Automated Identification of Bee Species in Biodiversity Information Systems

Volker Steinhage[1]

Abstract

Working within a joint national initiative on biodiversity information systems for insects, we describe our contribution to the research and software development of an automated identification system for insects, especially for bees. We describe the framework of the joint project as well as the innovative and interdisciplinary research and development approach both within our identification project itself and with respect to interaction within the overall joint project. Thus this paper presents technological aspects along with organizational aspects, both of which we believe to be important in achieving successful innovation in superior interdisciplinary projects of environmental research.

1. Biological Diversity and Biodiversity Informatics

The Convention on Biological Diversity, commonly referred to as the Biodiversity Treaty, was one of two major treaties opened to signatories at the United Nations Conference on Environment and Development (UNCED) in Rio in December 1992. The treaty defines biodiversity as "the variability among living organisms from all sources including, inter alia, terrestrial, marine and other aquatic ecosystems and the ecological complexes of which they are part; this includes diversity within species, between species and of ecosystems." Parties to the Biodiversity Treaty "affirm sovereign rights over the biological resources found within their countries, while accepting responsibility for conserving biological diversity and using biological resources in a sustainable manner", according to an International Union for the Conservation of Nature (IUCN) assessment of the treaty. The Federal Republic of Germany signed the treaty along with more than 170 other countries.

If we are to take our responsibilities as signatories of Biodiversity Treaty seriously, it becomes obvious that the conservation and sustainable utilization of biodiversity necessitates the mapping and assessment of the current status and of the risk of loss of biodiversity as well as the continual monitoring of biodiversity. These demands in turn require the gathering, storage, analysis, visualization and exchange of enormous amounts of data, which is essentially impossible without using modern

---

[1] Department of Computer Science III, University of Bonn

computer and communication technology as well as innovative computer science techniques.

To meet these specific demands for the mapping and monitoring of biodiversity, a new field in computer science has been established which is now termed Biodiversity Informatics. It is an innovative and interdisciplinary research field in computer science which employs modern information and communication technology for interdisciplinary integration of biological and geographical data and knowledge accessible worldwide.

2.   The Entomological Data Information System EDIS

Within the global biodiversity, insects represent by far the largest ecological group, having about 900,000 known species. Insects not only show an amazing diversity of species and shapes but also play a key role in all terrestrial ecosystems (especially as pollinators) and have economic importance both as pests and conversely as beneficial insects in agriculture and forestry as well as in pharmacy and medicine.

Within the so-called BIOLOG Programme for Biodiversity in Change which is funded by the German Ministry for Education and Research (Bundesministerium für Bildung und Forschung, BMBF), the joint EDIS project aims to build and establish a national **E**ntomological **D**ata **I**nformation **S**ystem. EDIS will permit Internet access and will comply with international standards and should therefore play a role as a national centre for mapping and monitoring and as the basis for international exchange of information concerning the biodiversity of insects.

The EDIS projects will deal in exemplary studies with the gathering, storage, documentation, processing, analysis, visualization, exchange and integration of information concerning the taxonomy, morphology, behaviour, distribution, ecology and molecular biology of certain groups of insects. In particular, as an important and innovative feature, EDIS will integrate new techniques for automated identification of insect species.

3.   The Automated Bee Identification System ABIS

Generally speaking, the efficient gathering of data is important for the construction and maintenance of every information system. In biodiversity research, a central part of the data gathering is the identification of the observed species. Due to the urgency of the need for diversity mapping and monitoring (engendered, for example, by the imminent, dramatic extinction of species) on the one hand and the huge amounts of data, the difficulties in taxonomy and the worldwide lack of experts on the other, the task of species identification becomes, in practice, the key role for usage of an entomological information system.

In the subproject ABIS, an interdisciplinary team of zoologists and computer scientists from the University of Bonn are working to develop an **A**utomated **B**ee **I**dentification **S**ystem ABIS for the automated gathering and identification of bee species. Furthermore the identification system ABIS will be integrated and applied as a tool for data gathering within the information system EDIS.

3.1 Image Analysis of Forewings

The key idea of our approach to ABIS is to identify bee species from images of their forewings. Using diffuse background illumination, the wings of the bees show a clear venation within a transparent skin. The structure of this venation is genetically fixed and therefore suited to the identification of the bee species. The forewings are clipped under a microscope slide and a digitised image is taken with a CCD camera connected to a frame grabber. Figure 1 shows the experimental setup of the prototype system (Steinhage et al. 1997) consisting of a laptop, video camera and a frame grabber. The resulting digitised image is shown in Figure 2.
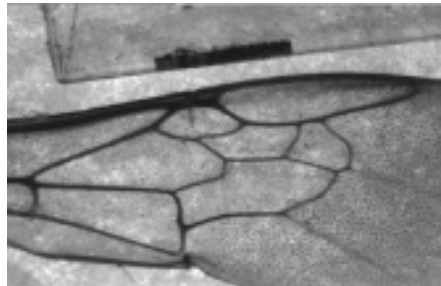

Figure 1: Setup of the ABIS prototype.

Figure 2: Digitized Image

3.2 Hierarchical Strategy for

The design of the method of species identification follows a step-by-step strategy which mirrors the taxonomy of bees in central Europe, where there are about fifty genera, two or three species complexes per genus, and between five and thirty species per complex.

Step 1: Identification of the genus. In this first step we employ image analysis processes to extract an initial substructure of the whole venation structure of the forewing image. The selected substructure has been found to be very robust against injuries, pollution and other causes of distortions to the wing and the image taking process. The initial substructure contains the marked venation cycles B, D1 and D2 shown in figure 3. From this extracted substructure, we select 15 numerical features such as distances, angles, form parameters etc.) which are then used for the identification of the genus by employing linear discriminant analysis (LDA).

Step 2: Identification of species complexes. The identification of the genus is used to index into a database of genera specific venation templates. These templates are subsequently employed to guide – in a knowledge-based way – more elaborate processes of image analysis to extract the parts of the complete venation structure which are more difficult to detect. From this structure we derive 50 numerical features which are used to identify the species complex. Again LDA has been shown to be suitable for this second classification step.

Step 3: Identification of the species. For this final step we need not only a more complex model of discriminant analysis but also an enlarged set of features. In addition to the 50 numerical features used in the second step, we also employ as features 240 image intensities taken from the geometrically normalized and subsampled forewing image (See figure 4.) Experiments have clearly shown that this combination of 50 derived symbolic features with 240 intensity features encodes much more species specific information than either part alone. Together with a nonlinear variant of the discriminant analysis technique (Roth & Steinhage 1999), the combined system has proved to be successful even in species identification tasks which have been judged by experts to be extremely difficult.
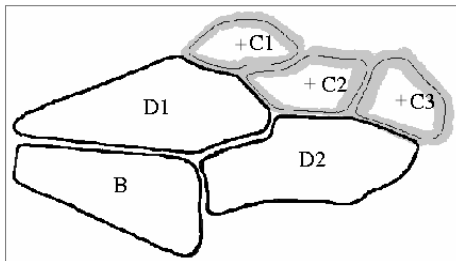

Figure 3: Extracted initial substructure (B, D1, D2) and predicted structures (C1,C2,C3).

Figure 4: Subsampled forewing image with 12 x 20 pixel resolution

3.   4. Results

We give two representative results of our prototype system. Figures 5 and 6 show the species identification for the genera *Bombus* and *Colletes*. Figure 5 shows the genus identification with two clearly separated cluster centres (top: the *Bombus* species, bottom: the *Colletes* species). Figure 6 shows the identification of two species complexes of the genus *Bombus*. For the identification of the species we used the enlarged feature set as well as the nonlinear discriminant analysis and identified the very difficult to distinguish species *Bombus sylvarum* und *Bombus veteranus* with a correct classification rate of 99.3 % using a leave-one-out cross validation procedure.

5. Discussion

Two other approaches to automated classifications of insects were given in Yu et al. 1992 and Weeks et al. 1997. The former extracts features of the venation in a manner similar to our approach but at a much lower level of automation. The latter uses no explicit model at all, taking into account only the pixel values of normalized images. Both of them investigate only ichneumonid wasps and no discussion of whether these methods are applicable to other insects, especially to bees, is given. To our knowledge, no further publications on this topic are available.
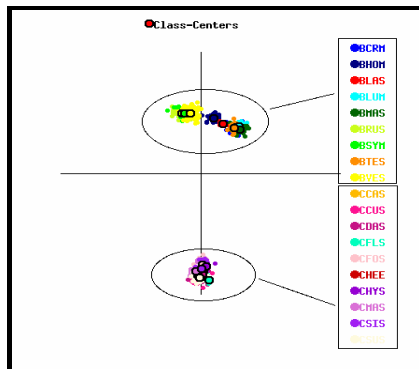
Figure 5: Genus identification.
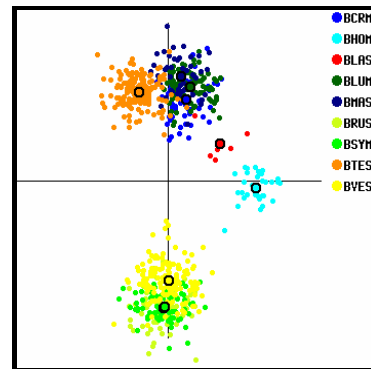
Figure 6: Identification of species complexes.

Our model-based approach is decisive for the robust extraction and interpretation of the venation structure. A purely uninformed approach, i.e. an approach based solely on low level image processing or analysis of pixel values, will fail in image areas with low contrast and other causes of distortions to the wing and the image taking process. For the first and second step of genus and species complex identification, the linear discriminant analysis has been shown to be a sufficient and efficient classification approach which provides the added benefit of visualization of the classification results. For the third and very difficult classification task of species identification the nonlinear approach of discriminant analysis is imperative. Detailed descriptions of the template-based image processing and nonlinear approach to discriminant analysis can be found Roth et al. 1999 and Roth/Steinhage 1999.

In summary: due to our hierarchical and model-based approach to image analysis and our efficient approach to nonlinear discriminance analysis our system works successfully with closely related species like that of the genera Andrena, Osmia and Bombus (Schröder et al. 1998). In the case of social bees it is also possible to separate casts and different populations. Furthermore we adapted our system to cope with reduced wing venation as in the case of stingless bees. The system was also

successfully applied to identify other insects with transparent wings (Hymenoptera) like wasp species of the genera Ceramius.

Future work: The ABIS sub-project will closely cooperate with the DORSA sub-project of the EDIS project. DORSA deals with the species identification of grasshoppers which can be identified by analysis of their audio signals. Both sub-projects intend to share software modules for classification and to develop a common geo-information system approach for integration, visualization and analysis of biological and geographical information. Furthermore, both sub-projects will cooperate with another sub-project, DAN-Tax, of EDIS which is working in the field of molecular biology.

4.   Acknowledgements

5.   References

Roth,V., Pogoda, A., Steinhage, V., Schröder, S. (1999): Pattern Recognition Combining Feature- and Pixel-based Classification Within a Real World Application. In: Mustererkennung 1999, 21. DAGM-Symposium, Bonn, pp. 120-129, Informatik aktuell, Springer.

Roth, V., Steinhage, V. (1999): Nonlinear Discriminant Analysis Using Kernel Functions. In: 13th Internat. Conf. on Neural Information Processing Systems, NIPS `99, Nov. 30 – Dec. 2, 1999, Denver (Colorado)

Schröder, S., Wittmann, D., Roth, V., Steinhage, V. (1998): Automated Identification System for Bees, Proc. XIII Intern. Conf. of IUSSI, Dec. 29, 1998 - Jan. 3, 1999, Adelaide/Autralia, p. 427.

Weeks P.D., Gauld I.D., Gaston K.J., O'Neill M.A. (1997): Automating the Identification of Insects: A New Solution to an Old Problem, Bulletin of Entomological Research (1997), 203-211.

Yu D.S., Kokko E.G., Barron J.R., Schaalje G.B., Gowen B.E. (1992): Identification of Ichneumonid Wasps Using Image Analysis of Wings, Systematic Entomology (1992), 389-395.