

Classification of Pollen Apertures Using Bag of Words

Gildardo Lozano-Vega^{1,2}, Yannick Benezeth²,
Franck Marzani², and Frank Boochs¹

¹ i3mainz, Fachhochschule Mainz, Lucy-Hillebrand-Strasse 2, 55128 Mainz, Germany
gildardo.lozano@fh-mainz.de

² Le2i, Université de Bourgogne, B.P. 47870, 21078 Dijon Cedex, France

Abstract. The taxonomical recognition of microscopic biological particles such as pollen and spores is relevant for medical and aerobiological applications. Focusing on an accurate and automatic vision-based pollen recognition system, we propose a method for classification of pollen apertures based on bag-of-words strategy, with the ability of learning new types from different taxa without the need of new algorithms. Results demonstrate suitable performance and ability to add new taxa.

Keywords: pattern recognition, classification, local jets, bag of words, apertures, palynology.

1 Introduction

The accurate recognition of micro-particles in the field of biology is an important task for the combat of diseases. Knowledge and forecasting of the concentration of airborne spores and pollen, for example, is useful for diagnosis, medication and prevention of diseases. Traditional counting methods consist of trapping airborne samples and counting particles manually. These methods are time consuming, involve costly and specialized labor which at the same time is highly susceptible to human error and inconsistency. An important application of recognition of pollen taxa is allergenic analysis, where it is desirable to identify pollen at genus level. The task faces a great complexity due to the high diversity, for example, pollen of the same taxonomic family usually share many similarities even when they belong to different genus branches. Their morphological and aerobiological characteristics are the key for taxa recognition, and are widely described within the palynological literature [1][2]. The main morphological characteristics include average size, shape, equatorial outline, aperture type and amount, exine and intine structure, and type of ornamentation. On the other hand, aerobiological characteristics provide information about the periods of pollination, specifically for each taxon. This knowledge is highly related with the levels of concentration of pollen in the air in a specific time slot.

One important discriminating characteristic of pollen is the type and amount of apertures. Apertures can be observed externally, and due to their considerable size with respect to the particle, most can be observed by means of typical

brightfield microscopes in typical lens magnifications ranging from 10x to 40x. Nevertheless, the appearance of apertures shows high variability due to both numerous taxa, and to the change of orientation of the particle with respect to the observer. Figure 1 shows examples of the same type of aperture seen from different perspectives.

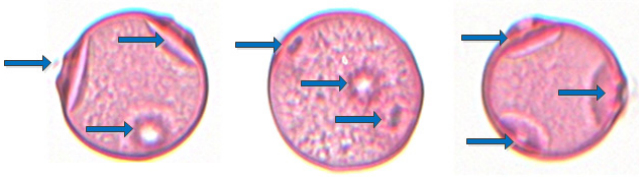


Fig. 1. Examples of aperture of birch pollen marked with arrows. Even though it is the same type of aperture, different appearances can be noticed due to change of the point of view.

The advent of computer vision has brought plausible techniques for fast and robust detection of visual characteristics in order to aid the identification of pollen. Those characteristics, for example shape, texture and color, are employed to describe globally the entire pollen particle in typical recognition systems. Only few works have focused on detecting specific particle structures such as apertures besides the global description of the particle. For instance, the ASTHMA project [3] created individual detectors of aperture, reticulum and cytoplasm. Although no much detail is provided, typical segmentation techniques were employed to detect those regions that were later characterized locally with shape and color features. In the study developed by Chen et al.[4], different methods specific to the pollen type and the point of view were developed in order to detect characteristic patterns formed by the aperture and colpus of birch and mugwort pollen. For circular appearances of pollen apertures, Hough transform and thresholding were used for detection. For the case of apertures located on the boundary of the particle, analysis on the intensity profile of the polar transformation of the particle image, like frequency spectrum and template matching, was conducted. In both cases, measurements on the detected structures were fed together to shape, textures or color features in a larger classifier of pollen taxa. No evaluation was done independently on apertures or other structures. Therefore there is not a clear baseline for comparison. The aforementioned strategies need a particular algorithm for each type of aperture, and in most cases, for different points of view. Consequently, both are too inflexible to learn new type of apertures because completely new algorithms would be needed.

We propose a method for classification of pollen apertures considering intra and inter-class variability. Moreover, it has the ability to add types of apertures from new taxa, without the need of designing a new specific detector. Flexibility is important in order to enable the classifier to learn new varieties of apertures from different points of views with the same method. The method relies on describing apertures in terms of common visual components in a similar way to the Bag of Words approach (BOW)[5]. BOW is a method originating from

strategies of text categorization and has been since then widely applied to image processing with several extensions and adaptations [6] [7]. However, to our knowledge, there is no previous application of BOW to microscopic objects. The core concept under the BOW strategy is the creation of a codebook of visual words which are estimated by computing local descriptors at strategic points within the input image. Descriptors with similar values are grouped together in clusters that represent the different visual words. Next, a histogram keeps the frequency of the visual words contained in the image. This particular histogram's bin arrangement characterizes the different regions of interest in the image. It is expected that regions belonging to the same category display a similar pattern in the histogram of words. Consequently, a classifier is employed to learn the different histogram patterns in order to define the classes.

The rest of the paper is organized as follows: In section 2, the method of creating the visual codebook and the differences to typical BOW are explained. In section 3, the setup of the experiments and results are presented. Conclusions about our method and outline of the future work are stated in section 4.

2 Application of Bag of Words to Aperture Description

In this section, we describe how to create the codebook of visual words and how the apertures are represented. The input of the method are image sections of pollen particles, referred in this text simply as regions regardless whether they are apertures or not. The method consists of the following steps: it begins with the subdivision of the regions into a grid of small patches. Next, local descriptors are computed for each patch in order to capture the particular arrangement of pixels, which are subsequently clustered together according to their similarity. These clusters are regarded as the visual words that build the codebook of patterns. Finally, regions are represented by creating their particular histogram of visual words. The overall scheme can be seen in figure 2.

2.1 Computation of Local Descriptors on Region Patches

Our BOW approach is based on the creation of a codebook that keeps the collection of possible visual patterns present in pollen particles. Consistency of the characterization of similar visual patterns is achieved by means of local descriptors. In order to learn these patterns, different regions of the particle must be observed. Contrary to typical BOW, our method divides a region into a grid of small patches of regular size, rather than detecting salient keypoints. Next, local descriptors are computed densely on each of the grid patches. A comprehensive survey of local descriptors can be found in [8]. We have employed the local-jets descriptor, which has demonstrated good performance describing shape, while also being independent from the detection of salient keypoints.

The local-jets descriptor is a differential descriptor designed by Schmid and Mohr [9] from the measures derived by Koenderink and van Doorn [10].

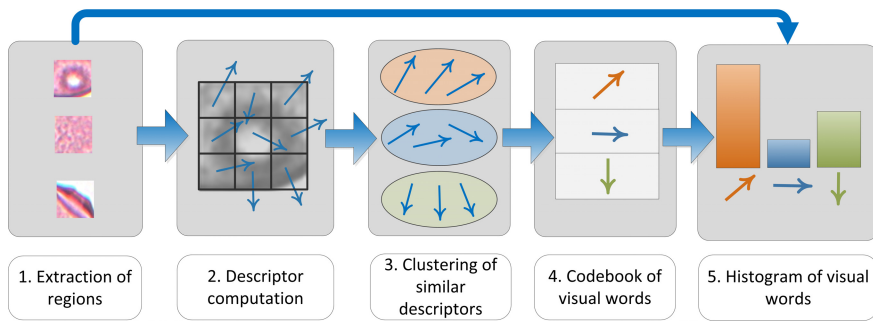


Fig. 2. Overall method of creating the codebook and representing the regions with visual words. 1. Regions are extracted from the particle. 2. Local descriptors are computed on a dense grid. 3. Similar local descriptors are clustered. 4. The codebook is formed with cluster representations. 5. Histograms of visual words is created for all regions.

A local-jet measure L_n is defined as the convolution of the image $I(x, y)$ with a series of Gaussian derivatives $G_n(x, y, \sigma)$ of different scale σ , that is:

$$L_n(x, y, \sigma) := (G_n \star I)(x, y, \sigma). \tag{1}$$

In this equation the Gaussian of order n yields to the local jet L_n . The local-jet measure uses derivatives up to 3^{rd} order. It is also computed at different scales ($\sigma = [0, 0.5, 1]$) in order to provide scale robustness. Finally, all local-jet measures at different scales and orders are concatenated to form a single local-jet descriptor.

Although the local-jets descriptor is not completely rotation invariant because there is no dominant rotation estimation, enough robustness is achieved by training the algorithm with views of the regions at different rotations. Complete scale invariance is not strongly required since the input images are assumed to have been processed with the same scale parameters, and additionally the size of the apertures does not change considerably throughout instances of the same type. However, the local-jets descriptor achieves needed scale robustness since the Gaussian derivatives are computed at several scales.

The training set is formed by grouping together all the local-jets descriptors computed from different regions. At this point, it does not matter whether the local-jets descriptors belong to an aperture or not, since we are interested only in learning all of the different possible patterns.

Because the training image size is such that it contains exclusively the aperture, most local descriptors are computed only on patches of the grid that are strongly informative. This allows to reduce the number of computed local descriptors, unlike typical BOW. A second difference to typical BOW is that our method keeps the position of the relative to the region in order to consider spatial information. Some strategies to achieve this, for example those based on pyramid matching, focus on the computation of abundant descriptors [11][12]. Instead, for our case of fewer regularly spaced local descriptors, we can simply

code the relative position of the patches in the grid because the grid has typically no more than 25 patches. A linearly spaced value is assigned to each position of the grid ranging from 0 to 1. The position code is finally added to the local descriptor as an extra dimension.

2.2 Codebook of Visual Words

Because pollen exhibits intra-class variability due to both its biological nature and changes of perspective, similar visual patterns have no identical local descriptor values. In order to identify categorical patterns, local descriptors need to be grouped according to their similarity. For this purpose, k-means clustering is employed where the number of clusters represents the number of learned patterns. This parameter is to be determined experimentally. The collection of clusters, interpreted as visual words, represents the different patterns in the codebook. Subsequently, they will be used as a reference for identifying patterns on unknown instances. Examples of patches with similar patterns for birch pollen are shown in figure 3.

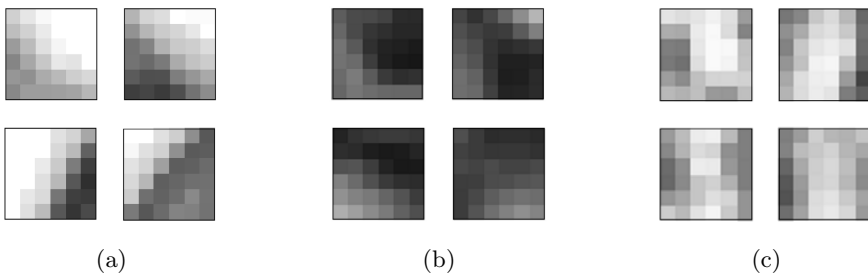


Fig. 3. Example of patches of birch pollen, grouped according to the assigned cluster. Those in the same cluster, show similar patterns.

Apertures and the rest of the regions can now be represented by their content of visual words. During the training phase, words in the codebook are extracted from images of both cases, aperture and not an aperture. A histogram indicating the frequencies of each visual word in the codebook is built for all regions. Given that input regions are initially labeled, a machine learning algorithm is employed to associate similar histogram patterns to the classes "aperture" and "not an aperture".

Different points of view affect directly the look of an aperture. A strategy to solve this problem is to split the task into as many binary classifiers as there are different views of the same type. Later, all classes corresponding to the same aperture are grouped together. A second strategy is to employ more complex algorithms, powerful enough to recognize different views as a single positive class and to differentiate them at the same time from the negative class.

Our method applies Support Vector Machines (SVM) for classifying regional histograms. In recent years, SVM have shown good results in binary classification problems by mapping data into a higher dimension where a linear separation is possible [13]. Later will be experimentally shown that this method is suitable for classification of apertures with multiple points of view. Nevertheless, since the classification stage is independent from the codebook, it is possible to interchange machine learning algorithms.

3 Experiments and Results

We tested our method on the classification of apertures of birch pollen (*Betulaceae Betula*). Images of pollen were scanned with a brightfield slide scanner Leica SCN400 with magnification 40x. Particles were stained with magenta dye. Images of regions belonging to pollen particles were sampled at specific locations. The birch dataset consisted of 184 regions of which, 92 showed an aperture and 92 did not show an aperture. The classification of an unknown region consists of extracting the visual words contained in the image, based on the previously learnt codebook, and building the corresponding histogram. Finally, the SVM model is applied in order to assign the test region into a class.

The size of the region is a parameter to be evaluated, and it is consistent in all the instances of the set. The size of a region is determined by two factors: the size of the image patch on which an individual local descriptor is computed, and the number of patches in which the grid of image region is split. In all cases, image patches and regions are square. The size of the region is always chosen as a multiple of the image patch size such that the grid fits exactly into the region. Therefore, a square region size is denoted as $N \times N \times P$, where N is the number of image patches that fit into any of the region's sides, and P is the number of pixels of the patch size on one side.

Input images were converted to grey levels and the intensity of the average background content was compensated throughout the image by subtraction. The whole set was employed to create the codebook. Then, the local-jets descriptor was computed on the subdivided regions. After computing the word histograms, a binary SVM classification was performed and evaluated using 10-fold cross-validation, labeling the "aperture" class as positive and the "not an aperture" class as negative examples.

The parameters, of both region size and the number of visual words in the codebook were tested for different configurations. The number of clusters, which directly determines the number of visual words, was varied within the range of 3 and 49. Similarly, the region size varied between 20 and 40 pixels for different combinations of $N \times N \times P$. Performance was evaluated by means of a Receiver Operating Characteristic (ROC) curve that compared the number of *hits* (recall) versus the number of *false alarms* (fall-out). Figure 4 shows the ROC curve using the best parameters: $4 \times 4 \times 6$ and 32 visual words. The best configuration has an accuracy of 95.8% employing a threshold equal to 0.60, having a confidence interval of ± 2.9 with a 0.95 confidence level. It was found that the variation of

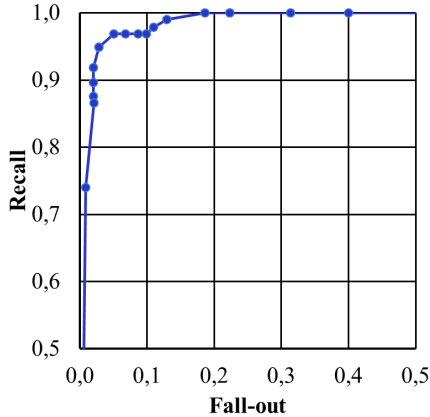


Fig. 4. ROC curve of the BOW performance of classification of birch pollen apertures using best parameters, region size of $4 \times 4 \times 6$ and 32 visual words with local-jets descriptor.

the number of visual words does not significantly change the performance. However, the optimal size of the region has to be found such that it fits the typical size of the aperture. Very small or large regions cause a decrease in performance.

We also experimented by substituting the local-jets descriptor. We used a Zernike descriptor, based on Zernike moments, which are particularly useful for describing shape [14][15] and have been applied widely in image retrieval tasks. The descriptor is rotation invariant, and translation invariance can be achieved by image normalization with respect to the centroid. By evaluating both the region size and the number of visual words in the codebook, we found the best performance using a size of $4 \times 4 \times 7$ and 5 visual words. The best configuration has an accuracy of 95.4% employing a threshold equal to 0.45, having a confidence interval of ± 3.0 with a 0.95 confidence level. Results confirm the best size for capturing information from apertures and the noncritical number of visual words. Therefore, it is expected that optimal region size changes for different types of apertures. Figure 5 shows the ROC for this configuration.

Tests were also done using the SIFT descriptor in our method [16], but did not produce satisfactory results. We attribute it to the fact that the discriminative power of the SIFT descriptor is effective when working together with the detection of plenty keypoints, rather than in a small grid.

Experimentally, it was confirmed that adding the extra dimension with the spatial information code improves classification results on average 4% compared to the case without it. This difference is statistically significant when compared to the confidence interval of 2.9% and 3.0% of the methods using local-jets and Zernike features respectively.

Finally, in order to test the ability of the method to learn apertures from new taxa, we evaluated our method on the classification of apertures of two types of pollen at the same time. An additional dataset of alder pollen (*Betulaceae*

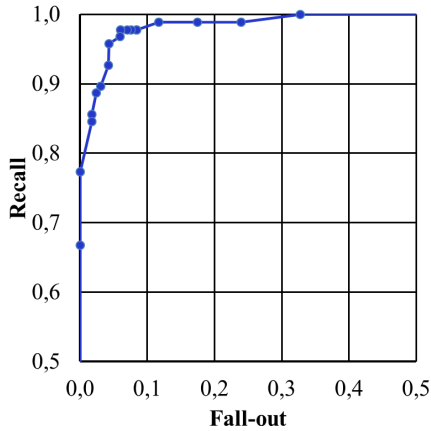


Fig. 5. ROC curve of the BOW performance of classification of birch pollen apertures using best parameters, region size of 4x4x7 and 5 visual words with Zernike descriptor.

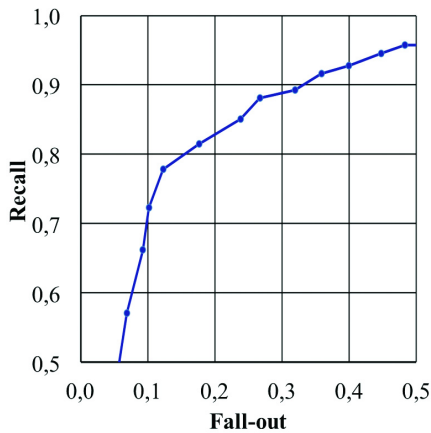


Fig. 6. ROC curve of the BOW performance of classification of alder and birch pollen apertures using 40 visual words with local-jets descriptor.

Alnus) was employed, which consisted of 146 unstained regions of which 73 showed an aperture and 73 did not show an aperture. Optimal region size was experimentally selected for each type, 4x4x6 for betula apertures and 2x2x20 for alder apertures. The performance ROC curve using the local-jet descriptor is shown in figure 6. The method adjusted very well to the newly added taxon, simply through learning the correct new region size. Best results were obtained when using more than 15 visual words. This reflects the need of the codebook for containing more patterns due to the addition of the new taxon. The main reason of the slight performance reduction compared to one-taxon classification is likely

that the alder samples were not previously stained with magenta dye. This sort of preparation of the samples is common in order to increase the contrast of the pollen characteristics. Since previous works focus on the final identification of the pollen types, there is no independent evaluation of the aperture detection. Therefore, there is not a clear baseline to compare our results with.

4 Conclusions and Future Work

We have presented a method for detecting apertures in pollen particles. It was tested on the aperture detection of two different pollen types. The results demonstrate that the method is robust and yet flexible enough to learn more types of apertures. Moreover, the presented BOW approach can be extended to process different types of specialized regions of different biological particles with minimal taxa-specific adjustments, for example in spores.

It was shown that local descriptors together with the modified bag-of-words approach are suitable for describing apertures by capturing the most significant pixel patterns, and are also able to identify them against regions of different contents. Despite the random orientation of pollen particles in the sample slide, the method showed robustness to changes of perspective and rotation. Finally, the spatial relationship among the patterns provided useful information with an improved classification performance.

With the goal of an automatic pollen recognition, future work should involve taxon-specific semantic information of pollen such as spatial location and amount of apertures. It is also planned to study additional identifying information present in different focal image planes (layers).

Acknowledgements. The authors are grateful for their financial support to the Bundesministeriums für Wirtschaft und Technologie in Germany through the 3PGM project under the program Zentrales Innovationsprogramm Mittelstand ID KF2848901FR1, to the Conseil Regional de Bourgogne in France and also to the Fond Europeen de Developpement Regional (FEDER). The authors are also grateful to the project partner Bluestone Technology GmbH for providing pollen images and to the Max Plank Institute for Chemistry for permitting the use of their facilities.

References

1. Erdtman, G.: An Introduction To Pollen Analysis. Chronica Botanica Company, USA (1943)
2. Hesse, M., Halbritter, H., Weber, M., Buchner, R., Frosch- Radivo, A., Ulrich, S.: Pollen Terminology. In: An Illustrated Handbook, Springer, Austria (2009)
3. Boucher, A., Hidalgo, P.J., Thonnat, M., Belmonte, J., Galan, C., Bonton, P., Tomczak, R.: Development of a semi-automatic system for pollen recognition. *Aerobiologia* 18(3), 195–201 (2002)

4. Chen, C., Hendriks, E.A., Duin, R.P., Reiber, J., Hiemstra, P., De Weger, L., Stoel, B.: Feasibility study on automated recognition of allergenic pollen: grass, birch and mugwort. *Aerobiologia* 22, 275–284 (2006)
5. Csurka, G., Dance, C., Bray, C., Fan, L., Willamowski, J.: Visual categorization with bags of keypoints. In: *Pattern Recognition and Machine Learning in Computer Vision Workshop, ECCV Grenoble, France*, pp. 1–22 (2004)
6. Wu, J., Tan, W.-C., Rehg, J.M.: Efficient and Effective Visual Codebook Generation Using Additive Kernels. *Journal of Machine Learning Research* 12, 3097–3118 (2011), Georgia Institute of Technology
7. López-Sastre, R.J., Tuytelaars, T., Acevedo-Rodríguez, F.J., Maldonado-Bascón, S.: Towards a more discriminative and semantic visual vocabulary. *Computer Vision and Image Understanding* 115, 415–425 (2011)
8. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27, 1615–1630 (2005)
9. Schmid, C., Mohr, R.: Local grayvalue invariants for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(5), 530–535 (1997)
10. Koenderink, J.J., Doorn, A.J.: Representation of local geometry in the visual system. *Biological Cybernetics* 5(6), 367–375 (1987)
11. Grauman, K., Darrell, T.: Pyramid matching kernel: Discriminative classification with sets of image features. In: *Tenth IEEE International Conference on Computer Vision, ICCV 2005*, vol. 2, pp. 1458–1465 (2005)
12. Lazebnik, S., Schmid, C., Ponce, J.: Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2006*, vol. 2, pp. 2169–2178 (2006)
13. Byun, H.-R., Lee, S.-W.: Applications of support vector machines for pattern recognition: A survey. In: Lee, S.-W., Verri, A. (eds.) *SVM 2002*. LNCS, vol. 2388, pp. 213–236. Springer, Heidelberg (2002)
14. Teague, M.R.: Image analysis via the general theory of moments. *Optical Society of America* 70(8), 920–930 (1979)
15. Vorobyov, M.: *Shape Classification Using Zernike Moments*. Technical Report. iCamp-University of California Irvine (2011)
16. Lowe, D.: Distinctive image features from scale-invariant keypoints. In: *IJCV* (2004)