2016 IEEE International Conference on Pattern Recognition (ICPR)

Pollen Recognition Using a Multi-Layer Hierarchical Classifier

Amar Daood
Department of Electrical and Computer Engineering
Florida Institute of Technology

Eraldo Ribeiro
Computer Vision Laboratory
Department of Computer Sciences
Florida Institute of Technology

Mark Bush
Department of Biological Sciences
Florida Institute of Technology

***Abstract*—We propose a method to recognize pollen grains using a two-stage classifier. First, texture classification categorizes the pollen grains into sub-groups. Then, a final classification of individual pollen types is done by segmenting the image int multiple layers of regions for each pollen image. The main novelty in our method is threefold: (1) Adopting two successive classification stages. (2) Combining hierarchical clustering and SVM algorithms to merge similar pollen types into sub-groups. (3) Adopting a layering approach prior to performing feature extraction. The combination of these aspects gives excellent results. We evaluated our method using 1,063 light-microscopy images of pollen grains from 30 species. The results show that: (1) the layering technique increases the classification rate by almost almost 7% over using the same features directly. (2) adopting two classification stage increases the classification rate by 6%. (3) the proposed system outperformed traditional techniques.**

I. Introduction

Pollen is a granular substance that carries male reproductive cells of plants. Well known for causing respiratory allergies, pollen is also key to a number interesting applications. For example, by analyzing fossil pollen from soil collected from the bottom ancient lakes, ecologists map past climate dated over thousands of years [1]. Pollen from archaeological sites give archaeologists clues about vegetation and climate [2]. Forensic scientists solve crimes by geo-locating pollen collected from crime scenes [3]. Pollen also plays a role in energy exploration as it can point the way to petroleum fields [4].

Most applications of pollen analysis, a study field called *Palynology*, require the counting and identification of pollen species. Currently, palynologists identify pollen visually, spending hours looking at pollen through microscopes while measuring visual attributes such as shape, texture, and ornamentation [5], [6]. Automating pollen-identification was first proposed by Flenley [7], and can drastically increase research throughput in Palynology and related areas.

Early automated approaches to pollen identification focused on measuring morphological and texture characteristics of pollen grains. Treloar et al. [1] measured grain's shape roundness, perimeter, and area to classify 12 types of pollen from scanning electron microscopy (SEM) images. This work was then extended by Li et al. [8] to include measurements of visual texture for characterizing pollen grains in light-microscopy images. Here, texture characteristics were based on gray-level co-occurrence statistics.

Newer approaches to pollen classification have used a combination of visual cues. For example, Lagerstrom et al. [9] used shape, geometry, and texture to classify 15 pollen types from light-microscopic images. Lagerstrom et al.'s measured 43 characteristics including histogram statistics, moments, grey-level co-occurrence, and multi-scale multi-orientation Gabor features. Marcos et al. [10] combined gray-level co-occurrence, Gabor features, local binary patterns, and moments to classify 15 pollen types. While there has been good progress towards the development of an automated system for pollen classification pollen, the problem remains largely unsolved. Additionally, most existing methods have been tested only on a few pollen types (i.e., from 5 to 17 types).

In this paper, we describe a pollen-classification method that uses various attributes of the grain in a multilayer decomposition of regions in the pollen image. Here, we use a hierarchical-classification scheme. Our method's first stage use texture attributes to pre-classify pollen grains into subgroups. The second stage further classify the pollen in each subgroup using a region-clustering technique (i.e., region segmentation). This segmentation decomposes the pollen grain into multiple layers from which we extract features that finally classifies pollen types individually. We tested our method using 30 types of pollen in a dataset of 1,063 images. Figure 1 shows one sample from each pollen type in our dataset. The proposed method achieved a classification rate of 94%, which is among the highest classification rate obtained in this problem.

II. Our Method

We propose a multi-layer hierarchical classifier to classify pollen grains. Recognition is done in two phases of classification. The first phase pre-classifies pollen into subgroups using texture features. The second phase decomposes the pollen images into multiple layers of regions using segmentation. Finally, features are extracted from each layer to classify pollen species individually within each subgroup.

A. Feature extraction

Our method's two-stage recognition process aims at decreasing the confusion caused by trying to classify a large dataset directly. The first stage divides the dataset into two subgroups using texture classification. The Leung-Malik filter bank [11] is used to describe the visual texture of the pollen

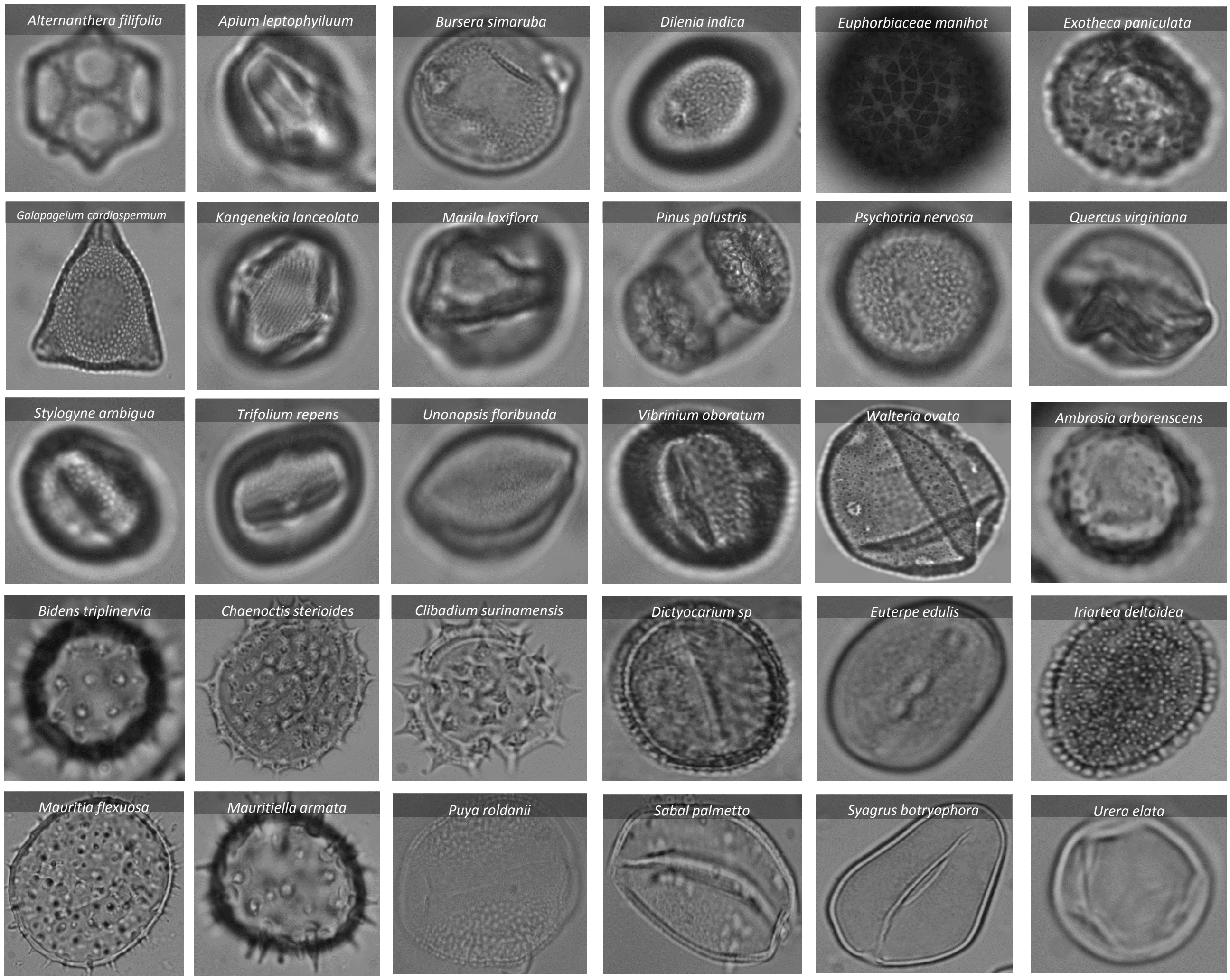

Fig. 1. A sample of each pollen type of our dataset. The dataset has 1,063 images of 30 pollen types, provided by the Florida Tech's Paleoecology Laboratory.

grains. This filter bank has 48 filters which are divided into two groups: 36 oriented filters and 12 circular filters. The oriented filters are created at 6 orientations, 3 scales, and 2 phases.

Pollen images are convolved with the filter bank to produce 48 responses for each image. Then, a local binary pattern histogram [12] is extracted for each response map to build a feature vector for each pollen grain sample. Figure 2 shows the block diagram of the feature extraction in the first stage. After creating the features vectors using texture information, we train a classifier using images of pollen-grain subgroups. Using hierarchical clustering and SVM algorithms, we divide the pollen grains into two groups. The first group includes 13 pollen types, and the second group contains 17 pollen types. After that, the individual pollen-grain types are classified within each subgroup. Prior to this individual classification, the quality of the pollen images is enhaced using histogram equalization, which increase contrast of image intensities.

Then, we cluster the pollen image into multiple layers of regions. Here, we use a modified version of the K-means algorithm. To keep layers of regions of similar pollen consistent, we sort the resulting clusters based on the gray-level intensity of their means. The final set of regions is given by:

$$R = \{L_1, L_2, \ldots, L_d\}, \tag{1}$$

where d is the number of layers and L represents an individual layer of a pollen-grain image, with $L_i = \{c_i, V_i\}$. Here, c_i is the cluster center of the i-th layer, and V_i are the pixels inside cluster i. We re-order set R according to the intensity of the cluster centers. This sorting process helps keep the order of the layers consistent, from darker to lighter regions. Feature extraction is done on each layer to create a feature vector for the pollen image. We use various features. Local binary pattern histogram and fractal dimension are used to describe each layer. In addition, gray level and histogram statistics are extracted and combined to create features. We calculate the fractal dimension of decomposed images using the Hausdorff algorithm [13]. Figure 3 shows a diagram of the feature extraction used in the second stage.

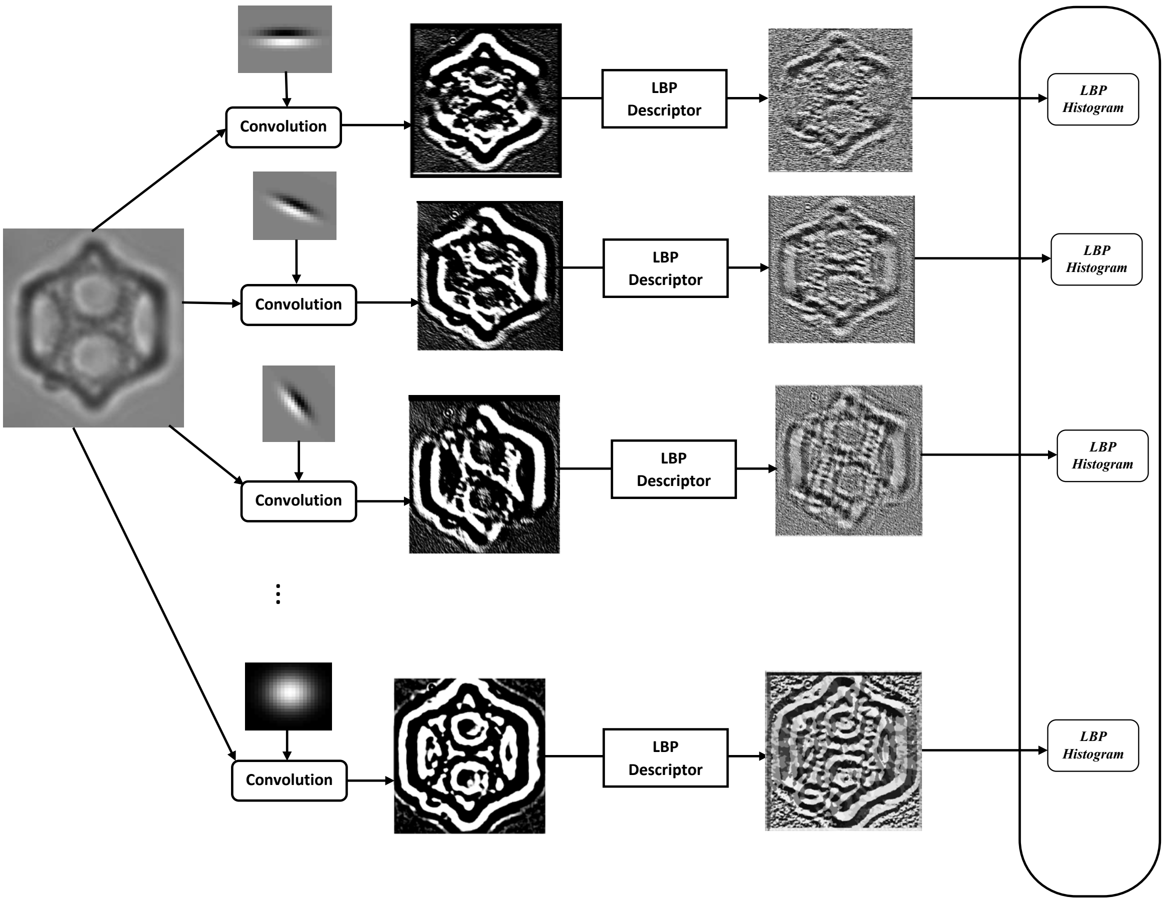

Fig. 2. Feature extraction in the first stage. Each pollen grain is convolved with the filter bank to create 48 responses. For each response, a local binary pattern histogram is extracted. These histograms form the final feature vector.

B. Group-merging technique

Two stages of classification are adopted in this work. The grouping procedure is implemented by combining an agglomerative-clustering algorithm and the SVM classification technique. First, each pollen-grain type is considered as one group. Then, we measure the similarity among the groups to combine the two closest groups together. Using a greedy approach, we re-measure the similarity of groups at each step to merge the closest two groups until all similar groups are merged together. This technique is similar to that of hierarchical clustering [14]. However, we choose the confusion matrix as a similarity distance. Algorithm 1 summarizes the steps of the group-merging technique.

By using the above grouping procedure, we can divide our pollen dataset into two subgroups. The first group contains 13 pollen types and the second contains 17 types. The major benefit of the grouping technique is revealed when we train the multi-class classifier according to ECOC technique. Instead of training a 30-type classifier, which needs 435 binary classifiers (Equation 3), we train two types classifiers.

Algorithm 1 Subgroup categorization from texture features

1: Apply convolution process between pollen grain images and the filter bank.
2: Create feature vectors by extracting local binary pattern histogram from each map response.
3: Define each pollen grain type as a subgroup.
4: Train SVM classifier to build texture classifier.
5: Compute confusion matrix to measure the similarity among the groups.
6: Check all the off-diagonal elements of the confusion matrix to find the two closest subgroups.
7: Combine the closest two subgroups in a new subgroup.
8: Repeat steps 4,5,6, and 7 until all the off-diagonal elements of the confusion matrix are zeros (i.e., there is no similarity among the merged subgroups).

C. Classification

Our method uses two types of classification technique: a single multi-class classifier approach, and an ensemble of muti-class classifiers. The initial results are implemented using

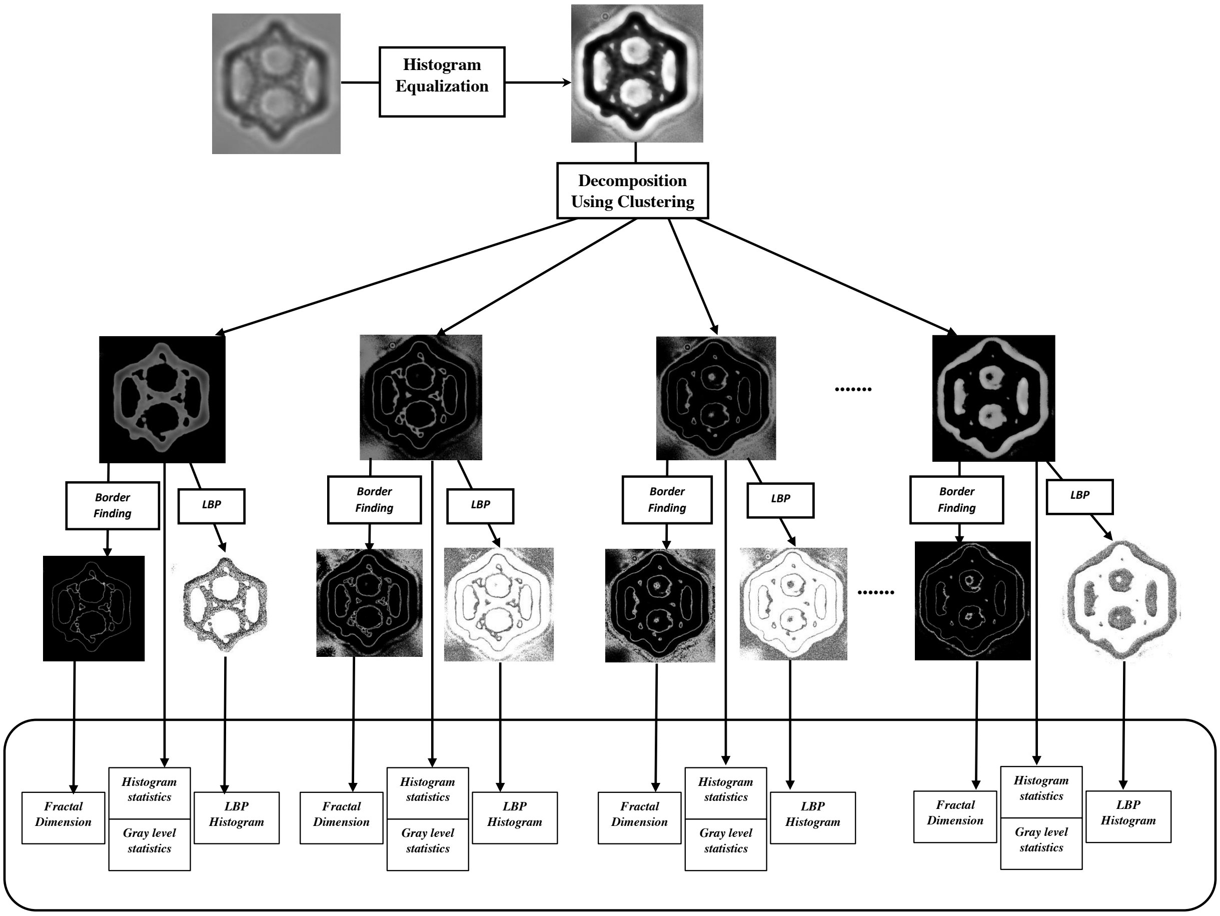

Fig. 3. The block diagram of features extraction method in the second stage. First, histogram equalization is used to to enhance the contrast of the image and then clustering process is performed to decompose the pollen grain into layers. Finally, we extract a feature vector from each layer.

SVM as a single multi-class classifier. Then, an ensemble-classifier technique is also implemented in this work.

a) Support vector machine (SVM).: Let D be a training dataset consisting of n samples of the form:

$$D = \{(x_i, y_i) | x_i \in R^p, y_i \in \{1, -1\}\}, \quad (2)$$

where x is a training sample, y is its class label, and p is the dimension of the samples [15]. SVM determines a hyperplane in high-dimensional space that classifies the data into two categories. This hyperplane is:

$$F(x_i) = w^T x_i + b, \quad (3)$$

where w and b are the hyperplane parameters, which are determined by finding the nearest samples to that hyperplane. These samples are the *support vectors* and the distance between them is the *margin distance*. The solution is a convex-optimization problem that finds the hyperplane that maximizes the margin between the two classes [16].

SVM is a binary classifier. To classify 30 classes of pollen, we adopt the error-correcting-output-code (ECOC) technique, which extends SVM to multiple classes. ECOC has two stages: coding and decoding. In the coding stage, we use a one-versus-all technique to build the codeword for each class. The number of binary classifiers trained with K class is:

$$N = K(K-1)/2. \quad (4)$$

In the decoding stage, we adopt a loss-based function to predict the class label by minimizing the sum of the binary losses of the trained binary classifiers [17], i.e.:

$$\hat{k} = \underset{k}{\operatorname{argmin}} \sum_{j=1}^{N} |m_{kj}| g(m_{kj}, s_j), \quad (5)$$

where $\hat{k}$ is the predicated label, m_{kj} is the element of the coding matrix, s_j is the score of the trained binary classifier, and g is the binary loss function.

b) Ensemble of classifiers.: Ensemble methods build a set of classifiers and then combine their prediction results [18]. Methods for combining classifiers include majority voting, and weighted majority voting. Sometimes a separate classifier is learned to find the final result. Accuracy and diversity are required conditions in the ensemble-of-classifiers technique to provide performance improvement [19]. The intuition behind the performance improvement is the diversity of the combined classifiers [18]. The accuracy depends on the classifiers that are used to construct the ensemble technique. Diversity can be achieved in many ways such as manipulating the training data and the feature space. In this work, we use the most popular ensemble techniques: Bagging, adaptive boosting, random forests, and stacking. Next, we summarize these techniques.

Bagging is the simplest method to re-sample the training data to train multiple base classifiers. Here, bootstrap re-sampling creates different copies of training data by randomly drawing a subset from the original data with replacement to achieve diversity [20]. Each subset of the training dataset is used to train a base classifier.

Adaptive boosting transforms the training data by assigning different weights to each training data subset. Multiple weak classifiers are learned. The samples that are misclassified using the first weak classifier are given large weights to increase their change to be more likely re-sampled in the next classifier [21].

Random forest uses a decision tree as a base classifier. Similar to the bagging technique, bootstrap is used to perform training data re-sampling. However, the main difference is that random forest uses a random subset of the features. Random selection of features is also called *bagging of features* [22].

Stacking achieves diversity by adopting different classification algorithms instead of choosing a single predictor as a base classifier of the ensemble. In addition to doing majority voting, a separate learner predictor is trained to combine the output of the trained classifiers [23].

III. Results

We divided our dataset into 75% for training and 25% for testing. The dataset was used for testing our method and also for comparing it with a number of approaches using our pollen dataset. These traditional approaches segment the pollen grains and perform features extraction directly to perform the classification process. We used the following features: histogram features (i.e., mean and variance of histogram), gray-level statistics (i.e., mean,variance, and entropy), geometrical features (i.e., area, perimeter, compactness, roundness, and aspect ratio based on minor and major axes), fractal dimension, gray level co-occurrence matrix (GLCM), moments invariant, Gabor features, histograms of oriented gradient (HOG) descriptors, and local binary pattern histogram (LBP). After performing feature extraction, we trained a support vector machine classifier based on these features.

Additionally, we compared our method with two approaches in the literature that combined multiple features: Marcos's method [10] and Silva's work [24]. Marcos combined gray-level co-occurrence matrix, Gabor features, local binary patterns, and discrete moments. Silva decomposed pollen grains into four layers using wavelets and then computed gray-level co-occurrence matrix to create features vectors. Table I shows the classification rates of this comparison.

To try to increase classification rates and analyze the suitability of the classification algorithms, we repeated the training step using different classification techniques. In addition to support vector machine, the random- forest classifier was used. Bagging and adaptive boosting were used to train multiple classifiers of the same type. Finally, stacking was implemented to train different classifiers such as support vector machine, K-nearest neighbors, linear discriminate, neural network, and decision tree. Majority voting was adopted to classify the test samples in the testing phase.

TABLE I
CLASSIFICATION RATES

Method	Classification (%)
Histogram features,Gray level statistics	70.97%
Geometrical features,fractal dimension	71.97%
Gray level co-occurrence matrix	51.34%
Moments invariants	44.59%
Gabor features	67.36%
HOG	62.34%
LBP	77.07%
Silva's Method	67.36%
Marcos's Method	78.92%
Histogram, gray-level statistics, fractal dimension, LBP	80.19%
Our proposed Method(using one stage)	**86.94%**
Our proposed Method(using two stages)	**93.32%**

TABLE II
CLASSIFICATION RATE USING DIFFERENT CLASSIFIERS

Classification algorithm	Classification Rate
Support Vector Machine	93.32%
Adaptive boosting	72.46%
Bagging of Decision Tree	92.08%
Random Forest	92.83%
Bagging of SVM	93.07%
Stacking of SVM,KNN,LDA,DT, and ANN	94.12%

To show that our method is an improvement over traditional approaches, we compared the multi-layer hierarchical technique with a feature-combination method that used histogram, gray-level statistics, fractal dimension, and LBP as features. This method achieved a 80.19% classification rate. After we applied a significance test, the P-value was 8.65×10^{-7} which rejected the null hypothesis. Additional classification metrics are shown in Table III including average of precision, recall, sensitivity, specificity, and F-score [25]. Figure 4 illustrates the recognition rate of individual species for both the proposed method and one using features combination.

TABLE III
EVALUATION MEASUREMENTS

Method	Precision	Recall	sensitivity	specificity	F score
Features combination	81.16%	79.68%	79.68%	99.31%	79.31%
Our Method	**94.58%**	**93.33%**	**99.33%**	**99.78%**	**93.59%**

IV. Conclusion and Future work

We proposed a method to identify pollen grains in images. Our method uses a two-stage classification approach. In the first stage, the method pre-classifies pollen species

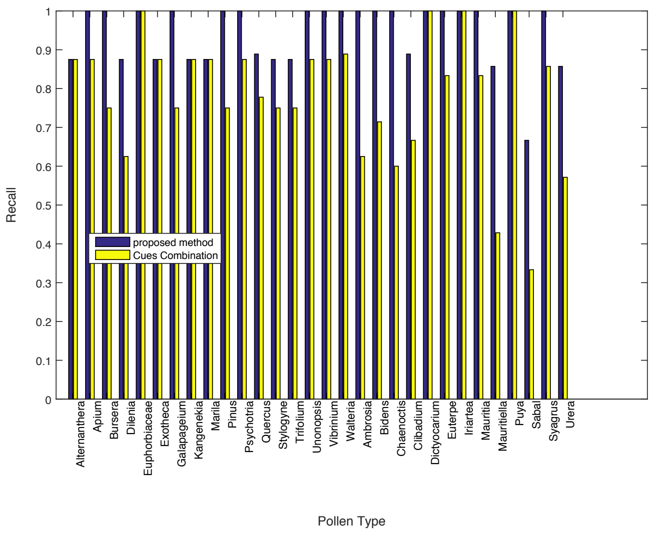

Fig. 4. Recognition rate for each species.

into two broad groups based on texture appearance. This pre-classification stage converts the large classification problem into two simpler subproblems. In the second stage, the method classifies pollen using a decomposition technique that creates multiple layers for each sample. A set of features were used to describe each layer to create features vector that represent pollen grain images. Experimental results showed that our method has superior performance over the traditional techniques. For future work, we plan to use different layer-decomposition techniques and add classification stages to create more subgroups.

Acknowledgments. The authors acknowledge support from National Science Foundation (NSF) grant No. 1152306.

References

[1] W. J. Treloar, G. E. Taylor, and J. R. Flenley, "Towards automation of palynology 1: analysis of pollen shape and ornamentation using simple geometric measures, derived from scanning electron microscope images," *Journal of Quaternary Science*, vol. 19, no. 8, pp. 745–754, 2004.

[2] K. A. Holt and K. D. Bennett, "Principles and methods for automated palynology," *New Phytologist*, vol. 203, no. 3, pp. 735–742, 2014.

[3] D. Mildenhall, P. Wiltshire, and V. Bryant, "Forensic palynology: Why do it and how it works," *Forensic Science International*, vol. 163, no. 3, pp. 163–172, 2006.

[4] C. Hopping, "Palynology and the oil industry," *Review of Palaeobotany and Palynology*, vol. 2, pp. 23–48, 1967.

[5] M. del Pozo-Banos, J. R. Ticay-Rivas, J. B. Alonso, and C. M. Travieso, "Features extraction techniques for pollen grain classification," *Neurocomputing*, vol. 150, Part B, no. 0, pp. 377 – 391, 2015.

[6] M. Rodriguez-Damian, E. Cernadas, A. Formella, M. Fernandez-Delgado, and P. De SaOtero, "Automatic detection and classification of grains of pollen based on shape and texture," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 36, no. 4, pp. 531–542, 2006.

[7] J. Flenley, "The problem of pollen recognition," in *Problems in Picture Interpretation. Clowes MB, Penny JP (eds) CSIRO: Canberra*, 1968, pp. 141–145.

[8] P. Li, W. J. Treloar, J. R. Flenley, and L. Empson, "Towards automation of palynology 2: the use of texture measures and neural network analysis for automated identification of optical images of pollen grains," *Journal of Quaternary Science*, vol. 19, no. 8, pp. 755–762, 2004.

[9] R. Lagerstrom, Y. Arzhaeva, L. Bischof, S. Haberle, F. Hopf, and D. Lovell, "A comparison of classification algorithms within the classifynder pollen imaging system," in *2013 INTERNATIONAL SYMPOSIUM ON COMPUTATIONAL MODELS FOR LIFE SCIENCES*, no. 1, 2013.

[10] J. V. Marcos, R. Nava, G. Cristobal, R. Redondo, B. Escalante-Ramirez, G. Bueno, O. Deniz, A. Gonzalez-Porto, C. Pardo, F. Chung, and T. Rodriguez, "Automated pollen identification using microscopic imaging and texture analysis," *Micron*, vol. 68, no. 0, pp. 36 – 46, 2015.

[11] M. Varma and A. Zisserman, "A statistical approach to texture classification from single images," *International Journal of Computer Vision*, vol. 62, no. 1–2, pp. 61–81, 2005.

[12] T. Ojala, M. Pietikainen, and T. Maenpaa, "Gray scale and rotation invariant texture classification with local binary patterns," in *Computer Vision - ECCV 2000*. Springer Berlin Heidelberg, 2000, vol. 1842, pp. 404–420.

[13] A. Costa, G. Humpire-Mamani, and A. Traina, "An efficient algorithm for fractal analysis of textures," in *Graphics, Patterns and Images (SIBGRAPI), 2012 25th SIBGRAPI Conference on*, Aug 2012, pp. 39–46.

[14] D. Mullner, "Modern hierarchical, agglomerative clustering algorithms," *CoRR*, vol. abs/1109.2378, 2011.

[15] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.

[16] T. Hastie, R. Tibshirani, J. Friedman, T. Hastie, J. Friedman, and R. Tibshirani, *The elements of statistical learning*. Springer, 2009, vol. 2, no. 1.

[17] S. Escalera, O. Pujol, and P. Radeva, "Separability of ternary codes for sparse designs of error-correcting output codes," *Pattern Recognition Letters*, vol. 30, no. 3, pp. 285 – 297, 2009.

[18] T. Dietterich, "Ensemble methods in machine learning," in *Multiple Classifier Systems*. Springer Berlin Heidelberg, 2000, vol. 1857, pp. 1–15.

[19] C. A. Shipp and L. I. Kuncheva, "Relationships between combination methods and measures of diversity in combining classifiers," *Information Fusion*, vol. 3, no. 2, pp. 135 – 148, 2002.

[20] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.

[21] R. E. Schapire, "A brief introduction to boosting," in *Proceedings of the 16th International Joint Conference on Artificial Intelligence - Volume 2*, ser. IJCAI'99. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1999, pp. 1401–1406.

[22] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[23] D. H. Wolpert, "Stacked generalization," *Neural Networks*, vol. 5, pp. 241–259, 1992.

[24] D. S. Da Silva, L. N. B. Quinta, A. B. Goncalves, H. Pistori, and M. R. Borth, "Application of wavelet transform in the classification of pollen grains," *African Journal of Agricultural Research*, vol. 9, no. 10, pp. 908–913, 2014.

[25] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, 2009.