Image Processing and Determination of Scorpion Dangerousness

**Abstract.** In the last decades the development of better antidotes for scorpion’s sting treatment has resulted in a decrement of the number of deaths worldwide. However, the application of an antidote is not always needed since there are few species of scorpions with venom that are dangerous to humans. The scorpion’s species and thus the venom type (poisonous and not poisonous to humans) can be determined by its physical characteristics. In this paper, an algorithm that distinguishes the scorpion’s species is presented using computer vision. The algorithm consists on obtaining 7 physical parameters and the 7 moments of Hu. A random forest classifier is then used to determine if the scorpion’s species is dangerous or not to humans. Experimental results are presented using photographs from living scorpions.

**Keywords:** Moments of Hu, Morphologic Classification, Random Forest, Scorpion.

1. Introduction

Scorpions can be found in most parts of the world. However, species which are poisonous to humans live in specific places considered high endemic zones. Mexico, USA, South America, North and South Africa, Middle East and India are high endemic zones where 2.3 million people on average are at risk. Annual deaths registered by scorpion’s sting in these high endemic zones are circa 3250 people. Moreover, scorpions are classified in 18 families with over 1500 different species. It is important to notice that only one, the Buthidae family, groups deathly poisonous scorpions [1].

In Mexico, over 200 different species can be found, from which only 7 can be considered dangerous to humans. From the Buthidae family, the species *Centuroides noxius* are located in arid zones like the states of Durango and Nayarit, while the species *Centuroides limpidus limpidus* exists in Jalisco, Guerrero, Morelos and Michoacán states. The Vejovidae family, a non-dangerous scorpion type, is one of the most commonly found in urban areas, like Pedregal de San Angel, in Mexico City. Example of each species is shown in figure 1. The number of cases registered of scorpion’s sting in Mexico is over 280,000 per year, while the average number of deaths in the past few years is of 70 people [2].

**Fig. 1.** Photograph of *Vaejovis mexicanus smithi* (left), *Centuroides limpidus limpidus* (center) and *Centuroides noxius* (right)

For several reasons [3], the recognition of patterns for classification of species was not used in the past. However, nowadays technology is improving and new algorithms can now be used to identify species, e.g. for insects [4]. Today’s techniques for image processing and patterns recognition are independent from translation, scale and rotation; this is independent from the size and resolution of images and angles [5].

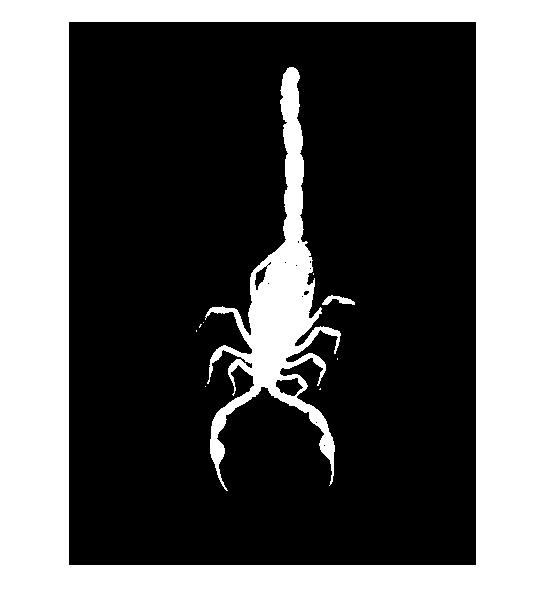
Additionally, classification techniques have become accessible to different applications since the processing time has decreased. Statistical approximations, machine learning and neuronal networks are some of the most commonly used classification methods today. The methodology of random forests was developed as an extension to classification and regression trees that allows the determination of the most relevant morphologic characteristics to classify the species. [6]

1. Algorithm

This algorithm for dangerous scorpion determination is divided in 3 main steps. First, digital photographs of dangerous and non-dangerous scorpions are filtered using image processing techniques in order to clean noise and obtain 10 images from different parts of the scorpion in black and white. Secondly, morphological characteristics of the scorpion’s images are obtained and stored. Finally, a Random Forest technique is used to classify the scorpions into dangerous and non-dangerous.

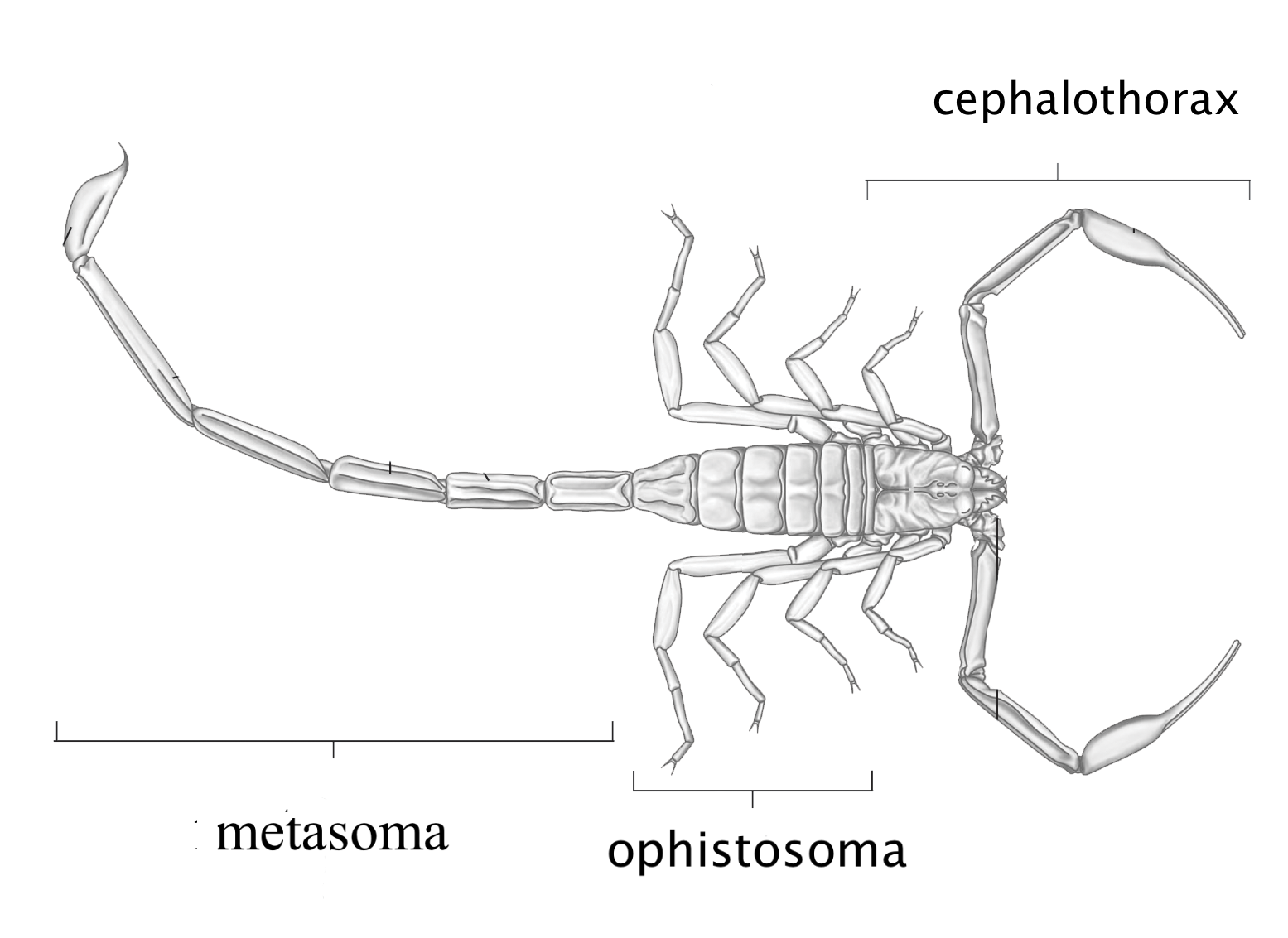
* 1. Image preparation

Photographs of living scorpions are taken over a white surface with light positioned in a way that produces almost no shadow and scorpions are fixed oriented vertically. Then, digital images are separated in their Red, Green and Blue components. Using the component’s histograms of the image, the boundary limits are obtained to convert the image to black and white. In this process, the background noise and remaining shadows are eliminated as shown in figure 2.



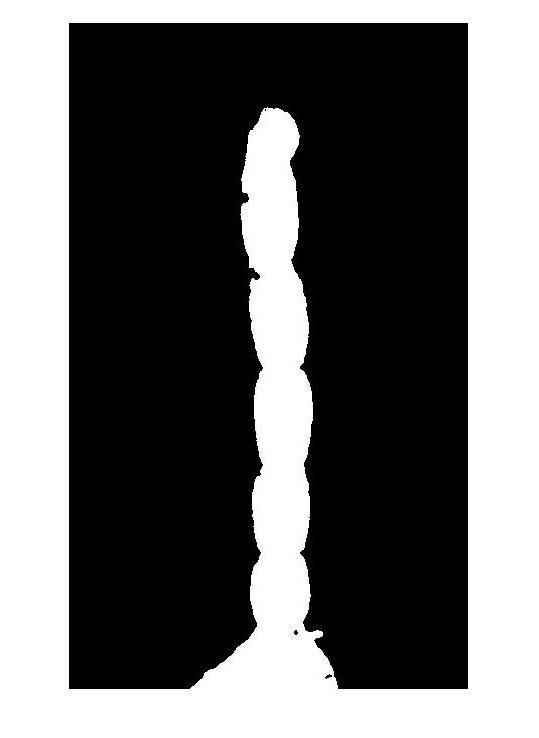
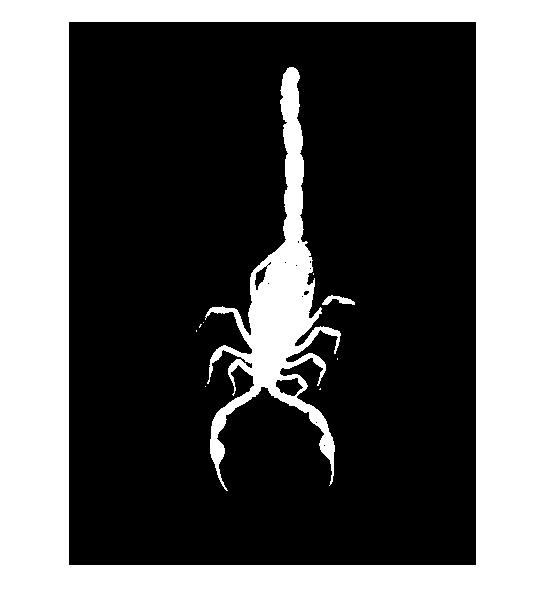
**Fig. 2.** Example photograph of a scorpion (left) and binary image after preparation (right)

The body of scorpions can be divided in 3 main parts. As shown in figure 3, the frontal part is the *cephalothorax* or *prosoma*, the central body part is called *ophistosoma* or *mesosoma* and the tail is the *metasoma*. The *prosoma* encloses the eyes, quelicerae, pedipalps, chelae, the claws and four pairs of legs. The *mesosoma* consists of seven segments which includes the back and the womb. The *metasoma* contains the last six segments; from them, the first five correspond to the dorsal fins and the last segment named telson is the one containing a pair of venomous glands and the stinger.

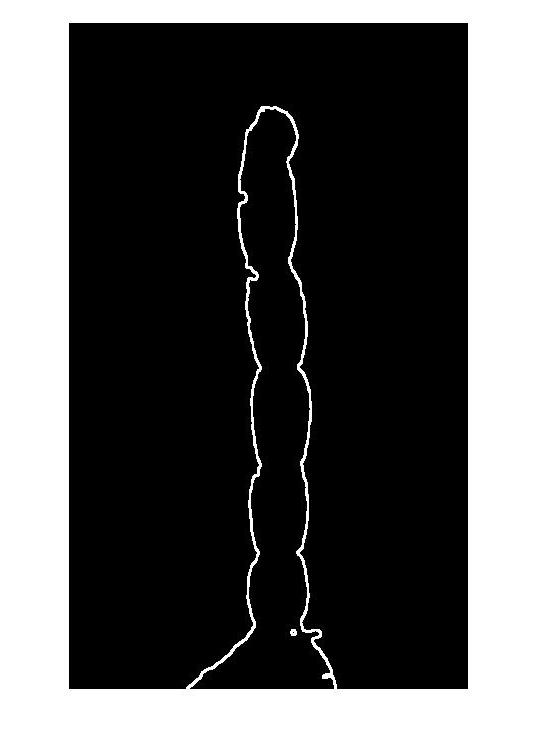
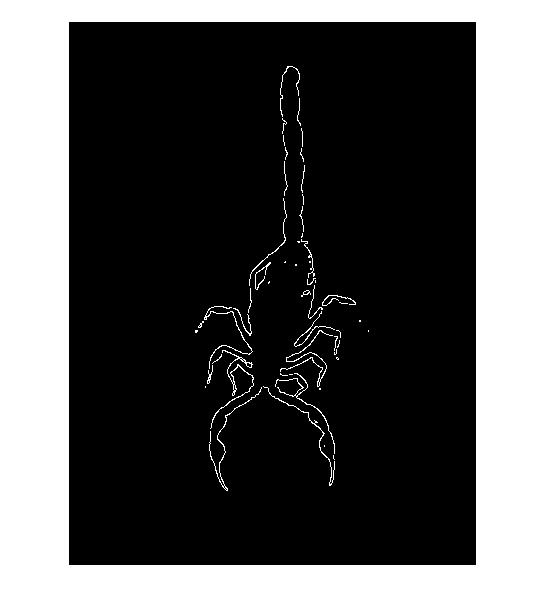


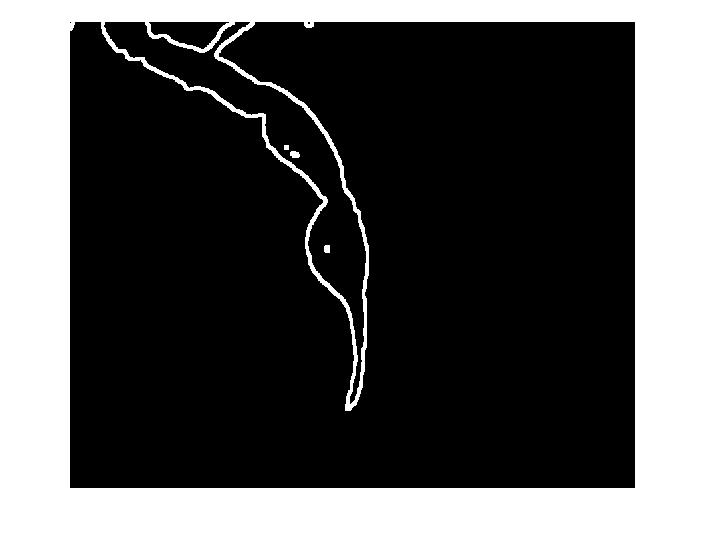
**Fig. 3.** Scorpion’s body parts

In order to perform a refined analysis of the morphological characteristics, 10 binary images are obtained for each scorpion as shown in figure 4. The previously cleaned full body image is used and other 4 sub-images are obtained using a sectioning algorithm. These 4 sub-images are: left and right part of the prosoma (claws), the mesosoma (center body) and the metasoma (tail). Additionally, a Canny filter is used to obtain the border of each of these 4 sub-images and the full body image.









**Fig. 4.** Example of the 10 binary images result after image preparation phase

* 1. Obtaining morphological characteristics

Important morphological characteristics, which distinguish the scorpion’s species, can be observed in any of the 10 previously preprocessed images. In order to make a good morphological analysis, the 7 invariant moments of Hu [5] and 7 physical parameters are obtained from each of the 10 images. The result is stored in an array of 140 values representing the observed and filtered morphological characteristics of each individual scorpion’s image.

First, the geometric moments are defined with the formula (1), where *R* is the set of pixels belonging to the image and *f*(*x*, *y*) is the binary value of the pixel with coordinates (*x*, *y*). The sum of the indices *p* and *q* is the order of the moment *mpq*.

(1)

In order to obtain an invariant geometric moment regarding the location of *R* within the image, the origin of the image’s coordinate system is moved to the image’s center of gravity .

(2)

Furthermore, the centralized geometric moment is normalized in order to guarantee invariance regarding the size of *R*.

(3)

Finally, the seven invariant moments are described by the following formulas:

(4)

(5)

(6)

(7)

(8)

()

(10)

Moreover, seven additional physical parameters are obtained. The *min\_distance* and *max\_distance* are defined as the minimal and maximal distance from the center point to the contour.

(11)

(12)

()

()

()

()

()

* 1. Classifier

The final phase of the algorithm consists on building a determination tree which will classify if a scorpion is or not dangerous. Random forests are an ensemble learning method for classification that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes output by individual trees. Using the 140 previously obtained parameters; random forests are formed to classify the scorpion’s species and therefore know if it is dangerous or not.

1. Implementation and Experimental Results

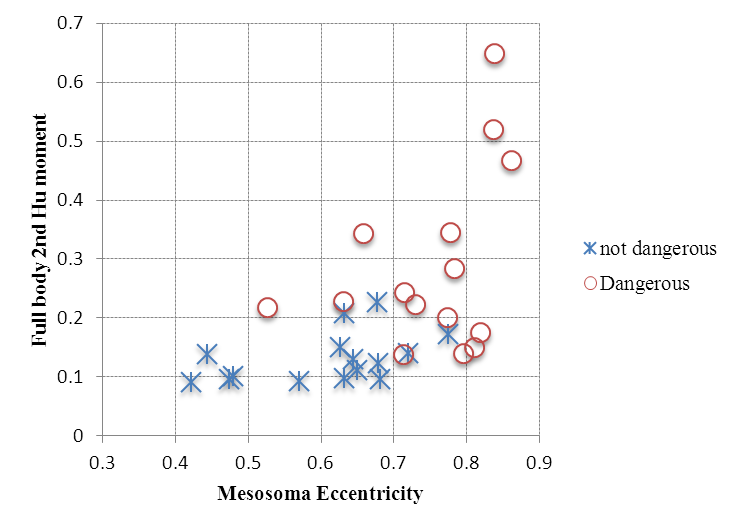
Photographs of 48 different living scorpions from the species *Centuroides noxius, Centuroides limpidus limpidus* and *Vaejovis mexicanus smithi* where taken in a semi-controlled environment. Scorpions were placed inside a bowl with a white sticky surface to hold them down. Photos were taken using the camera of a popular mobile devise. Picture’s resolution is 3264 x 2448 pixels and the camera was set on autofocus for simplicity.

All image processing algorithms were performed using MATLAB ® 2011B © and its Image Processing Toolbox. Functions like bwmorph and regionprops were used to filter and calculate the values of the physical characteristics from the scorpion’s photograph. Example results for dangerous and non-dangerous scorpions are shown in table 1.

**Table 1.** Example values of the 14 characteristics for the full body image.

|  |  |  |
| --- | --- | --- |
| Properties | Dangerous | Non dangerous |
| 1st moment | 6.59 x10-1 | 4.87 x10-1 |
| 2nd moment | 3.09 x10-1 | 9.18 x10-2 |
| 3rd moment | 1.49 x10-1 | 9.75 x10-2 |
| 4th moment | 6.69 x10-2 | 1.51 x10-2 |
| 5th moment | 6.67 x10-3 | 5.74 x10-4 |
| 6th moment | 3.67 x10-2 | 4.55 x10-3 |
| 7th moment | 3.86 x10-4 | 2.37 x10-5 |
| sphericity | 2.83 x10-1 | 4.83 x10-1 |
| lobation | 5.66 x10-1 | 9.66 x10-1 |
| rectangularity | 1.42 | 1.31 |
| compactness | 3.93 x10-2 | 4.39 x10-2 |
| elongation | 2.83 x10-1 | 4.83 x10-1 |
| roundness | 3.19 x10-2 | 5.03 x10-2 |
| eccentricity | 9.59 x10-1 | 8.76 x10-1 |

All 140 variables provide information about the morphological difference between the dangerous and non-dangerous scorpions. For example, figure 5 shows the location of variables *Mesosoma Eccentricity* against *Full body 2nd Hu Moment* in a plane. From the plot, it is possible to infer that, in general, dangerous scorpions have higher *mesosoma eccentricity* and higher *Full body 2nd Hu moment* than non-dangerous scorpions. However, this assumption is not conclusive and a proper classification must be performed.



**Fig 5**. Example of values from two different variables plotted in a plane

The classification process was done using R Project 2.15.3 Security Blanket with the randomForest library version 4.6-7. Due to the short amount of photo samples available, several classification methods could give similar results. However, when large samples are utilized, the random forest algorithm usually performs better than other methods.

Table 2 shows the dependence between the classification error and the number of trees. The classification error for 300 trees is about 10%. For this example, trees were form using 9 out of the 140 variables randomly. There are two types of errors produced. Non-dangerous scorpions were wrongly classified as dangerous in 13% of the times. However, dangerous scorpions were incorrectly classified as non-dangerous only in 6% of the cases.

**Table 2.** Random forest error of classification error dependence to the number of trees.

|  |  |
| --- | --- |
| Number of trees | Classification error |
| 1 | 35.29 % |
| 3 | 22.22 % |
| 10 | 21.25 % |
| 30 | 20.83 % |
| 100 | 14.58 % |
| 300 | 10.42 % |

1. Conclusions and Future Work

Computer vision can be used to determine the species of different scorpions. Dangerous and non-dangerous species can be identified through their physical characteristics. All 14 proposed characteristics provide information to help distinguish among scorpion’s species, but not with the same precision. Further analysis can be performed in order to identify the most differentiating characteristics.

The random forest method showed 10% accuracy during the classification process. Better results can be achieved if a larger number of photographs samples are used.

Finally, the long-term objective of this research is to develop an application for mobile devices including all the algorithms so that its user could take a photo of a scorpion and determine with a high probability if it is dangerous or not to humans.

1. References
2. Chippaux, J.P., Goyffon, M.: Epidemiology of scorpionism: a global appraisal. Acta Tropica 107(2):71–9 (2010)
3. Dirección General de Epidemeología de la Secretaría de Salud, <http://www.epidemiologia.salud.gob.mx/>
4. Gaston, K.J., O'Neill, M.A.: Automated species identification: why not? Philosophical Transactions: Biological Sciences, vol. 359, no. 1444, pp. 655-667. (2004).
5. Yang, H., Liu, W., Xing, K., Qiao, J., Wang, X., Gao, L., Shen, Z.: Research on insect identification based on pattern recognition technology. In: 2010 Sixth International conference on Natural Computation, pp. 545-548. (2010)
6. Hu, M.-K.: Visual Pattern Recognition by Moment Invariants, IRE, Transactions on Information Theory, pp.179 -187 (1962)
7. Breiman L., Random Forests, Machine Learning, v.45 n.1, p.5-32, (2001)