Software to identify Butterflies species through their images

Juanita de los Ángeles García Chanagá, Nina Alejandra Mesías Barrera.
Mechanical and Mechatronics Engineering Department. Universidad Nacional de Colombia. Bogotá D.C., Colombia.

Background

In Colombia there are almost 3272 butterflies species, distributed in 6 families and 22 subfamilies; due to the high variety, the taxonomic determination and identification of these insects is a first step to acquire information about each species. It's important to say that the presence of a species in one specific area is an indicator of the environment and that the recognition of the existing biodiversity is essential to preserve it.

As students of the undergraduate program in Mechatronics engineering, and for the final project of artificial intelligence area, it was developed a software that classifies an image of a butterfly, among 5 species (*Phoebis philea philea, Dismorphia crisia foedora, Leptophobia philoma intermedia, Melete leucanthe leucanthe, Pseudopieris nehemia luisa*), that belong to the family *Pieridae*. This project was proposed to be the first possible phase in the development of software, useful to a larger number of species and higher taxa.

Materials and methods

The classification software was developed in C++ language and through the image processing library OpenCV, supported on programming platform Qt Creator; both, OpenCV and Qt Creator are free access.

The process that performs the software consists of two phases: The first one is the features extraction of the butterfly image, the image is uploaded by the user; this phase allows to synthesize the information contained in the image, for it being the most important for the classification, this synthesis consists to turn the features in numeric vectors.
In the other hand, the second phase defines the species that is represented by the features group extracted from image; this step is reached through of the artificial neuronal networks (ANN), an artificial intelligence tool who from set of training samples previously classified, chooses the conditions that should be met by a new element for it being considered part of one or other category. Therefore for the application of the ANN it was necessary to collect images of the selected butterflies' species (30 photographies by species) and extract the features of each one of them. The figure 1 summarizes the intern process that is done by the program to identify the butterflies.

Figure 1
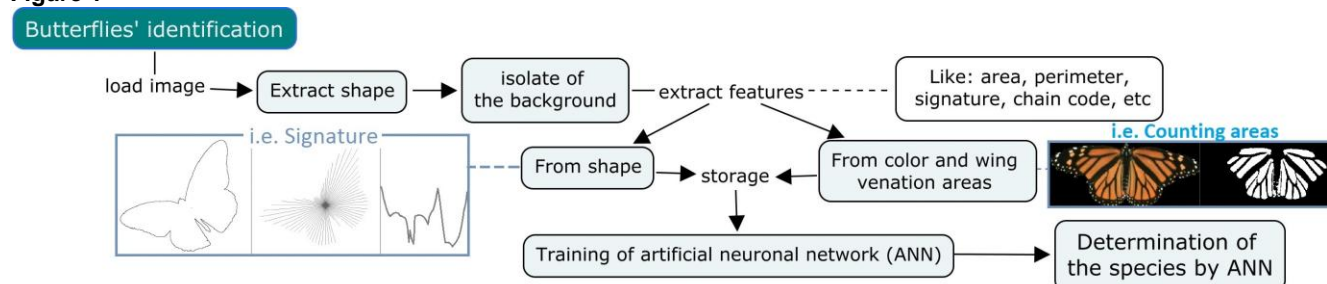
Internal process done by the software to identify butterflies species.

Results

The proofs done with the software, delivered an effectiveness of 92% (with photographs taken in similar conditions and with less than 20 images), it was observed less effective in the identification on the species *Leptophobia philoma intermedia*.
It was evidenced on the software's evaluation process, that the main factors that affect the result are the photograph's angle and the brightness. On the other hand the definition and the contrast affect to appropriate separation between the butterfly and the background of the image.

Conclusions

The development program is an interdisciplinary proposal that shows favorable results and that being expanded and deepened, could facilitate the butterfly species identification from photographic images, becoming a support tool for amateurs and experts in the field. Require improved for future versions of the program: the GUI (graphical interface) that communicates the user with the program, the sturdiness of the program and reconsider the characteristics that are extracted from the butterflies according to the particularities of the selected species.

Acknowledgements

References

1. Bradsky G, Kaehler A: **Learning OpenCV**. *O'Really*. United States of America, September 2008.
2. Andrade G, Campos L, González L, Pulido H: **Santa María: Mariposas alas y color. Guía de Campo.** *Natural Sciences Institute, Faculty of Sciences, Universidad Nacional de Colombia-Sede Bogotá*, September 2007.