Ecological Informatics xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

A reference process for automating bee species identification based on wing images and digital image processing

Fabiana S. Santana [a,*], Anna H. Reali Costa [b], Flavio S. Truzzi [b], Felipe L. Silva [b], Sheila L. Santos [a], Tiago M. Francoy [c], Antonio M. Saraiva [b]

[a] Centro de Matemática, Computação e Cognição, Universidade Federal do ABC, Avenida dos Estados, 5001, Santo André, SP CEP: 09210-580, Brazil
[b] Escola Politécnica da Universidade de São Paulo, Av. Prof. Luciano Gualberto, travessa 3, 158, São Paulo, SP CEP: 05508-970, Brazil
[c] Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, Rua Arlindo Béttio, 1000, São Paulo, SP CEP: 03828-000, Brazil

ARTICLE INFO

Article history:
Received 1 March 2013
Received in revised form 29 November 2013
Accepted 1 December 2013
Available online xxxx

Keywords:
Reference process
Digital image processing
Machine learning
Bee wing classification
Taxonomic classification

ABSTRACT

Pollinators play a key role in biodiversity conservation, since they provide vital services to both natural ecosystems and agriculture. In particular, bees are excellent pollinators; therefore, their mapping, classification, and preservation help to promote biodiversity conservation. However, these tasks are difficult and time consuming since there is a lack of classification keys, sampling efforts and trained taxonomists. The development of tools for automating and assisting the identification of bee species represents an important contribution to biodiversity conservation. Several studies have shown that features extracted from patterns of bee wings are good discriminatory elements to differentiate among species, and some have devoted efforts to automate this process. However, the automated identification of bee species is a particularly hard problem, because (i) individuals of a given species may vary hugely in morphology, and (ii) closely related species may be extremely similar to one another. This paper proposes a reference process for bee classification based on wing images to provide a complete understanding of the problem from the experts' point of view, and a foundation to software systems development and integration using Internet services. The results can be extended to other species identification and taxonomic classification, as long as similar criteria are applicable. The reference process may also be helpful for beginners in this research field, as they can use the process and the experiments presented here as a guide to this complex activity.

1. Introduction

The well-being of human populations of the world depends, among others, on the so called ecosystem services. These services can be defined as the benefits people obtain from ecosystems and they can be divided into four categories: supporting systems, such as nutrient cycling and seed dispersal; provisioning services, such as food, water, minerals and others; regulating services, such as pollination and carbon sequestration; and cultural services, such as ecotourism and recreational experiences. That is why the worldwide loss of biodiversity has been a major concern at least since the last century, and it was recognized as such by the United Nations (Brundtland, 1987). It brought to attention that population, energy, industry, food security, and human settlements, are connected to the loss of species and genetic resources. Different factors may cause biodiversity depletion, such as destruction and reduction of natural habitats, pollution of ground water and of the atmosphere, and global warming (Sala et al., 2000).

Pollination is a key ecosystem service that represents a very important tool for environment conservation (Daily, 1997). A serious decline of the pollinator populations is being noticed since the end of the 20th century (Buchmann et al., 1997). It has recently been suggested that the main causes to this decline are linked to the rapid expansion of human activities, such as loss and fragmentation of natural habitat, aggressive agricultural practices, pathogens and climate changes (Potts et al., 2010; Schweiger et al., 2010). In particular, bees are major pollinators; therefore, their preservation helps to promote biodiversity conservation.

Identification of bee species is part of their preservation. It may be obtained by studying, processing, and analyzing images of bee wings (Bueno et al.; Giannini et al., 2011). Several studies have shown that features extracted from patterns of wing venation are good discriminatory elements to differentiate among species of insects (Francoy et al., 2008, 2009; Nielsen et al., 1999; Schröder et al., 2002; Weeks et al., 1997). The use of geometric morphometrics is very helpful, since it applies a set of computational techniques for analyzing shapes with applications in species identification, in genetic variability, among others.

* Corresponding author. Tel.: +55 11 3091 9088.
E-mail addresses: fabiana.santana@gmail.com (F.S. Santana), anna.reali@usp.br (A.H.R. Costa), flavio.truzzi@usp.br (F.S. Truzzi), f.leno@usp.br (F.L. Silva), sheila.leal@ufabc.edu.br (S.L. Santos), tfrancoy@usp.br (T.M. Francoy), saraiva@usp.br (A.M. Saraiva).

1574-9541/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ecoinf.2013.12.001

However, automated species identification is a particularly hard problem and the entire process related to image acquisition, plotting of landmarks, feature extraction and statistical analysis is still mostly manual, demanding a considerable amount of time. Therefore, the development of auxiliary tools for bee identification is of great importance. They can reduce the work load for the small number of taxonomists that spend a long time identifying species for non-taxonomist scientists. Morphometrics, DNA barcode and cuticular hydrocarbons are important alternatives to assess biodiversity, enriching the information obtained in alpha taxonomy (Francisco et al., 2008).

This paper presents a reference process for automating bee identification based on their wing images. Reference processes provide a complete understanding of the problem from the experts' point of view (Bass et al., 2003). In Internet-based software systems, as is the case with most biodiversity databases and tools, system development and integration are greatly improved by using SaaS, Software as a Service, and Cloud Computing (Fox and Patterson, 2012). Reference processes represent an adequate approach for the development of systems in this scenario because the steps and the communication among different steps are clearly identified.

The reference process is divided into four main sub-processes, namely image acquisition, digital image processing, classification, and validation. Image acquisition is the step in which wing images are obtained and stored. Live bees or dissected bee wings may be used in this step. Digital image processing is the second step, its purpose is to recognize landmarks and to extract relevant and discriminatory features so that bees can be classified. In the classification step, classifiers have the extracted features as input and they decide which species each wing image belongs to. The basic criteria for initial classification are defined by experts in the field. Though they do not always agree about which criteria should be applied in each case, an approach must be chosen in order to perform the classification sub-process properly. The last step is validation; it comprehends the interactions with users, researchers and experts to decide whether the criteria for classification should be revised, or classification results are acceptable, or if further analysis should be performed.

Morphological analyses are performed in the process. They help with bee identification by defining and extracting homologous landmarks and relevant and discriminatory features (such as shape, color and patterns) from the wing images, so that bees can be classified. Landmarks are extracted from Cartesian coordinates plotted at intersections between veins on the bee wing.

The reference process is presented in BPMN, Business Process Modeling Notation (White and Miers, 2008). BPMN is a standard to simplify the understanding of businesses. It provides a graphical notation to facilitate performance and collaboration studies, allowing the identification of gaps and potential improvements.

BPMN permits bridging the gap between business process design and implementation. BPMN is an agreement between multiple modeling tool vendors to use the same notation for the benefit of end-users, and it is supported by many of the largest software companies in the world (BPMN Information Home, 2013). It is also applied by several companies to model their businesses, being one of the most accepted notations by the community (Intalio Business Process Management System Home, 2013). The education community and the aerospace and defense, business service, and construction and operations industries are among their main users.

A BPMN-based process can also be automatically converted into BPEL, Business Process Execution Language. This feature is very desirable for integrating software services on the Internet because it addresses the integration of REST-compliant web services and other web services compliant to W3C standards, which are the majority of services available on SaaS-based systems in the cloud. In order to properly apply BPMN, in this paper bee classification is considered as the business.

The application of reference processes to this problem certainly contributes to its clarification while the knowledge gained with this work allows improvements in the development of software methods and packages for bee identification and classification. The main steps of the reference process were implemented in order to assess its viability and the viability of digital image processing to this problem. The results are presented as case studies, showing its effectiveness. If similar computational techniques can be applied, the approach presented here may be extended to the identification and taxonomic classification of other species, as long as their own characteristics are well defined, extending the contributions of this work. Finally, as it is presented, the reference process may be very helpful for beginners in this research field, as they can use the process and the experiments presented here as a guide to this complex activity.

2. Reference processes and BPMN

The main purpose of this paper is to establish the foundations to construct software systems for automated bee identification based on wing images and digital image processing. A recommendation to achieve this challenge is to design reference architecture for this domain (Bass et al., 2003). A reference architecture can be defined as a set of engineering and design principles within a specific domain. It defines the structure of the system, the responsibilities of system components, and templates, among others, mapping the main functionalities that must be addressed by the software system. Reference architecture offers a basis for developing specific software architectures; because the main issues of the domain are already identified and dealt with, the architect only needs to invest time and effort in the adequacy of the reference architecture for each specific implementation.

Mapping the detailed information contained in an existing standard or model is necessary to define a reference architecture (Bass et al., 2003). This can be obtained by a reference process. Reference processes document and formalize current practices and the experts' knowledge about a subject (Bass et al., 2003), proving to be a reliable guide in the identification of existing gaps and failures, and of extra work and improvements needed in the informal processes adopted (Santana et al., 2008). They represent important tools for the software architect because they have the ability to capture and to summarize the knowledge in a relatively simple way, as an ordered sequence of business activities and supporting information (White, 2004). They can also be applied to construct specific processes for each application. In service oriented architectures, it is usually recommended to design processes to identify the system issues, such as communication requirements (Huhns and Singh, 2005). Since this work considers SaaS and Cloud Computing as a natural part and the future of biodiversity information systems, a reference process is recommended to achieve the desired purposes.

In order to present the reference process, a notation is required. BPMN is a standard developed to provide a flow-chart based notation understandable by all business users (White, 2004), including business analysts, technical developers responsible for implementing the technology, and business people. Business Process Modeling (BPM) comprises the activities of representing, analyzing and improving the process of an enterprise, in order to achieve better levels of quality and efficiency.

A software package to design processes in BPMN is called BPMS, Business Process Management System. There are several of these systems available that implement the same notation (BPMN Information Home, 2013; Bizagi, 2013). The differences among BPMS packages are mainly related to functionalities other than design, such as performance evaluation and process analysis. Since the purpose here is only to design and to introduce the process, most of them would be suitable. The Bizagi Process Modeler™ was chosen to design the reference process for bee classification presented in this work because it is a free, standard-based, BPMS, which would thus be easily accessible by the biodiversity community.

Another advantage of BPMN is the existence of resources to map the graphical representation of a process to a services instantiation and

Please cite this article as: Santana, F.S., et al., A reference process for automating bee species identification based on wing images and digital image processing, Ecological Informatics (2014), http://dx.doi.org/10.1016/j.ecoinf.2013.12.001