AUTOMATED IDENTIFICATION AND RETRIEVAL OF MOTH IMAGES WITH SEMANTICALLY RELATED VISUAL ATTRIBUTES ON THE WINGS

Linan Feng and Bir Bhanu

Center for Research in Intelligent Systems
University of California, Riverside

ABSTRACT

A new automated identification and retrieval system is proposed that aims to provide entomologists, who manage insect specimen images, with fast computer-based processing and analyzing techniques. Several relevant image attributes were designed, such as the so-called semantically-related visual (SRV) attributes detected from the insect wings and the co-occurrence patterns of the SRV attributes which are uncovered from manually labeled training samples. A joint probabilistic model is used as SRV attribute detector working on image visual contents. The identification and retrieval of moth species are conducted by comparing the similarity of SRV attributes and their co-occurrence patterns. The prototype system used moth images while it can be generalized to any insect species with wing structures. The system performed with good stability and the accuracy reached 85% for species identification and 71% for content-based image retrieval on a entomology database.

***Index Terms*—** Entomological image identification and retrieval, semantically-related visual attribtues, attribute co-occurrence pattern detection

Fig. 1. (a), (c), (e), (g): the samples of the moth wings. (b), (d), (f), (h): four types of semantically-related visual attributes: eyespot, central white band, marginal cuticle and snowflake mosaic.

1. INTRODUCTION

The vast number of digital insect images is a great challenge for entomologists and taxonomists who routinely conduct manual identification, archiving and retrieval of the specimen images. Several computer-aided systems for insect identification have been developed and evaluated in the last two decades, including the computer-based classification system for orchard insects [1], species-specific pattern recognition system on insect wings [2] and digital autimated identification system (DAISY) [3]. The development of these systems adopts many image processing and pattern recognition algorithms, such as local binary pattern (LBP), principal component analysis (PCA), linear discriminant analysis (LDA), artificial neural networks (ANNs), and support vector machine (SVM). Other efforts have been made using content-based image retrieval (CBIR) techniques to find relevant images for a query image based on visual similarity. The prototype systems for retrieving insect images include the "Butterfly Ecology" [4] and "But2Fly" [5].

The above systems provide a number of attractive functions, however, drawbacks have been revealed in several aspects. First, most systems focus on analyzing low-level features such as color, texture and shape that essentially have no semantic meaning. These features lose the effectiveness in describing images when they have complex visual and semantic contents. Second, a majority of the systems work in a fully automated way without any human intervention. However, unsupervised systems often suffer failures due to the large intra-species visual variation and the small inter-species visual differences. Third, most of the existing systems focus on extracting more effective and efficient individual feature descriptors from images, however, the co-occurrence property among the features has drawn very limited or no attention.

Unlike other identification and retrieval systems, our proposed system is based on the detection of semantically related visual (SRV) attributes. The SRV attributes are defined as a paticular pattern of elements on the moth wings as shown in Figure 1. Compared to the low-level visual features, the SRV attributes are a higher level of semantics. Some of the semantic information contained in these SRV attributes is namable such as "eyespot", "band", etc. While most of them do not have a concise verbal description, they still show clear visual patterns that can be easily distinguished from other random low-level features. Our system detects and learns SRV-attributes in a semi-supervised way. A training image set is proposed that contains human labled attributes. A probabilistic model is built on the training set as the core of the SRV-attribute detector and is used to infer attributes for other unlabeled database images. The training set also provides hidden co-occurrence information between SRV attributes. The co-

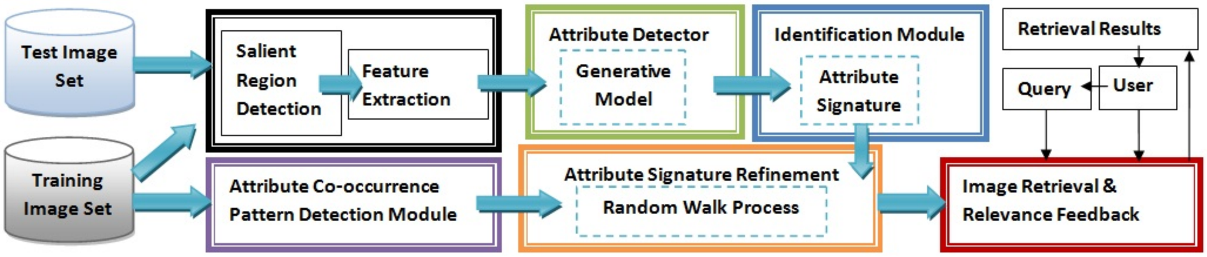

Fig. 2. The proposed moth image identification and retrieval system framework.

occurrence patterns are detected as contextual constraints to further improve the identification results.

2. TECHNICAL APPROACH

Our automated identification and retrieval framework is described in Figure 2. Our system detects the SRV attributes in a semi-supervised way. Textual labels related to the SRV attributes are given by human experts only to a small subset of the database images. The labeded subset is used as the training set to train a probabilistic model for building up the inference relationship between image regions and the SRV attributes. The model is further used as the attribute detector for the majority of unlabeled images. The species identification and retrieval are based on the detected SRV attributes. We give the details about each parts in the following sections.

2.1. SRV Attribute Detector

Consider a subset of training patches $Y = \{y_i, i = 1, 2, ..., m\}$ containing the manually labeled SRV attribtues $X = \{x_j, j = 1, 2, ..., n\}$. Let the visual feature descriptor of patch y_i be $F_i = \{f_k, k = 1, 2, ..., d\}$ in R^d with entries containing the Grey Level Co-occurrence Matrix [6] and Scale-Invariant Feature Transform [7] features. The SRV attribute detector is trained by using a generative approach based on probability theory. Suppose we have an unlabeled image patch $\widetilde{y}$ extracted by a hierarchical salient region detection approach proposed in [8], the SRV attribute detector will assign an attribute x to the patch based on its visual feature descriptor $\widetilde{F}$. The best solution acquired from the probability theory is to assign the attribute that has the largest posterior probability score $p(x|\widetilde{F})$. In the training process, we model the joint distribution of the patch attributes and the visual feature descriptors $p(x, F)$. Based on Bayes' therorem, we can get an alternative way to derive $p(x|\widetilde{F})$:

$$p(x|\widetilde{F}) = \frac{p(x, \widetilde{F})}{p(\widetilde{F})} = \frac{p(\widetilde{F}|x)p(x)}{\sum_{j=1}^{n} p(\widetilde{F}|x_j)p(x_j)} \tag{1}$$

As the sum in the denominator takes the same value for all the attribute categories, it can be viewed as a normalization factor over all the attributes. Equation (1) can be simplied into:

$$p(x|\widetilde{F}) \propto p(x, \widetilde{F}) = p(\widetilde{F}|x)p(x) \tag{2}$$

We further estimate the posterior probability by computing the attribute prior probabilities $p(x)$ and the likelihood $p(\widetilde{F}|x)$ separately. The advantage of this generative model is that it can actually augment a small amount of labeled data with a large amount of unlabeled data.

Algorithm 1: SRV attribute signature generation

Input: Input test image $\widetilde{I}$.
Output: SRV attribute signature S.
Method:
1. Extract the salient regions $\widetilde{Y} = \{\widetilde{y}_i, i = 1, 2, ..., \widetilde{m}\}$ by [8] as the candidate patches containing SRV attribtues.
2. **for** $i = 1 \rightarrow \widetilde{m}$ **do**
3. extract the visual feature descriptor $\widetilde{F}_i$ for $\widetilde{y}_i$
4. **for** $j = 1 \rightarrow \widetilde{n}$ **do**
5. estimate the likelihood $p(\widetilde{F}_i|x_j)$ by
   - p$(\widetilde{F}_i|x_j) = \frac{1}{n_j} \sum_{k=1}^{n_j} \frac{exp\{-(\widetilde{F}-F_k)^T \Sigma^{-1} (\widetilde{F}-F_k)\}}{\sqrt{2^{n_j} \pi^{n_j} |\Sigma|}}$

   n_j is the number of patches in the training set that contains the attribute x_j, and the kernel is parametrized by Σ which is the covariance matrix of the feature vectors in Y_{x_j}
6. estimate the prior probability $p(x_j)$ by
   - p(x$_j$) $= \frac{\mu \delta_{x_j,m} + N_{x_j}}{\mu + N_m}$

   μ is the smoothing parameter estimated from the training and validation. $\delta_{x_j,m} = 1$ if x_j occurs in y_m and 0 otherwise. N_{x_j} is the number of training patches that contain x_j and N_m is the total number of training patches.
7. $p(x_j|\widetilde{F}_i) = p(\widetilde{F}_i|x_j)p(x_j)$
8. **end for**
9. **end for**
9. $S = (\frac{1}{\widetilde{m}} \sum_{k=1}^{\widetilde{m}} p(x_1|F_k), ..., \frac{1}{\widetilde{m}} \sum_{k=1}^{\widetilde{m}} p(x_j|F_k))$

To estimate the likelihood, we assume that the feature vectors of the labeled patches are generated from some underlying multi-variate density function $p_{\widetilde{F}(\cdot|x_j)}$ and we apply a non-parametric kernel-based density estimate for the distribution $p_{\widetilde{F}}$. For each patch extracted from the test image, we have a vector of posterior probabilities with each entry denoting a SRV attribute. And for a test image that contains several patches, we average the vectors of posterior probabilities to get a final vector called *SRV attribute signature* for that image. The signature is essentially a semantic description of an image. The algorithm for generating the SRV attribute signature is summarized in Algorithm 1.

2.2. Co-occurrence Pattern Detection

If a group of individual SRV attributes always occur together, we argue that an underlying co-occurrence pattern may be composed of these attributes. We further argue that these co-occurrence patterns can help improve the detection of individual SRV attributes as ground-truth meta-knowlege. If we represent the SRV attributes by the nodes in a network, and represent their co-occurrence correlations by the connecting edges, the patterns can be determined as the densely connected clusters of nodes in the network. We further address the problem of co-occurrence pattern discovery by the hierarchical community structure detection algorithm proposed in [9].

Algorithm 2: Refinement of the SRV attribute detection

Input: The SRV attribute signature $\{\dot{S}_1, \dot{S}_2, ..., \dot{S}_N\}$ of Image I and the co-occurrence patterns discovered by the approach in [9].
Output: The refined SRV attribute signature of image I.
Method:
1. **for** $i = 1 \rightarrow N$ **do**
    2. **for** $j = 1 \rightarrow N$ && $j \neq i$ **do**
        3. calculate the distance D_{x_i,x_j} between attribtue x_i and x_j by eq (3)
        4. **do** the refinement iteration:

$$\dot{S}_i^t = \alpha \sum_j \dot{S}_j^{t-1} \cdot D_{x_i,x_j} + (1-\alpha) \cdot \dot{S}_i \quad (3)$$

until $\dot{S}_i^t$ converges at iteration T
5. **return** $\dot{S}_i^T$

The co-occurrence patterns are utilized for boosting the performance of attribute detection by performing a random walk process [10] over the patterns. We define the distance between two attributes x_1 and x_2 as

$$D_{x_1,x_2} = \frac{2 \times \# \, of \, CP\{x_1, x_2\}}{\# \, of \, CP\{x_1\} + \# \, of \, CP\{x_2\}} \quad (4)$$

where $\# \ of \ CP\{x$ is the number of co-occurrence patterns that contain attribute x. Suppose we have N SRV attributes in all and initially the probability of attribute x_i detected in Image I is S_i given by the detector, in the *t*-iteration the probability is refined by Algorithm 2.

α in equation (4) is a weight parameter and we set it to 0.3 empirically. The above algorithm can strengthen the occurring probabilities of the attributes in the same patterns and weaken the isolated ones. The refinement iteration will stop until it converges at a certain point. The controlling parameter is determined by using the training and validation sets. The iterations of random walk process [10] can further improve the SRV attribute detection results.

2.3. Automated Identification and Retrieval

2.3.1. Automated Species Identification Scheme

For the automated identification process, we first derive the SRV attribtue signature for each training image. The attribute signature of a training image I is defined as $S^{|I|}$ with each element $s(x_j) \in \{0, 1\}$ and $s(x_j) = 1$ when image I has patches labeled with attribute x_j and $= 0$ otherwise. We further divide the training images into species categories. The attribute signatures of all the images in the same species category are averaged for each entry. The value of each entry x is in the interval $[0, 1]$, we set the final cell value to 1 if $x \geq 0.1$ and to 0 otherwise. We call the final vector for each species category the *prototype signature*.

The test specimen is identified by comparing its signature with the prototype signature from each known species. The distance between the signature and the prototype is calculated based on the Euclidean distance. The specimen is finally classified into the species with the closest distance. If some known species have the same distance to the test image, we keep all the species labels for that image, and let the image retrieval system give the final decision after several retrieval sessions on the image species.

2.3.2. SRV Attribute Signature Based Image Retrieval Scheme

In our system, we provide a browsing function in the user interface, and the user is allowed to browse all the images in the database and submit queries. Image similarity is compared based on both the low-level visual features and the SRV attribtue signatures. Each image is represented by a low-level visual feature vector F and a high-level SRV attribute signature S, for a query image Q and a database image D, the distance between two the images is evaluated by the Euclidean distance over the feature vectors and the Earth Mover's distance [11] over the attribute signatures:

$$Dist(Q, D) = \mu D_{Euc}(F_Q, F_D) + (1-\mu) D_{EMD}(S_Q, S_D) \quad (5)$$

where μ is the adjusting parameter between the two distance measure and is determined by the long-term cross-session retrieval precision history. If the precision for a particular Query is increased when more importance is put on the feature distance, then μ is adjusted to a larger value, otherwise it becomes smaller. The Earth Mover's distance [12] can be viewed as a measure of the least amount of work needed to transfer one signature into the other, a unit of work in the process is evaluated by the ground distance.

3. EXPERIMENTS

We report the following experimental results: (i) performance of moth species identification from attribute signatures; (ii) performance of image retrieval with relevance feedback.

3.1. Image Source

To meet the need for practical moth species identification, we used 4,530 specimen images in our experimental dataset covering 50 species across 8 families: Apatelodidae, Hesperiidae, Lasiocampidae, Mimallonidae, Notodontidae, Saturniidae, Sphingidae and Thyrididae. The original images were collected from northwestern Costa Rica provided with permission by Dr. Dan Janzen `http://janzen.sas.upenn.edu/caterpillars/database.lasso`. All the patches were manually labeled with SRV attributes as the ground-truth for evaluation.

3.2. Species Identification Results

For each species category, we use 10-fold cross validation to evaluate the performance. The images in one species category

Table 1. Accuracy vs. the number of attributes.

	Accuracy range	
Number of attributes	*Lower bound*	*Upper bound*
50	0.2518	0.4653
100	0.2768	0.4917
150	0.3269	0.5513
200	0.3483	0.5982
250	0.4163	0.6232
300	0.4625	0.8532
350	0.4451	0.8421
400	0.4312	0.8367
450	0.4215	0.7635

are randomly divided into 10 subsets. In each testing round, one subset of the images is used for validation and the rest are used for training. The training and validation processes are repeated 10 times and the results are averaged to produce the mean and standard deviation of the identification accuracy. We evaluate the automatic identification results of the test images by comparing their SRV attribute signature with the ground-truth labels.

We compared our SRV attribtue based identification scheme with two other automated insect identification systems proposed in [13] that adopts a discriminative kernel on color and shape features and [13] uses the texture features. Figure 3 summarizes the mean and standard deviation of the averaged identification accuracy of the three systems on randomly selected 10 species computed by the 10-fold cross validation. The number of manually labeled SRV attributes in our experiment setting is 300. The range of the mean accuracy of our system is between 0.4625 and 0.8532. Our system outperforms the other two systems to a large extent in the mean accuracy, and it has smaller values of the standard deviations. We adjust the number of attributes and show the corresponding accuracy range variations in Table 1. As we can observe, the best accuracy occurs when the number of attributes approximates 300. We set the number of attributes to 300 for the rest of the experiments described below.

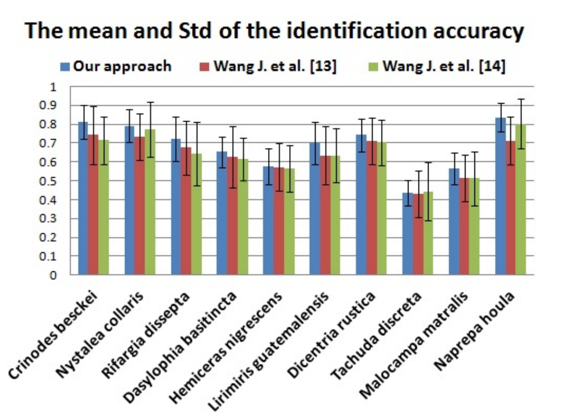

Fig. 3. The comparison of the identification results.

Table 2. The comparison of retrieval accuracies averaged over all the species category in different iterations.

	Mean Accuracy	
Iteration	*Our approach*	*CBIR approachin* [14]
1	0.5258	0.4412
2	0.5712	0.4634
3	0.6459	0.5243
4	0.6835	0.5472
5	0.7133	0.5632

3.3. Image Retrieval Results

To test the effect of our SRV attribute based retrieval scheme, we divide the entire dataset into a training subset containing a randomly sampled 40% of the images and a test set containing the rest. We learn the image attribute signature for each test image by using the probabilistic attribute detector model. We simulate the retrieval process launched by human users by generating automated queries using each of the database images. The retrieved images are ranked according to their degree of similarity to the query. For each query, we refine the retrieval results by executing relevance feedback in five iterations. Note that the ground-truth labels are used only for the training and experimental evaluation purpose, they are not assumed to preexist in the learning for test images and the retrieval process. Therefore, the proposed retrieval system is able to deal with the "never-seen" database images.

We compared the results of our retrieval scheme with the content-based image retrieval approach proposed in [14]. Table 2 compares the mean accuracies across all the species in five retrieval iterations. As we can observe, both approaches have a retrieval accuracy increase when more iterations are launched and more relevance feedback from the users are collected. Our approach reaches the highest accuracy of 0.7133 at the fifth iteration which outperforms the other approach greatly. The results demonstrate the excellent performance of our system in finding similar moth images in a higher level of semantics.

4. CONCLUSIONS

In this paper we presented an automated moth image identification and retrieval system based on wing SRV attributes. We proposed to learn the SRV attributes, which are assigned by human experts, with a probabilistic model. The model is based on the association between low-level features extracted from the image patches and the high-level SRV attributes. We designed image attribute signature to represent the semantic image contents and the automatic identification is based on the comparison between the prototype signature of the known species with the signatures of the unknown specimen. Experiments demonstrate our system provides good reliability.

5. ACKNOWLEDGMENT

This work was supported in part by NSF grants 0641076 and 0727129.

6. REFERENCES

[1] C. Wen, D. E. Guyer, and W. Li, "Local feature-based identification and classification for orchard insects," *Biosystems Engineering*, vol. 104(3), pp. 299–307, 2009.

[2] T. M. Francoy, D. Wittmann, M. Drauschke, S. Muller, V. Steinhage, M. A. F. Bezerra-Laure, D. D. Jong, and L. S. Goncalves, "Identification of africanized honey bees through wing morphometrics: two fast and efficient procedure," *Apidologie*, vol. 39(5), pp. 488–494, 2008.

[3] M.A. Oneill, I.D. Gauld, K.J. Gaston, and P.J.D. Weeks, "Daisy: an automated invertebrate identification system using holistic vision techniques," in *Inaugural Meeting of the BioNET-International Group for Computer-aided Taxonomy*, 2000.

[4] S.H. Lai, "Color retrieval of butterfly images," Tech. Rep., Department of Computer Science and Information Engineering, National Taiwan University, 2002, Master thesis.

[5] F. Liu, Z.R. Shen, J.W. Zhang, and H.Z. Yang, "Automatic insect identification based on color characters," *Chinese Bulletin of Entomology*, vol. 45, pp. 150–153, 2008.

[6] C. C. Gotlieb and H. E. Kreyszig, "Texture descriptors based on co-occurrence matrices," *Computer Vision, Graphics, and Image Processing,*, vol. 51(1), pp. 70–86, 1990.

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

[8] S. Gould, T. Gao, and D. Koller, "Region-based segmentation and object detection," in *Advances in Neural Information Processing Systems*, 2009, pp. 655–663.

[9] L. Feng and B. Bhanu, "Semantic-visual concept relatedness and co-occurrences for image retrieval," in *ICIP*, 2012.

[10] W.H. Hsu, L.S. Kennedy, and S-F. Chang, "Video search reranking through random walk over document-level context graph," in *Proceedings of the 15th international conference on Multimedia*, 2007, pp. 971–980.

[11] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth movers distance as a metric for image retrieval," *International Journal of Computer Vision,*, vol. 40(2), pp. 99–121, 2000.

[12] H. Ling and K. Okada, "An efficient earth mover's distance algorithm for robust histogram comparison," *IEEE Transactions on Pattern Analysis and Machine Intelligence,*, vol. 29(5), pp. 840–853, 2007.

[13] J. Wang, C. Lin, L. Ji, and A. Liang, "A new automatic identification system of insect images at the order level," *Knowledge-Based Systems,*, vol. 33, pp. 102–110, 2012.

[14] J. Wang, L. Ji, A. Liang, and D. Yuan, "The identification of butterfly families using content-based image retrieval," *Biosystems Engineering,*, vol. 111, pp. 24–32, 2011.