Local Appearance Feature Based Classification of the Theraphosidae Family

Ákos Utasi
Computer Automation Research Institute, Hungarian Academy of Sciences
Kende u. 13-17, H-1111 Budapest, Hungary
akos.utasi@sztaki.mta.hu

Abstract

This paper addresses the problem of feature based classification of spider images. In this work we focus on the Theraphosidae family (known colloquially as Tarantula), usually comprising very large and often hairy arachnids. This family contains approximately 120 genera and over 900 species. With the help of breeders and other experts we created a carefully validated dataset containing images of seven species, which are the most widely available commercially. In the classification scheme we followed a bag of visual words approach. First, from the images we extracted local color descriptors, then we created a vocabulary to represent each image as a document of visual words. Finally, we trained and evaluated three different classifiers to predict the class labels of the test set.

1. Introduction

Classification of animals is typically a very difficult task in machine vision applications mainly because of their deformable bodies and the properties of natural scenes where animals are usually observed in. These difficulties are particularly present when spiders are observed. In general they have a large number of deformable body parts resulting in a significant number of poses and in a high self occlusion rate. On the other hand, these animals are usually observed either in their natural living environment, or in the vivaria of hobbysts. In such circumstances the scenes might have different lighting conditions, the vegetation can partially occlude the body parts of the animal or might cause cast shadow. The vivarium can also contain other decoration elements such as artificial vegetation, branches, barks, water bowl, or stones. The species of the Theraphosidae family are either terrestrial or arboreal. The substrate where the species of the 1st group are living can also be different *e.g.* potting soil, peat moss, or vermiculite. Arboreal spiders have usually longer, leaner legs and they are living on trees, shrubs. Due to such difficulties the segmentation of the spider from the environment is usually infeasible. Moreover, the so-called sexual dimorphism phenomena makes the recognition of several species even more difficult. In such cases females and males can have a very different appearance in color or pattern. On the contrary, some species (one or both genders) are visually indistinguishable. To the best of our knowledge no database exists which takes into account the above characteristics, and the public spider galleries are usually unreliable, having poor or completely missing annotation. Therefore, with the help of breeders and other experts we created a dataset, that currently contains approximately 510 carefully validated images of seven species.

The rest of the paper is organized as follows. First, in Sec. 2 we present our database. In Sec. 3 we discuss the local appearance features we use for classification. Three different classification methods are presented in Sec. 4. Finally, Sec. 5 summarizes our experimental results.

2. Database

Prior to creating our database we contacted experts and breeders to collect the most common Theraphosidae species that are commercially available. For the species on this list we marked those that are affected by the sexual dimorphism phenomena. Since female and male spiders of these species have significantly different appearance we collected the images of both gender separately, and for species not affected by the phenomena we aggregated the images of the two genders. Images originate from the gallery of breeders, from the `http://www.arachnida.hu` and `http://www.arachnoboards.com` public websites, from the members of the *Madárpóktartók FaceKlub* facebook group, and from the author of the present paper. As the classifier training process requires large quanti-

Figure 1. Examples of the seven species in our dataset (from left to right): Acanthoscurria geniculata, Brachypelma smithi, Chromatopelma cyaneopubescens, Grammostola pulchra, Poecilotheria regalis (female), Psalmopoeus irminia (female), Pterinochilus murinus.

ties of data, in our experiments we selected a subset of the database where the number of images of the examined species exceeded 65. The resulting dataset contained seven species[1]. Table 1 shows the size of the dataset, while Fig. 1 presents some example images.

3. Feature Extraction

The coloring of the Theraphosidae family shows a wide variety of combinations. Some species have similar color combinations, however the layout/pattern can be completely different. In a significant number of cases the area of the background part (*i.e.* the substrate, vegetation, etc) is larger than the area occupied by the spider in the image, and also some of the spiders' colors are similar to the color of the background. These difficulties make the use of a global color or textural descriptor impractical. For example most of the body of both the Grammostola pulchra and the Psalmopoeus irminia female species have black tint (see Fig. 1), but the latter species has distinctive orange stripes at their legs, which has a color very similar to dry substrate or to pots used as decoration in vivaria.

Instead, we extract local descriptors at keypoint locations. We believe that capturing the appearance patterns visible on the main body parts, *i.e.* the prosoma (cephalothorax) and the opisthosoma (abdomen), and on the legs of the spider provides a more robust description than a global descriptor. In [8] scale invariant feature transform (SIFT) [7] was used to recognize individual salamanders in a large database by matching the extracted features of the images. In our work we use an extension of SIFT to detect keypoints and to extract features. In Color SIFT [3] the features are extended with color gradients and chromatic description. The descriptor carries the intensity related information in the first vector of 128 bytes, and the orthogonal chromatic information in the second and third vectors. In the preprocessing step prior to keypoint detection we resized the images to have a maximum dimension of 640 (width) and 480 (height) to limit the number of possible keypoint locations, and we also perform a simple histogram equalization. Finally, for keypoint detection and feature extraction we used the Color SIFT implementation[2] of the authors of [3]. The output of the keypoint detection step is demonstrated in Fig. 2 with the center of keypoint locations superimposed.

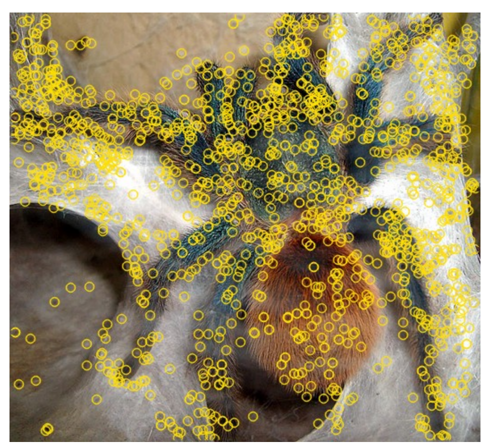

Figure 2. Color SIFT features are extracted at keypoint locations.

Class	Species	Images
1	Acanthoscurria geniculata	75
2	Brachypelma smithi	75
3	Chromatopelma cyaneopubescens	73
4	Grammostola pulchra	74
5	Poecilotheria regalis (female only)	72
6	Psalmopoeus irminia (female only)	66
7	Pterinochilus murinus	73

Table 1. Theraphosidae species in our database

[1] The dataset and the extracted Color SIFT features are available at `http://web.eee.sztaki.hu/~ucu/spiders`

[2] Binaries are publicly available at `http://staff.science.uva.nl/~mark/downloads.html`

4. Classification

After extracting Color SIFT features we followed the Bag of Words (BoW) representation, *i.e.* similarly to [5] we treated each image as a document of visual words. First, we created a codebook (vocabulary) from the extracted features using *k*–means clustering. Using the resulting codebook we mapped each feature vector into a codeword by finding the closest cluster seed. Thus each image is represented by a set of visual words selected from a fixed size codebook $\mathbf{V} = \{W_1, \ldots, W_K\}$, where K denotes the size of the codebook. Having the documents created from the images we apply multi-class supervised learning methods on a labeled training set. Finally, the trained classifier is used to predict the class label of the elements of the unlabeled test set.

The first two classifiers, discussed in [5], model the occurrences of visual words in documents, and are widely adopted for feature based image classification problems. Instead of directly modeling these probabilities the third method introduces latent topics over words to describe the documents. We think that this property is advantageous in our case, and can improve the classification efficiency by capturing the topics of decoration elements, natural scenes, and different species.

4.1. Naïve Bayes Classification

In Naïve Bayes classification each word $w \in \mathbf{V}$ in a document $\mathbf{D} = \{w_1, \ldots, w_N\}$ is assumed to be independent of other words, and the $p(C)$ class prior and the $p(W|C)$ class specific codeword conditional distributions are known. Having these two information available we apply Bayes's rule to select class C with maximal posterior probability, which is defined as

$$p(C|\mathbf{D}) = \frac{1}{Z} p(C) \prod_{w_i \in \mathbf{D}} p(w_i|C) , \tag{1}$$

where Z is a normalizing constant. To estimate the $p(W|C)$ conditional probabilities of codewords we used a Laplacian smoothing technique similarly to [5]. Thus denoting by $N(t,j)$ the number of times a given codeword $W_t \in \mathbf{V}$ occurs in document $\mathbf{D}_j$ we get

$$p(W_t|C) = \frac{1 + \sum_{\mathbf{D}_j \in C} N(t,j)}{K + \sum_{s=1}^{K} \sum_{\mathbf{D}_j \in C} N(s,j)} . \tag{2}$$

4.2. Linear Support Vector Machine

A support vector machine (SVM) [4] in a binary classification problem constructs a hyperplane, which has a maximal distance from to the nearest data point on each side, *i.e.* it has a maximal margin. The parameters of the hyperplane are $\mathbf{w}$ and b, and the classification of an input vector $\mathbf{x}$ is defined as

$$f(\mathbf{x}) = \operatorname{sgn}\left(\mathbf{w}^{\mathsf{T}}\mathbf{x} - b\right) . \tag{3}$$

In our case the input vector $\mathbf{x}$ is a codeword histogram with K bins determined from the $N(t,j)$ counter defined in Sec. 4.1. In case of linearly not separable datasets we can extend the SVM with an additional C soft margin parameter to penalize misclassified samples. Thereby SVM finds the trade-off between a large margin and a small misclassification error. Finally, for our multi-class classification problem we apply a one-vs-all approach, *i.e.* we train a separate SVM for each Theraphosidae species, where the feature vectors extracted from the images of the given species represent the 1st class of the optimization problem, and images of the other species are used to generate the feature vectors of the 2nd class. To obtain the $\mathbf{w}$ and b SVM parameters we used the LIBSVM library[3], and we followed the guidelines of [5] to set the soft margin parameter.

4.3. Supervised Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) [2] is a topic model, where the words of a given document are assumed to arise from a set of latent topics, which are considered as unknown distributions over a vocabulary. In LDA the topic distribution is assumed to have a Dirichlet prior. LDA and most other topic models are unsupervised. Supervised LDA (sLDA) [1] is an extension to LDA for supervised learning problems. In sLDA the documents are jointly modeled with an additional response value, and the goal is to find topics that will best predict the response for unlabeled documents. In our experiments we used the sLDA implementation of C. Wang[4] both for model training and label prediction.

5. Experiments

In our experiments we selected 50 training samples for each class of Table 1. Then the extracted Color SIFT features were quantized by a codebook with $K = 512$ codewords. Finally, we trained all three classifiers with the resulting documents, and we used them to predict the labels of the remaining images. As the sLDA implementation requires two input parameters, first we evaluated the accuracy of this technique. The first parameter

[3]Source code available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm/`

[4]Source code available at `http://www.cs.princeton.edu/~chongw/slda/`

	$T=24$	$T=48$	$T=96$	$T=144$	$T=192$
$\alpha=0.2$	51.90%	63.29%	73.42%	74.05%	72.78%
$\alpha=0.4$	56.33%	64.56%	72.15%	70.89%	73.42%
$\alpha=0.5$	51.90%	68.35%	70.89%	67.09%	68.35%
$\alpha=0.6$	53.80%	62.66%	67.09%	68.35%	67.72%

Table 2. Average accuracy of the sLDA classifier for various α Dirichlet parameters and for different number of T topics

Class	NB	SVM	sLDA
1	16.00%	48.00%	**72.00%**
2	48.00%	40.00%	**52.00%**
3	86.96%	**100.00%**	86.96%
4	45.83%	62.50%	**79.17%**
5	22.73%	45.45%	**72.73%**
6	50.00%	50.00%	**75.00%**
7	56.52%	**86.96%**	82.61%
Avg.	46.20%	62.03%	**74.05%**

Table 3. Accuracy of the Naïve Bayes, linear SVM, and sLDA classifiers on 7 species, assuming K=512 codewords

is the number of latent topics T, while the second one is the topic Dirichlet distribution parameter α. The meaning of T is straightforward, and α is normally less than 1, the smaller value we choose the sparser topic distributions we get [6]. In our experiments α value was limited to $\alpha \in \{0.2, 0.4, 0.6, 0.8\}$. Table 2 shows the average accuracy of sLDA for various α and T values. We can observe that in general a very low T value results in poor performance, and as we increase the number of topics the classification accuracy increases. We can also see that very large α values result in relatively low accuracy, since we get very smooth topic distributions with lower discriminative power.

As a next step we compared the accuracy of the Naïve Bayes (NB), the linear SVM, and the sLDA techniques. Table. 3 shows the results for the the 7 classes along with the average accuracy. We can see that the sLDA based classifier clearly outperforms the other techniques in most cases. Finally, in a second experiment we increased the codebook size to $K = 1024$ and evaluated the methods with the same train and test images. We obtained very similar results: the average accuracy rate of the NB classifier was 46.84%, SVM achieved 71.52%, while sLDA improved to 77.22%. We think that the relatively modest results are mainly due to the challenging environment (*e.g.* decoration, vegetation) where the keypoint detector produces a very large number of keypoint locations, in many cases larger than the number of keypoints located on the spiders' body. This might cause an inaccurate codebook representation, and a less discriminative classifier.

6. Conclusion

In this paper we analyzed three different classification techniques, the Naïve Bayes classifier, the linear SVM and the sLDA topic model, using a new carefully validated image dataset containing seven Theraphosidae species. The classification of these animals is very challenging, mainly because some species have similar color combinations, and a high occlusion rate, and various lighting conditions might be present in their living environment. According to our experiments the sLDA based classifier outperformed the other methods and achieved higher classification accuracy. In the future we are planing to increase our dataset with images of other species, and we will also investigate other preprocessing techniques and other image features to further improve the performance.

Acknowledgment. The author would like to thank Mária Gombásné-Gudenus, László Gombás, Tamás Kokavecz, Imre Bengerno, Attila Nagy, Sándor Szilvási, István Fekete, Alex Rácsik, and Zoltán Kovács for their contribution in data collection and annotation. This work was supported by the Hungarian Scientific Research Fund under grant number 80352.

References

[1] D. M. Blei and J. D. McAuliffe. Supervised topic models. In *Proceedings of the 21st Annual Conference on Neural Information Processing Systems*, 2007.

[2] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[3] G. J. Burghouts and J. M. Geusebroek. Performance evaluation of local colour invariants. *Computer Vision and Image Understanding*, 113(1):48–62, 2009.

[4] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

[5] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray. Visual categorization with bags of keypoints. In *Proceedings of the Workshop on Statistical Learning in Computer Vision, ECCV*, pages 1–22, 2004.

[6] J. Huang. Maximum likelihood estimation of Dirichlet distribution parameters. Technical report.

[7] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[8] C. M. Yang and S. Ravela. Deformation invariant image matching by spectrally controlled diffeomorphic alignment. In *Proceedings of the 12th International Conference on Computer Vision*, pages 1303–1310, 2009.