es

Key words: digital imaging, soil fauna, specimen preparation, microscopy

This chapter will try to make some generalisations on current approaches to the imaging of small specimens. By that, we mean those with sizes lying typically between that of microphotography and that of macrophotography. These constraints do not necessarily mean a requirement for taxon-specific techniques. Rather, mainstream imaging procedures can be successfully used, although it is advisable to consider certain characteristics of the taxa in question. A case, or rather 'bookcase', study (i.e. different taxonomic groups) depicting our laboratory's procedures will act as a basis for discussing several issues.

Assemblages of the largest specimens can be pictured, though rarely, with SLR cameras fitted with a good macro lens. However, mesofauna are usually imaged under the stereomicroscope or by transmitted light microscopy. Preserved specimens are typically mounted on slides and imaged at low power under optical or scanning electron microscope, and details are captured at higher magnifications. Surfaces, and superficial features, are targeted more often than internal structures, and optical techniques are much more common for contrast improvement than chemical stains.

The main drive should be directed at ensuring a good scientific usability of the images. Thus, compromises will be needed where feature visibility, enhancement, and fidelity conflict with aesthetic qualities.

Soil Mesofauna

Soil mesofauna is the most diverse component of the soil ecosystem. Animals living among the litter and inside the microscopical crevices of the soil have a fundamental role as processors and translocators of the organic matter that ends up forming the humus. Many taxa are represented, including several orders of insects and their larvae, as well as Myriapoda, Crustacea, Thysanura, Tardigrada, and others. But three of them (Acari, Collembola, and Nematoda) dominate in terms of numerical abundance and diversity. A typical soil sample from a Mediterranean forest may contain several hundreds of different species, many of which have yet to be described. The least diverse of the three mentioned groups, Collembola ('springtails'), includes more than 7,600 known species: more than all mammal species, or three quarters of the known number of bird species. About fifty new species are described each year. And there are about twice as many known free-living species of Nematodes. It is suspected that the number of undiscovered species is very high in these groups (there are about two orders of magnitude less invertebrate taxonomists per invertebrate species than plant taxonomists per plant species), which warrants a high probability of many new type series needing to be imaged per year.

The wide spectrum of soil mesofauna means that a variety of handling and preparation procedures, as well as several observation and imaging techniques, are applied even within a group. Thus, the proposed "bookcase" study will have several tiers at times, where techniques differ for specific groups.

Case Study

Taxon Group: Main Soil Mesofauna Components (especially Phylum Nematoda; Phylum Arthropoda: Subclass Acari, Class Collembola)
Institution: Department of Zoology and Ecology, University of Navarra, Spain
Persons Responsible: Enrique Baquero and Rafael Jordana
Number of Person Hours per year devoted to imaging: Varies
Number of Images captured and stored each year: Varies

Purpose of Image

In most cases, images are for scientific use. Whole specimens and parts thereof are depicted that have taxonomic significance. Images are produced as they are needed to illustrate textual descriptions for taxonomic research papers; to build an iconographic bank for visual identification keys; to exchange information and queries amongst researchers; to detect and research taxonomically relevant features; to measure specimens or their features for taxonomical and ecological purposes; and to serve as basis for scientific drawings.

Target Audience

Mainly scientific, especially systematists/taxonomists, ecologists and field biologists.

Form

Originally intended for research, documentation and paper publications, conforming to scientific literature standards. However, copies of the primary pictures are also post-processed differently (i.e. resampled, resized and/or enhanced) for web use.

Choice of imaging technique

The imaging techniques to be used in soil mesofauna are, of necessity, varied both because of the large size range of the specimens and their taxonomic features, and their different tegumentary characteristics.

Stereomicroscopy provides the lowest level of detail, but is often the choice for living, large specimens. It allows colour pattern, which would otherwise be lost, to be retained (see fig. 2) and provides "natural" aspects of the specimen. However, even at maximum magnification it cannot reveal great detail for most mesofauna (see fig. 9 for a typical acari). Artificial enlargement of the image, i.e. by using a dense CCD, may aid visualization but perhaps not resolution, as the latter is fraught with other problems such as narrow depth of field and chromatic aberration, discussed elsewhere in this volume (ARIÑO & GALICIA, "Taxonomic-grade images").

Choice of imaging technique (cont'd.)

Optical microscopy is the technique used most often for soil mesofauna. The usual set of objectives in most microscopes allows for almost the full range of features to be observed, from whole specimen to minute details less than one micron across. Good images, however, are notoriously difficult to obtain for the following reasons:

- *Specimens are observed by transmitted light but are mounted complete. Attenuation and dispersion of light is intense. Much tissue lies between the focal plane and the surface that obscures the observation. Also, many taxonomic features, such as setae, may lie along the Z-axis, making observation at a single focal plane pointless.*
- *Observations are usually made by juggling the focus knob to make a mental picture of the objects in 3D space, but the effect of this dynamic play can only be approximated by automontage techniques.*
- *As magnification increases, the definition of the image relative to the field size decreases as it approaches the dimension of the Airy function. High magnification images show poor detail.*
- *The frequent lack of contrasting features, staining, or other light-absorbing characteristics in the specimens forces the use of contrast enhancement techniques on the optical system. These almost always mean that the width of the light beam entering the objective has to be reduced, which has the effect of reducing the numerical aperture of the lens and, consequently, the image sharpness. There is a trade-off between the gain in feature visibility by contrast enhancing techniques and the loss of detail, which is normally much biased towards the former.*
- *The same contrasting techniques tend to increase the dynamic range of the images to the point where the sensors can have many pixels saturated, whereas other remain below response threshold. Once the full dynamic range is achieved, additional contrast visible to the eye is lost.*

Scanning electron microscopy (SEM) overcomes many of the above limitations but introduces others. It dramatically extends the depth of field above optical microscopy, and pushes the resolution power several orders of magnitude beyond (see fig. 7). But it can only be used on surfaces, under high vacuum, and requires special preparation of the specimens. It is often not possible to spare a type for this special preparation, which cannot be undone. In all, however, SEM should be attempted whenever possible to fully characterise a type.

Microarthropods	Nematoda
Collectively known as 'microarthropods', Acari (mites) and Collembola (springtails) along with other groups such as Thysanura or Tardigrada are extracted from soil samples by Berlesse-Tullgren funnels or high-gradient behavioural methods. Heptane flotation is our method of choice for soil samples already preserved by neutralised formalin (see BELASCOAIN et al., 1998 for a detailed description), although we have also used other flotation methods. In all cases, specimens are ultimately transferred to ethyl alcohol-based fixative fluid. Suction micropipettes, micropins, or single-bristle pencils are used for handling.	Behavioural methods for nematodes include wet extraction on water funnels (FLEGG & HOOPER, 1970). Formalin preservation is done after killing the specimens by heat (HOOPER, 1986). Centrifugation in sugar gradient of already-preserved samples (CAVENESS & JENSEN, 1955) is the physical alternative, often carried out on samples from which microarthropods have been previously extracted by heptane flotation method (BELASCOÁIN et al., 1998). Specimens are generally not stained but treated with Nesbitt solution to improve their transparency, as in this case internals are taxonomically important. All handling involving individual transfers between treatments or onto slides is done with micropins or single-bristle pencils under the stereomicroscope.
Acari	**Collembola**
Many Acarina, such as Oribatida, possess strongly chitinized, dark integuments that prevent observation by transparency at different depths. Thus, they are often bleached, usually with chloral hydrate or Nesbitt solution, to observe features at different planes by transmitted light.	Colour pattern can be a taxonomic character for some groups of springtails. As colours may fade in time if animals are preserved in the usual ethyl alcohol medium, pictures can be taken from unmounted specimens from the series at this stage during the type description process.
Dead specimens are almost invariably observed and imaged at this stage under the dissecting microscope while submerged in low-grade ethyl alcohol, unless selected for scanning electron microscopy (SEM) mounting.	

Table 1. Isolation procedures for soil fauna.

Wish-list

- to have all specimens from the type series optically sliced and digitized at several magnifications;
- to have a fine series of al least one specimen optically digitized with enough z-precision and contrast so as to be able to produce a workable, deconvoluted 3D-model;
- to have at least one specimen from each series imaged by scanning electron microscopy; and
- to have an image bank of all known and relevant taxonomic characters from the pooled type series.

Capturing

Handling Specimens

As most types are preserved, images can rarely be obtained from live soil mesofauna belonging to the type series. Extraction of live specimens is achieved by behavioural methods, with specimens usually collected on or in water, or taken to the laboratory together with their substrate or nutritious material (plants, fungi) when feasible. Imaging is then done directly under high-power macro lenses for masses of the largest specimens on their natural medium, or under the dissecting microscope as single organisms. Non-living extraction can be subsequent to behavioural methods, or is effected directly by physico-chemical procedures. Whichever mode of extraction is adopted, specimens are mounted on slides for detailed observation, and frequently for final storage as well. Table 1 describes the handling methods used for various taxa.

Mounting Specimens / Specimen Preparation

Methods depend on the type of observation procedure to be used (optical or electron microscopy) and the taxon type. Table 2 explains the possible mounting choices.

Optical microscopy	
Specimens, often dehydrated by the Seinhorst procedure, are mounted on microscope slides in any of several mounting media (glycerin, glycerogelatin, Hoyer fluid, etc.) according to taxon type, and covered with a glass cover. These mounts are permanent or semi-permanent and constitute the basic storage and handling unit.	
Microarthropods	**Nematoda**
Pre-mounting triage of the specimens under the dissecting microscope generally allows the preparation of slides of single species that can be accessioned once. Mapping the slide by establishing grid co-ordinates, or by encircling the specimen with a pen, is useful on the smallest mites, as these are often too small to be seen by the naked eye.	Mapping the slide by establishing grid co-ordinates, or by encircling the specimen with a pen, is most useful on nematodes, as they are not visible to the naked eye. The storage unit usually contains several specimens from various species; thus, the type series may be spread across different slides and each slide can contain both type series and other material. Generally, several accessions refer to the same storage unit.
Scanning electron microscopy	
Specimens are cleaned in an ultrasonic water bath from residuals clinging to their surface. The adjustment of timing and power is delicate, as many thin structures such as setae may become loose.	
Microarthropods	**Nematoda**
Microarthropods are dehydrated in ethyl alcohol series and dried by critical point drying (CPD) with carbon dioxide. Anti-static cloths are used to prevent dehydrated specimens from “jumping”.	Nematodes have been treated by various methods, including critical point drying (CPD) with carbon dioxide and DMP method (WEYDA, 1992) followed by resin inclusion.
Specimens are mounted on aluminium SEM stubs with two-sided sticky tape. Large microarthropods can also be glued to the tip of micropins in order to separate them from the stub’s surface. Temporary storage is effected inside a dehydrating jar.	
Specimens are finally metal-coated by evaporated gold or gold-palladium.	
Final storage is effected for each stub separately inside a plugged glass vial partially filled with dehydrating resin	

Table 2. Mounting procedures for soil fauna.

Capture Devices

A number of combinations and techniques are in use (Table 3).

Optics / Digitizers	Macro	Optical Microscopes			SEM
	Nikkor AF Micro ED (7)	Zeiss DV4 (8)	Leica MZ6 (9)	Olympus BX50 (10)	Zeiss DSM 940 A
Nikon D100 (1)	(A)				
Nikon Coolpix E995 v1.6 (2)		(B)	(C)	(E)	
Canon Powershot S45 (3)			(D)	(E)	
JVC KY-series (4)				(F)	
Point Electronic DISS 5 (5)					(G)
Agfa Arcus 1200 (6)	Digitising of chemical prints and slides				
HP Scanjet 5550c, 8200 (6)	Digitising of chemical prints				

(1) Single Lens Reflex (SLR) digital camera; 6.1 megapixels 23.7 x 15.6 mm CCD

(2) Digital camera with integrated 8 – 32 mm (38 - 152 mm as 35mm equivalent) aspherical lens; 3.2 megapixels 7.2 x 5.3 mm CCD

(3) Digital camera with integrated 7.1 – 21.3 mm (35 – 105 mm as 35mm equivalent) aspherical lens; 4 megapixels 7.2 x 5.3 mm CCD

(4) 752-line video camera with ½" CCD sensor without lens; linked to Scion LG3 frame grabber through single (green) channel

(5) Digital sensor for SEM, feeding directly from SE- and RE-detectors at preamplifier output for maximum 16k (x) by 16k (y) pixels

(6) High resolution desktop scanners

(7) 70-180 mm (105 – 240 mm as 35mm equivalent) f/4.5-5.6 macro lens

(8) Stereomicroscope with 0.63x-4x aspherical lens system

(9) Stereomicroscope with 0.63x-4x f/13 lens

(10) Optical microscopes with phase and differential contrast systems

(A) Masses or colonies of specimens on their substratum

(B) Unmounted specimens shot through eyepiece tube with adapter

(C) Unmounted specimens shot directly through phototube

(D) Unmounted specimens shot through phototube with optical (eyepiece) adapter

(E) Fixed specimens on slides shot through eyepiece tube with mechanical adapter

(F) Fixed specimens on slides shot through phototube with optical (eyepiece) adapter

(G) Carbon or carbon/gold-coated specimens

Table 3. Combinations of imager and optics in use at the laboratory.

In addition, a number of chemical 35mm, 6” by 9”, and 4” by 5”-Polaroid holders were used and still some are on various microscopes. Prints, slides or negatives are scanned with the appropriate adapters.

Optical microscopy images

See ABRAMOWITZ & DAVIDSON, 2003, for an excellent tutorial on basic microscopy techniques. Much of what is discussed here in terms of general optics comes from this tutorial.

Although heavily chitinized animals such as Oribatida can be observed with ordinary bright field techniques, high-magnification images of most soil mesofauna, which seldom absorb light, benefit from contrast enhancement by dark field, phase contrast or differential contrast microscopy at the time of capture.

In bright and dark field techniques, condenser settings, including its numerical aperture (‘condenser diaphragm’) and position, and field diaphragm settings, do not differ from the ones that would afford good observation contrast at the chosen magnification without the appearance of diffraction artefacts. It should be noted, however, that the amount of light should be better regulated by the use of filters, or the shutter speed of the camera, as opposed to the other alternative of manipulating the voltage of the filament. (It is not regulated by condenser or diaphragm settings, which influence the sharpness, numerical aperture and contrast ratio). Changing the voltage also changes the colour temperature of the image.

Phase contrast converts differences in refractive index into differences in phase of the light passing through the object, in such a way that the transmitted light and the diffracted light originating at the object can interfere destructively. The final result is that the image of the object acquires amplitude contrast, showing features darkened against a bright background. Good phase contrast images require a correct alignment of the phase annuli of condenser and objective, and benefit also from adequate selection of the amount of light illuminating the object.

Fig. 1. *Ypsilonellus similis*, female holotype, shot with four microscopic techniques. Clockwise from top left: ordinary brightfield; darkfield; differential interference contrast; phase contrast. Observable features vary among techniques. Field width 0.23 mm; pixel size 0.11 µm. White 5000°K light, unfiltered, Olympus BH50 microscope (except darkfield: Olypus Vanox); 40x objectives. Nikon E995 on phototube. Post-processing: colour balance; image resizing to SVGA; contrast enhancement and gamma adjustment. JPEG format.

Lighting

Images taken with SLR cameras are shot, whenever possible, with natural sunlight. Otherwise, either an electronic ring flash or a set of cold, full spectrum (three-phosphor) fluorescent lamps at 5000ºK are used. See ARIÑO & GALICIA, in this volume, for details.

Objects under the stereomicroscope are illuminated depending on whether or not colour capture is necessary. Colour patterns are generally lost during the fixation procedure, and should be captured beforehand if taxonomically important. In this case, halogen-tungsten

filament dichroic lamps at high temperature setting (i.e. 5500°K) are used. Light is spread over the object by optic fibre channels. This produces a rich spectrum that is additionally filtered for excess red and infrared by light green glass.

Optical microscopy images (cont'd.)

The main limitations of phase contrast imaging in soil mesofauna arise from the thickness of the specimens. The technique works better with thin slices, as phase contrast from planes above or below affect the current focal plane. Also, bright artefact halos appear surrounding the image details that may saturate the sensor, erasing captured detail.

Differential interference contrast (DIC) virtually eliminates these halo artefacts. Contrast in DIC is produced by evaluating the length gradients of the optical path through the specimens resulting from density differences (the rate of change in the direction of wavefront shear), whereas in phase contrast the different densities that result in different optical path lengths yield different light magnitudes (denser objects appear darker). The optical gradients in the specimens are converted into intensity gradients. Polarized light and prisms are used, and the light spectrum is altered in such a way that specimens appear both with a shadowing that confers a 3D appearance and bright, selectable interference colours. The numerical aperture reduction is less than that of phase contrast, yielding better detail; but specimens having bi- or multi-refringent features such as mineralised parts cause interference with the polarized light, affecting resolution and creating artefacts.

Another unwanted effect of DIC is that since the tri-dimensional effect observed does not correspond to the actual 3D geometry of the specimen, its reconstruction by deconvolution techniques is hampered.

See fig. 1 for an image comparison between contrast enhancement techniques.

When colour is not important, objects are illuminated with 23-W PL-type compact fluorescent lamps in reflective aluminium mounts. These mounts are also used for routine work. Transillumination is used concurrently if appropriate for the specimen being pictured.

Shadows are often reduced by using at least three light points (optic fibres), six-point concentric lamps, or large-surface lamps (fluorescent and compact) and diffusive elements. See ARIÑO & GALICIA (this book), figure 18 (left), for a typical macroscopical stage with stereomicroscope, two fiberoptic lighting systems (ring and spotlights), camera mount and workstation.

Fig. 2. Live *Orchesella* (above, halogen white spotlight) and after decoloration (below, diffused compact fluorescent light, colour-corrected). Image pixel is 2 μm. Taken with Coolpix mounted on Zeiss stereomicroscope. Cropped from the original field width (4 mm).

Resolution

Camera images are invariably taken and stored at their maximum size and resolution, except for SEM digital images that are produced at 1 Mpx (square 1k by 1k pixels) size. Images from D100 are 3028 x 2018 pixels; Coolpix produces 2048 x 1536 pictures, and Powershot

Size Range and Resolution

The size of soil fauna can spread across five orders of magnitude. The largest epiedaphic Collembola are about 10 mm long, although typically sizes range from 0.5 to 5 mm. Edaphic acari range from 0.2 to 1.5 mm, and free-living Nematoda from 0.1 to 2 mm in length and about one-tenth wide. Taxonomic features are generally much smaller but often can be of greater interest than the image of the whole animal. Microarthropod sensory setae can be less than 10 µm in diameter, and Nematoda taxonomic features or Collembola cuticle details may be less than one micron across. Thus, it is important to select an image size that is commensurate with the object being imaged (figs. 4 and 5), which in turn may influence the choice of observation technique (Table 3).

The degree of detail achieved will depend on the macro- or microscopic technique selected and the relative size of the pixel in the image. Conventional optical microscopy using white light has a resolution limit of about 250 nm (the Rayleigh limit, one-half the average wavelength). Although near-field scanning optical microscopy (NSOM) can increase the resolution power to λ/60 through the avoidance of the diffraction limit, in practice it does not seem to suit the heavily tridimensional nature of specimens. Laser scanning confocal microscopy could also achieve resolutions below the Abbe's limit in some fluorescence-dependent novel implementations, but seems perhaps little suited to fixed specimens where subcellular features are not of particular interest.

However, the diffraction limit is seldom reached in soil fauna. Even with a perfectly focused image plane, artefacts, chromatic aberration, dust, dispersion in thick specimens, or the necessary aperture reduction in several techniques, reduce the image definition. Numerical aperture (f-stop) in macro or stereomicroscopy must be used judiciously in a delicate trade-off between image definition at the focal plane, which increases with larger apertures, and depth of field, which increases with smaller apertures.

yields 2272 x 1704 pixels. Chemical prints are scanned at variable resolution, from 300 dpi minimum for plain 4x5 Polaroid plates up to a maximum of 4800 dpi for 35mm slides. The frame grabber captures video output from the G channel at 752 x 512 pixels. These resolution settings try to capture as much detail as it is necessary to characterise the specimen adequately, without excess, unnecessary information.

Size Range and Resolution (cont'd.)

Relative pixel size is the measure that the image pixel represents in the real object plane, and is directly dependent on the number of pixels into which the image is divided. The image of a one-millimetre line that completely and exactly fits the frame and is one thousand pixels across, will have an one-micrometre relative pixel size. Denser sensors (more megapixels) imaging the same field of view render smaller relative pixel sizes.

The interplay between resolution power at the optical system, and pixel size, can be worked out in terms of the digital image "sampling" the contrast of the analog image (SPRING et al., 2004). Features in the focal plane can be thought of as alternate areas of light and shadow. The smallest possible separation of these areas corresponds to the diameter of the Airy disk, which in turn is a representation of the diffraction limit. The Nyqvist theorem can thence be invoked by considering the Airy disks as sine waves to be sampled in the spatial domain instead of the temporal domain. It can be boiled down to having to sample the image with a "probe" which is at least twice as small as this resolution limit in any one dimension (2.5 times is a practical figure in microscopy), the 'probe' being the pixel size, to allow for a faithful reconstruction of the analog image on the sensor array. ampling at smaller frequencies (e.g. with a pixel size comparable to the Rayleigh limit) may introduce aliasing artefacts masquerading as real features.

Thence, we postulate that ***digital images cannot show details finer than the resolution limit of the technique used, or twice the relative size of the pixel, whichever is greater.*** *Figure 4 can aid in this selection for optical techniques. At high optical magnifications, having a large number of pixels in the CCD becomes irrelevant and 1 Mpx cameras may suffice.*

Although SEM images can capture huge amounts of data in a single frame, in practice shots at various magnifications are taken. See figure 6 for a comparison between optical and electron microscopy on mesofauna taxa.

Software

Capture software is dependent on the camera being used. Normally, cameras or internal camera software are directly controlled from workstations and their local storage facilities are disabled, the signal being transferred via USB or RGB to the workstation. These are in most cases Compaq EVO D510 with Windows NT4 Workstation or Windows XP Pro operating systems.

- Nikon D100: Nikon Capture 3.0.0, USB transfer.
- Canon Powershot S45: Canon Utilities RemoteCapture 2.6.0 with WIA controllers, USB transfer.
- Nikon Coolpix E995: Images are captured directly on the camera CF card and read with a CF reader. Image control is done on the camera. Image monitoring is done on a SONY 15" Trinitron monitor fed by PAL composite video from the camera.
- JVC analog camera/Scion LG3 FrameGrabber: Scion Image for Windows Beta 3B.
- Point Electronics image system for SEM: DISS-5 for image capture and DIPS for processing.

Scale

Images from macro lens normally need an explicit scale, as the distances are variable. The scale is photographed at the same focal plane as the focused plane of the specimen, or, more generally, on the plane where the specimen lies. Pixel width is deducted from the scale by measuring it at post-processing. For fixed lenses, such as those of microscopes, a pre-tabulation of the objectives by a microscopic scale suffices for all pictures provided the objective used is recorded and the image is not resized.

Specimens that are imaged in 3D (i.e. SEM images of whole specimens at some angles, see fig. 7) cannot be scaled correctly at all points without parallax correction (see ARIÑO & GALICIA, this book).

Scales can only approximate dimensions, which must be calculated from landmark placement.

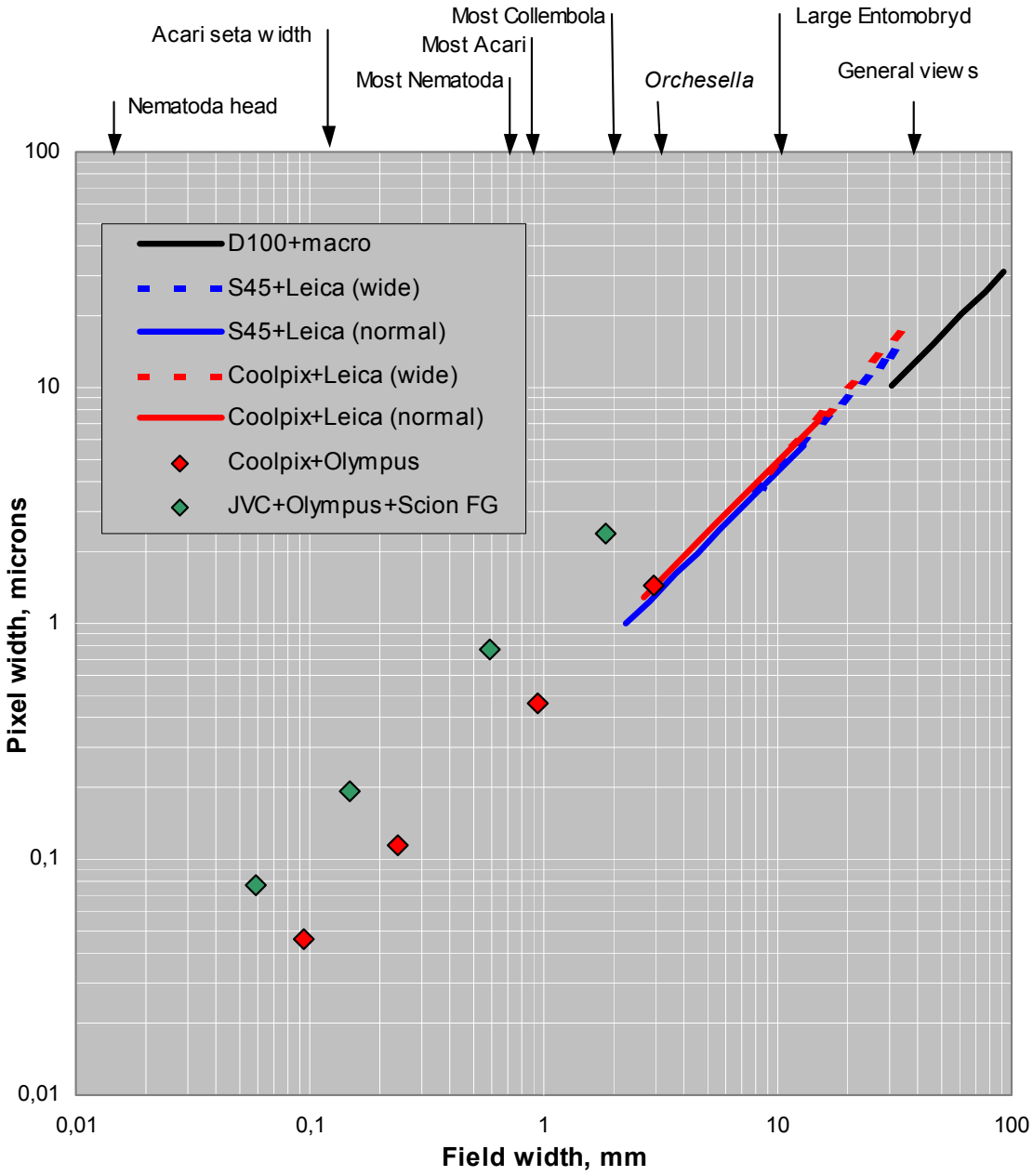
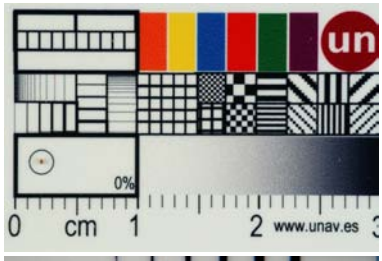
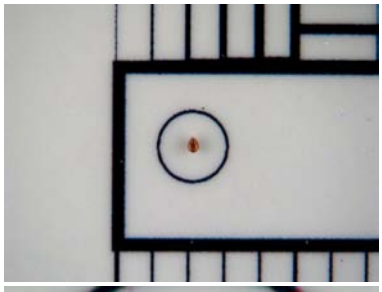


Fig. 4. Image width and ultimate pixel width for various combinations of imager and optics used in the case study. Note logarithmic axes.



Nikon D100 + 180 mm macro, 30.8 mm across, pixel 10.25 µm



Canon S45 + Leica stereomicroscope 10x, 10.5 mm across, pixel 4.64 µm

Canon S45 + Leica stereomicroscope 40x, 2.26 mm across, pixel 0.99 µm

Nikon E995 + Olympus BX50 microscope 100x, 1.03 mm across, pixel 0.5 µm. Insert corresponds to next image at current magnification.

Nikon E995 + Olympus BX50 microscope 400x, 0.23 mm across, pixel 0.11 µm (just below resolution limit of conventional optic microscopy).

Nikon E995 + Olympus BH50 microscope 1000x, 93 µm across, pixel 45 nm. Detail is limited by the microscope, not the camera.

Fig. 5. Images at various scales pertinent to mesofauna. Animals depicted are an oribatid mite (Acari) and a nematode.

Wish-list

As imaging of these animals involves a variety of observational techniques, future developments and ideal setups vary greatly.

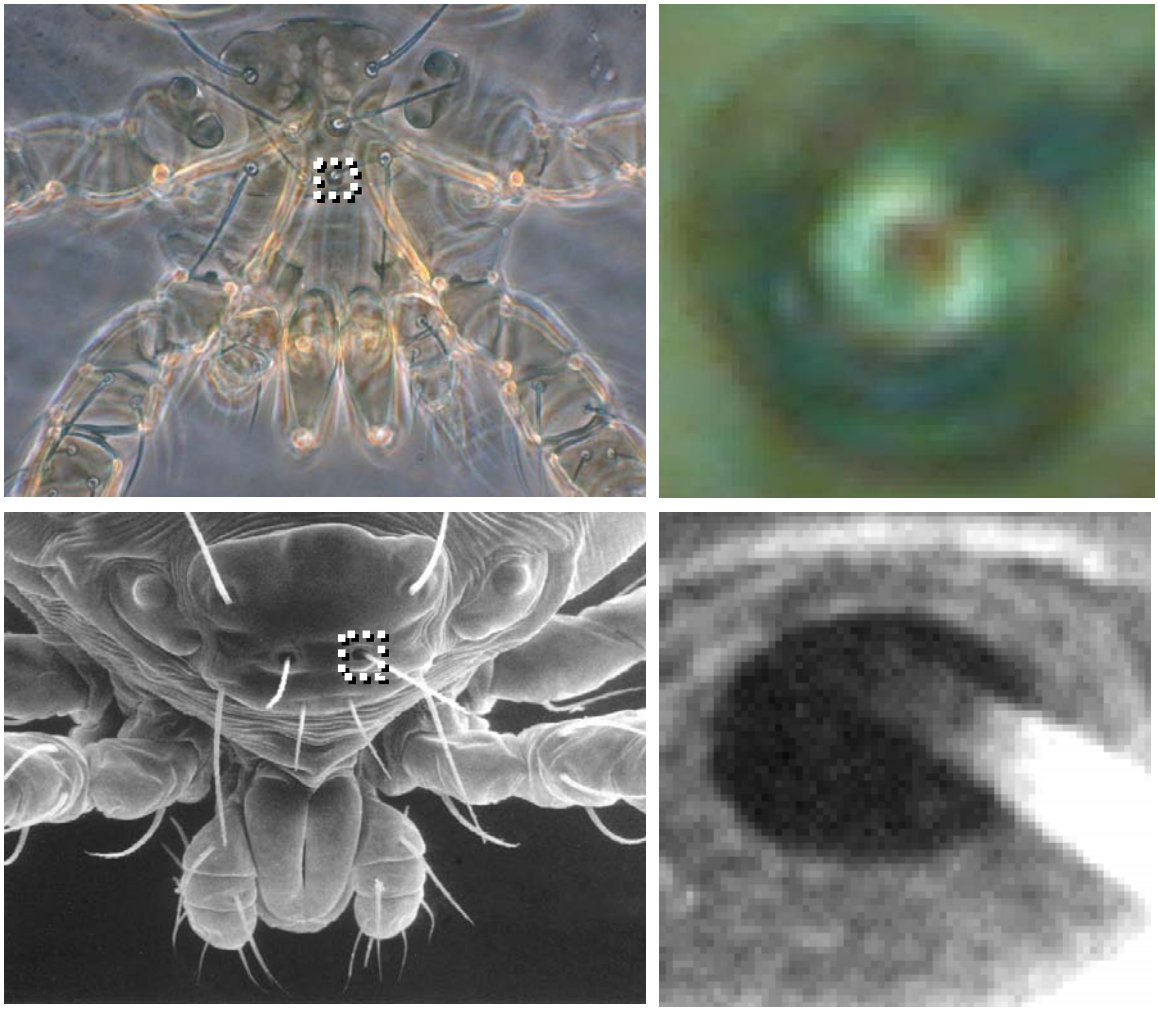


Fig. 6. *Polydiscia deuterosminthurus* (types), parasitic acari from a Collembola. Above: Nikon Coolpix E995 on Olympus BX50, 40x objective, phase contrast. Image 0.21 mm across, pixel size 101 nm. At right, blow-up of marked ROI, 66 pixels across. The smallest discernible features (bothridium) appear to be about five pixels (ca. 500 nm) across, which is congruent with the Abbe limit of the technique. Below: Same region captured on Zeiss DSM 940A SEM, Polaroid plate scanned with Agfa Arcus 1200. Image 0.15 mm across, pixel 74 nm. At right, blow-up of same ROI, 86 pixels across. Resolvable details are about 2-3 pixels (150-210 nm) for similar magnification.

Stereomicroscopy could benefit from fully planapochromatic optical trains with extreme high quality lenses. At typical maximum magnifi-

cation of 40x at the microscope, chromatic aberration may easily appear near the fringes of the image if the camera is used together with its own lens at full zoom in order to enlarge typically small specimens. Lens-less cameras with large CCD sizes coupled with high-magnification ocular lenses could greatly reduce the effect, at the cost of less definition in terms of pixel size.

SEM specimen preparation

Scanning electron microscopy images are becoming a very useful tool for the taxonomy of soil fauna. The ability to produce images at a very wide range of magnifications, with very fine detail and a large depth of field, enables the examination of taxonomically important surface features impossible by light microscopy. The technique, however, is still quite expensive and complicated compared with light microscopy. This prevents its routine use, but not its use for imaging type specimens, which, naturally, should warrant enough investment.

Notwithstanding special cases (live specimens at very low magnification in the so-called "environmental SEM"), successful SEM images can only be obtained if the full process of sample preparation and observation is done carefully. Clean (often ultrasonicated) specimens must be prepared to withstand high vacuum both during coating and during observation. Except for heavily chitinized taxa such as Oribatida, which can be dried directly, or morphologically compact animals having relatively impermeable cuticles, such as some Nematoda, that can be prepared by internal inclusion into resin, most soil fauna specimens have very fragile or soft teguments, and direct exposure to vacuum collapses them both from turgescence loss and from surface tension effects that appear when the water evaporates. The specimen must be completely dehydrated before entering the vacuum chambers; it could otherwise explode and/or prevent or delay the high vacuum to be achieved. But this must be achieved while ensuring that the specimen does not lose its morphological or taxonomic detail. A good, canonical histological fixation of soft specimens significantly contributes to good imaging.

A motorised, automatic stage would also ease the task of obtaining the frames of the stack. Although accurate Z-movement can be achieved by using the vernier markings of the focus knob, and auto-

montage algorithms are more robust against differences in slice position than deconvolution algorithms, a precise, repeatable positioning could improve the final performance.

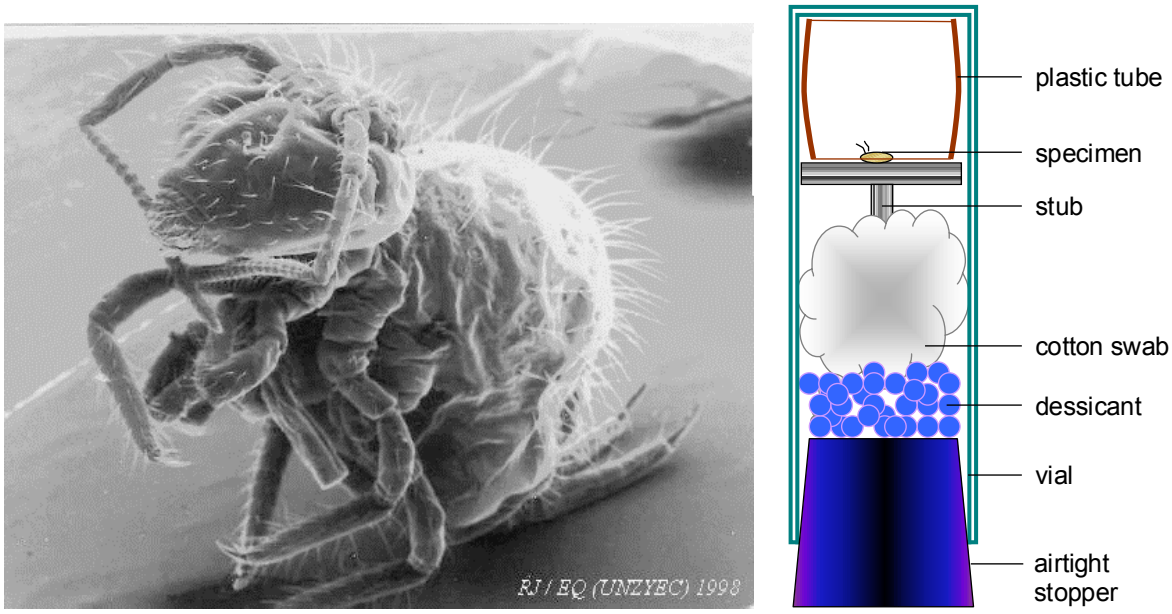


Fig. 7. A specimen of the Collembolan *Sminthurus viridis*, placed on the SEM stub at the angle adequate for imaging the feature of interest. Length of the animal about 1.5 mm. At right, individual desiccation and storage chamber for this type of mounts.

Alternate microscope techniques could be explored for new, better, or complementary imaging. Currently, observation and imaging of specimens is mostly dependent on contrast-enhancement techniques at the microscope: phase contrast and differential interference contrast. A move to confocal microscopy could potentially render well-focused images or 3D models of the internal structures of the animals, perhaps bridging conventional microscopy and scanning electron microscopy. It remains to be seen whether these techniques, which are best suited to fluorescence microscopy, could be widely used in the taxonomic field (where most imaging is done on already preserved and often mounted thick specimens) at an advantage over scanning electron microscopy. At least, image processing based on deconvolution techniques related to the ones used in these two fields might potentially be investigated in order to remove the strong brightness artefacts that badly hamper current automontage algorithms.

SEM specimen preparation (cont'd.)

Members of most taxa must undergo histological dehydration through the usual alcohol series. Critical point drying usually follows this process. The specimens' alcoholic fluids undergo a new change into liquid CO_2, either directly or through an intermediate change in acetone. This is done inside a pressure chamber at 40 bar. Once all alcohol (or acetone) has been thus removed, the chamber is heated (and pressure increased correspondingly) until the critical point of carbon dioxide is surpassed. At this moment, liquid becomes gas by phase transition, and no internal pressure or surface tension develops in the specimens. The chamber is flushed slowly, maintaining the gaseous phase, until atmospheric pressure at which point it can be opened. It is of paramount importance to prevent the gas phase reverting to liquid while flushing.

At this point, specimens are extremely light and are very delicate and sensitive to static charges. A static-free environment helps avoiding the specimens from "jumping" and being lost. Free air manipulation must now be kept to a minimum, and specimens must be stored inside dehydrating chambers. This allows for permanent archiving of the specimen itself, rather than just the pictures, as it was common in the past.

Dry animals are glued to an aluminium stub in a delicate operation performed under the stereomicroscope. Good observation can be guaranteed only if the interesting features are well exposed to subsequent coating and scan beam, and do not lie towards the stub surface. Often a small pole, such as the tip of an entomological pin or a deformable aluminium tape, is planted on the stub and the specimen is glued to the tip to ensure the widest possible coating and observation angle. The specimen is coated with a 8-20 nm layer of gold or gold-palladium in a sputter-coater. The thickness of this layer can be controlled, and contributes to the quality of the image. A thick layer can obscure details, but can also help the specimen to withstand thermal damage by the electron beam and improve conductivity. Although coated specimens are less prone to damage by environmental humidity, they are also stored in individual desiccation chambers (see fig. 7).

Another approach worth exploring would be any novel development in NSOM-based imaging for very small features. Currently, the

need for the probe to stay within a few nanometres of the sample seems to prevent its use with mounted or 3D complex specimens.

If storage space and time were not an issue, current SEM systems could produce extremely detailed slow scans up to 256 Mpx. However, other scan parameters such as voltage, coating thickness and beam alignment can affect image quality much more than the mere pixel size.

Working with Images

Storage Formats

Two basic image formats are used. Master images are stored as they are captured, in raw format (.NEF, .CRW, .TIFF) without any postprocessing, and copied onto removable media (CD-R and DVD-R). Secondary (working) images are produced from masters, at various levels of compression and resolution depending upon the intended use, and stored online. Table 4 shows the formats according to their source and purpose. Typical file sizes are about 4-9 MB for compressed masters, 1-2 MB for working files, and 400 KB, 100 KB, 40 KB and 10 KB for web varieties.

Software

Working software falls into three categories: post-processing, calculations, and storage/databasing management.

- Post-processing (cropping, enhancing, colour correction, filtering, automontage) is done with standard image packages such as Photoshop®, or specific-purpose analysers such as Image-J or CombineZ.
- Calculations on images are done with morphometrics packages such as MORPHEUS.
- Bulk managing and processing (batch resizing, trans-formatting, compressing, renaming), as well as examination, is done with an image database application (ThumbsPlus) or image managers (IrfanView). Metadata about images are captured by database managers directly from the database files created by image database (ThumbsPlus) or by EXIF extractors (ImageMagick).

Images are mounted generally with document editors in the context of their intended output or production document (Word, PowerPoint, Publisher, Dreamweaver, Flash, Acrobat).

Digitizer	Master	Working	Publication	FTP download	Web-inline	Catalog thumbnail (large)	Catalog thumbnail (small)
Nikon D100	NEF	Compressed TIFF	TIFF	95% JPEG	80% JPEG resized to SVGA (800x600)	70% JPEG resized to QVGA (320x240)	GIF resized to 1/8 (160x120)
Canon Powershot S45	CRW	Compressed TIFF	TIFF	95% JPEG	80% JPEG resized to SVGA (800x600)	70% JPEG resized to QVGA (320x240)	GIF resized to 1/8 (160x120)
Nikon Coolpix E995	TIFF	Compressed TIFF	TIFF	95% JPEG	80% JPEG resized to SVGA (800x600)	70% JPEG resized to QVGA (320x240)	GIF resized to 1/8 (160x120)
JVC + LG3	BMP	GIF	TIFF	95% JPEG	80% JPEG	70% JPEG resized to QVGA (320x240)	GIF resized to 1/8 (160x120)
DISS	TIFF	Compressed TIFF	TIFF	95% JPEG	80% JPEG resized to SVGA (800x600)	70% JPEG resized to QVGA (320x240)	GIF resized to 1/8 (160x120)

Table 4. Storage formats for the images produced. NEF: raw sensor output from Nikon DSLR; CRW: raw sensor output from Canon DC.

Linking to/within Databases

Images are not included in a database, but kept in directories under standardised file names including the accession number of the items. Two separate file systems are used: a NTFS namespace for the raw and working files, and a SMBFS that an Apache webserver uses for serving web-versions of the images.

Metadata from images, including pointers to the actual files, are included in a database that links to the specimen database of the Museum through the specimen's ID which is part of the file name of each image.

Imaging in SEM

Virtually any taxonomic work with the scanning electron microscope requires imaging. This instrument was originally designed to capture the image on film by slow scan, and the imaging techniques are well known. Good manuals exist for general SEM.

A good SEM image in taxonomy should show clearly the features of interest and not lack sharpness or quality, should have no extraneous noise, and should not show deformation or distortion of the specimen. Many factors may lead to the undesirable effects. A partial list would include inadequate photographic settings (focus, focal depth, contrast, brightness); improper accelerating voltage, probe diameter and current intensity, or astigmatism correction, equivalent to bad lighting conditions in optical microscopy; improper or off-centred objective aperture; instabilities on accelerating voltage or gun emissions (inadequate heating of filament); discharge of detector and column interior charge-up which may induce image drift; improper positional relation between specimen and detector or specimen tilting; excessive photomultiplier gain; or mechanical vibration.

These glitches are applicable to all types of objects. However, some factors seem to be particularly relevant in soil fauna, such as deformation of a specimen during its preparation, avoidable by careful dehydration and handling procedures (especially after critical point drying); dirtiness; charge-up of specimen surface, which may result from uneven metal coating due to the very complex spatial structure (hairs, bristles, scales) allowing for poorly coated spots that may break the electrical continuity of the specimen; or electron beam damage, due to local heating. Poorly transmitting heath, specimens with inadequate coating frequently do not resist damage caused by the well-focalized, high-voltage, intense beams that allow for fine detail or focus. It is often necessary to reduce voltage and current, and/or to scan rapidly with a shorter exposure time, in order to avoid damage to the specimen. This is of paramount importance when imaging type series specimens. But all these operations have the effect of reducing the overall quality of the image, resulting in poorer definition, sharpness and contrast, as well as more noise. Ideal images from these animals, thus, are usually below the standards for more robust insects.

Enhancing Images

Original ('master') images are always stored without any enhancing or post-processing, in raw (i.e. direct sensor output) format, normally off-line because of their large size. Several masters are taken and stored from one object, although only selected ones proceed to the working stage. From each master labelled as working, copies are produced that are enhanced if necessary, as long as this enhancement does not degrade the scientific validity of the image These include:

- Cropping. Image is cropped to eliminate irrelevant regions and reduce online size, especially when using fixed focal length optics such as microscope objectives that capture fixed-width fields.
- Generally, grey scale images are expanded to the full dynamic range by histogram equalisation. In addition, gamma values are adjusted subjectively.
- Colour correction is often not necessary, as monochrome or green-channel images are widely used.
- Multi-image stacks can be occasionally deconvoluted or FFT-filtered to remove random noise. However, filters are used very parsimoniously. They may enhance differences in detail, but also introduce artefacts: all filters change the brightness or colour values of each pixel. Conversely, some filters may erase details altogether. There is no given rule as to which filters produce a better view of already-existing features; trial and error (and careful cross-examination of the resulting image against its master) are necessary for each image being enhanced.
- Some images, especially SEM scans, are artificially coloured to highlight interesting features. This is done on 24-bit RGB copies of the raw, greyscale TIFFs. The specimen's background is negatively masked and the image is assigned a given hue, without any change of lightness. A wand-type selection tool is then used to encircle the ROI, and its colour is changed to a contrasting (or complementary) one, by inversion or any additive or XOR filter. Finally, the masked background is often filtered with edge-sensitive median filter and assigned a third, neutral hue. Figure 8 shows an example. It should be noted, however, that we generally publish the original, greyscale pictures in scientific papers

and only colour them for more popular outlets or web site posting.

- Publication images are finally resized and compressed according to the intended destination (web, paper).

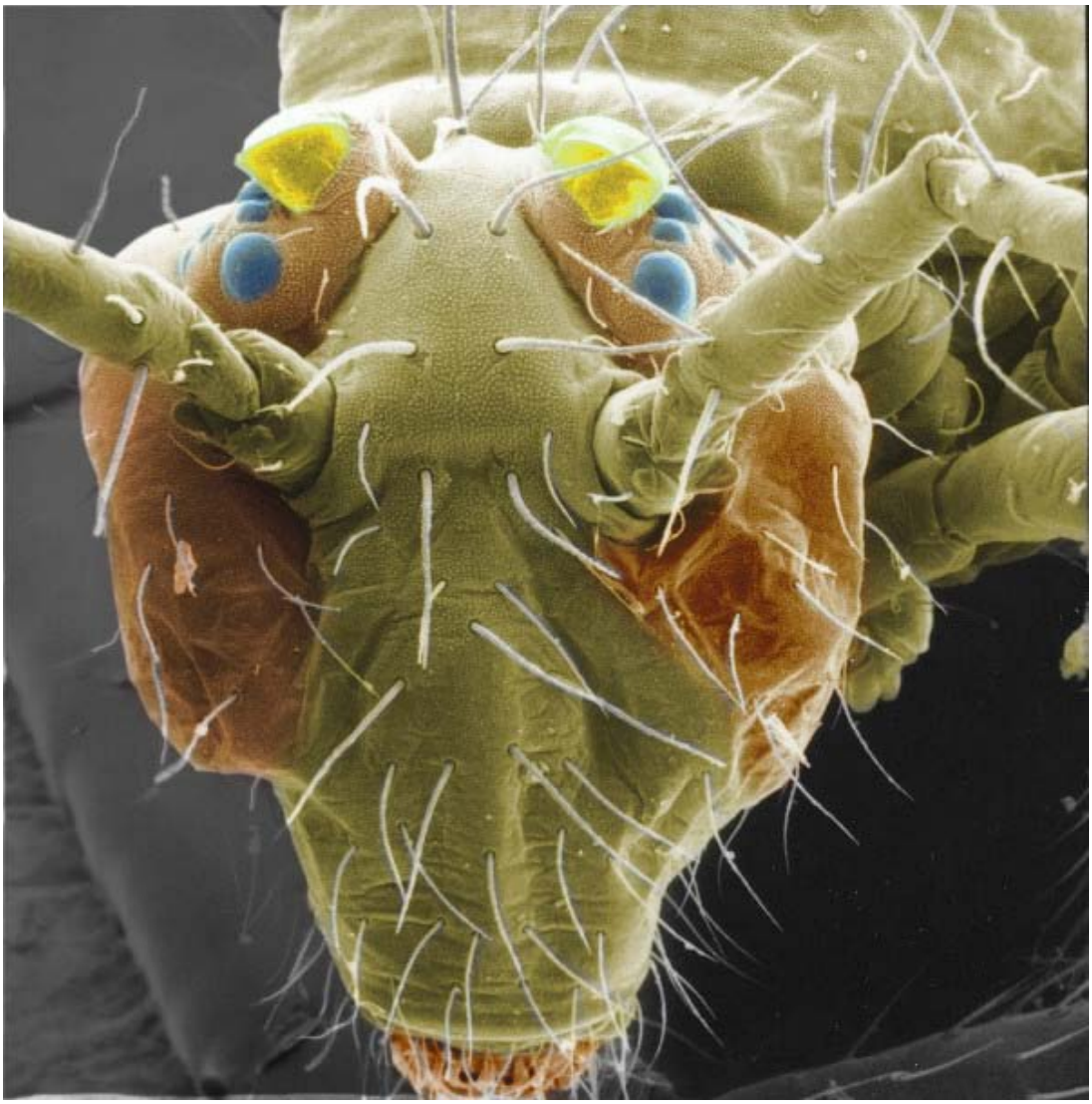


Fig. 8. *Vesicephalus europaeus* ARDANAZ & POZO, 1985 enhanced to show features of interest, suitable for popular publication. Masking and darkening have neutralized the background, and the specimen has been colorized to further detach it. The newly discovered photosensitive vesicles are shown in yellow to mark their location on the head relative to the known eyes (in blue). Scanned from a Polaroid plate. Image width is 1018 pixels spanning 0.73 mm (pixel width 0.71 µm).

Automontage

As automontage algorithms are not standard, different programs (ImageJ Stack Focuser class; CombineZ v.4.6, or AstroStack 3) are used. A stack is produced of several (in practice, up to a dozen) images from evenly distributed focal planes, i.e., fixed positions in the focus vernier knob. Images in the stack are checked for alignment, magnification and orientation, and matched and corrected if necessary by landmark placement. Several patch sizes between 4 and 25 pixels are tested, and the best overall result is selected.

This procedure is restricted to a few cases ideally suitable for the technique; in most cases, specific focal planes are preferable as images, for individual details are often more important than the overview. There are certain limitations to the technique that are very apparent in soil mesofauna specimens, especially related to the relative position of taxonomic features.

Quality Control

On-screen inspection of the master images against the visual counterpart at the microscope, and comparing the masters and the final deliverable images, are still the best QC check in use. Comparisons need to be made especially regarding the accuracy and visibility of taxonomically relevant characters. In automounted images, QC also involves correct alignment of patches. With SLR or compact cameras using their own lenses (D100, S45, E995) EXIF data are checked in order to ensure a correct focal length match between the image and that of the tabulating ruler.

Naming Images

Current naming conventions for images include the accession number of the specimen and various prefixes and suffixes denoting species, optics, magnification, and feature. These data are also included as metadata in the database.

Automontage in soil mesofauna

Montage techniques attempt to select well-contrasted patches from a stack of optical slices, generally by running Sobel filters, and assigning them a given depth in 3D space. A particular set of pixels in the image has to be chosen from the available slices, according to either its inferred depth or its relative contrast. Patches are stitched together in the final image, which is essentially a mosaic formed from well-focused pieces chosen from the slices forming the stack. Refinement at the position of the stitches, somewhat sophisticated in high-end packages such as Syncroscopy's AutoMontage, can result in a smooth, almost seamless image if the specimen is adequate for this technique.

High-magnification images of soil fauna are often inadequate for automontage. The imaging techniques used, especially phase contrast and differential contrast, produce very large lighting and position differences between focal planes, easily confusing this algorithm.

In addition, most soil mesofauna specimens are observed by transparency at different focal planes. Phase contrast and DIC allow for good focusing and separation of features: the ones on the focal plane being observed, i.e. the underside of the specimen, are clearly visible whereas features in the opposite side of the comparatively thick specimen remain totally blurred and are not discernible. After automounting, patches from all focal planes collapse into one, and features on one side 'seem to appear' along others in the opposite side. This can be confusing, especially in animals where the relative position and placement of the hairs and bristles (the chaetotaxy) is important for their characterisation. See fig. 9 for an example.

However, high-contrast specimens imaged with standard microscopy at medium or low magnifications can be automounted after some preparation of the master images, especially size matching and alignment, often by selecting a small portion (e.g. from dorsal view to middle view) of the stack (see fig. 10).

Metadata Information

Metadata should include all relevant data about the picture object and circumstance that are not to be found in other tables of the database such as the specimen table or taxon table. A partial field list includes:

- ID: Filename; GUID.
- Object: Accession number; Series number of the specimen (in multi-specimen accessions); Aspect, if applicable (i.e. dorsal, ventral); Target feature or part.
- Optics: Type of scope; Lens; Objective; Zoom; Contrast mode; Aperture; Voltage (SEM); Eyepiece; In-camera zoom.
- Camera: Type of camera; Mount position; EXIF Table Camera Fields.
- Take: EXIF Table Take Fields.
- Image: Size; Pixel Width; Original Colour Depth; Gamma.
- History: Parent filename (for derived images); Series number (for stack frames); Original size; Post-processing sequence (filters, adjustments, feature enhancing, colour correction).
- IPR: Photographer; Date of take; Post-processor; Date of change; Copyright date; Release policy; Permissions.

IPR policy for Images

There is no common IPR policy for images, although some rules apply. Copyright is always enacted for the photographer. Images for publication on paper are released to the publisher as mounted copies of the masters and IPR retained in all cases. Downsampled images released through the web server retain IPR but their copy or reposting is allowed, provided that the IPR notice is kept and a backlink is included to the original location, for scientific or educational purposes. Any commercial usage of any image is subject to a written agreement and fee or royalty payment.

Wish-list

Some good automontage software packages run into the K$s (i.e. AutoMontage by Syncroscopy). Although less expensive, or even free, software (i.e. AstroStack, CombineZ, ImageJ) can attain similar effects, the routine usability and throughput of the more expensive programs seems better. Any automontage program, however, cannot be used successfully in many microscopy images of the types adequate for soil animals without prior extensive retouching and masking of the frames, which must be done by hand. Thus, adequately

trained staff are probably the most interesting investment in the production of quality images.

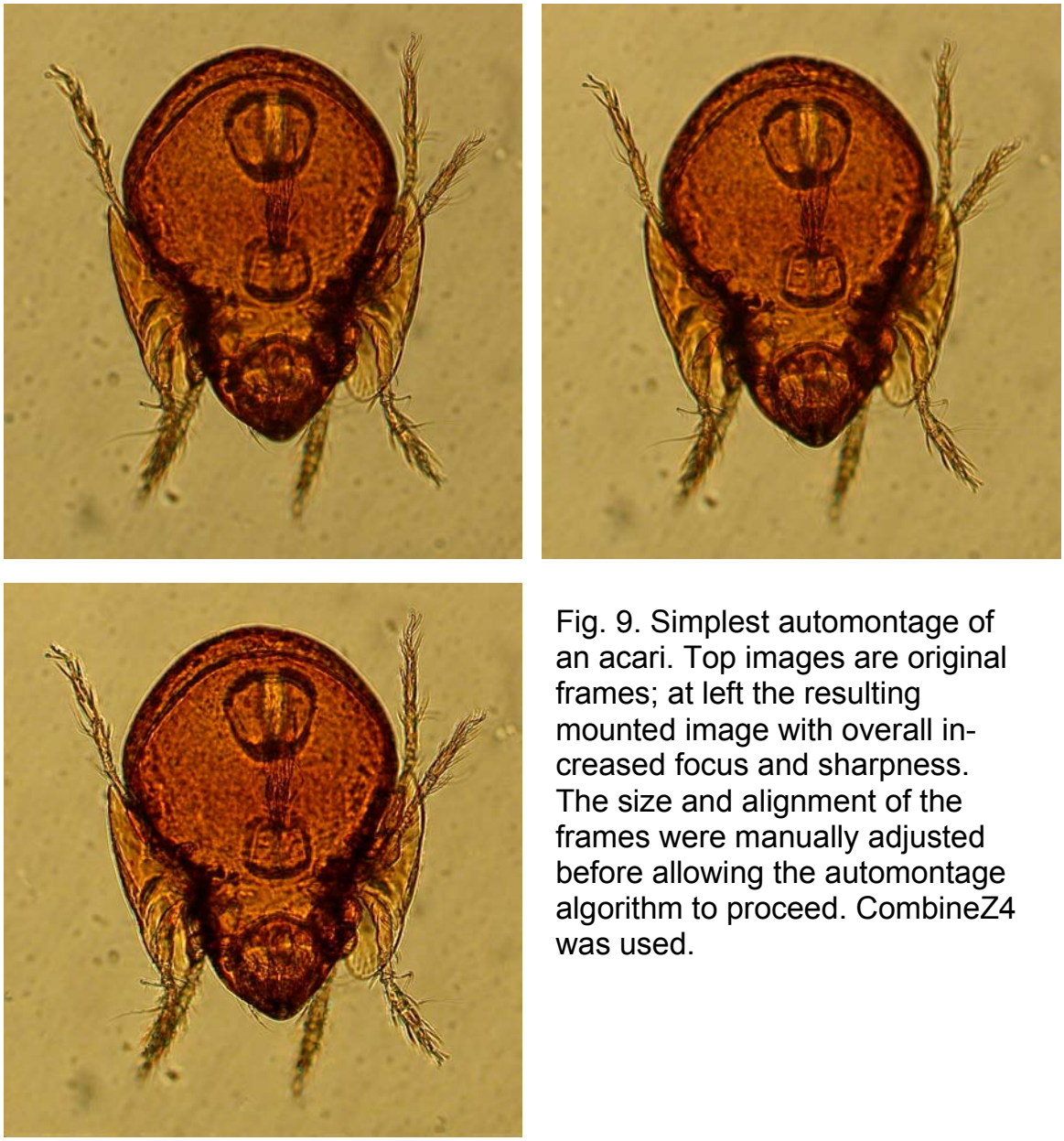


Fig. 9. Simplest automontage of an acari. Top images are original frames; at left the resulting mounted image with overall increased focus and sharpness. The size and alignment of the frames were manually adjusted before allowing the automontage algorithm to proceed. CombineZ4 was used.

Another immediate development should probably deal with image format. The relatively new JPEG-2000 standard allows a much more versatile image management in a distributed environment (see MORRIS, this book). Both the ability to define and hyperlink regions of interest within the picture, and the “lossless” option of this format, could lead to a new concept of image storage and annotation.

An affordable image server having the ability to produce and serve on-demand both scaled-down versions of the full images and full-quality ROIs based on that format could also be of much interest.

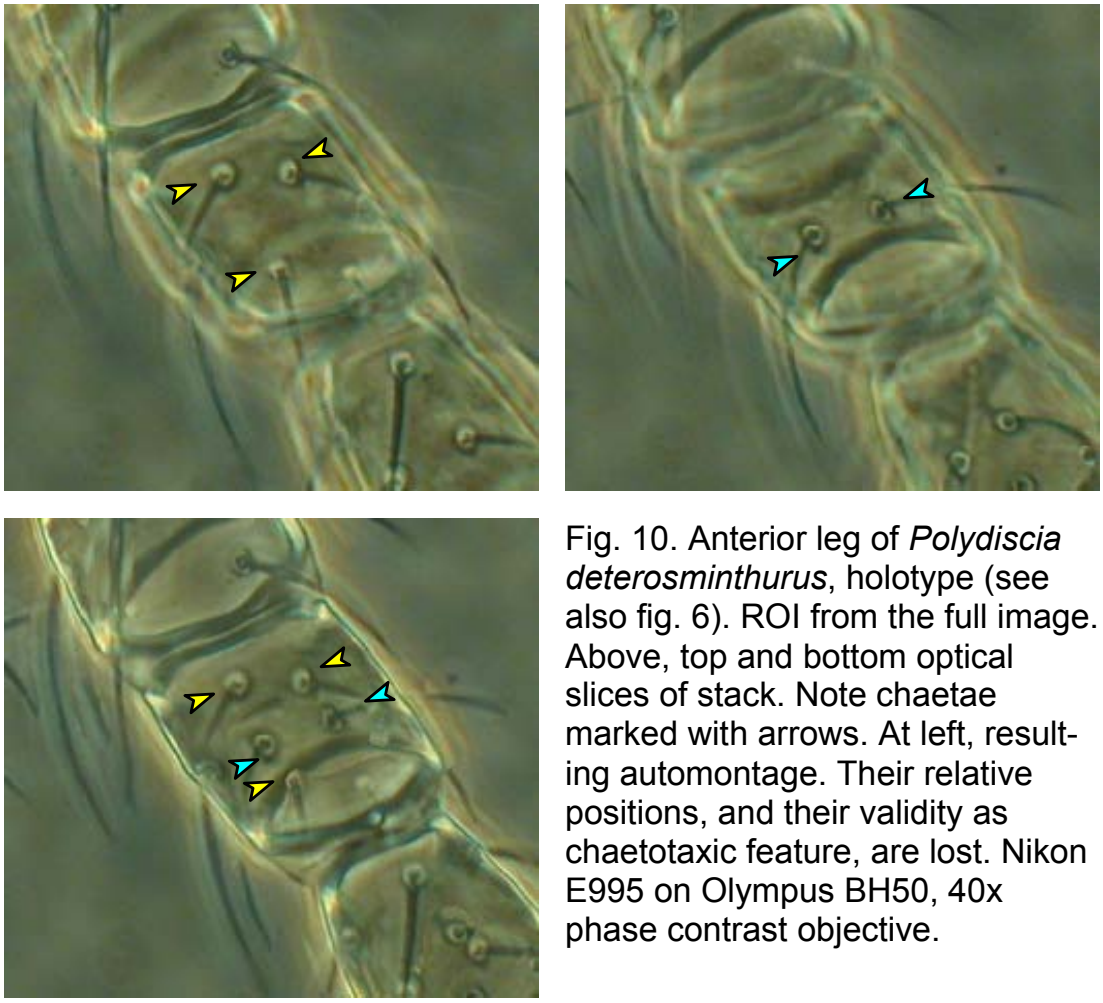


Fig. 10. Anterior leg of *Polydiscia deterosminthurus*, holotype (see also fig. 6). ROI from the full image. Above, top and bottom optical slices of stack. Note chaetae marked with arrows. At left, resulting automontage. Their relative positions, and their validity as chaetotaxic feature, are lost. Nikon E995 on Olympus BH50, 40x phase contrast objective.

Maintaining databases

Master databases are kept in-house, although copies are integrated within the general server system of the Institution. In general, photographers are the researchers themselves and they are also in charge of entering the metadata into the database. However, design, maintenance, backup and migration of the databases is done on a more comprehensive, general level by the systems administrator and DB manager.

The master DB resides in an NTFS and requires authorised access.

Back up

Master databases are fully and automatically copied by programmed tasks twice a day via LAN into four alternating backup servers. At any moment, an original and four copies with a half-day lag between them are available on-line in order to prevent logical damage, breakdown or error propagation. Discretional copies are also made irregularly to a fifth backup server before major changes. Copies of the masters for web publication are maintained in a RAID system that also follows its own daily copy routine. Database status and access is logged daily. Deletions and additions are journalled, and changes are ID- and time-stamped. Additionally, permanent, weekly snapshots of the entire modified tables are produced and stored both online and off-line. Current off-line backup method is CD-R, although MFM and ZIP were also used. Weekly CDs are stored outside the premises of the laboratory in order to prevent loss by local, physical disaster. Every six months a full copy of the DB is burned onto removable media and sent to a safe place in a distant city in order to prevent loss from natural disaster.

Images themselves are also stored at a central NTFS file server, and copied daily into two alternate backup servers. Every file that is created or modified is also copied into CD-R and stored outside. Master, raw images are also copied into DVD for primary storage after working copies have been produced.

Storage capacity

Currently the system reserves 160 GB from each of the primary and backup servers, but this includes the ongoing imaging of specimens as well as that of type series. Off-line storage is essentially unlimited. At the time of writing the off-line accumulated storage mass (including all filetypes) is 237 CD-R in compressed format.

Maintaining links

A policy of dead link search is enacted irregularly as a part of the QC check.

Linking to other systems

The intradepartmental network is linked to the general, institutional network through routers. A copy of the database and files is fed to a separate file server that includes a web server for outside access. In addition, partial copies of the databases are fed to the GBIF node in Madrid, Spain.

Wish-list

The overhead of maintaining and cleaning-up the databases and image files increases with the size and number of files. Dedicated staff, including DB managers, would be a good, though not cheap, investment.

Acknowledgements

We wish to acknowledge the help and assistance of MAITE MARTÍNEZ ALDAYA, JUANI CRUCHAGA, Mª ARÁNZAZU IMAZ BARANDIARAN, Mª LOURDES MORAZA, DAVID GALICIA and JAVIER OSCOZ at the Department of Zoology and Ecology and Museum of Zoology of the University of Navarra. We are also much indebted to MALCOLM SCOBLE, Natural History Museum, for editorial and language comments.

This work has been partially supported by the Research Plan of the University of Navarra.

References

ANONYMOUS: *A Guide to Scanning Microscope Observation*. JEOL Corp.

ANONYMOUS: Point Electronics DISS 5 Product Sheet. http://www.pointelectronic.de/english/products/diss5_english.htm [accessed 1.IX.2005].

BELASCOAIN, C., ARIÑO, A.H. & JORDANA, R.A. (1998): A new integrated extraction method for microarthropods and nematodes from de same soil samples. – *Pedobiologia*, **42**: 165-170.

CAVENESS, F.E. & JENSEN, H.J. (1955): Modification of the centrifugal-flotation technique for the isolation and concentration of nematodes and their eggs from soil and plant tissue. – *Proceedings of the Helminthological Society of Washington,* **22**: 8789.

CROMDEY, D.W. (2004): Digital Imaging: Ethics. http://swehsc.phamacy.arizona.edu/exppath/resources/pdf/Digital_Imaging_Ethics.pdf [accessed 1.IX.2005]. The University of Arizona.

FLEGG, J.J.M. & HOOPER, D.J. (1970): Extraction of freeliving stages from soil. In: SOUTHEY, J.F. (ed.): *Laboratory methods for work with plant and soil nematodes*.

HADLEY, A. (2004): CombineZ Version 4.6. http://www.hadleyweb.pwp.blueyonder.co.uk [accessed 1.IX.2005].

HAYDEN, J.E. (2000): The Ethics of Digital Manipulation in Scientific Images. – *Network Journal of Illustration,* **5-4**: 11-18.

HOOPER, D.J. (1986): Handling, fixing, staining and mounting nematodes. In: SOUTHEY, J.F. (ed.): Laboratory methods for work with plant and soil nematodes. – *Commonw. agric. Bureaux, Techn. Bull*., **2**: 5-22.

SPRING, K.R., RUSS, J.C., PARRY-HILL, M.J., FELLERS, T.J., ZUCKERMAN, L.D. & DAVIDSON, M.W. (2004): Digital Image Sampling Frequency. In: DAVIDSON, M.W. & al.: *Molecular Expressions: Optical Microscopy Primer. Digital Image Processing.* – National High Magnetic Field Laboratory and Florida State University Research Foundation. http://www.microscopy.fsu.edu/primer/java/digitalimaging/processing/samplefrequency/index.html [accessed 1.IX.2005].

STEKELENBURG, R.J. (2001): *AstroStack 3 Manual*. – Innostack.

UMORIN, M. (2002): Stack Focuser (ImageJ Module). http://rsb.info.nih.gov/ij/plugins/stack-focuser.html [accessed 1.IX.2005].

WEYDA, F. (1992): Simple desiccation method for scanning electron microscopy using dimethoxypropane. In: BAILEY G.W., BENTLEY J. & SMALL, J.A. (eds.): Proc. EMSA/MAS/MSC/SCM, p. 760.

Digital Imaging of Beetles (Coleoptera) and Other Three-Dimensional Insects.

ALEXANDER RIEDEL

Staatliches Museum für Naturkunde Karlsruhe, Erbprinzenstr. 13, D-76133 Karlsruhe, Germany. E-mail: riedel@smnk.de

Key words: Digital Imaging, Coleoptera, insects

Introduction

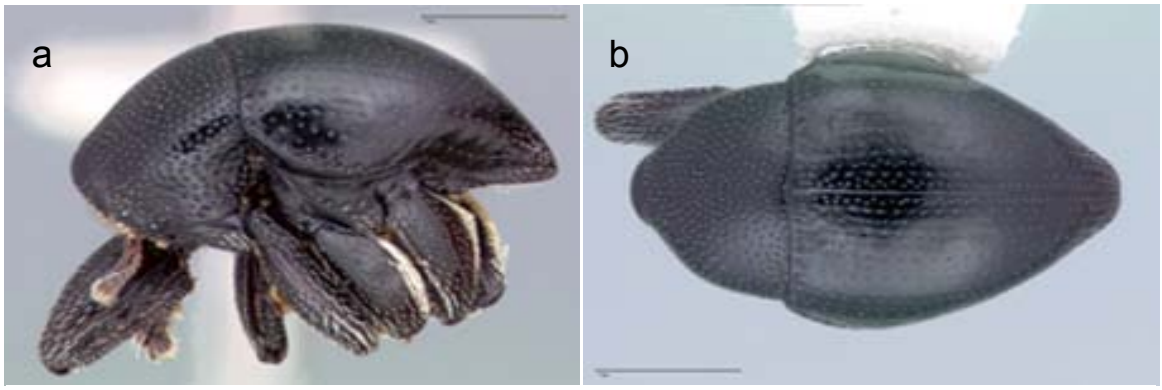
This chapter focuses mainly on imaging techniques for beetles. Usually, beetles impose special problems on the photographer, such as a highly reflecting surface and / or a body that is too convex for the depth of field available. In this respect they are quite different from the rather two-dimensional (when set) Lepidoptera but similar to other groups of insects, such as ants, certain Heteroptera, or some Diptera. The general aspects of photography of such insects will be covered here; however, only the beetles will be treated in depth.

Beetles exhibit a considerable range of body size. Members of the Ptiliidae may be as small as 0.25 mm, whereas *Titanus giganteus* (Cerambycidae) may exceed a length of 160 mm. Thus, the photographic techniques recommended will need to vary with the subjects. As I have mainly worked on beetles of intermediate size I will concentrate on describing the imaging techniques employed for beetles between 1 mm and 10 mm. This size-range forms the bulk of beetle diversity. For photographing beetles of truly microscopic size (such as some Ptiliidae), some useful hints may be drawn from other chapters of this manual, such as “Imaging Soil Mesofauna: The Land in Between” by ARIÑO, BAQUERO & JORDANA.

Mounting of specimens and choice of the background

There are two different traditions of how a beetle is mounted. Most Anglo-Saxon Coleopterists prefer to pin beetles, generally down to a relatively small body size. Specimens too small for pinning are mounted on of cardboard points, with the legs hanging more or less unarranged below the body (figs. 1a-b). In continental Europe the preference is to glue specimens onto rectangular pieces of cardboard. Legs and antennae are arranged symmetrically around the body, so that a rather natural appearance of a sitting beetle is achieved. This method is usually applied to specimens smaller than 20 mm, but occasionally larger beetles are prepared this way. Collectors of beetles as large as *Carabus* may use this method of card-mounting to avoid piercing the specimen with an insect pin and to prepare them in an aesthetically appealing manner. Furthermore, this method provides increased protection to the specimen because its appendages are protected by the card to which it is attached. The point-mounters however, reject this method because the card conceals the entire ventral surface of the beetle. Moreover, it requires tedious manipulation, that may damage the specimen. Both procedures have their merits, and it is impossible to recommend one over the other.

The different mounting methods have implications for imaging and post-processing. While it is possible to choose any kind of background for a pinned or a point-mounted specimen, no such choice is possible for a card-mounted beetle once attached. Correct exposure and imaging of delicate hairs and setae of a specimen will



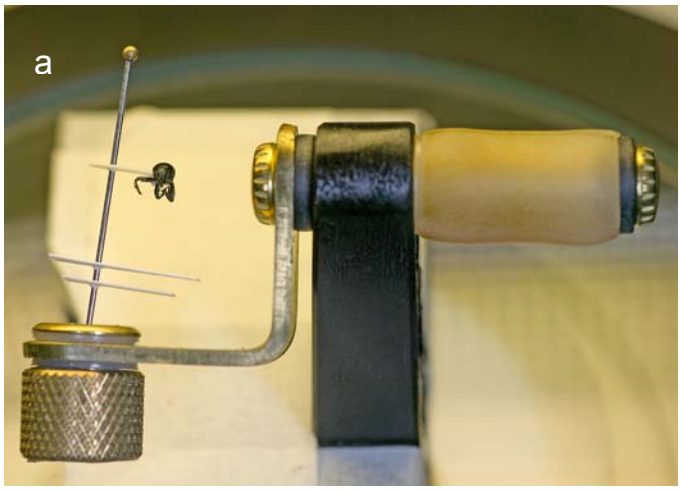
Figs. 1 a-b. Example of a point-mounted weevil (Curculionidae, *Idotasia* sp.) in lateral aspect (a) and in dorsal aspect (b).

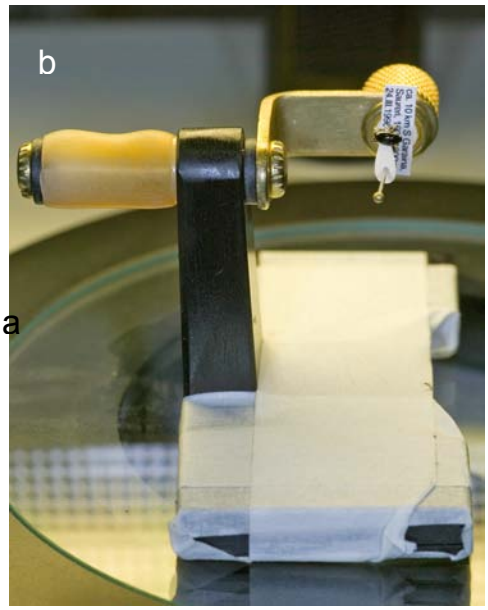
be easiest if it is photographed against a neutral grey exposure and imaging of delicate hairs and setae of a specimen will be easiest if it is photographed against a neutral grey background. Thus, if pinned or point-mounted beetles are the subject, it is recommended that they are photographed against a neutral-grey card (fig. 1). On the other hand, the white background of a card-mounted specimen can be turned to advantage. It is quite easy to ‘cut’ a beetle's image from its white background during digital post-processing. And, if the beetle is mounted adequately, its dorsal aspect will reveal more characters than a specimen that is point-mounted. Furthermore, it is much easier to position it into a precisely horizontal plane. Thus a comparison of characters and body-proportions is much easier when based on images of card-mounted specimens.

The photographer usually has little choice between the two methods. Museums in the US largely hold pinned and point-mounted specimens while those in continental Europe have the specimens mounted on white cards. If the purpose of the imaging project is to create a large scale image-library, manipulation of the specimen should rarely be considered a viable option, especially if type specimens are involved. The situation is different if a specialist taxonomist is undertaking imaging as part of revisionary research. In such a situation the specimens will have to be handled anyway and it is appropriate to use the preferred method of preparing the specimen.

My personal approach is to point-mount the majority of the specimens I am working on but I mount one or two specially selected and carefully cleaned specimens of a series on relatively large white cards for the specific purpose of imaging. These card-mounted specimens can be photographed in an aesthetically pleasing manner that makes more characters visible than the point-mount method could.

A card-mounted specimen is usually photographed dorsally and it is sufficient simply to use a piece of Styrofoam for holding the insect pin. Point-mounted or pinned specimens can be photographed from various aspects. For this purpose, it is very useful to place them on a microscope stage (fig. 2) that allows their rotation around three spatial axes.





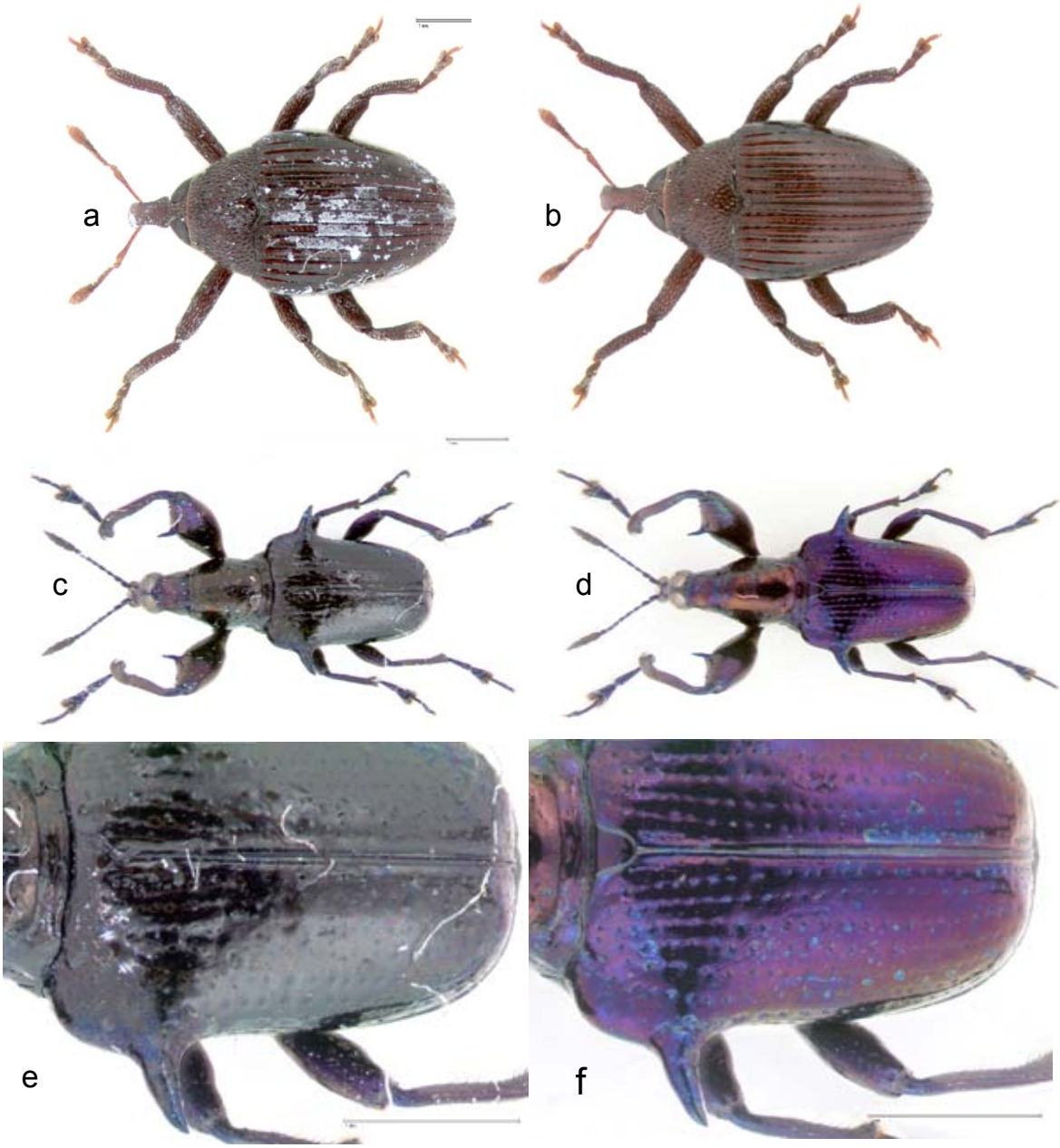
Figs. 2 a-b.
A point-mounted specimen on a microscope stage moved into different positions.

Cleaning specimens

Unless cleaned by the collector, certain beetles may still be covered with original layers of soil, wood dust etc. even after years in a collection. Also, stored specimens inevitably collect dust - even in dust-proof drawers. Grease may emerge from the body of a stored dry beetle and accumulate in layers (often thick) on its integument, thus concealing microsculpture and colour. In some cases, white crystalline substances accumulate on the beetle's surface.

The question arises as to whether specimens should be cleaned before imaging. Here again, we face a dilemma. Excessive cleaning may severely damage a specimen. Also, potential information on the environment of the beetle (in the case of adhering soil / wood dust) will be removed during the cleaning process. Yet some specimens may be covered by such quantities of dirt, dust or grease as to render them hardly recognizable. Two examples illustrate this, each one a weevil covered by white exudate (fig. 3a) and grease (figs. 3c, d), before and after the cleaning (figs. 3b, d, f).

A hands-off approach is recommended for large scale imaging projects, unless the specimen is in a barely recognizable condition. In such a case the curator should decide on adequate measures. In smaller scale projects, such as revisions, where the scientist is handling the specimens, it is recommended that specimens are



Figs. 3 a-f. Examples of beetles before and after cleaning with organic solvents. (a-b) unidentified weevil with white exudate that appeared after 12 years of storage; before and after cleaning. (c-f) specimen of *Euops yali* with grease covering the surface after 13 years of storage; before and after the cleaning.

carefully cleansed of dust and grease if necessary. This should be done with a very fine artist´s brush and an organic solvent such as ethanol or ethylacetate. The legs and antennae should not be touched without first relaxing the entire specimen since these structures break all too easily. The same comments apply to beetles with hairy or setose integuments. Oil from the body of glabrous

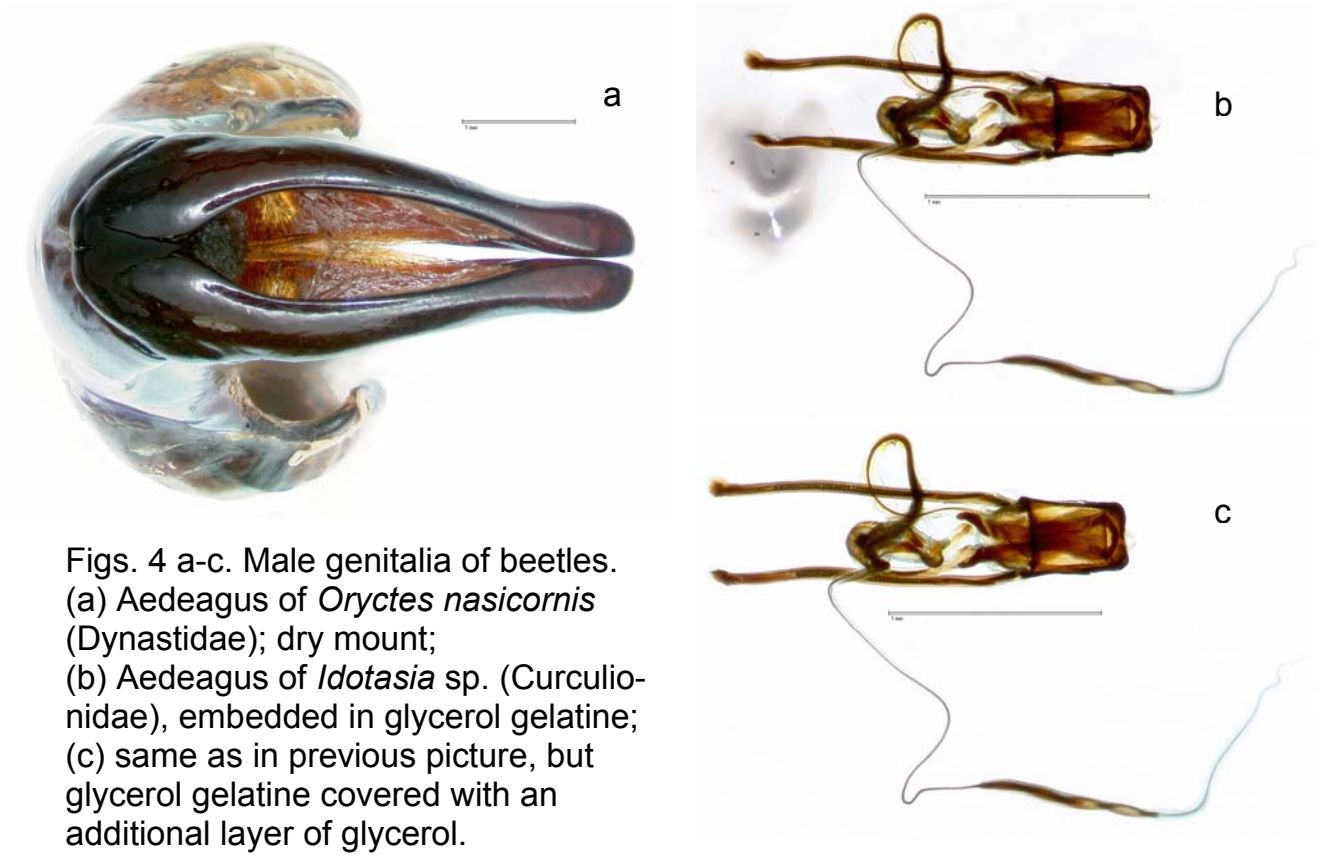
specimens should be carefully removed from dry, card-mounted specimens. The effects of successful cleaning are immediately obvious in specimens that exhibit metallic colouration. Even when just a thin layer of oil is removed from the surface, the change is dramatic with the original appearance restored.

Preparation of beetle genitalia for photography

The diagnosis of many beetle species relies heavily on characters of the genitalia. It may, indeed, be more important to photograph the relevant genital structures than to provide an illustration of the beetle's habitus. There follows therefore some comments on the final preparation of the genitalia for photography. Appropriate techniques for extracting and preparing genitalia depend significantly on the taxon in question. An expert for the specific group should be consulted for advice and help. Dissecting specimens for the extraction of the genitalia should only be done by persons with extensive experience.

In some groups the genitalia are heavily sclerotized and the diagnostic characters are mainly external (e.g. members of the Dynastidae). In this case, the genital can be mounted on cardboard and photographed in dry condition. Lighting and imaging techniques are more or less the same as applied for photographing the entire insect (fig. 4a).

In other cases also internal structures are relevant, i.e. structures of the endophallus that are within the body of the aedeagus. The same applies if membranes cover parts of the genitalia. These membranes while transparent in water, will conceal the view when they are dry. In most such cases it is necessary to clear, in KOH, the genitalia of adhering tissue and examine them in glycerol. A specialist in the relevant taxon should be consulted for precise instructions on dissection and staining. In most cases transmitted light microscopy is best for observing and illustrating the relevant characters. If the size of the specimen is not too great, a compound microscope is preferable to a dissecting microscope. Unless the structures are sufficiently flat for a cover-glass to be used, they are best examined on cavity slides. The main problem that arises when taking photographs of ob-



Figs. 4 a-c. Male genitalia of beetles.
(a) Aedeagus of *Oryctes nasicornis* (Dynastidae); dry mount;
(b) Aedeagus of *Idotasia* sp. (Curculionidae), embedded in glycerol gelatine;
(c) same as in previous picture, but glycerol gelatine covered with an additional layer of glycerol.

jects stored in glycerol is the difficulty of fixing them in a desired position. This is especially true for genitalia of a curved shape, with appendices etc. The best way to overcome this problem is the use of glycerol gelatine. There are various recipes, but the most suitable seems to be the one given by KISSER (1934), which contains less water than the others. The beetle genitalia are fully prepared and stored in glycerol. A small quantity of glycerol gelatine is then placed in the cavity of a slide and heated until it melts and the genitalia are placed into this drop; it is left to cool down while the genital are arranged and held in the desired position with forceps and / or pins. When the glycerol gelatine has coagulated the instruments should be carefully removed (fig. 4b). If necessary, the mount should be covered by a quantity of glycerol to create an even surface (fig. 4c). Such a mount can be photographed most easily and is stable enough to create stacks of images. Later, the genital can be removed easily from the block of glycerol gelatine by placing it into water for some time.

In some cases structures of the endophallus are so complex that even the optical properties of glycerol are inadequate to resolve all details. In such cases another medium should be chosen, such as Canada Balsam or Euparal. The latter is most suitable if the structures are flat and a full mount with a cover glass is made. Unfortunately, this medium dries very slowly, so the objects may be drifting for quite some time. Canada Balsam dries faster and preparations without a cover glass are sufficiently stable after 2-3 days. The position of the object can be carefully corrected during the hardening process and thus it is also suitable for curved objects that are otherwise hard to position. Care must be taken to ensure that the Canada Balsam is not acidic, otherwise it may adversely affect the embedded genital.

Photography of larger beetles

A few general comments should suffice on digital photography of beetles larger than ca. 15 mm. They are usually pinned and comparable in size to butterflies so the chapter on Lepidoptera should be consulted for additional information. A standard digital camera seems most appropriate for the purpose of photographing such species; the best being a digital SLR camera equipped with a 50 to 100 mm macro lens. The camera should be attached to a reprostand or a focusing rail. At the time of writing, the Canon EOS 20D would appear to be a good choice with its 8.2 Megapixel sensor. Another interesting option with this camera is the Canon macro lens MP-E 65, for the range of natural size to 5X magnification. This lens bridges the gap between macro- and micro-photography. The MP-E 65 will work only for specimens smaller than 20 mm. For larger specimens an ordinary macro lens is preferred. Lenses of longer focal length give a greater distance between insect and camera, which is helpful for manipulation. But, it should be kept in mind that the distance necessary to fit a large beetle into the frame might exceed the height of the reprostand or the tripod. Therefore, the decision as to whether a 50mm lens or a 100mm lens is used will depend on both the setup and the size of the beetle. A universally appropriate recommendation cannot, therefore, be given. Unlike the

situation in butterflies, getting even illumination can be a challenge for beetles, particularly those many species with a smooth, highly reflective surface. In such cases, indirect, diffused lighting must be applied. The use of softboxes attached to the lights / strobes is one option for these larger specimens.

Photography of medium-sized and small beetles

The majority of beetles measure between 1 mm and 10 mm. Usually, a digital camera attached to a high-end dissecting microscope will be the best approach to this size-range. However, the use of microlenses combined with camera bellows (fig. 5) is a little-known alternative. This equipment will be especially attractive in those cases where a digital SLR camera is available but not a dissecting microscope. Micro lenses, such as the Zeiss Luminar, or the Leica Photar can be connected to most digital SLR cameras through suitable adapters. With the Leica Photar 25mm/f2.0 magnifications of 5X to 22X can be achieved with an optimum performance at 6.6X. The camera must be mounted on a precise focusing rail. Such a combination will require some practise to work with and is certainly not as easy to handle as is a camera mounted on a microscope. The quality of the resulting images, however, can be outstanding, so such an arrangement is an option worth considering. This technique has been applied by H. SCHILLHAMMER (Naturhistorisches Museum Wien) and samples can be seen at http://www.pbase.com/rovebeetle/mostly_beetles.

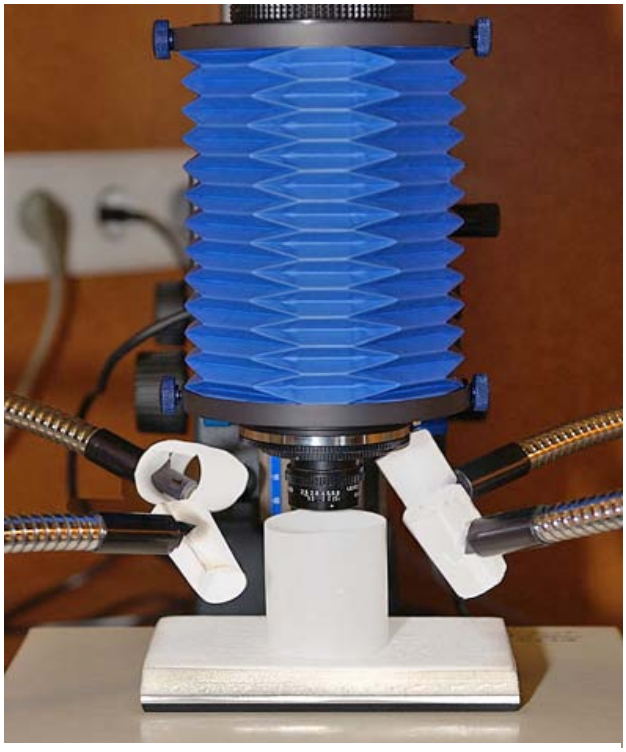


Fig. 5. Setup of camera bellow and microlens. Courtesy of H. SCHILLHAMMER (Naturhistori-sches Museum, Vienna).

The depth of field in photographs increases with decreasing (smaller) aperture (higher aperture number); but, at the same time, lens artefacts caused by diffraction increase, becoming more pronounced when the magnification increases. For magnifications higher than 1.5X, apertures smaller than 8.0 or 11 should be avoided when using an SLR camera. The same is true of a dissecting microscope where the iris should be kept open. In digital photography of still objects the depth of field should be expanded by using adequate computer software (See below “Montage software”).

Some general aspects of digital photography of insects are covered by ASHWORTH & FOGARTY (2003); thus in the following, only some additions to the cited publication are given and some idiosyncrasies of beetle-imaging are described in greater detail.

Capture device

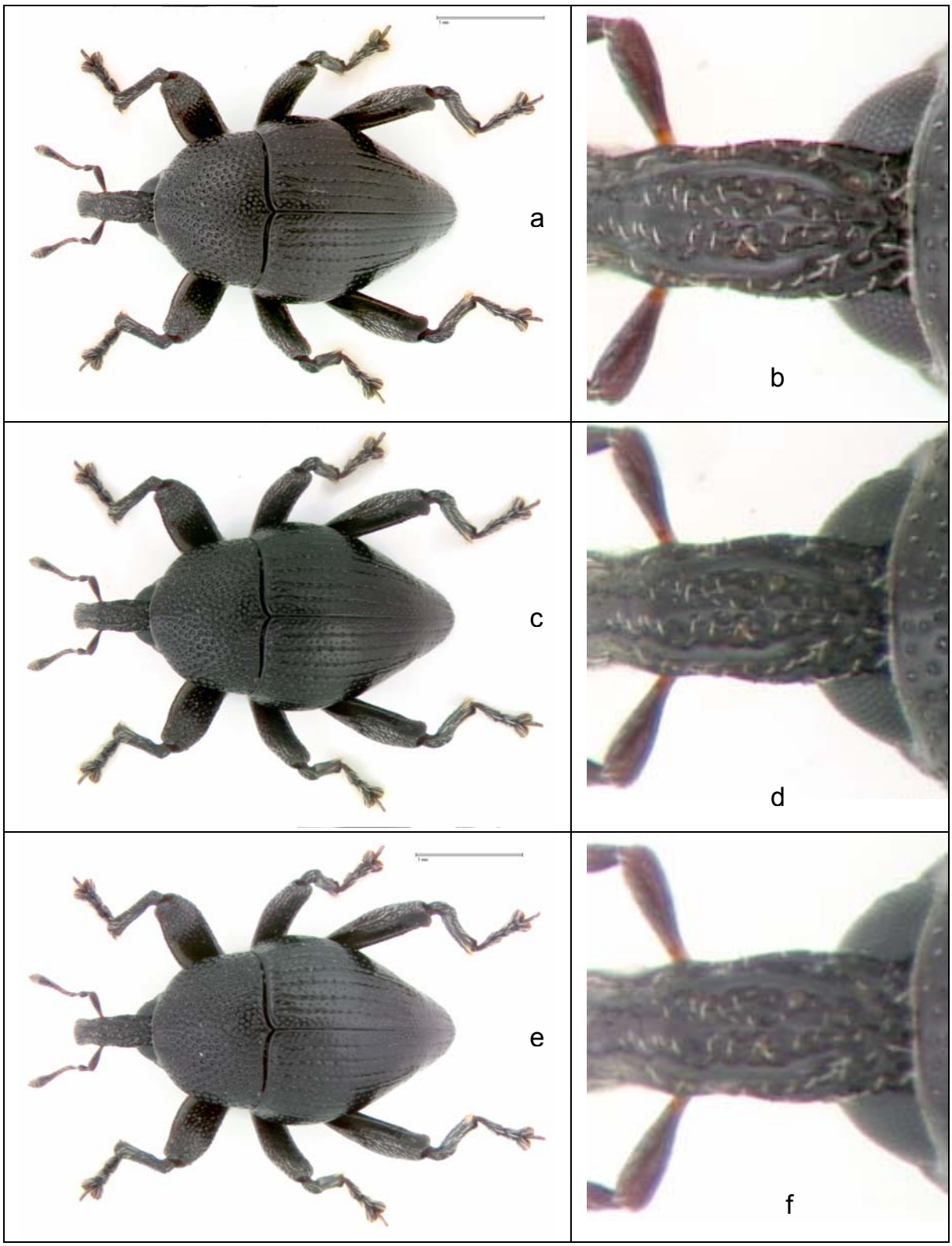
Digital camera technology is currently in a state of rapid improvement and change. It is difficult, therefore, to give specific recommendations that will be valid for more than a few months. Important general aspects are of course image quality and the availability of a live image on the computer screen, so that precise focusing is possible. I am currently using a JVC KY70, as recommended by ASHWORTH & FOGARTY (2003). Another good option is a Leica digital camera such as the DFC 320 (3,2 MP) or the DFC 480 (5,0 MP). With this system image data are transferred to the computer via fire-wire or USB2. Thus, it also provides a live image, just like the JVC KY70. The main drawback of such a Leica camera is that Syncroscopy does not provide the driver software to operate it directly from their Automontage© program. This option is only provided for cameras that are distributed through Syncroscopy, such as the JVC KY70. However, it is only of minor inconvenience to create and save the images with different software than that which is used to create a montage-image.

A mistake to avoid is concentrating on the image quality of the camera, while neglecting the quality of the optical system. The weakness in optical performance of a given microscope is not

necessarily obvious when looking through it. Test images of the same subject should be made and compared. It is also important to realize that some expensive products of manufacturers with an excellent reputation may not necessarily produce high-quality results in a given setup. Each option should be carefully tested for the purpose of digital imaging. Test images (figs. 6a-f) taken with the same camera but with different optical systems illustrate the effects of the optical system on the resulting image. One important point is a suitable video adapter to ensure that the full microscope image reaches the camera sensor. Attaching a 0.5´ sensor camera without a suitable video adapter will result in higher magnification - but concurrently in poor image quality of a subject photographed at the same scale (figs. 6e-f). Apochromatic lenses should be used whenever possible. As a rule of the thumb, lenses with a shorter zoom range will have better contrast and resolution compared with those with a longer zoom range. And, lens systems with a shallow depth of field will have a higher resolution than those with a longer depth of field – deciding on which to use will depend on the availability of the Automontage© software. Most will agree that the Leica Z6 Zoom-system provides outstanding image quality and is surely worth trying (figs. 6a-b). A significant advantage of the construction of the Z6 is that it possesses only one single optical axis, eliminating any shifting of images (parallax) through changes in the focal plane. Some other macroscopes / dissecting microscopes may work equally well, especially if they are equipped with an axial carrier, that allows the user to shift into one single optical axis. In any case, it should be borne in mind that it may not be the fault of the digital camera if the resulting images lack contrast or resolution. The optical system in front of the camera deserves some attention for the best results.

Colour adjustment

Colour management is a complex subject, and no comprehensive treatment will be given here. However, a few points should be noted. Whatever camera is used a whitebalance should be performed, and when any changes are made to the lighting a fresh whitebalance is necessary. With the whitebalance, the camera is calibrated, so that it “knows” at what wavelength the colour white is being transmitted into



Figs. 6 a-f. Comparison of different optical systems using the same camera (JVC KY70U) and the same weevil specimen (*Idotasia* sp.); left: overview; right: image detail. (a-b) Leica Z6 plus 0.63X video adapter; apochromatic lenses; (c-d) Zeiss SV11 plus 0.5X video adapter; achrochromatic lenses; (e-f) Zeiss SV11 without video adapter; achrochromatic lenses.

the camera.

To optimize settings it is critical that the monitor is properly calibrated, for otherwise, wrong colour-casts may be introduced into the picture based on a false image perception through an incorrectly calibrated monitor. Monitor calibration tools, such as the “Spyder” have become affordable in recent years and are strongly recommended.

Sometimes, the auto-whitebalance function of digital cameras is far from being perfect. The auto-whitebalance of the JVC KY70 can be used as a starting point, but the white-balance must be fine-tuned using the manual setting. If the JVC KY70 is operated through Automontage©, it offers two different options of whitebalance adjustments. One is through the software, *i.e*. through the „Adjust Camera Settings“-Window. The latest version of Automontage© offers a “tool” that can greatly assist in finding a suitable whitebalance (fig. 7). The values shown for each of the three colour channels should be about equal.

Another option for setting whitebalance is through the camera menu that has to be set through the pin-buttons of the camera. The latter option is critical if the following problem is observed: depending

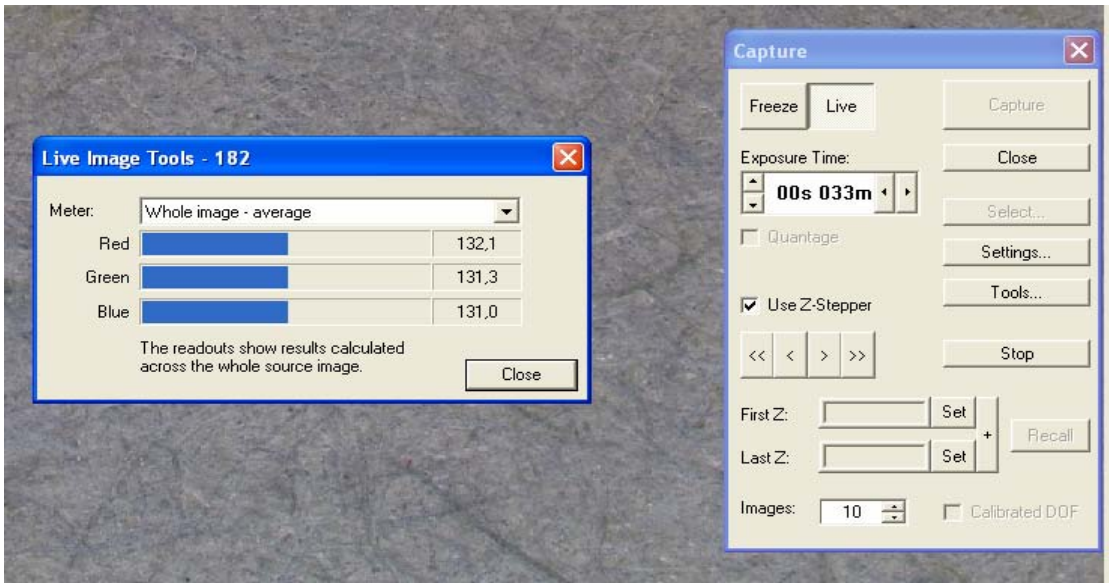


Fig. 7. Live Image Tool of Automontage© being used for improving a whitebalance based on a grey card. The values for the Red, Green and Blue channels should be about the same.

of what microscope / adapter combination is used, the field of view may show an uneven colour cast, despite even and neutral lighting. That means, on one side there is a red cast, towards the middle it becomes neutral, and on the opposite side there is a bluish or greenish cast. It is not possible to get rid of these colour casts by a regular whitebalance. However, if we navigate through the camera menu to the point “whitebalance” we can activate the „Shading mode“ of the camera by filling in RGB values in the “Adjust” submenu. I found this effect most pronounced when working on a Leica Diaplan compound microscope. To compensate, I need to put in the following values: R: -50; G: 30; B: -30. When using the Leica Z6 macroskop, different values are required: R: -15; G: 0; B= -20. It requires quite some attention to fine-tune these settings.

During postprocessing a correct whitebalance should be verified and if necessary adjusted. If parts of the image include white or neutral grey cardboard, then their RGB values should be about the same for each channel. In any case, the photographer should get as close as possible to a perfect whitebalance before imaging. For removing colour casts of images already taken see below under “Postprocessing”.

Lighting

Most beetles are a photographer´s nightmare from the viewpoint of lighting. They are smooth and highly reflecting. Thus, if a strobe or a cold-light source is used without any attempt to "soften" this harsh light, the beetle will show the blown-out reflection of the light source and the remainder of the insect will be a dark blot without any detail. The only solution to this problem is to spread the surface area of the light source and to diffuse it. This can be accomplished in various ways. When imaging larger beetles with a macro lens, the use of softboxes is an option. When working with a microscope cold-light sources are commonly chosen to provide a strong enough permanent light without burning the specimen. These light sources can be equipped with various glass-fibre arms. The most popular options are a ring-light to attach in front of the microscope lens, and a goose-neck light guide with two branches. The ring-light gives good results -

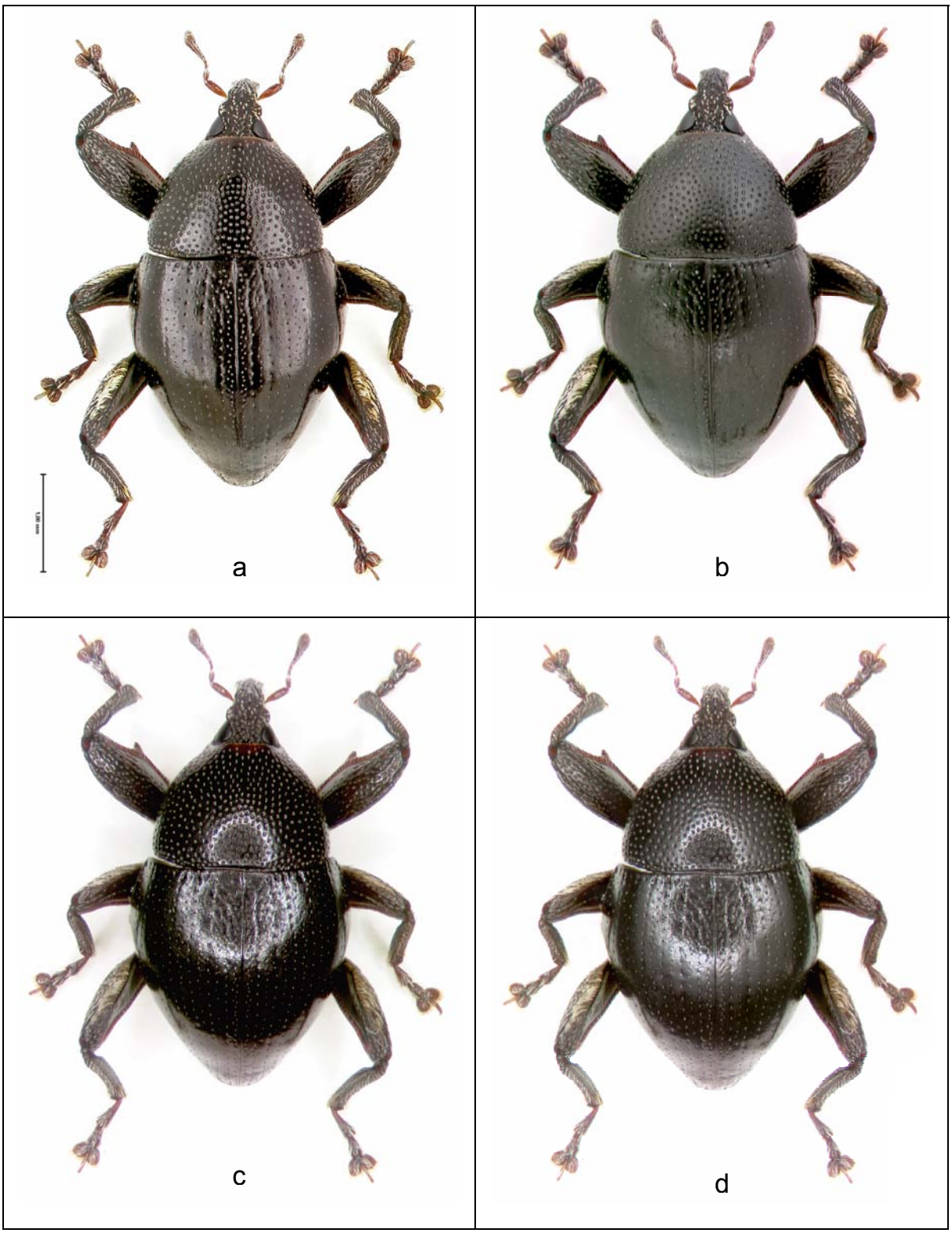
provided the beetle is non-reflecting. If the beetle is somewhat polished, a ring-shaped reflection with a "black hole" in the centre will appear on the beetle's dorsum (fig. 9c). This artefact can be alleviated by using a vellum cylinder that fits between the lens and the ringlight, so that the light shines through the vellum (fig. 8d). The vellum should reach down as far as possible and surround the beetle. Although this arrangement makes handling the specimen a bit cumbersome, the illumination of surface details is fairly good. The main draw-back is still the artificial ring-shaped reflection (fig. 9d). This method may be useful for scaled or non-reflecting beetles, but for shiny specimens there are better methods.

A technique described by LONGINO (2002) gives a more "natural" image: the ringlight is replaced by another set of double goose-neck fibre optic lights (fig. 8b) and the vellum cylinder with a Styrofoam cup with a cut off bottom (fig. 8c). A Styrofoam cup with two double goose-neck lights gives better results for diffuse lighting than does a set-up with a vellum cylinder (fig. 9b). With this method it is critical that the light from the fibre optics does not shine directly at the specimen, but only onto the Styrofoam wall above it. From there it is bounced and effectively diffused. The four light beams may enter the cup from above, or through small openings in the cup's upper third. This technique has been employed successfully for an imaging project of ants (C. KLINGENBERG, Staatliches Museum für Naturkunde, Karlsruhe).

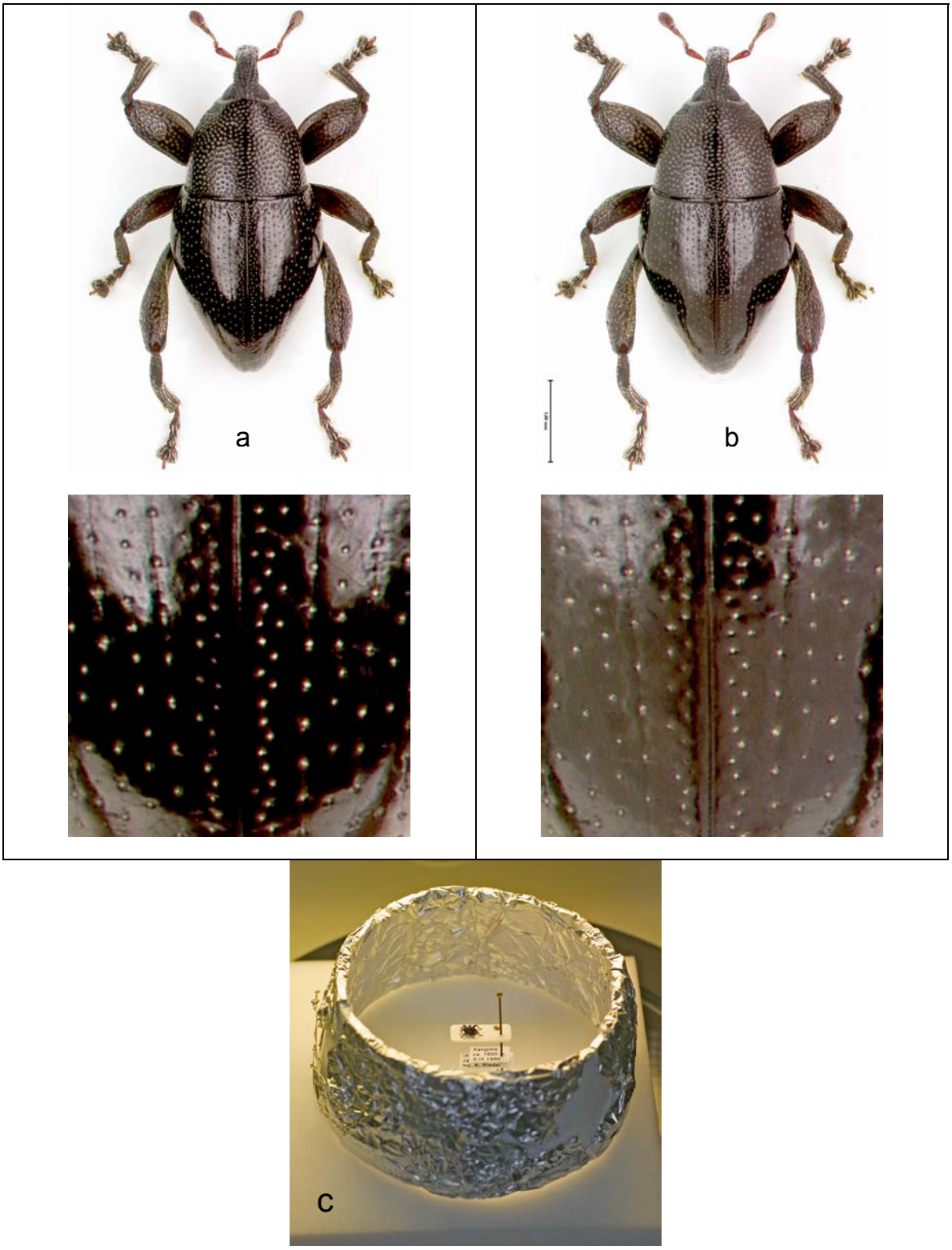
BUFFINGTON et al. (2005) developed the styrofoam-cup-technique further by the use of a construction made of two Styrofoam soup bowls; they dubbed this light chamber “the spaceship”. Good results were also obtained for small, shiny Hymenoptera. I have not tested this technique personally.

Probably, the method best suited for shiny beetles with delicate surface structure is the use of fluorescent desk lamps (fig. 9a). The set-up (fig. 8a) is both cheap and provides excellent illumination. Two facing desk lamps are arranged around the specimen. Such lamps are usually provided with a silvery reflector about 22 cm long. The precise properties of the lighting may vary with the reflector chosen. The Philips PL-11 W light tube, with a length of about 20 cm

Figs. 8 a-d. Different lighting setups. (a) Leica Z6 and a pair of energy-saving desk lamps; (b) Leica Z6 and setup of four goose-neck fiber lights and a styrofoam cup; (c) detail of styrofoam cup; (d) setup of ringlight and vellum cylinder.



Figs. 9 a-d. Example of a card-mounted weevil (Curculionidae, *Idotasia* sp.) in dorsal aspect illustrating different ways of lighting. Length from tip of elytra to apex of rostrum 4.0 mm. (a) with a pair of energy-saving desk lamps; (b) with four goose-neck fiber lights and a styrofoam cup; (c) with ringlight, without vellum cylinder; (d) with ringlight, with vellum cylinder.



Figs. 10 a-c. Example of a card-mounted weevil (Curculionidae, *Idotasia* sp.) in dorsal aspect illustrating the effect of a reflector of aluminium foil. In both images a pair of fluorescent desk lamps was used for lighting. (a) without the reflector; (b) reflector in use; (c) the reflector used in (b). Note that more fine detail is retained in shadow areas.

yielded good results. Shorter and brighter light bulbs are available, but the lighting is more even with the longer ones, which are preferred despite the added difficulty of positioning them around the specimen. Shadowed areas on the specimen can be better overcome by placing a shallow reflector around the specimen (figs. 10a-b). A reflector can be made by cutting a ring from a Styrofoam-cup and wrapping it with aluminium foil (fig. 10c).

The method selected depends on the characters of the specimens and the purpose for which the image is required. And it should be noted that that the resulting images may differ significantly depending on the method applied.

Montage software

Digital imaging combined with the appropriate computer software has overcome the limitations of depth of field for photography of museum specimens. There exist a number of computer programs that receive stacks of images, examine each image for areas that are in focus, and then stitch all the in-focus components together to make one single perfect image. Examples are CombineZ5©, Astro-Stack©, and Stack Focuser©. Some of these are freeware, many others cost little. Unfortunately the excellent Automontage© by Syncroscopy, the market leader comes with a considerable price-tag. The cheap programs have serious limitations at present, so they are not discussed in greater detail herein. However, some of the montage programs are still being developed and improved and it is worth tracking them in the hope of getting a good yet cheap solution. The following comments are intended to provide a few practical hints on Automontage©. A more comprehensive treatment can be found in ASHWORTH & FOGARTY (2003) and in the user manual of Auto-montage©.

Automontage© can be used for all stacks of images: either stacks taken manually (i.e. triggered image by image) with any camera, or for stacks taken by automatic capture (i.e. where a z-stepper-device moves the focus and triggers the camera). I found that the use of a z-stepper is only satisfactory if the microscope is equipped with a motor-focus. The construction of a microscope with manual focus

and an attached motor is problematic since the connection between the focus knob and the motor can slip. If a microscope with a manual focus is available only, it is easiest to move it by hand and to trigger the camera manually.

With a fully automatic setup and an average-sized beetle, I capture typically 40-70 steps per montage-set. I could not detect any improvement in image quality if the precision-option is used for scan-montage, thus I make use of the standard settings of “Speed” and “Fixed” (fig. 11). Usually, there will appear some artefacts in the montage-image, i.e. “dead areas” that are not resolved by the montage process. One way to eliminate these artefacts is to enlarge the “patch size”. However, with larger patch size, fine details will be lost from the image, so I prefer to perform the scan montage with a patch size of 10 and remove the resulting artefacts in a subsequent step. The “edit brush” of Automontage© is a useful tool to remove “dead areas” from a montage image by cloning the respective areas from an appropriate single image (fig. 12). The problem of such artefacts can be further limited if a stack of images is divided into two parts. Partial stacks can be processed individually and subsequently montaged manually in Photoshop. This is especially useful if extremities of the insect are sticking out and overlap with other parts. The full stack should be divided into two parts, one with the extremity, the other with the remaining body.

Automontage© has sometimes a problem to identify maximum sharpness of a contour in a stack of images. It can be useful to check such edge areas, *e.g*. the sides or the declivity of the elytra for maximum sharpness. If necessary, the montage image can be improved manually with a small sized “edit brush”.

My personal settings as I use them at present can be seen in fig. 13. I apply the “contour mode” (which provides a sharpening) to some degree to assist the software in calculating a precise image. But, in general it should be kept at a low level to avoid early over-sharpening of the image (see below “Post-processing).

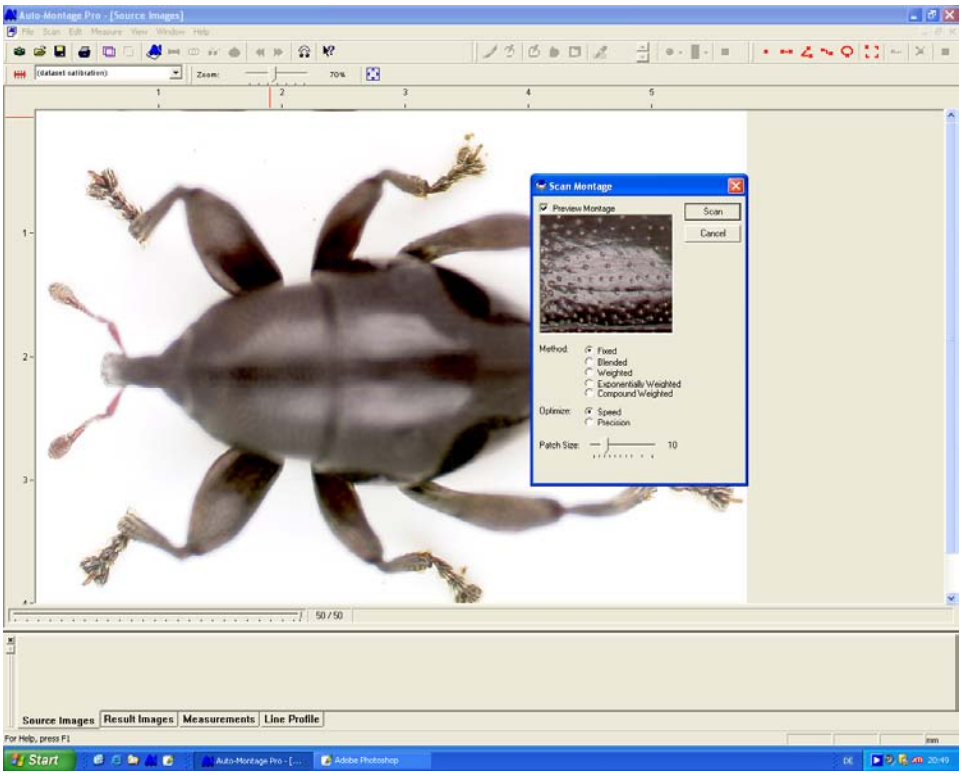


Fig. 11. Screenshot of the Scan montage window of Automontage©. Settings as they were used by the author for the previous images.

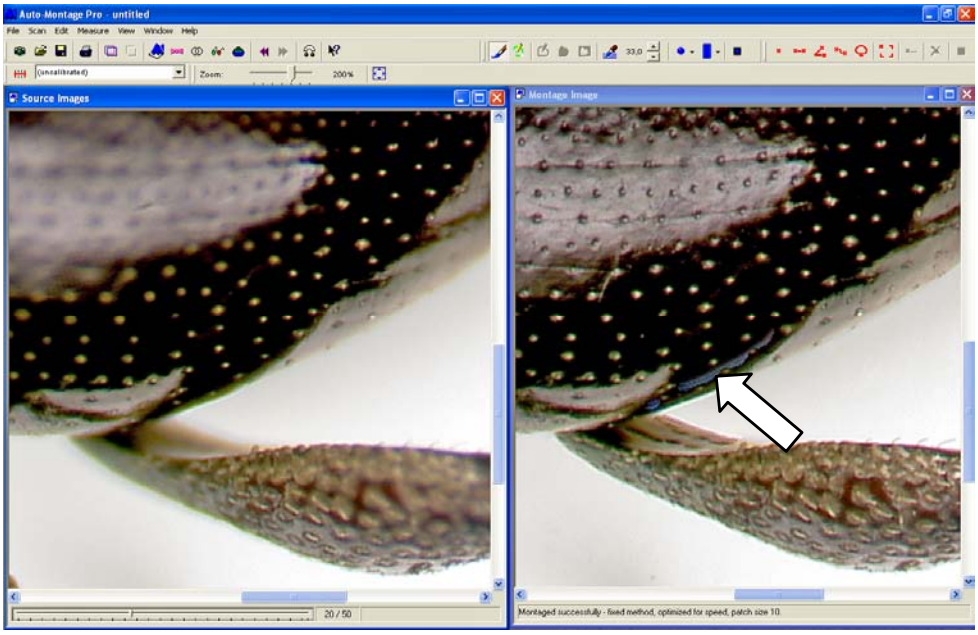


Fig. 12. Screenshot of Automontage© during the editing process of the montage image. On the left a window with a stack of 50 original images. On the right the montage image. Note the artefacts marked with the arrow. They can be removed by cloning the respective area from one of the stack images.

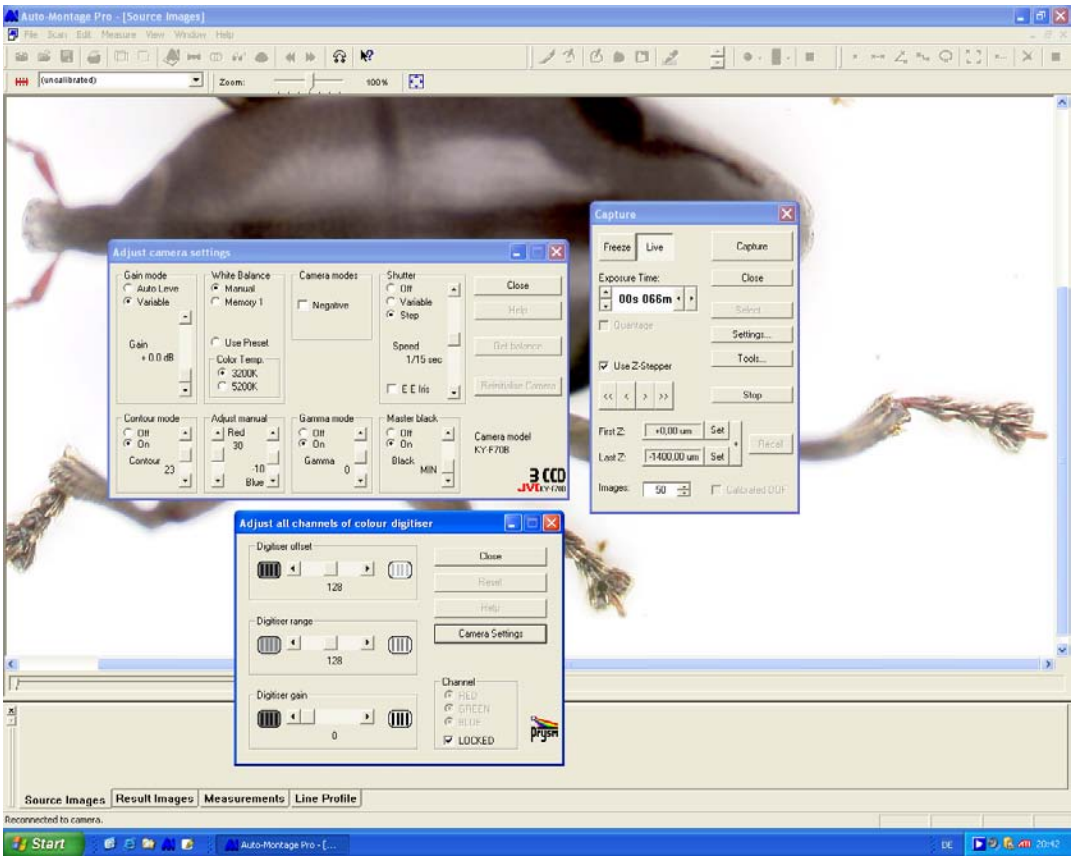


Fig. 13. Screenshot of with the main setting-windows of Automontage©. Settings as they were used by the author for the previous images.

Scale

Since most specimens are very small it would be hard to place an original scale bar next to each specimen that is recorded as a part of the image. Thus, a scale bar needs to be added during post-processing of the image. Automontage© offers convenient options for calibrating and drawing the scale bar. The first option is to use fixed magnifications for which the corresponding scales can be selected from the menu once calibrated. This is convenient, especially if the images of a given taxon are at the same scale. However, this method rarely takes full advantage of the frame of the camera . Personally, I prefer the method of calibrating the scale for each image individually. After taking a montage set of images the specimen is removed and a sheet of millimetre-square-grid paper is placed under the microscope. The magnification must not be changed at this stage, the focus may need some adjustment. The

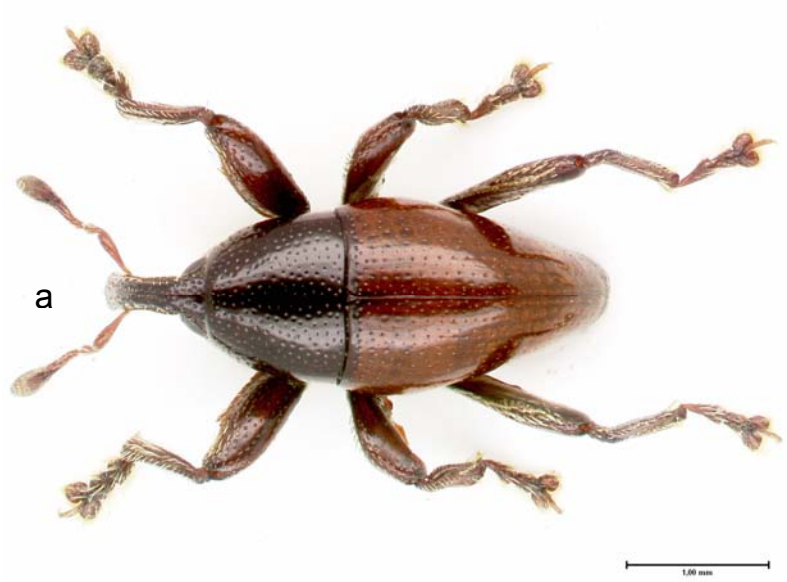
entire length of the frame or a major portion of it is measured. After closing the capture menu, a scale is calibrated in Automontage© using this distance, *i.e*. a line is drawn over the full length of the frame. Automontage© will calculate a suitable scale bar based on this value and place it at a chosen position. The "engrave scale bar"-option must be activated when exporting the montage image as a TIF file from the montage data set.

Post-processing of images

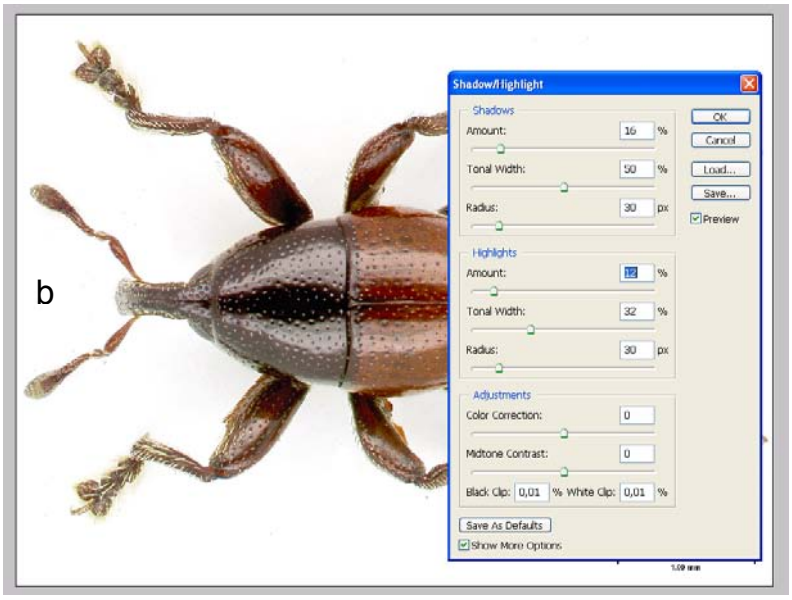
An image can be greatly improved to meet its purpose by post-processing with programs such as Adobe Photoshop© or Corel Photopaint©. A few general hints are given here that will apply to all image processing programs. Some tools of Adobe Photoshop are described in greater detail because this software is most widely used today.

Many digital cameras still suffer from a narrow dynamic range. This means that the tonal range of an image from light to dark may not be covered by the camera without clipping some information. This problem occurs if a black beetle mounted on a white piece of cardboard needs to be photographed. Sometimes, there is only the choice between a correctly exposed body and overexposed legs; or, correctly exposed legs and an underexposed body that may not reveal any detail. There are two different ways to resolve this problem.

The first option is to take a picture with a slightly underexposed body and with overexposed legs and antennae. Subsequently, these artefacts are corrected using the “Shadow/Highlight” option of Adobe Photoshop© (Version 8 or higher). Possible settings and the resulting effects are illustrated in figs. 14a-c. Generally, it is easier to handle a somewhat overexposed image (unless the areas are completely blown out) since the brighter tones contain more data than the darker ones. Usually, underexposed areas contain more digital noise, too. So, unlike the situation in conventional photography where it is easier to handle somewhat underexposed images, the opposite is true for digital images.



Figs. 14 a-c.
Use of the “Shadow/ Highlight”-tool of Adobe Photoshop©.
(a) A card-mounted weevil (Curculionidae, *Idotasia* sp.) in dorsal aspect; before the procedure.



(b) example of the settings being used.

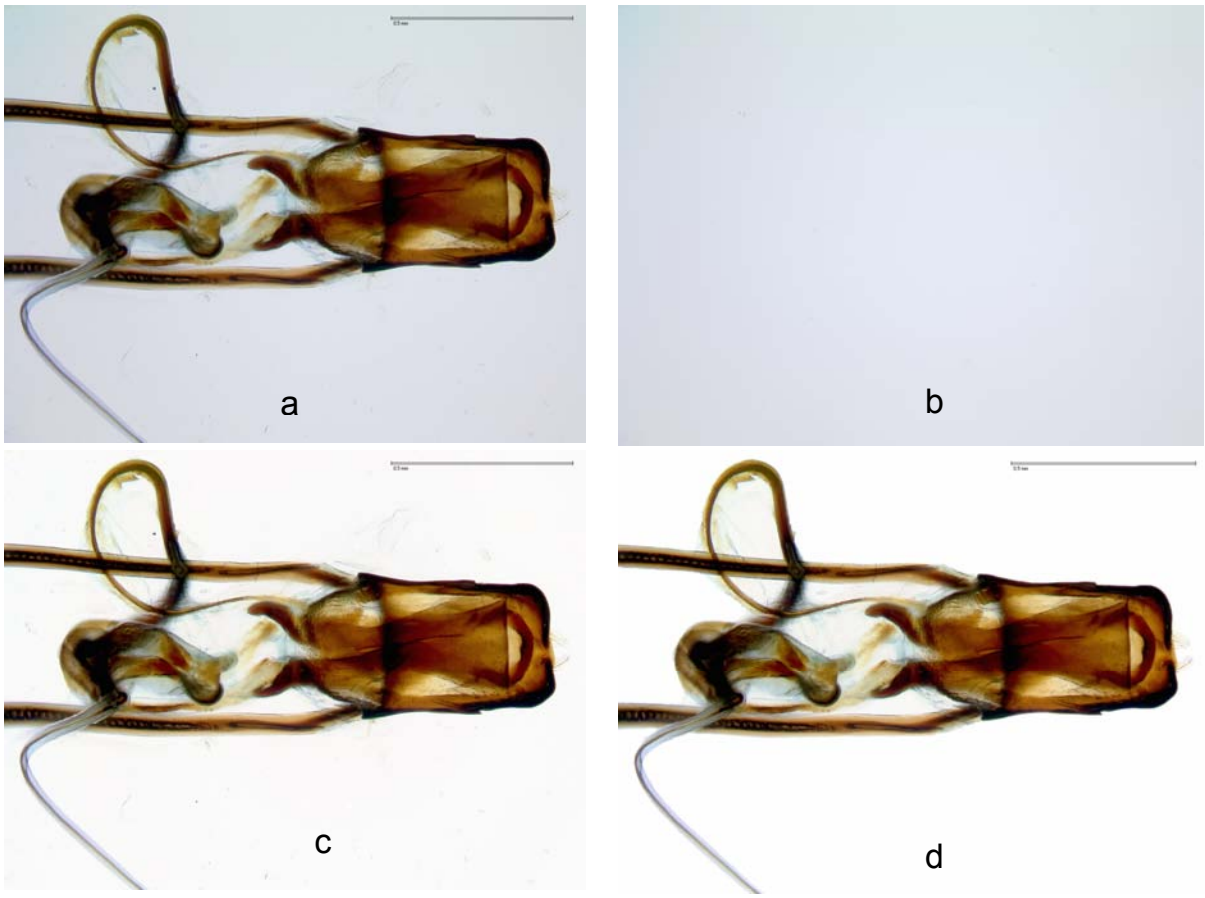
(c) resulting image; more detail can be seen in shadow areas; the tarsi and the antennae are darker and look more natural.

The second option is to take two images, with different exposure settings. Subsequently, these can be merged into one perfect image. This may require some practise , but is in fact quite simple: the two images should be opened on two separate layers and brought into perfect congruence. This can be done by switching between the layers and checking if the beetle "shifts" during the switch. Then, the eraser tool is used to remove parts of the image lying above, so that the better parts of the image lying below shine through. Various degrees of feather and transparency can be applied to the eraser for smooth transitions. Finally, both layers are merged into one image.

Other problems occur when parts of beetles, such as genitalia are photographed with a camera attached to a compound microscope. Images taken with transmitted light microscopy sometimes suffer from vignetting or some shading of the background. This problem can be overcome easily with a technique used by S. SCHMID (Zoologische Staatssammlung, Munich): Along with each photograph of a subject (fig. 15a), a second one is taken of the background (fig. 15b). For this purpose the subject is moved slightly to the side so that it is just outside the frame and a second picture is then taken. Magnification, illumination etc. must not be changed. The picture of the background can then be used in Adobe Photoshop to cancel out uneven lighting by the following steps:

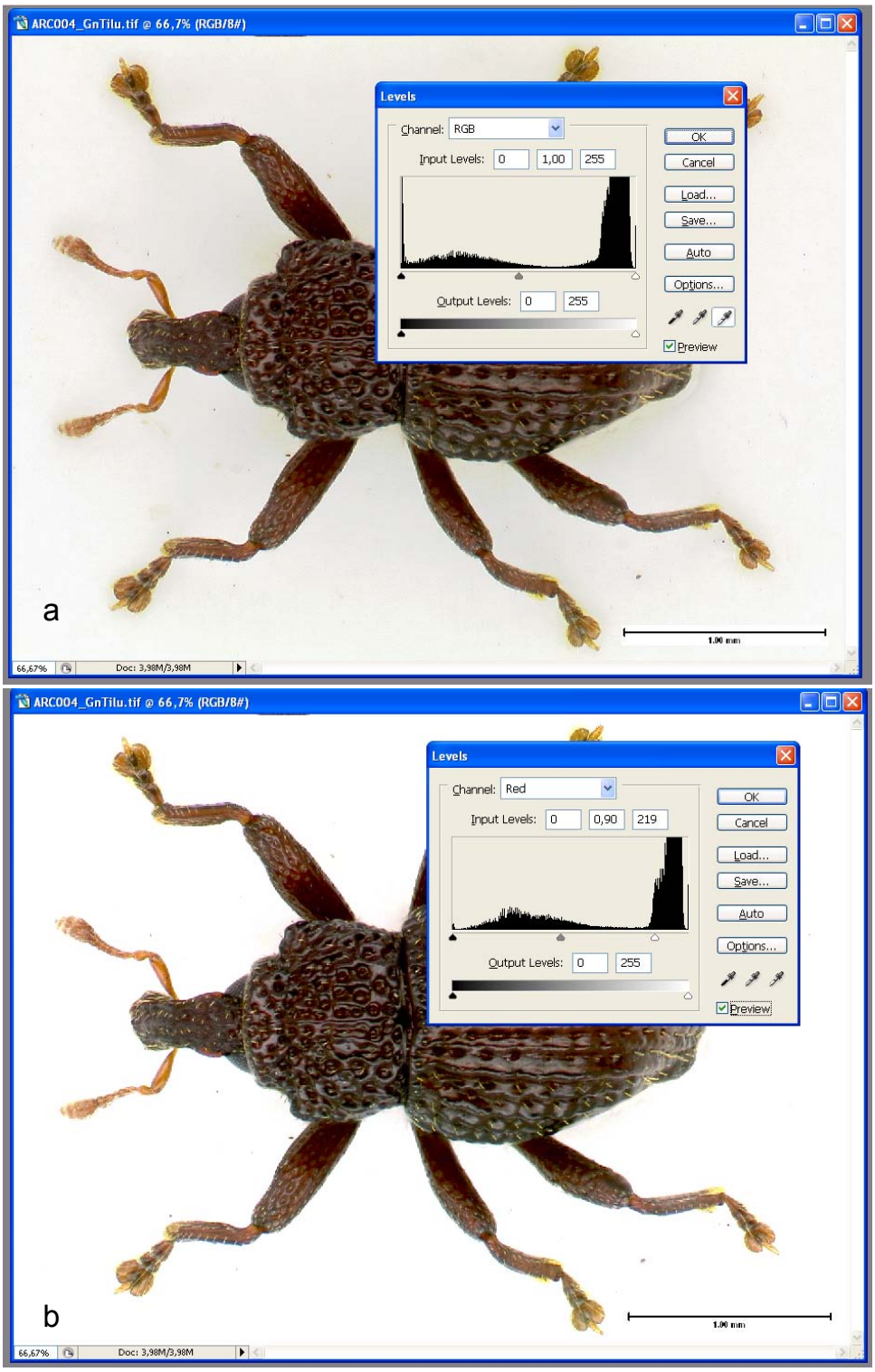
1. Open the image of the subject in Adobe Photoshop.
2. Create a new layer on top of it.
3. Open the background image, and invert it by using the Ctrl-I-shortcut.
4. Select everything (Ctrl-A) and copy it (Ctrl-C). Then close the file of the background-image.
5. Add the copied image into the new layer (Crtl-V).
6. Chose in the menu of the "Layers"-window "Color dodge". The effect should be visible now (fig. 15c).
7. Combine the layers.

Finally, contrast and tonal levels of an image should be adjusted, so that the entire tonal range is made use of, i.e. that white is white and black is black. It is a common mistake to use the "adjust contrast - brightness" options of such programs. Information could be irrever-



Figs. 15 a-d. Aedeagus of *Idotasia* sp. (Curculionidae), same as in fig. 6c; explanation of postprocessing procedure. (a) image of the subject; (b) image of the background; (c) background image was used to cancel out shading of background in Adobe Photoshop; (d) final, further improved image.

sibly lost from the image without this being noticed. A much better way is to make use of “levels” (figs. 16a-b) or of the “tone curve”. Slight colour casts can also be removed in this way. Usually the picture will contain some portion that should be white, such as the cardboard to which the specimen is glued. Such an area can be used to set a “Whitepoint” with the eyedropper-tool of the “levels”-menu in Adobe Photoshop. The tonal distribution of the image will be adjusted in a way that brings the selected area to a pure white. Such adjustments can also be made manually by pulling the slider for the shadows to the right, and the one for the highlights to the left. Clipping certain tones can be monitored precisely by pressing the “Alt” and “Ctrl”-Keys simultaneously.



Figs. 16 a-b. Use of the “Levels”-tool of Adobe Photoshop©. (a) A card-mounted weevil (Curculionidae, *Idotasia* sp.) in dorsal aspect; before the procedure. (b) preview of the resulting image; the white eyedropper was used to set a whitepoint and brighten up the background; subsequently, a remaining reddish cast was removed by changing the midtone-value of the red channel.

Colour casts can be corrected by selecting the respective colour channel in “Levels” and then adjusting the slider for the mid-tones. For example, if there is a bluish cast, the blue channel is selected and then the value of the mid-tones is somewhat decreased. Similar corrections can be done using “tone curves”. For an in-depth treatment of such techniques the relevant literature should be consulted, e.g. EVENING (2005).

Finally, some degree of unsharpen mask can be applied. However, this must be done very carefully. As the settings differ depending on the final purpose of the image (*e.g*. print size), no general recommendation can be given. It is a good idea NOT to apply any sharpening effects to an original image, not even to a copy that has been edited and adjusted with lots of efforts. Sharpening should be applied as the very last step, and the sharpened image should be saved as a separate file. Sharpening may result in artefacts such as visible halos on edges and this should be avoided in any case. Certain plug-in programs for Adobe Photoshop are available. Most of them achieve better results than the regular “USM” of Adobe Photoshop. I found that “Focal Blade” is useful for sharpening while keeping artefacts at a low level.

Acknowledgements

I wish to express my thanks to M. VERHAAGH, J. BIHN and C. KLINGENBERG (all Karlsruhe) for help in optimizing the photography unit of the section of entomology at SMNK. Years ago, P. NASKRECKI (Cambridge) introduced me to digital photography in general and to the Automontage system in particular, which helped me greatly in getting started with these techniques. I am grateful to S. SCHMID (Munich) for his hints on the Photoshop techniques to reduce background colour casts in microscopic images. J. REIBNITZ (Stuttgart) provided an important clue on the use of desktop lights for resolving fine surface structures in digital images. A. NOBILE (San Francisco) contributed hints on effectively using the Automontage-sofware. And last, but not least thanks are also due to A. H. STANICZEK (Stuttgart) for providing information on Leica digital

cameras and to H. SCHILLHAMMER (Vienna) for information on the use of micro-lenses.

References

ASHWORTH, S. & FOGARTY, J. (2003): A Guide to Digitizing Insect Collections Using MANTIS and Following the Protocols of the Harvard MCZ Entomology Type Image Project. http://insects.oeb.harvard.edu/MCZ/PDFs/Guide_to_digitizing.pdf [accessed 2005]

BUFFINGTON, M. L., BURKS, R. A. & MCNEIL, L. (2005): Advanced techniques for imaging parasitic Hymenoptera (Insecta). – *American Entomologist,* **51** (1): 50-54

EVENING, M. (2005): *Adobe Photoshop CS2 for Photographers. A professional image editor's guide to the creative use of Photoshop for the Macintosh and PC*. Focal Press. i-xxviii, 1-675.

KISSER, J. (1934): Bemerkungen zum Einschluß in Glycerin-Gelatine. – *Zeitschrift für wissenschaftliche Mikroskopie und mikroskopische Technik,* **51** (3): 372-374.

LONGINO, J. T. (2002): http://www.notesfromunderground.org/archive/vol912/methods/pictures.html [accessed 10.Xii.2002]

SCHILLHAMMER, H. (2005) http://www.pbase.com/rovebeetle/mostly_beetles [accessed 1.VII.2005]

Digital Imaging of Butterflies and Other Lepidoptera

More or less "flat" objects?

CHRISTOPH L. HÄUSER, JOACHIM HOLSTEIN & AXEL STEINER

Staatliches Museum für Naturkunde Stuttgart, Rosenstein 1, D-70191 Stuttgart, Germany. E-mail: chaeuser@gmx.de, holstein@naturkunde museum-bw.de, a-steiner@web.de

Key words: Digital imaging, Lepidoptera, butterflies, moths, "light box".

Introduction

In butterflies and moths (Lepidoptera), wing pattern and colouration often show important distinguishing characters. For their preservation in collections, Lepidoptera specimens are pinned vertically straight through the thorax and then dried with wings outspread in a 90° angle to the pin on a setting board. When ideally set, the wings should thus all be fixed in one plane so that one is dealing with an almost flat, two-dimensional object. Depending on the original state of preservation of the specimen, the preparatory skills applied, and particularly under conditions of high levels of humidity, however, in many collection-based specimens the wings can end up at various angles to the body, e.g., hanging downwards or directed upwards in a V-shaped profile.

As part of a longer term effort to mobilize collection based biodiversity data from museum collections for a national contribution to the Global Biodiversity Information Facility (GBIF), the authors have been coordinating a database project to establish a web based information system for butterflies (Global Species Register for Butterflies / Global Butterfly Information System - GART/GloBIS; see HÄUSER et al. 2003a, 2004, http://www.lepidat.org). One of the GART/GloBIS project goals was to document and database primary

butterfly type specimens from larger museum collections in Germany. As this meant processing large numbers of type specimens at several institutions, we attempted to test and develop standards both for equipment and procedures to ensure identical working conditions when documenting specimens under different local environments. The experiences gained over the course of this project were combined with the development of standards for the ENBI project, and the techniques devised by other workers on Lepidoptera and similar groups of insects.

Lepidoptera range in size from tiny to large with wingspans varying between 3 and 300 mm. As our current project focuses on butterflies, this chapter considers techniques to digitally photograph medium-sized specimens with the smallest examples having a wingspan of about 1 cm. Techniques to photograph smaller specimens as well as microscopic preparations are not dealt with in this chapter, and require differently adapted special setups and equipment.

General problems and requirements for photographing Lepidoptera

With colour representation being an important aspect, two main factors always to be considered when photographing Lepidoptera specimens are lighting and background. When examining illustrations in the printed literature or on websites it is immediately apparent that a wide range of different techniques are being used, while attempts to optimize and standardize Lepidoptera photography date back to the era of black-and-white photography (DOS PASSOS 1949, KOYLER 1965). The examples shown in figs. 1-4 are meant to illustrate some of the more frequently encountered problems and challenges.

When colour photographs of specimens are to be compared with any degree of accuracy, a standardized background and light source has to be used. Black or white backgrounds are considered unsuitable in representing the extremes of the grey-scale. For the GART/GloBIS project we experimented with photographic grey cards (18% grey) as offered by KODAK and other manufacturers but found

Fig. 1. A black background can be disadvantageous for darker specimens. Here a diffuse light-source has been used but there is not sufficient light (underexposed image). Similarly, a white background is not optimal for brightly coloured specimens. – Noctuid moth (*Catocala* sp.), from SARGENT (1982).

Fig. 2. Black background with over-exposed specimen. A unidirectional light-source has been positioned too low in front of the specimen resulting in undesirable shadows. The wings of set insects are never perfectly flat but reflect wing veins and folds. A diffuse or multidirectional light-source is preferable. – Noctuid moth (*Euxoa* sp.), from CALLE (1982).

Fig. 3. A bright background is usually better for both dark and pale specimens. However, unidirectional light causes shadows which can make the wing shape difficult to recognise. Nymphalid butterfly (*Agrias aedon pepitoensis* MICHAEL, syntype). From: http://insects.oeb.harvard.edu/mcz/FMPro?-DB=Image.fm&-Lay=web&-Format=images.htm&Species_ID=16648&-Find)

Fig. 4. Even the use of several flash-lights does not always eliminate shadow when a pale background is used. A ring flash positioned too close to the specimen has a similar effect. Noctuid moth (*Euchalcia stilpna*), from HACKER & RONKAY (1992).

them as generally being too dark. A pale grey plastic sheet produced as a divider for ring binders (Herlitz Article No.: 05961107; EAN No.: 4008115961106) proved to be both neutral in colour and to possess a surface structure that does not cause reflection. This material was adopted as the standard background for photographing butterfly and other Lepidoptera specimens throughout the GART/GloBIS project.

If for later image processing it is intended to digitally crop or "extract" the specimen from the image, it is advantageous to use a background with a colour which is not present in the specimen itself. Some photographers prefer a blue background (fig. 5) or a similarly intense colour for aesthetic reasons. As with colour representation in general, there will always be a subjective element in choosing the most suitable or "best" background colour for Lepidoptera.

Specimen setup

As Lepidoptera are usually pinned, the easiest way of placement is to pin the specimen directly onto the background, but then care should be taken to avoid visible pinholes (e.g., fig. 4). If the underside is to be photographed difficulties arise because now the pinhead has to be fixed somehow to the background, e.g. by placing it in a pellet of clay, plasticine or a similar medium that will hold the pin with the specimen. As most specimens are pinned with one third of the pin above the body and two thirds below, this method also results in different distances between the specimen and the background for upper- and the underside photographs. As camera distance in many cases equals flashlight distance, this results further in different background brightness for upper- and underside photographs. There are methods to overcome this problem, e.g. by placing the specimen on an elevated pane of glass so that the visible background is at a greater distance from the camera and differences in lighting become less apparent (NAUMANN 2001).

In the GART/GloBIS project a relatively simple setup is used which eliminates the need to pin or firmly attach the specimen to the background, and which ensures an identical distance between specimen and background for both upper- and underside pictures.

The method also allows for a much faster and more efficient handling of specimens, especially when larger numbers are to be photographed. Instead of attaching the specimen via its pin, the specimen is placed with its wings on two parallel threads of very thin fishing line (finest type available with 0.06 mm diameter), which are stretched across a wide frame or between two elevated supports. We currently prefer to use a long and wide U-shaped metal frame (see fig. 9) and to fix the fishing line tied into a loop on both sides of this frame with small magnets, which allows to easily adjust the threads whenever needed. Most small- to middle-sized Lepidoptera are sufficiently lightweight to remain perpendicular if suspended in this way. Specimens with a large and heavy abdomen, however, sometimes require additional support to keep them from toppling over.

Fig. 5. Placement of a specimen on two strands of fishing line (oblique view). Depending on light and viewing angle these supports can be almost invisible (*Colias* sp., Pieridae).

Against a pale grey background, the fishing line is often invisible in the picture (see figs. 7-8). If still visible, the resulting two fine white lines can usually be eliminated easily later from the image. If handled carefully, there are no adverse effects on the specimens except for the occasional loss of a few scales. Thus, only the threads need to be wiped from time to time as some wing scales will adhere to them during the course of work.

Lightweight objects such as locality labels, bar codes or scale bars can also be placed on the fishing lines to be photographed together with the specimens (figs. 7-8). If specimens do not have their wings in a flat position or are not pinned in a satisfactory way with an approximate angle of 90 degrees to the body axis, however, this setup will also face problems in facilitating a perfect picture (figs. 6-8).

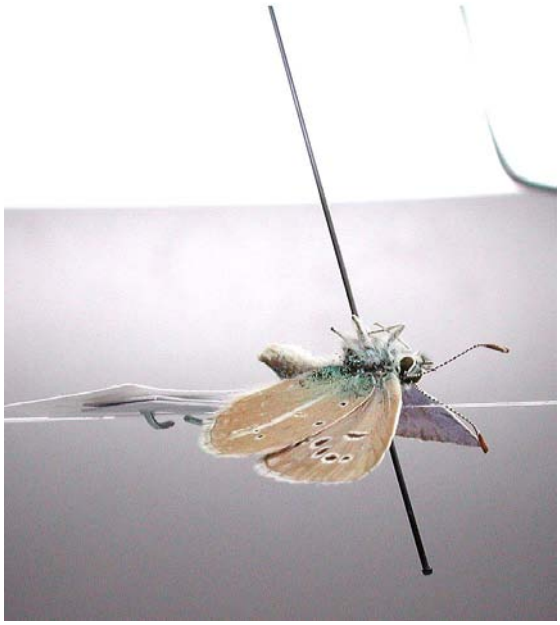


Fig. 6 (right). A problematic specimen. The wings are not in one plane and the specimen is pinned at an irregular angle (*Agrodiaetus* sp., Lycaenidae).

Figs. 7-8 (below). The dorsal and ventral view photographs of the same specimen. Note that fishing line supports are barely visible (*Agrodiaetus* sp., Lycaenidae).

Light source

A large variety of lamps and flashlights has been used in Lepidoptera photography. To eliminate shadows both on the wing surface of the specimen (fig. 2) and on the background (fig. 3) a diffuse or multidirectional light source is clearly the preferred option. However, iridescent colours produced by the reflection of light in the wing scale microstructure can generally be captured better with unidirectional light sources (e.g. fig. 3).

One of the most serious problems in the photography of spread Lepidoptera specimens is the presence of shadows cast by the specimen itself onto the background. There should be as little shadow as possible – ideally none at all – so as not to distract from or to diffuse the actual specimen. A medium-sized ring-shaped lamp is the best way to ensure an almost shadow-free image. For live insects, a conventional ring flash mounted onto the camera lens is a good option. In the case of pinned and spread specimens, however, a ring flash often causes a diffuse shadow around the specimen due to its small diameter (fig. 4).

The ring lamp solution: The best option for producing shadow free pictures of spread Lepidoptera seems to be a circular light source that has a larger diameter than conventional ring flashes. Circular fluorescent light-tubes are available commercially in several sizes, and are not very expensive. These tube lights, however, mostly produce either the extremes of the usual "warm" spectrum tending towards yellow or rather "cold" bluish light, which both lead to serious problems in colour representation. As a special high-end product, however, a fluorescent light tube with a so-called “full spectrum” with a light temperature of 5500 K is also available (VITA-LITE, DURO-TEST), which comes very close to natural daylight. Although considerably more expensive than common fluorescent light tubes, we found this lamp an ideal solution both for allowing shadow free images, and good colour representation.

Initially, this lamp was chosen because the GART/GloBIS project had been planned in the late 1990ties, when the photographic documentation was still expected to be carried out using conven-

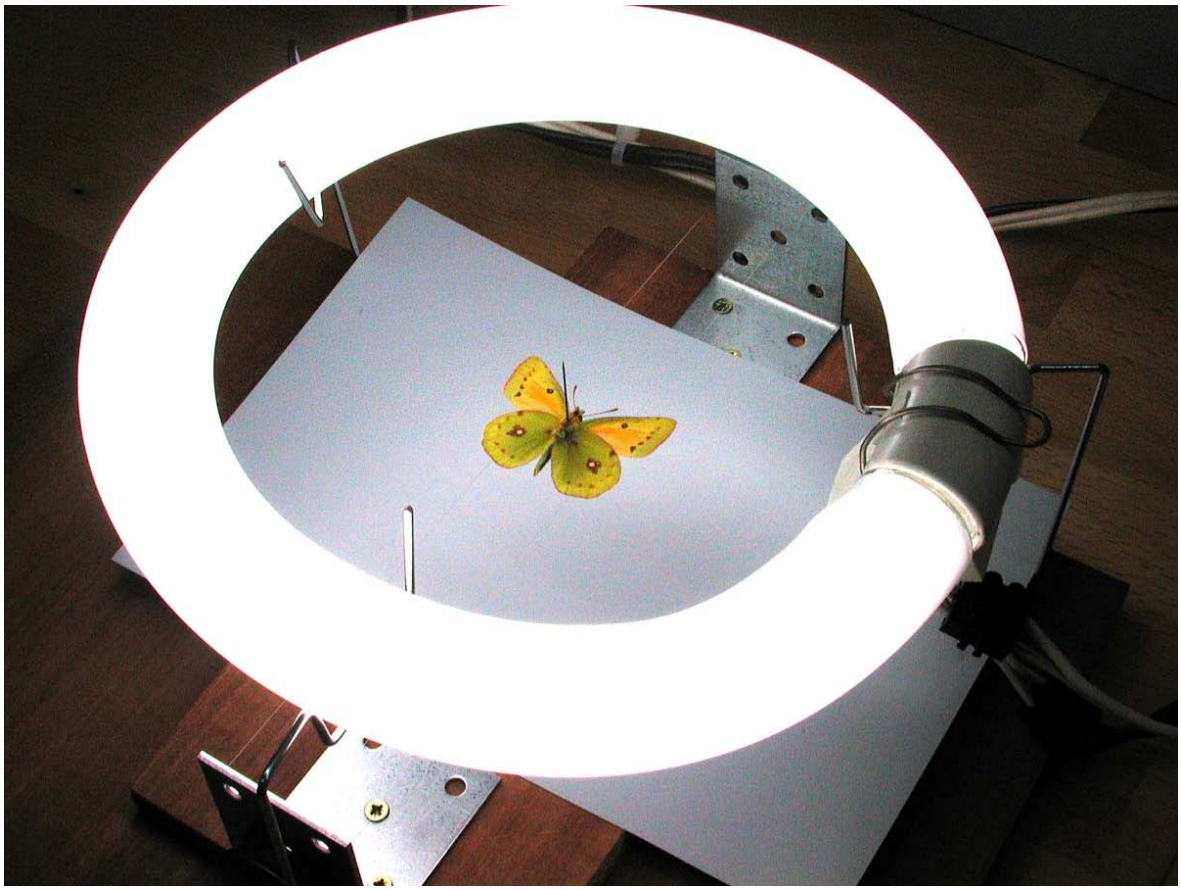


Fig. 9 (above). The circular light tube and a specimen set up for photography, shown without container for clarification. The supporting threads of fishing line are stretched between the upturned ends of the metal frame. Background papers can be exchanged easily.

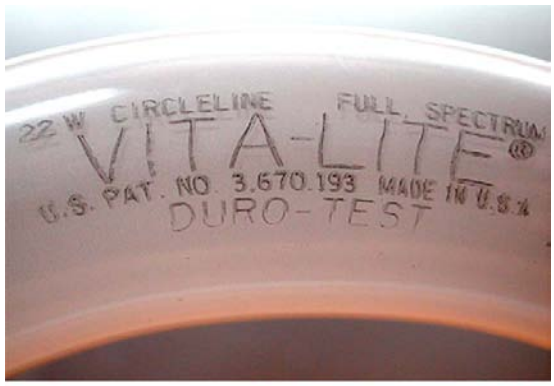


Fig. 10 (right). Lamp detail.

tional cameras with colour slide film. With the increasing availability of medium-end digital cameras for the general market, these soon proved to be sufficient for the purposes and requirements of the project. We first used the NIKON coolpix 990, later the NIKON coolpix 995, and the NIKON coolpix 5700. All these camera models possess a white balance function which no longer makes a full spectrum lamp a requirement for good, natural colour representation. A full spectrum light, however, is still very helpful whenever using

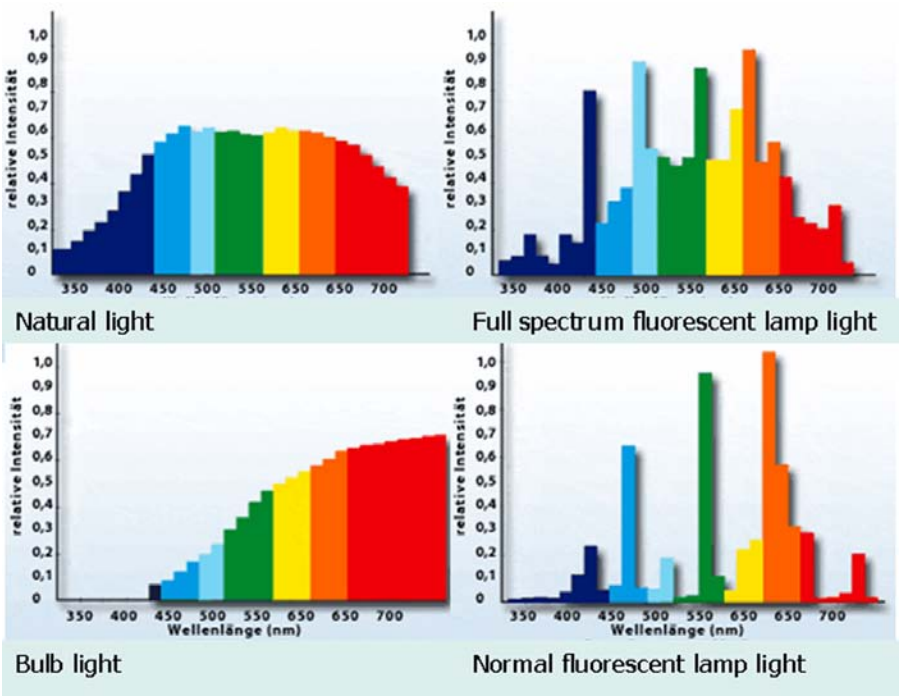


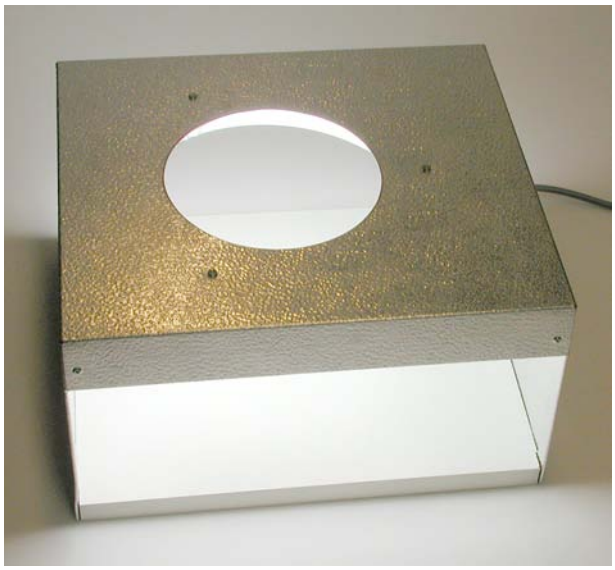
Fig. 11. Comparison of color spectra of natural light, bulb light and two types of fluoresent lamps.

cameras with colour film or when studying and comparing specimens with the naked eye.

The "light box" developed for the GART/GloBIS project: To facilitate the standardized processing of large number of specimens for the GART/GloBIS project for a longer term, these components were integrated into the construction of a durable special "light box", which since has been turned into a commercial product. The main function of the light box is to provide a stable, permanent setup for the circular fluorescent tube to be left for specimen photography at a central location within the collection for continuous use.

To prevent the lamp light from directly entering into the lens of the camera (and also into the photographer's eyes) the light tube is mounted inside the top of a rectangular aluminium box (36 x 33 x 20 cm) with a round opening on top which is slightly smaller in diameter than the circular light tube (figs. 12-14). The aluminium frame has a structured surface to reflect and further diffuse the light, and the box is open on two sides to allow the specimen setup to be moved easily across the bottom of the box. An electronic ballast built into the box provides for flicker-free light, and a switch allows the box to be plugged in without the lamp to be turned on all the time.

Fig. 12 (right). The ring lamp mounted inside the "light box".



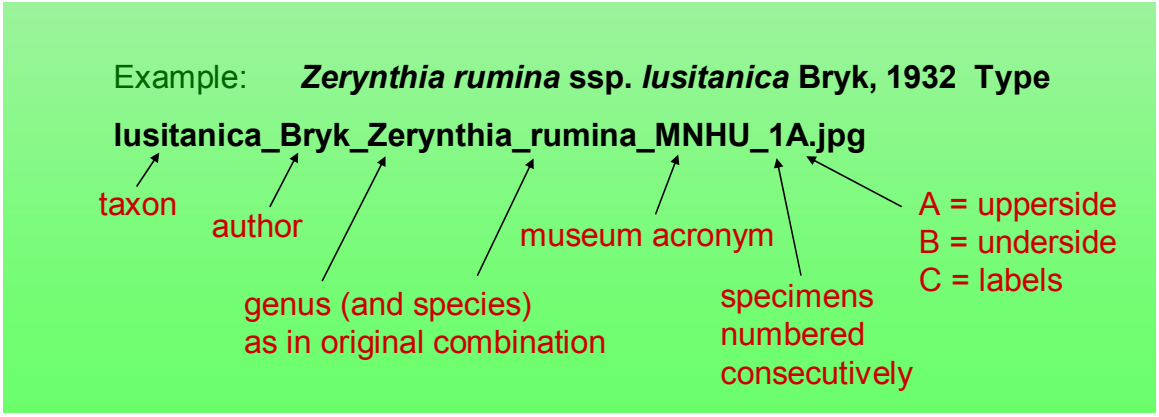
Figs. 13-14 (below). Working with the "light box".

The “light box” was initially designed by WOLFGANG ECKWEILER (see ECKWEILER 2001), and is currently produced and commercially marketed by the company of FRITZ WEBER (Stuttgart) (www.fritz-weber-entomologiebedarf.de/21904.html). The company now also offers a collapsible model for easy transportation, in which the two side boards can be folded inside protecting the lamp and resulting in a size of 39 x 33 x 7 cm.

With this setup and equipment, specimen photographs can be taken freehand which significantly accelerates the progress of work when photographing larger number of specimens. When using camera settings requiring longer exposure times or for other special purposes, a tripod mount is still advisable and can be added easily to the general setup.

Naming image files

When photographing larger number of specimens, the need for an efficient administration of the resulting images requires some consideration in naming the individual image files. As at the beginning of the GART/GloBIS project plans for globally unique identifiers (GUIDs) or other international standards for image file names were still in their infancy, we developed our own standards for image file names for the purpose of the project. Apart from the basic requirement for creating unique names for all files, file names should preferably also contain information about the image content, for which the taxon name was thought to be most useful in case of type specimens. Accordingly, file names of specimen images for the GART/GloBIS were created along the following scheme, which is currently still applied for the project:



Final considerations

The experiences gained during the GART/GloBIS project with the setup and standards for photographing butterfly specimens presented here have been overall very positive. So far, about 5,000 type specimens in more than a dozen museum collections have been documented, which resulted in more than 16,000 image files. The images have been used both for digital applications such as web-based databases (www.lepidat.org; http://www.biologie.uni-ulm.de/systax/) as well as printed end products (e.g., HÄUSER et al. 2003b, 2004), without yet facing the need to change any of the applied standards or techniques. Meanwhile the "light box" has been acquired by several larger museums and is available for use by re-

searchers in London, Berlin, Bonn, Dresden, Frankfurt, Karlsruhe, München, and Stuttgart.

Despite all efforts towards optimized standards, however, there will always be cases to which they cannot be applied and where creative, sometimes instant solutions are called for. This is certainly the case with many damaged specimens which must be handled as little as possible, but sometimes also with specimens preserved in an unusual fashion (figs. 15-16).

Figs. 15-16. Problematic cases: Specimens like these 18th century butterflies that are severely damaged or set/pinned in very unusual fashion cannot be accomodated by fishing line supports.

Fig. 15 (above). Type of *Papilio aurota* FABRICIUS, dermestid-eaten.

Fig. 16 (right). Type of *Hesperia flaccus* FABRICIUS, "pinned" on a thorn.

When further considering colour standards it should also be noted that it will practically be impossible to achieve perfectly identical colours for all printed products and on all computer screens due to

different hardware properties and varying screen settings (see also MORRIS on "Colour Management", in this volume).

Acknowledgements

Financial support from the European Commission for ENBI Work-package 6 is gratefully acknowledged. The GART/GloBIS project was supported through grants from the German Federal Ministry for Education and Research (BMBF, Förderkennzeichen 01LI0204), and additional support was provided by the Ministry for Science, Research and the Arts of the State of Baden-Württemberg (MWK).

We thank WOLFGANG ECKWEILER, CHRISTOPH ESCH, and NORBERT HIRNEISEN who have made major contributions to the GART/GloBIS project. Our thanks are also due to those who have tested and used our equipment and worked with the databases and images: DANIEL BARTSCH, ANJA CIESLAK, RALF DEICHSEL, MARION GSCHWENG, ELKE KLASS, GUY KUATE, VLADIMIR LUKHTANOV, SANDRA MUSTAFA, BIRGIT OETTINGER, JULIA REIN, ANJA SCHINDLER, JULIA SCHMALFUß, STEFAN SCHMIDT, CHRISTIAN SCHULZE, ANNETTE STUMPF.

References

CALLE, J. A. (1982): *Noctuidos Españoles.* – Boletín del Servicio de Defensa contra Plagas e Inspección Fitopatológica, Fuera de Serie, **1**: 1-430.

ECKWEILER, W. (2001): Schattenfreie Fotografie von Insekten mit Hilfe einer Ringleuchte. – *Nachrichten des Entomologisches Vereins Apollo, N.F.* **22** (3): 136.

HACKER, H. & RONKAY, L. (1992): Taxonomic news in the genus Euchalcia Hübner, 1821 (Lepidoptera, Noctuidae, Plusiinae). Part I. – *Esperiana*, *3*: 457-471.

HÄUSER, C., HOLSTEIN, J. & STEINER, A. (2003a): GART – The Global Species Register Butterflies (Globales Artregister Tagfalter). In: Federal Ministry of Education and Research (BMBF) & German Aerospace Center, Project Management (DLR - PT) (eds.): Sustainable use and conservation of biological diversity – A challenge

for society. International Symposium, 1-4 December 2003, Berlin. Symposium Report, Part A. Deutsches Zentrum für Luft- und Raumfahrt, Projektträger des BMBF (DLR-PT), Bonn; pp. 204-205.

HÄUSER, C. L., BARTSCH, D., HOLSTEIN, J., & STEINER, A. (2003b): The Lepidoptera type material of G. A. W. HERRICH-SCHÄFFER in the Staatliches Museum für Naturkunde, Stuttgart. – *Stuttgarter Beiträge zur Naturkunde, Serie A,* **657**: 1-78 (24 col. pls.).

HÄUSER, C. L., HOLSTEIN, J., & STEINER, A. (2004): Das globale Artregister Tagfalter – GART: Ein Web-basiertes Informationssystem. – *Mitteilungen der Deutschen Gesellschaft für allgemeine und angewandte Entomologie,* **14**: 145-148 (3 figs.).

NAUMANN, C. M. (2001): Schattenfreies Fotografieren von präparierten Insekten. – Entomologische Zeitschrift, **111** (6): 187-188.

SARGENT, T. D. (1982): *Legion of Night. The Underwing Moths.* Amherst (University of Massachusetts Press). XIII + 222 pp.

KOYLER, J. M. (1965): An inexpensive apparatus for photographing mounted specimens. – *Journal of the Lepidopterists' Society*, **19** (4): 212-214.

DOS PASSOS, C. F. (1949): The Photography of Types of Lepidoptera. – *Lepidoptera News*, **3** (4-5): 41-42.

Imaging Type Specimens of Fishes at the Natural History Museum, London.

PHIL HURST[1] & DARRELL SIEBERT[2]

[1] Photographic Unit, The Natural History Museum, Cromwell Road, London, UK SW7 5BD. E-mail: p.hurst@nhm.ac.uk
[2] Zoology Dept., The Natural History Museum, Cromwell Road, London, UK SW7 5BD. E-mail: d.siebert@nhm.ac.uk

Key words: digital imaging, radiography, fishes, type specimens.

Introduction

The imaging of type specimens of fishes poses no special technical problems beyond good practice for imaging fish specimens generally, which come in all sizes and shapes. Some are small – less than a centimetre in length – and a few are longer than two metres. Variation between the extremes of size is continuous but most are in the range of 5.0 to 30.0 cm in length. Naturally, some taxa consist entirely of small species and others solely of very large species. Fish specimens vary considerably in shape, from elongate, sometimes extremely so, to flattened dorso-ventrally or laterally, or simply what most people would recognise as fish-shape (fusiform).

Even though possibilities for distinguishing between or among fish species are as great as the possible number of combinations of their features, most fish species can be recognised or distinguished from each other simply by looking at them, especially if the observer is experienced or possesses a trained eye. Consequently, in most instances, the job of presenting images that allow recognition of a fish species is a relatively simple task, but it can be very difficult for very similar or cryptic species and it is impossible for species that differ only in internal anatomy, behaviour, or at a molecular level.

Specimens may be of any age, from as old as several centuries when fish specimens first began to be kept, to very recent. The

manner in which they were initially prepared can have been very different too. Some will have been prepared as dried skins sewn onto card or glued to an herbarium sheet, some will have been stuffed and mounted, some will have been prepared as a dry skeleton, others will have been preserved in salts of various kinds, but most will have been fixed with spirit or formalin. Preservation subsequent to initial preparation can have varied a great deal too but most often fish are preserved in an ethanol solution of about 70% concentration. Imaging techniques need to be tailored to suit how a specimen was prepared and has been preserved. Thus, procedures for photographing wet specimens must include ensuring they do not dry out to a point where fragile structures become desiccated and damaged, or radiography might need to take into account that salt crystals have precipitated throughout the tissues of a specimen.

Specimens may be in a poor state. Initial preparation or subsequent preservation, which certainly will have had an effect on the condition of a specimen, may have been inadequate, in any imaginable manner. As a result they may be so soft as to collapse under their own weight or contorted, sometimes to an extreme degree, because they were jammed and bent into containers too small for their length. They may have lost most, even all, colouration as a result of having been exposed to natural or other light high in UV for long periods, or have been preserved in fluids that contain compounds that are inimical to colour pigments. Specimens may have become decalcified because the fluid environment in which they were fixed or subsequently preserved may have mobilised minerals from bony or calcified structures.

Recommendations and Special Considerations for imaging type specimens

If imaging type specimens poses no special technical problems, their status as types does impose a few extra considerations. These considerations stem from the fact that type specimens are the name bearers of taxa and are unique in place and time, and are, therefore, irreplaceable. The most obvious consideration is that an extra margin of duty-of-care is necessary. If poor practice results in damage to, or

Fig. 1. BMNH 1984.11.15.1 *Hypergastromyzon eubranchus* ROBERTS, 1991. Holotype. Fishes which differ significantly from fusiform shape benefit greatly from presentation of multiple views, in this case dorsal, lateral, and ventral. Copyright NHM.

loss of, a type specimen, another specimen cannot be substituted in its place – that type specimen is damaged or lost forever. For example, it is recommended that a wet type specimen be imaged immersed rather than 'dry', if imaging is going to take long enough for the specimen to even marginally dry out. The unique-in-place-and-time and name-bearer natures of type specimens impose another condition. No matter how aesthetically unappealing a type specimen might look because it is in poor condition or was inadequately or inappropriately prepared or conserved, it cannot be substituted by a more photogenic specimen. Since types are name bearing specimens, it is also recommended that additional views of them are provided to communicated the maximum amount of information

possible, especially if the species is not fusiform in shape. For example, specimens with significant dorso-ventral flattening need dorsal and ventral views in addition to a lateral view to capture essential aspects of shape (Fig. 1).

Additional types of images, for instance radiographs, should also be presented where available.

Fish Specimen photography at the NHM

Most fish specimen photography at the Natural History Museum, London (NHM) takes place in dedicated studios in the Photographic Unit (Fig. 2) and follows a standard methodology. Since the overwhelming majority of fish specimen photography in the museum is of specimens preserved in spirit, we will concentrate on techniques appropriate for alcohol preserved subjects.

Fig. 2. Senior author at work in his studio in the Photo Unit, NHM. Copyright NHM.

Digital capture is used almost exclusively, because it offers many advantages, including instant feedback. A digital single lens reflex camera (DSLR) coupled with a dedicated macro lens is ideal. The Photographic Unit uses systems where the camera is tethered directly to a computer. Shutter speed, aperture, and the final firing of the shutter are all controlled from the computer interface, such that the camera need never be touched after focusing. A delay of two seconds is set between the mirror lifting (in the camera) and actual activation of the shutter to reduce risk of vibration to an absolute minimum.

A DSLR provides the accuracy of framing and focus required for scientific specimen imaging. The camera should have a sensor of at least 6 mega pixels to capture the detail necessary, even if the image is to be reproduced at a small size. A macro lens provides a flat field with accurate focus across the frame. A lens of 50-60 mm range is ideal for most circumstances, but for smaller specimens or for close up detail a macro lens of 90-100 mm range is better because of the additional working distance, which allows room to light the specimen well. It is possible to use a ‘standard’ lens with extension tubes for fish specimen imaging but the level of detail, especially at the edges, may not be as good.

The camera and lens should be supported on a vertical copy stand with easy adjustment of camera height to allow for quick and easy framing of specimens of different sizes. A robust, sturdy support is essential to avoid any vibration. This camera stand should be placed on a solid table or bench, preferably on a concrete floor, again to prevent any vibration of the camera during imaging.

Electronic flash lighting is employed to avoid the specimen drying out. Continuous tungsten lighting can be used but should be turned on for the absolute minimum length of time – only for focusing and actual exposure. Even when using electronic flash, modelling lights should be restricted to just focusing. They should be left off at all other times.

A laboratory jack greatly assists focusing, because the combination of vertical positioning of the specimen with the jack and focusing

through the lens generally achieves more accurate results than by using the lens alone.

Methods for Digital Imaging of Fish Specimens

It is customary that fish specimens are oriented facing left for imaging (Fig. 1). Fins may be carefully spread to reveal shape, colouration, and meristics, but only if it can be accomplished without damaging the specimen. For some specimens fins can be gently 'massaged' into position but for others they can be held in place with suitably placed pins. Black entomological pins and a sheet of plastizote are excellent for this purpose. Pins are never placed through the specimen.

Specimens are normally placed over black velvet, which provides a clean black background whilst allowing clear subject definition. If contrast between a specimen and a black background is insufficient, a carefully balanced, backlit white background, avoiding shadows, can be used.

A suitable scale is positioned 'below' the specimen, by which we mean the scale appears to be below the belly of the fish when viewing the image. We prefer a white on black centimetre/millimetre rule five centimetres in length, but this is a matter of taste and can even be replaced with another form of scale during image manipulation subsequent to image capture. If measurements are to be taken from the image the scale is positioned, ideally, in the same horizontal plane as the plane of focus.

Small specimens should be imaged immersed in a transparent dish of suitable size. The base of the dish is lined with black velvet and the specimen covered with spirit to a depth of 2 to 3 mm. Immersion reduces, or avoids, specular highlights and entirely circumvents the specimen drying out.

Specimens too large to fit in a suitable dish, or if circumstances require it, are imaged 'dry', that is, out of fluid. It is paramount to avoid letting specimens dry out. This means working quickly, but carefully, and returning a specimen to its storage jar for a while if it

shows signs of drying. Signs that a specimen is drying out are ‘dulling’ and ‘clouding’ of the scales.

Normal practice is to use a single light source that illuminates from the top left of the specimen. Photographic reflectors are used to illuminate the areas of shadow. Where reflectors do not provide sufficient illumination a second and third light source can be employed, but always keeping the balance of lighting to the top left. Sometimes mirrors, instead of reflectors, are used to reflect light into the shaded areas.

Another way of lighting, and for some the preferred method, is to construct a white light tent around the specimen. This has the advantage of providing soft lighting that emphasises detail on the specimen and it also minimises or avoids ‘hot spot’ reflections.

A major advantage of digital capture is that a preview of the effects of the lighting setup and exposure are easily obtained by test exposures, without the cost and time involved in exposing onto film. Images from the camera are transmitted directly to the computer, where they are viewed to assess whether or not any adjustments are needed. The test exposure is assessed for detail in the highlights as well as in shadows. When making adjustments it must be remembered that it is always possible to regain shadow detail but loss of highlight detail is irrecoverable.

If exposing onto film is requisite, settings premised on readings of the incident light or from an 18% grey card should be used.

Handling of Digital Image Files

Where possible, images are captured in RAW format for maximum image quality. Adobe Camera Raw (ACR), a plug-in available for Adobe Photoshop, is the preferred software for handling RAW image files. This plug-in can open images directly into the main application for final editing and conversion to more accessible TIFF or JPEG formats. Colour balance, exposure and contrast should be set as close as possible to an optimal image when loading the RAW image file into Photoshop. This can be achieved for colour balance (or white

balance) by photographing an 18% grey card under the same conditions in which the specimen is placed. These settings are then transferred to each image. Once an image has been opened in Photoshop, subsequent adjustments to an image should be kept to a minimum. Typically, the only adjustments needed should be to levels to improve contrast, tonal range, and a small amount of sharpening to counteract the anti-aliasing filter present on all digital cameras. When converting to a TIFF or JPEG format the best image possible should be produced to avoid destructive editing after to conversion.

Radiography Techniques for Fish Specimens

The outcome of radiography of fish specimens depends on many factors, including equipment available. Local practice must be tailored accordingly. At the NHM fish specimens are radiographed with a soft x-ray source. The source tube is mounted on a swing arm that can be adjusted vertically, enabling efficient exposure of almost any specimen. Exposure time (in seconds), voltage (in kV), and amperage (in mA) are controlled by the operator at a control panel.

A high quality, fine grain, double sided industrial film is used at the NHM. The film is loaded into a light-tight, paper film holder and placed on a bench under a heavy sheet of clear polythene, beneath the radiation source. Specimens are then arranged on the plastic sheet covering the film holder. Contorted specimens may be taped to the plastic sheet to straighten them during exposure. Fine grain film is required for small specimens and because fishes possess many small, thin bones. Film with emulsion on both sides is extremely useful. It can be over-exposed, then, if required, the emulsion scraped from one side of the developed film to produce a radiograph of good quality. This is accomplished by placing a drop of water on the film where the emulsion is to be removed. The emulsion absorbs the water and is easily lifted. Removal of the emulsion from one side of the film is especially useful in those circumstances where adequate exposure of dense, bony elements results in more delicate structures, such as finrays, being over-exposed. One side of the film can be scraped to remove the emulsion from over the finrays to

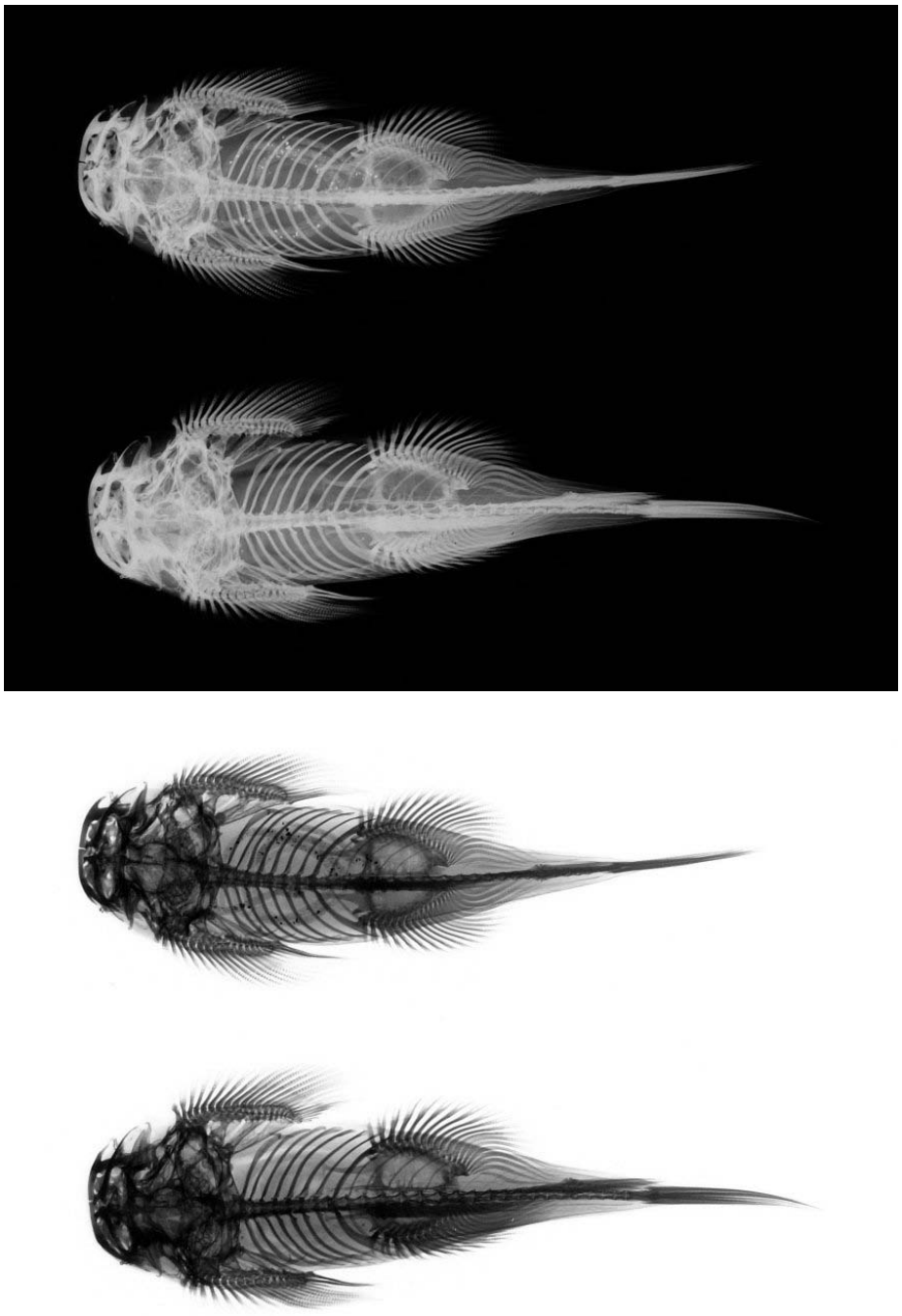


Fig.3. BMNH 1982.3.29.111-112. *Gastromyzon contractus* ROBERTS, 1984. Paratypes. Negative and positive digital images of a radiograph of fish specimens, dorsal view. Copyright NHM.

reveal them in fine detail while leaving the film intact over denser structures, retaining good resolution for them too. Specimens that have been decalcified, or are or simply poorly ossified or mineralised in the first instance, may not make good subjects for radiography.

In most circumstances six variables affect the outcome of radiography of fish specimens: appropriateness of the film, thickness of the specimen, distance between the radiation source and the specimen, exposure time, voltage (in kV), and amperage (in mA). Experience in controlling the later four variables results in consistent, high quality radiographs. For most work distance between the x-ray source and the specimen (approximately 40cm for the set up at the NHM), exposure time (usually 120 seconds in our work), and amperage are kept constant. Most often, then, voltage is the variable adjusted to accommodate specimens of different thicknesses. Large, thick specimens may require adjustment of all four. For specimens that vary substantially in thickness from one place to another, for instance between thickness at the point of the shoulder girdle and thickness at the caudal peduncle, it can be useful to mask the thinner area. Heavy card laid over the thinner area reduces its level of exposure.

Once radiographs have been obtained, they are rendered as digital images through digital photography. The radiograph is positioned on a light box, masked if necessary, and photographed. The digital image file can then be manipulated with standard image software to present a negative print, positive print, or both.

Summary

Producing high quality images of fish specimens requires good equipment and personnel with knowledge, skill, and experience. It is also time consuming and requires organisation if the task is to image many specimens. If specimens are of a similar size and imaging requirements do not vary much between them, they can be imaged at a rate of 10-20 per hour. If setup has to be adjusted substantially between specimens it can take significantly longer. At the NHM it is not unusual to image several dozens of specimens for a single job. A job that size might take several days. The NHM has over 2,960 database records for primary type specimens of fishes in its collection, and over 9,780 database records for primary and secondary type specimens of fishes. A task to image all type specimens of fishes at the NHM, or just the primary type specimens,

would take many months, and be very costly in terms of personnel time.

Acknowledgements

Over many years our colleagues in the Photographic Unit and Zoology Department of the NHM have generously engaged in discussion concerning the imaging fish specimens – we have benefited greatly from these discussions. Specimen imaging by HAROLD TAYLOR, Photographic Unit, NHM.

3D Imaging for a Virtual Museum

Bird Type Specimens of the Zoological Museum Amsterdam

HUUB VELDHUIJZEN VAN ZANTEN, EDWIN VAN SPRONSEN & RUUD ALTENBURG

ETI BioInformatics, University of Amsterdam, Mauritskade 61, NL-1092 AD Amsterdam, The Netherlands. E-mail: info@eti.uva.nl, http://www.eti.uva.nl

Key words: 3D Imaging, birds, type collection, digital imaging, virtual museum, Quicktime VR.

Introduction

In this chapter we will elaborate on the solutions we have chosen for digitising three-dimensional type specimens. By means of rotational object photography these specimens were made available through the internet for interactive manipulation. A computer mouse is all that is needed for rotating them and even magnifying them larger than life-size. For a while, it has been the intention of the ornithology department of the Zoological Museum of the University of Amsterdam to photograph its bird type specimen collection (ROSELAAR & PRINS 2000) in such a way as to allow them to be used as recognisable reference material. One or several photographs were not good enough to serve this purpose. By using Quicktime VR software, photographs from a lot of different angles were combined to present an interactive, video-like display. In the near future this visualisation of the type collection will be linked to the on-line collection database of the museum, which at present is already available through GBIF. In this way, not only scientists but also laymen will have full access to the often vulnerable collection items. After the rotational object photographs were made available on the internet, they drew a lot of attention from both the professional and

the popular press. Apart from this popularisation, the three-dimensional presentation decreases the need for taking vulnerable and rare specimens out of their storage when professionals need them for examination. Any scientist can study the type specimens on-line and decide whether travelling to Amsterdam for seeing the real specimen is necessary.

Fig 1. *Diphyllodes (Paradisea) gulielmi III* Currently considered a hybrid between *Diphyllodes magnificus* and *Cicinnurus regius,* the unique reflective plumage can not be fully shown in a few photographs.

Taxa Group:	Aves (Bird types)
Institution:	ETI / (ZMA)
Persons Responsible:	Huub Veldhuijzen van Zanten (photography), Edwin van Spronsen (technique and software), Ruud Altenburg (website)
Number of Images captured:	36,000
Website:	http://ip30.eti.uva.nl/zma3d/

The need for digitalisation

International agreements for conserving biodiversity have increased the focus on museum collections and the need for making their content available to a worldwide audience. The Dutch GBIF node subsidises institutions in the Netherlands for putting their collections on the internet. As a proof of concept, bird type specimens were used. Because of their rarity, scientific value and vulnerability, type specimens are no longer sent to other institutions for study purposes. Transport is also restricted by CITES regulations and sometimes by health considerations, e.g. in the case of birds during the recent bird flu epidemic.

Intended for a broad spectrum of interested users, both scientific and layman, the bird types of the Zoological Museum of Amsterdam are presented on the web as interactive virtual reality objects using QuickTime VR files. With a specially designed setup, birds have been digitally photographed from various angles; the images were subsequently processed by computer to create fluently rotating movements of the objects on screen. Using the mouse, the 3D files of bird type specimens can be rotated on the computer screen and in this way the characteristic plumages of the various bird species can be studied from all angles. The files are a result of a demonstration project in which an efficient method was developed to make photographic images of birds – as if they were real objects. Furthermore, it is no longer necessary to touch and handle the often vulnerable and rare objects.

Capturing images

To create a stable and uniform setting for illumination of the object to be digitized, a photo studio was created with a rotational platform with a robotic arm, a black background, 2 or 3 studio lamps and an Apple computer connected to a digital camera and the rotational platform.

A simple applescript controls the camera and the rotation platform. It tells the platform to turn the object table by 10 degree intervals, then take a short rest to dampen vibrations and take a picture with

the last setting used on the camera, then rotate again etc. 36 photographs taken at 10 degree intervals provide a fluent movement in the final movie. One full rotation takes less than 10 minutes. By also taking series of pictures at different inclinations from plus 45 degrees to minus 15 degrees, a good overview was created especially for mounted specimens.

Fig 2. Complete setup for rotational object photography

The platform was placed on a set of tables to create a good working height. We preferred manual focus over autofocus, because the latter was sometimes confused by the black background and because all photographs in the series required the same treatment in order to create a fluent animation.

Handling Specimens

Handling was done by the curator of the bird department. The specimens remained in the same building as we were able to move our entire studio in a small van. Labels were carefully removed and replaced in strict order of photography. The specimens were exposed to a bright light source during photography for a maximum of 0.5-1.5 hours.

Mounting Specimens / Specimen Preparation

Bird skins were placed on a needle. As the skins never were sewn tight completely, placement of the needle next to the tail was never a problem. Cotton-wool stuffed birds sometimes were more difficult to prepare for photography.

Fig 3.
Example of a mounted specimen (*Crypturellus tataupa tataupa*, ZMA 4962).

Capture Device

Our semi-automated photo studio requires a room with dimensions 4 x 3 x 3 m. On a couple of firm tables stands a motorized turntable to support the object being photographed. A digital camera is connected to an adjustable cantilevered swing arm. A software program on a Macintosh computer controls the following hardware:

- The turntable that rotates the object in the horizontal plane, typically in 36 steps at 10 degree intervals.
- The swing arm that tilts the camera in the vertical plane, typically in 5 steps at 15 degree intervals.
- The shutter of the camera.

Images captured by the digital camera are transferred automatically to the computer. Then each set of images was ported to a server and to another computer for storing and processing. They were batch-processed in Photoshop. This means the adjustments performed on one photo were recorded and stored as an action in Photoshop, then this action was used in order to apply the same adjustments to all photographs in the same folder. The edited images were then processed in QuickTime VR Authoring Studio software, which is used to generate an interactive movie suitable for web downloads. Processing 180 images per object and creating a movie can be done simultaneously while at the same time a new object is photographed.

Special care was taken to create a totally black background. With a special black velvet of 150 x 300 cm, most light and shadows were absorbed. This reduced also the pixel noise and the final file size by 50%. The illumination consisted of two or three studio lamps of 650 W with a frosted lens. Soft boxes around the lamps were tried. When overheating, the automatic off-switch sometimes resulted in a series of darkened images. Soft boxes therefore proved unreliable for this task and were abandoned.

White foam was used to create extra reflection from below as most light came from above. Through software we added delay times to eliminate vibration after a turning event. This was more important for larger objects.

Total time needed for photography was 35-45 minutes for 180 images. If only a single rotation is needed this could be done in less than 10 minutes.

In this project a total of 3 QTVR files for each object were made for different quality web connections.

Hardware
Nikon D1X camera body, Nikkor (manual) lenses 28 mm, 24 mm and 20 mm.
Kaidan Magellan M-2500 robotic arm for Rotational Object Photography.
Firewire IEEE 1394; connection cable between camera and computer.

The Nikon D1X camera is currently the only Nikon camera body that can be operated by a computer with applescript software. As this camera is discontinued, the Kaidan company needs to adjust its software.

Lighting
2 (3) Photo lamps 650 W PF 813.

Resolution
2000 x 1312, 7.51MB, object cropped to 728 x 464 pixels.

Software
Mac OS 9 (including AppleScript, part of the OS)
Connected to Kaidan rotational platform and arm with Nikon D1X
Nikon Capture 2
eMCee Motion Control Software
Adobe Photoshop 7.0
QuickTime VR Authoring Studio

Scale
No scale bar was used in the photographs, because measurements of stuffed skins are not very reliable in the first place. Also, the change of perspective can frustrate the use of a scale bar as this is only accurate in one particular plane. However, we are currently pursuing ways of dealing with this problem by adding scale bars with depth correction that will yield correct dimensions in any plane and from any angle.

Working with Images

Storage Formats

After a quality check we concluded that the camera setting with compressed files as .JPG was good enough for our goal. The resolution of 2000 x 1320 has enough detail for our maximum image size of 724 x 464. The total storage capacity needed for this project was with compressed files ca. 40 GB. A camera setting with RAW images would need at least a 10 fold storage capacity and much faster hardware.

Enhancing Images

Recorded images are stored as medium size JPEG, 0.7-1Mb.
Nikon Capture also stores a separate data file for each picture in which all setting information is stored (with the exception of information about some older lenses that cannot be recognized by the camera's electronics).
All image processing (e.g. cropping, scaling, filtering) was done with Photoshop “action” batch files.
Special care was taken to reduce noise in the background, this saved up to 50% in the file size! Due to accumulating dust on the background the file size was becoming increasingly larger. The background was not completely black anymore and therefore the file size grew out of proportion.

Quality Control

Three phases:
1. Calibration and standardization of light conditions the Nikon D1X with the white balance setting in the Preset function, photograph a white sheet of paper, record this setting, then take a picture of a Kodak colour card and compare the results.
2. Processing of image files in Photoshop, saturation sometimes needed adjustments especially in light green colours (e.g. Eclectus Parrot).
3. Quicktime VR files checking the results, centering movements could only be checked when all images were imported into a QTVR movie.

Naming Images

A filename consists of an automatically generated file number and is stored into folders with a batch number (in order of photography) and a unique collection number.

Metadata Information

Nikon Capture software is able to store all relevant data for each image. This data remains independent and can be stored for later reference. This is only valid for autofocus lenses.
In the QTVR files Metadata is stored in Movie properties, containing, name of type, current taxon name, collection number, remarks, copyright, photography.

Software

Adobe Photoshop 7.0
Quicktime VR authoring studio
Microsoft Excel

An Excel-sheet was prepared for calculating the number of images when the first number was entered. This was particularly important when creating different QTVR files with less vertical series of images. Three types of files were made with 180, 108 or 36 images.

Wish-list:

Image quality is not used optimally. To obtain a useful file size we limited the maximum file size to approximately 10MB. The original file resolution of 2000 x 1312 was reduced to 728 x 464 pixels. In the future the original files can be used to update the QTVR files to larger computer screens, as they are getting larger we could even use a height of 1024, instead of 728 pixels.

Maintaining the website

Linking to/within website

For this project, an HTML template was prepared for each taxon. Three versions of the prepared QTVR movies were placed on the server and the appropriate data (screenshot from each movie and text describing each specimen) were added to the HTML page for each taxon. Links opening the QTVR movies were coded completely

in JavaScript. By opening the movies in such a way, it is more difficult for inexperienced visitors to download the movies themselves. All separate HTML pages were connected to each other using a central index page and next/previous buttons.

IPR policy for Images
Legal notice is included as link inside open window.

Back up
As the site was written entirely in static HTML pages, backing it up is as simple as writing the appropriate folder on the server to a DVD. During the development phase, the site was added to the regular backup cycle (to DDS3 tape).

Storage capacity
As there are three QTVR movies for each taxon (about 800 K, 1.5 MB and 9 MB each), the webserver requires a large hard disk.

Maintaining links
Links are checked automatically by the web development software (Adobe GoLive).

Linking to other systems
Linking to other websites will be easy if these use a standard procedure for each taxon. The appropriate page on the website should be accessible by an ID in the URL (GET procedure). As the website is prepared in static HTML, linking the other way round also should not pose many problems. Storing the data in a database simplifies exchangeability.

Recently the ZMA collection databases were brought online. As the QTVR filenames were named after the collection number of the specimen, it was relatively easy to link them to the collection database. This way files are also available when searching the online ZMA birdcollection database for species name or for all type specimens. See http://145.18.162.60/zmawebsite/search_in_collection.php?collection=Aves

As this specimen data on this website is stored in this database, it can also be queried from the GBIF site. The connection between the databases is established using a 'database wrapper'.

Wish-list:
For the ZMA site, static HTML pages were prepared for each taxon. For large scale projects it is advisable to prepare a database with all relevant info. If specimen data are stored in a database, a website can be prepared with much less effort. Storing data in a database also has the advantage that the data can be searched.

Results

The ZMA collection databases are online. Promoting this new application of the QTVR technique by a press-release helped to become a national front page news item and achieve an internet top-100 ranking which together resulted in 7000 visitors in 3 weeks.

Discussion

We expect this prototype to be transferred to an easy to use international access point to all biodiversity information and it should be able to be searched through the GBIF data portal.

Creation of a website is one thing, visitors and use of the website depend on publicity and ease to be found with internet search engines. We found out that it helps tremendously if the page is linked through popular sites. Therefore we recommend spending time to analyse connection and promotion possibilities on other websites. Write press releases, contact scientific internet magazines, specific interest sites and popular weblogs.

Last but not least, illustration of the museum databases with QuickTime Virtual Reality files can help to promote and educate about the importance of collection material for the international community.

News

- A follow-up started February 2005 with extinct birds from the collection of the National Museum of Natural History, Naturalis.
- A nomination by Archives & Museum Informatics for Best of the Web Award 2005 in the category Research Site.

Acknowledgements

We thank MAARTEN KAPELLE and TINDE VAN ANDEL for their role in NLBIF and funding of this project. Our thanks are also due to TINEKE PRINS and CEES ROOSELAAR at the ZMA Museum for their elaborate work on the collection data and the supply of type material, to GIDEON GIJSWIJT for hardware support at ETI, to HANS VAN BRANDWIJK, JAN VAN ARKEL and LEO MALLIE for photographic tips and supplies.

References

LESLIE, M. (ed.) (2005): Images. A bird in hand. – *Science*, **307**: 21.

ROSELAAR, C. S. & PRINS, T. G. (2000). List of type specimens of birds in the Zoological Museum of the University of Amsterdam (ZMA), including taxa described by ZMA staff but without types in the ZMA. – *Beaufortia,* **50**: 95-126.

Taxa Group: Plants (Alismataceae, Bignoniaceae, Cactaceae, Clusiaceae, Onagraceae, Rutaceae, Simaroubaceae, and *Croton* (Euphorbiaceae))
Institution: National Botanical Garden of Belgium, Belgian Biodiversity Information Facility
Person Responsible: PIET STOFFELEN, ALAIN VANDER VELDE, PATRICIA MERGEN, FRÉDÉRIC WAUTELET
Number of Person Hours per year devoted to imaging:
Number of Images captured and stored each year:

Purpose of Image

Target Audience: Botanists
Form: Custom website, DarwinCore and ABCD mapping to GBIF

Capturing

Handling Specimens:
Mounting Specimens / Specimen Preparation: vouchers, herbarium sheets
Capture Device: Scanner Epson XL 1600
Lighting (including light spectrum): Not applicable
Resolution (dpi): 600
Software: Espon driver
Automontage: None
Scale: 1x1

Working with Images

Storage Formats: TIFF (bzipped)
Software: Access through web browser, using home made Perl scripts and ImageMagick on Linux Server
Linking to/within Databases: filename is barcode number
Enhancing Images: Cropping on line (intranet through web browser), using home made Perl scripts and ImageMagick
Automontage: None
Quality Control: verification of focus, that all label text are present, verification of compressed files
Naming Images: barcode used, if second image for same barcode -> barcode_a, barcode_b, ...
Metadata Information: filename, date of scan, who has scanned, on which computer, scanner, size of file...
IPR policy for Images: all images are copyright reserved

Maintaining databases: internal in the Botanical Garden
Back up: on high capacity hard disk at the Botanical Garden
Storage capacity: BeBIF webserver : ~500GB
Maintaining links: with database
Linking to other systems: In development

Taxa Group: Earwigs (Dermaptera)
Institution: Staatliches Museum für Naturkunde, Stuttgart (SMNS)
Person Responsible: FABIAN HAAS
Number of Person Hours per year devoted to imaging: 1
Number of Images captured and stored each year: 0-200, occasionally 2000 if there was a SYNTHESYS grant to visit institutions

Purpose of Image

Target Audience: Public and Scientists, no special focus on type material, though that is welcomed
Form: own Website and sometimes paper publications

Capturing

Handling Specimens: myself, each specimen separately
Mounting Specimens / Specimen Preparation: usually pinned and /or glued to card board and pinned to a small photographing stage made from foam
Capture Device: Olympus E100, E20 or Nikon D70, whatever was available
Lighting (including light spectrum): standard flashes
Resolution (dpi): I am using max. pixels available with camera
Software: Adobe Photoshop, GraphicConverter to deal with many images in the same way and at once
Automontage: no
Scale: about live size, or 1:2 to 2:1

Working with Images

Storage Formats: TIFF, JPEG
Software: Adobe Photoshop, GraphicConverter to deal with many images in the same way and at once
Linking to/within Databases: yes, by file name after Naming Images (see below)
Enhancing Images: yes but minimal: sometimes cropping, contrast/brightness and sharpness
Automontage:
Quality Control: yes
Naming Images: Institution_Year_FileNamefromCamera
Metadata Information: EXIF, not currently used
IPR policy for Images: images usually copyrighted by the museums

Maintaining databases

Back up: yes, by tape drive and CDs, no DVD yet
Storage capacity: unlimited
Maintaining links: yes
Linking to other systems: no

Taxa Group: Fruit Flies (Diptera, Tephritidae)
Institution: Natural History Museum (London) & Royal Museum for Central Africa (Tervuren)
Person Responsible: I.M. WHITE & M. DE MEYER
Number of Person Hours per year devoted to imaging: variable
Number of Images captured and stored each year: variable

Purpose of Image

Target Audience: Images were originally intended for use in an electronic multi-entry key (CABIKEY) for facilitating identification. Target audience was researchers interested in identification of fruit flies. Current target audience is larger since intended for part of website with taxon and specimen related information (contained in relational database) for more general public (quarantine officers, applied entomology, etc.). Hence this is not part of a routine digital imaging programme but a specified action for a specific objective.
Form (e.g. Website, GBIF portal, paper publications): See above: initially on CD-ROM, now envisaged for website and GBIF porta

Capturing

Handling Specimens: Use of mounted collection specimens
Mounting Specimens / Specimen Preparation: use of mounted collection specimens on pins, no further preparation
Capture Device: JVC KY-55BE video camera mounted on a Zeiss Stemi SV 11 APO stereo microscope. Images captured via Synoptics Prysm colour framegrabber card
Lighting (including light spectrum): Shott KL 1500 electronic fibre optic unit (light spectrum unknown)
Resolution (dpi): c. 70 d.p.i. (screen resolution)
Software: Automontage
Automontage:
Scale: Variable

Working with Images

Storage Formats: TIFF converted to JPG
Software: Edited with Paint Shop Pro 3.1
Linking to/within Databases: To be determined
Enhancing Images: Original images (768x576) were cropped to the required detail and then scaled to make them fit within a

640x480 pixel frame, and subjected to a light sharpening (wings to a heavy sharpening). A few images were colour adjusted to remove a blue shift in their colour balance (due to a mal-adjusted camera).

Automontage:

Quality Control:

Naming Images: Each image is given caption and unique taxon ID number which is also used in relational database

Metadata Information:

IPR policy for Images: Images remain copyright of Natural History Museum where images were taken. Agreement between Natural History Museum and Royal Museum for Central Africa that images can be used for above mentioned website as long as copyright is indicated on each individual image.

Maintaining databases

Back up: On CD-Rom

Storage capacity: BeBIF webserver : ~500GB

Maintaining links: With database

Linking to other systems: In development

Taxa Group: Parasitic Hymenoptera (Insecta)
Institution: Department of Entomology, University of California, Riverside
Person Responsible: R. BURKS and M. BUFFINGTON (John Heraty Laboratory)
Number of Person Hours per year devoted to imaging: 1040
Number of Images captured and stored each year: 1500

Purpose of Image

Target Audience: Primary: Systematic Hymenopterists; Secondary: Biological Control Community
Form: Web: Personal pages Photographic key to genera of Eulophidae: http://cache.ucr.edu/%7Eheraty/Eulophidae/index.html; public digital image databases (e.g. www.MorphBank.com); digital publications (e.g. Zootaxa). Currently we are focused on making publicly available images of all vouchers specimens used in molecular phylogenetics.

Capturing

Handling Specimens: Specimens collected in a variety of fashions, preserved in ethanol, then mounted for examination if not needed for DNA extraction.
Mounting Specimens / Specimen Preparation: Specimens are removed from ethanol and allowed to dry. If the integument of the specimen is thin (e.g. Chalcidoidea), the specimens are run through an ethanol gradient to 100%, then through two washes of HMDS for final drying (Heraty and Hawks 1998); harder specimens (e.g. Cynipoidea) were simply air dried. Specimens are mounted on either white or gray card points or cards, and held in place with shellac gel; the mounted specimen, in turn, is then pinned with an insect pin. Pinned specimens are held in place for imaging via plasticine modelling clay in a watch glass.
Capture Device: JVC KY-70 3 CCD digital camera fitted to Leica MZ16 zoom lens via simple c-mount in focusing rig. We also use a Zeiss Axioskop-2 compound microscope primarily for slide specimen imaging, but we have also adapted this scope for imaging whole mount specimens.
Lighting (including light spectrum): This is the primary focus of our imaging research. Areas of extreme contrast, whether due to glare or a black specimen on a white card mount, results

in loss of image data and invariably produces chromatic aberration. Correct lighting, in conjunction with correct background colour, result in specimens whose areas of extreme contrast are significantly reduced. This results in a sample whose light properties are favourably balanced according to what the digital camera is capable of recording (after the camera has been white balanced). The light source of choice for our work has been the Fostec fibre optic illuminator with diaphragm control. The diaphragm control is important for maintaining correct colour temperature while adjusting light intensity. BUFFINGTON et al. (2005) summarizes much of our work to date. We found that through the development of light boxes, we can disperse the incident light sufficiently to eliminate glare from very small (e.g. 0.75mm), very shiny, black insects. This results in an image that contains the most data (i.e. correctly resolved morphological features) and optimizes postproduction through Automontage and/or Photoshop. We have also experimented with varying the background colour below the specimen to minimize the effects of extreme contrast. The problem we encountered here was a loss of resolution around the edges of specimens (e.g. along wing margins or the edges of legs). An 18% grey card (used by professional photographers) has worked the best in most situations. This type of background creates a neutral background midway between black and white.
Resolution (dpi): 1360 x 1024 pixels, full color
Software: Automontage v. 5.0
Automontage: Fixed depth, patch size 10-40.
Scale: 0.5 – 5.0 mm (adult size)

Working with Images

Storage Formats: BMP, TIFF
Software: Adobe Photoshop, Adobe Illustrator
Linking to/within Databases: Cumulus 6.0
Enhancing Images: Adobe Photoshop
Automontage: Yes
Quality Control: none
Naming Images: none
Metadata Information: Stored on MorphBank or on MorphoNet.
IPR policy for Images: none

Maintaining databases

Back up: multiple copies of all images on multiple computers; CD backups; HD backups.
Storage capacity: 200 GB
Maintaining links: Manually
Linking to other systems: MorphBank

Acknowledgements. Everyone in the Heraty Lab has contributed in some way in helping to develop these techniques. We would also like to acknowledge GT Vision for their continued support in our efforts. Funding was provided by NSF PEET grants BSR 9978150 (awarded to Heraty and Pinto, UC Riverside) and BSR 9629515 (awarded to Woolley and Wharton, Texas A&M).

References

BUFFINGTON, M. L., BURKS, R. A.& MCNEIL, L. (2005): Advanced techniques for imaging Parasitic Hymenoptera (Insecta). – *American Entomologist,* **51**: 50-54.

HERATY, J. M. & D. HAWKS, D. (1998): Hexamethyldisilazane: chemical alternative for drying insects. – *Entomological News,* **109**: 369–374.

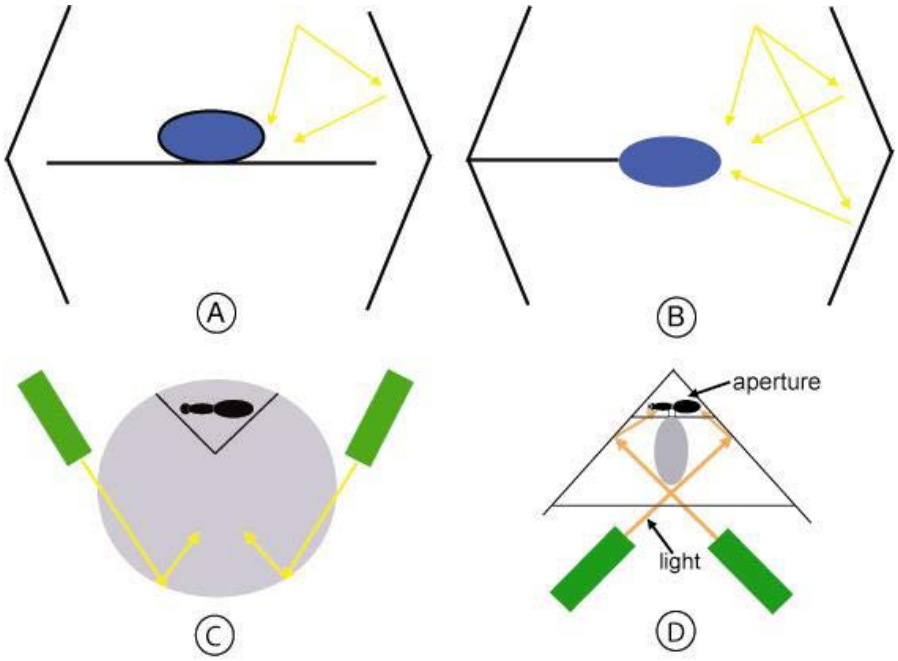


Figure 1

Fig. 1 A-C demonstrates how a light chamber works to disperse light in a typical stereoscope application. ‘A’ and ‘B’ are side views of the chamber and demonstrate how the incident light reflects upon entering the chamber; ‘C’ is a top view of the chamber. The bottom of this chamber is open, and the entire chamber fits around the specimen after the specimen has been properly aligned for imaging. Figure 1D is a schematic representation of a smaller lighting chamber designed for a compound microscope application.

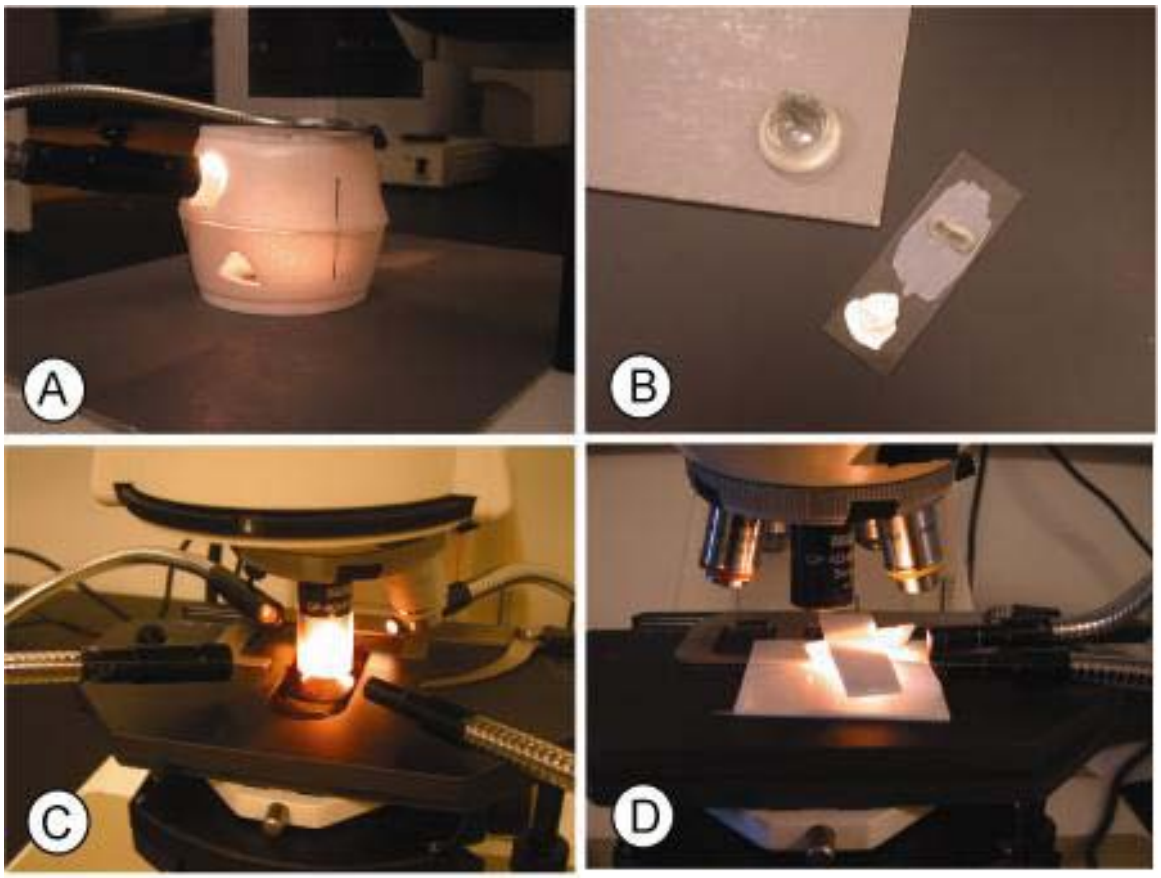


Figure 2

Fig. 2 illustrates the various types of equipment used for imaging of Microhymenoptera. Styrofoam chambers for use with the stereoscope (Fig. 2A) and compound scope (Fig. 2D) are easy to construct and very inexpensive. Fig. 2b shows the photo grey card, primer paint on a microscope slide and plasticine clay in a watch glass that are routinely used for controlling light and specimen positioning. Fig. 2C shows how a ring of Mylar film can be used with illuminator wands to produce a cheap and effective light chamber around the microscope objective lens.

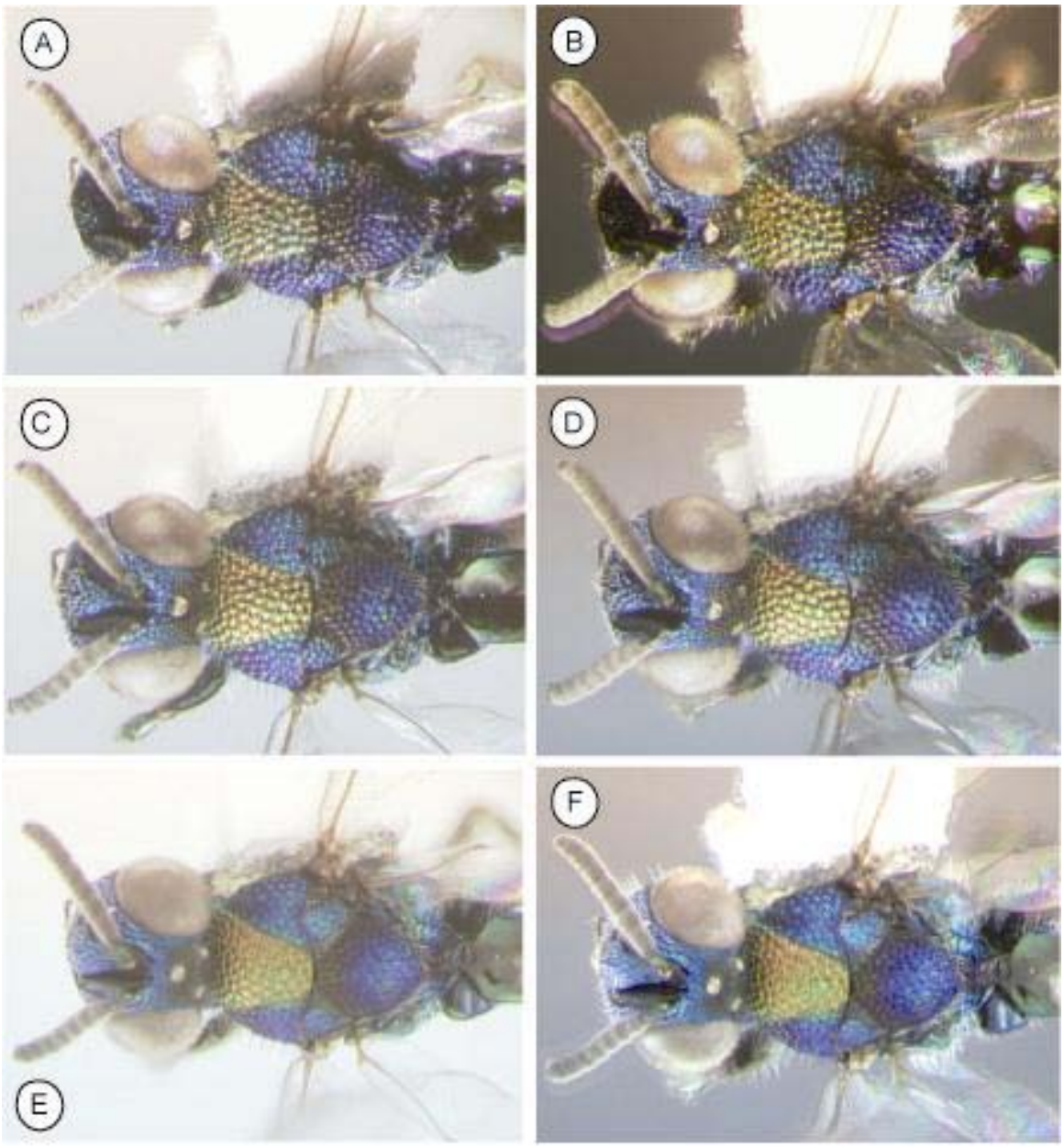


Figure 3

Figure. 3 demonstrates how differently the same specimen can be resolved by the same camera under different lighting schema. A, naked light, white background; B, naked light, 18% grey card background; C, Mylar filtered light, white background; D, Mylar filtered light, 18% grey card background; E, light chamber dispersed light, white background; F, light chamber dispersed light, 18% grey card background.

Taxa Group: Butterflies (1) (Lepidoptera: Papilionidae, Pieridae, Lycaenidae)
Institution: Staatliches Museum für Naturkunde Stuttgart, Germany
Person Responsible: CHRISTOPH HÄUSER, AXEL STEINER, JOACHIM HOLSTEIN, projects GART/GloBIS
Number of Person Hours per year devoted to imaging: variable
Number of Images captured and stored each year: variable

Purpose of Image

Target Audience: Entomologists (especially taxonomists interested in studying the type specimens), biologists, general public.
Form: Website (http://www.s2you.com/platform/lex/globis/).

Capturing

Handling Specimens: Specimen setup see HÄUSER et al. in this volume.
Mounting Specimens / Specimen Preparation: Specimens used are already pinned and spread; so no further preparation is necessary.
Capture Device: NIKON Coolpix 950 / 990 / 5700
Lighting (including light spectrum): 22 W Ringlight "VITA-LIGHT fluorescent full spectrum" (5500 K), mounted in reflector-box (details and images see HÄUSER et al. in this volume).
Resolution (dpi): Nikon 5700: 2560 x 1920 pixels, 300 dpi, 24 Bit
Software: none.
Automontage: no.
Scale: variable.

Working with Images

Storage Formats: JPG
Software: Adobe Photoshop
Linking to/within Databases: Image management in Filemaker by using links to image files.
Enhancing Images: Colour adjustment: before photosessions the white balance function of the camera is used measuring the white field on a Kodak colour scale. - Brightness is adjusted by reference to the grey sheets used as background. - Images are slightly sharpened (unsharp masking function; settings: intensity

50%, radius 1.3 pixels, threshold 1). Images are cropped around the specimen and labels.
Automontage: no.
Quality Control: visually.
Naming Images: see HÄUSER et al. in this volume.
Metadata Information: embedded EXIF metadata.
IPR policy for Images: All images remain copyright of the institutions holding the specimens.

Maintaining databases

Back up: Original and processed images on CD/DVD, processed images also on local museum server.
Storage capacity:
Maintaining links:
Linking to other systems: Data and images are available via SysTax (http://www.biologie.uni-ulm.de/systax/) and are being implemented into GBIF.

Taxa Group: Butterflies (2) (Lepidoptera: Papilionidae & Charaxinae)
Institution: Royal Museum for Central Africa, Tervuren, Belgium
Person Responsible: UGO DALL'ASTA
Number of Person Hours per year devoted to imaging: already done
Number of Images captured and stored each year: not applicable

Purpose of Image

Target Audience: Zoogeographers
Form: Custom website, DarwinCore and ABCD mapping to GBIF

Capturing

Handling Specimens: FRANS DESMET
Mounting Specimens / Specimen Preparation: Not applicable
Capture Device: Coolpix 990
Lighting (including light spectrum): TL ring of 20 cm diameter
Resolution (dpi): FINE - JPEG images according to Coolpix standard
Software:
Automontage: no.
Scale: Depends on the wingspan of the butterfly.

Working with Images

Storage Formats: JPEG
Software:
Linking to/within Databases:
Enhancing Images: No.
Automontage: No.
Quality Control: No.
Naming Images: Internal numbering of the Coolpix 990.
Metadata Information: on separate Word (doc) file.
IPR policy for Images: None.

Maintaining databases: periodic upgrade from museum using Access database.

Back up: offsite server and on CD-ROM.
Storage capacity: BeBIF webserver : ~500 GB.

Maintaining links: with database.
Linking to other systems: In development.

The butterfly part of the ENBI database was started as a project for documenting the Albertine Rift which is an African hotspot of biodiversity. The Albertine Rift is situated in the eastern Democratic Republic of Congo and adjacent countries and is widely accepted to be one of the top three biodiversity hotspots and areas of endemicity in Africa. It holds two UNESCO World Heritage Sites. Given the continued political and economic crisis in the region, it is under enormous threats. Despite its possibly being the top biodiversity hotspot in Africa, its status could not even be properly assessed in a global study due to data deficiencies (MYERS et al., Nature, 2000 (403): 853-858).

The Belgian Federal Scientific Institutions hold vast collections of the region, e.g. resulting from surveys in the National Parks (then under Belgian colonial rule) and more recent expeditions. These data, when made accessible, are an important tool in decision making on the selection of priority regions for conservation.

For the butterfly part of the Albertine Rift database two groups were chosen: the Papilionidae and the Charaxinae. Of these, all specimens of the RMAC collection labelled from the Democratic Republic of the Congo, Rwanda, Burundi, Uganda, Kenya and Tanzania were included in the database. During a second phase, also data from Papilionidae of the National Museums of Kenya, Nairobi, Kenya, were added. The database itself is Access Microsoft based and was designed especially for this purpose by entomologists.

The photographs were taken so as to document all taxa (species and subspecies) present belonging to the family/subfamily. Photographs were even included of taxa occurring outside the Albertine Rift, but present in the collection. Consequently, all Papilionidae and Charaxinae present in the MRCA are documented. Both sexes were photographed, recto and verso, which means a maximum of four photographs per taxon.

Taxa Group: Geometrid Moths (Lepidoptera: Geometroidea: Geometridae)
Institution: Zoologische Staastssammlung München (ZSM), Germany
Person Responsible: AXEL HAUSMANN, SVEN ERLACHER
Number of Person Hours per year devoted to imaging: Variable
Number of Images captured and stored each year: Variable, approx. 800

Purpose of Image

Target Audience: Entomologists, especially taxonomists interested in studying the type specimens; biologists; general public.
Form: Website:
http://www.biologie.uni-ulm.de/cgi-bin/query_all/query_all.pl?lang=d&pr=gbif-e1
http://www.biologie.uni-ulm.de/systax/daten/index_e.html
http://www.s2you.com/platform/lex/globing

Capturing

Handling Specimens: Specimen setup see chapter on butterflies by HÄUSER et al. in this volume.
Mounting Specimens / Specimen Preparation: Specimens used are already pinned and spread.
Capture Device: NIKON Coolpix 990, 4500.
Lighting (including light spectrum): 22 W Ringlight "VITA-LIGHT fluorescent full spectrum" (5500 K), mounted in reflector-box (details and images see HÄUSER et al. in this volume.
Resolution (dpi): NIKON 4500: 2272 x 1704 pixels, 300 dpi, 24 bit.
Software: None.
Automontage: No.
Scale: Variable.

Working with Images

Storage Formats: JPG.
Software: Photoshop Elements 2.0
Linking to/within Databases: Image management in Filemaker by using links to image files.
Enhancing Images: Colour adjustment: before photosessions the white balance function of the camera is used measuring the

white field on a Kodak colour scale. - Brightness is adjusted by reference to the grey sheets used as background. - Images are slightly sharpened (unsharp masking function; settings: intensity 50%, radius 1.3 pixels, threshold 1). Images are cropped around the specimen and labels.
Automontage: No.
Quality Control: Visually.
Naming Images: see HÄUSER et al. in this volume.
Metadata Information: Embedded EXIF metadata.
IPR policy for Images: All images remain copyright of the institutions holding the specimens.

Maintaining databases

Back up: Original and processed images on HD and DVD.
Storage capacity: 80 GB.
Maintaining links: 85http://www.zsm.mwn.de/lep/research2.htm#GlobInG
Linking to other systems: Data and images are available via SysTax at Ulm University (http://www.biologie.uni-ulm.de/systax/daten/index_e.html) and are implemented into GBIF (http://www.biologie.uni-ulm.de/cgi-bin/query_all/query_all.pl?lang=d&pr=gbif-e1).

Taxa Group: Pisces: Cichlidae (Cichlids)
Institution: Vertebrate Section, Africa Museum, Tervuren, Belgium
Person Responsible: Dr. JOS SNOEKS
Number of Person Hours per year devoted to imaging:
Number of Images captured and stored each year:

Purpose of Image

Target Audience: Zoologists, Part of the Albertian Rift Project, not only type specimens, includes also pictures alive observed individuals
Form: Custom website, DarwinCore and ABCD mapping to GBIF

Capturing

Handling Specimens:
Mounting Specimens / Specimen Preparation:
Capture Device: Nikon Coolpix
Lighting (including light spectrum):
Resolution (dpi): 600
Software: None
Automontage: None
Scale:

Working with Images

Storage Formats: TIFF
Software:
Linking to/within Databases:
Enhancing Images: None
Automontage:
Quality Control:
Naming Images:
Metadata Information:
IPR policy for Images:

Maintaining databases

Back up: on CD-ROM
Storage capacity: BeBIF webserver : ~500GB
Maintaining links: with database
Linking to other systems: In development