International Journal of
Intelligent Systems and Applications in Engineering

ISSN:2147-6799 www.ijisae.org Original Research Paper

The Principal Component Analysis Method Based Descriptor for Visual Object Classification

Zühal Kurt[1], Kemal Özkan[2], Şahin Işık[3]

Accepted 5th June 2015 DOI: 10.18201/ijisae.08060

Abstract: In the field of machine learning, which values / data labeling or recognition is done by pattern recognition. Visual object classification is an example of pattern recognition, which attempts prompt to assign each object to one of a given set of object classes. The basic elements of the process of pattern recognition, feature extraction, feature selection and classification. The complexity of feature selection/extraction is, because of its non-monotonous character, an optimization problem. The process of feature extraction, pattern characteristic feature is eliminated and the acquisition of a certain amount of irrelevant information is provided dimensionality reduction. In the fields of machine learning and statistics, feature selection algorithms are known the choice of variable selection or additional subset of variables. For the most part of visual object classification methods use bag of words model for image representation with image features. In this method, patches extracted from images are described by different shape and texture descriptors such as SIFT, LBP, LTP, SURF, etc. In this paper we introduce a new descriptor based on weighted histograms of angle between two vectors of local based PCA transform. We compare the classification accuracies obtained by using the proposed descriptor to the ones obtained by other well-known descriptors on Caltech-4 and Coil-100 data sets. Experimental results show that our proposed descriptor provides good accuracies indicating that PCA based local descriptor captures important characteristics of images that are useful for classification. When we described image representations obtained by PCA based descriptor with the representations obtained by other detection of keypoints, results even get better suggesting that tested descriptors encode differential complementary information.

Keywords: *Feature Descriptor, Feature Extraction, Visual Object Classification, Principal Component Analysis, Bag Of Words Model.*

1. Introduction

Visual object classification can be described as the work of assigning an image one or multiple labels corresponding to the presence of a visual object class. It is an important work, and a well-done visual object classification system may significantly improve the performances of other major computer vision applications such as image retrieval and object localization. The primary difficulty in object classification is because of the large intra-class variations and viewpoint changes in all of the categories. In addition to this, lighting-scale changes, complex background, occlusion and presence of noise in the images make the problem even difficulty.

Most of the recent state-of-art object classification methods use bag of words (BoW) models which is first used for text classification. After extension of this model to the visual object classification by Csurka et al. [1], such representations have been widely used for both object classification and localization [3, 9, 19, 20, 21]. The BoW model behaves on each image as an ordered collection of representative patches. Therefore, it needs sampling a set of patches from the image, computing descriptor vectors for each patch, quantization of descriptors, and accumulating histograms or signatures of patch appearances based on this quantization to obtain the final image representation. Then, resulting image feature histograms are supporting a classifier to identify the label(s) of the image category. Even though BoW models disregard spatial relationships between the features, they astonishingly work well for object classification because of the high discriminative power of some words. They also have good capacitance to occlusions, geometric deformations, and illumination variances.

There are basically three primary implementation issues in BoW: how to sample patches from image, how to describe them (descriptor selection), and how to quantize the resulting descriptors. This is also known as codebook generation. Patches are typically illustrated from the image at many different positions and scales, either densely [3,6], randomly [5], based on the output of some kind of salient region detector [1,20], or based on the output of segmentation algorithms [7,8]. Then selected patches are determined by using different descriptors. On one hand, the descriptors extracted from patches should be invariant to variations owing to the image transformations, lighting variations and occlusions, which are unrelated to the categorization. Also, they must provide enough information to distinguish between the object categories. Among these, histogram based descriptors have become very popular owing to their performance and efficiency. Many of these are based on oriented image gradients, including SIFT [4], SURF [15], Histograms of Oriented Gradients (HOG) [17], Generalized Shape Context [16]. Others are based on local patterns of qualitative gray level differences, including Local Binary Patterns

[1,*] *Mathematics - Computer Sciences Department, Eskişehir Osmangazi University, Eskişehir/Turkey*
[2,3] *Computer Engineering Department, Eskişehir Osmangazi University, Eskişehir/Turkey*
** Corresponding Author:Email: zkurt@ogu.edu.tr*

(LBP) [11], and Local Ternary Patterns (LTP) [18]. The resulting descriptors are then clustered to achieve visual words (dictionary). Descriptors extracted from image are assigned to visual words based on some similarity measure, and the final image vector representations are obtained by accumulating histograms of occurrences of each visual word. Hence, identification of such visual words is important from two direction: Firstly, it supplies some robustness against descriptor variations since similar patch descriptors are assigned to the most similar visual word. Secondly, it supplies a fixed length representation vector for images with different sizes.

Different quantization algorithms are proposed to address quantization process. Among these, quantization algorithms based on k-means clustering [1,19], mean shift [3], hierarchical clustering [21], randomized trees [2] are a few to name. At last, the image feature vectors achieved from quantization algorithm are fed to classifier such as Nearest Neighbor, Naïve Bayes or Support Vector Machines (SVM) classifiers to determine the label(s) of the visual object category.

In this study, we focus on descriptors that are used to represent the image patches and propose a new descriptor based on weighted histograms of angle between two vectors of local based PCA transform. The remainder of the paper is organized as follows: In Section II we describe the proposed descriptor. In Section III experimental results are given. Lastly, our conclusions are presented in Section IV.

2. THE METHOD

A. Motivation

PCA is a vector-based approach. The aim in this PCA method, large size and interrelated vectors, in the form of small unrelated vectors is to perform a dimensionality reduction. In using image processing applications, the digital image data is realized with the representation of the vector format. PCA is the way to obtain most necessary information on a group characteristic features of an image. And then the other way, given this descriptive name concatenated set of images as a linear expression is based on the principle. By receiving the projection of the feature vector in the drawings in the descriptive vectors smaller size are obtained. By receiving the projection of the feature vector in images the descriptive vectors that are obtained smaller size.

First of all, interest points are detected from the image with DoG detector. After finding points of interest to the radius of points of interest (scaling parameter) by the right, the left, above and below, will expand three times the scaling parameter is drawn a square frame. This frame, as in SIFT algorithm is divided into 4x4 cells.

In order to find the row differences of the entire frame, mutual vector is subtracted from the averaged data matrix. The result provides the variance ratio in the direction of x axis. In order to figure out the variance ratio in the vertical axis, the row vectors of the frame matrix around the key point is used for the same process. Therefore, while calculating the difference matrix of Sift descriptor by taking the derivative, in this study difference matrices are calculated by adopting a method based on PCA in a vectorial dimensions. In addition to that, depending on the directivity angle of the difference matrix obtained by the rows and the columns, histograms are found in 8 different boxes. Thus, 128 dimensional (4x4x8) descriptors based on PCA are acquired. The acquiring of this 128 dimensional (4x4x8) descriptors can be seen in the Figure 2.

Vertical and horizontal variance ratios obtained by adopting methods based on PCA on an original image are given in the Figure 1. In the figure 1. (b) and (c) respectively, the vertical and horizontal variance ratios obtained by adopting methods based on PCA of the image is given. In the figure 1. (d) and (e) respectively, variance ratios in the direction of x axis and y axis obtained by applying gradient is shown. By comparing the images (b) and (c) with the images (d) and (e), it is likely to say that the method based on PCA and the method based on gradient provides similar results.

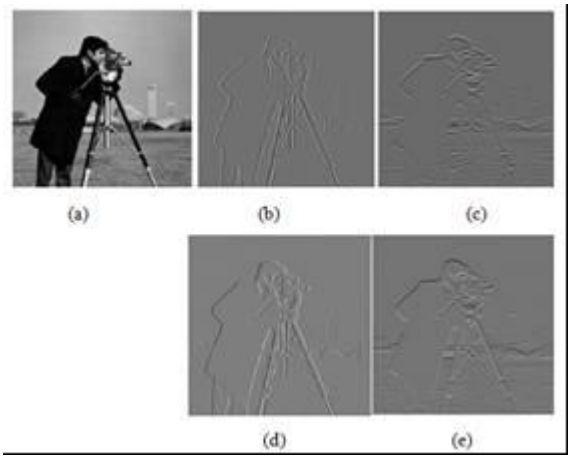

Figure 1. a) Original image **b)** Horizontal variance ratio obtained by adopting methods based on PCA **c)** Vertical variance ratio obtained by adopting methods based on PCA **d)** Variance ratio on the x axis obtained by methods based on gradient **e)** Variance ratio on the y axis obtained by methods based on gradient.

In the Figure 2, in order to be able to explain the descriptor based on PCA, a certain area of the image is examined. This area is divided into 2x2 instead of 4x4, in order to be able to present the cells in the area clearly. First, the variance ratios of the area based on PCA are calculated. Secondly, those ratios are weighted according to their directivity angles and the histogram is drawn. As can be seen from the figure 2, 8 different directivity angles are considered, therefore, the histogram is acquired by weighting the directivity angles. Lastly, putting together the 8 dimensional histograms obtained by each cell, 128 dimensional descriptive vector is formed.

Thus, the training step is accomplished. The next step is testing. Test vectors are using for attributes in the training vectors, and aims to make the class assignment is to perform a classification.

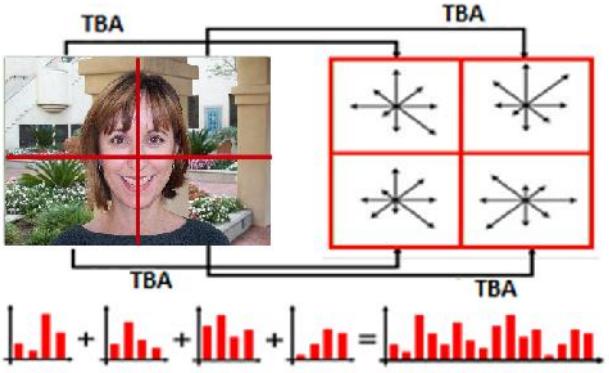

Figure 2. Computing PCA based descriptor. Image patch is first divided into 2×2 cells in this example. PCA components are extracted for each cell, and the histogram of each cell is created by accumulating weighted votes to angle bins. The final descriptor is formed by concatenating the histograms of all cells.

B. Bag of Words Based Object Classification With Principal Component Analysis Based Descriptor

BoW based visual object classification is illustrated in Fig. 3. In this method, patches are sampled from images at many different positions and scales by using different sampling techniques. This is followed by extracting fixed-size features from the patches by using various descriptors such as SIFT, SURF, LBP, LTP, etc. The resulting patches from all training images are then clustered to obtain visual words. During image representation, descriptors extracted from an image are assigned to visual words based on some similarity measure, and the final feature vector is obtained by accumulating histograms of occurrences of each visual word.

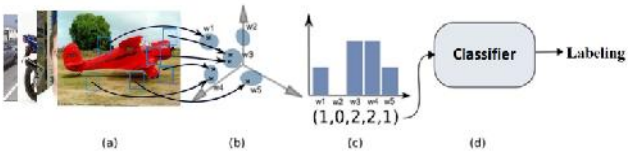

Figure 3. Visual object classification based on bag of words model.

3. EXPERIMENTS

We tested the proposed descriptor on two visual object classification databases: the Caltech-4 and Coil-100. BoW model is used for representation of images, and we compared the proposed PCA based descriptor to SIFT, LBP, LTP, and SURF. We also combined the image representations obtained by FT descriptor with the representations obtained by other descriptors to judge whether the descriptors complement each other in the sense that they capture different information which are useful for classification. We divided the patches into 4x4 cells, thus the dimensionality of the PCA based descriptor is 128. We used nonlinear Support Vector machine (SVM) classifier for classification in all experiments, one-against-one procedures are used to extend the binary SVM classifier to multi-class classification. The Gaussian kernel $k(x, y) = \exp(-\| x - y \|^2 / \pm)$ is used as a kernel in nonlinear SVM in all experiments.

Figure 4. Some image samples from Caltech-4.

A. EXPERIMENTS ON THE CALTECH-4 DATABASE

The Caltech-4 database [22] includes images of objects belonging to 4 visual categories; airplanes, cars, faces, and motorbikes. There are respectively, 1074, 526, 450, and 826 images per category. Each class includes images of highly variable object poses under different lighting conditions with complex backgrounds as shown in Fig. 4. We used two techniques for sampling patches: dense sampling and sampling based on DoG (Difference of Gaussians) interest point detector [4]. All tested descriptors are computed by using the same patches. The descriptors extracted from training images are clustered by k-means clustering method and cluster centers are considered as visual words forming the visual vocabulary. The size of the visual vocabulary is set to 1,000. To build image representation, each extracted descriptor is compared to the visual words and associated to the closest word. The final image feature histogram vector is L1 normalized.

Table 1: Classification rates (%) for different descriptors on the Caltech-4 dataset.

Descriptors	Classification Rates for DoG points	Classification Rates for Dense Sampling
SIFT	$81.36^{\pm 5.9}$	$96.94^{\pm 0.9}$
SURF	$75.09^{\pm 3.4}$	$96.35^{\pm 1.5}$
LBP	$86{,}54^{\pm 2.8}$	$94.74^{\pm 3.8}$
LTP	$84.35^{\pm 2.5}$	$95.30^{\pm 3.2}$
PCA	91,23∓3,38	97,08∓2,24

Figure 5. Forty objects chosen from Coil-100 dataset.

B. EXPERIMENTS ON THE COIL-100 DATABASE

The Coil 100 database [23] includes images of objects belonging to 100 visual categories. There are 72 images for each category. The size of each image is 128x128 pixels. Each class includes highly variable poses of the same object under same lighting conditions. For our experiments, we chose 40 classes (shown in Fig. 5) from the database.

The images are easy in the sense that the background is uniform. Here we only used dense sampling since dense sampling produced better results on Caltech-4 images. As in the previous case, we set the visual vocabulary size (hence, the dimensionality of the image feature vectors) to 1000.

Table 2: Classification rates (%) on the Coil 40 dataset.

Descriptors	Classification Rates for Dense Sampling
SIFT	$89.45^{\pm 2.5}$
SURF	$87.88^{\pm 1.9}$
LBP	$88.36^{\pm 1.3}$
LTP	$89.26^{\pm 3.3}$
PCA	93,54∓3,23

The classification rates are given in Table 2. Among all tested descriptors, our proposed PCA based descriptor provides the best accuracy followed by SIFT. Similar to the previous case, when we combine the image representations obtained by using different descriptors, accuracies get higher.

4. Conclusions

We have introduced a new descriptor using PCA method of an image patch for visual object classification based on BoW models. In contrast to the traditional gradient based descriptors using the principal components, we emphasize on the variance ratio on the x and y axis. The proposed descriptor is formed by accumulating weighted histograms of angles of variance ratio based PCA. Our initial results on small data sets are very encouraging, and as a future work we are planning to test our

proposed descriptor on more challenging data sets including many classes. We will also try different overlapping/non-overlapping cell sizes, different bin sizes, and different normalizations of histograms during descriptor computing to improve the classification accuracies further. Lastly, we are planning to extend the descriptor to be able to use it in object detection tasks.

References

[1] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, C. Bray, *"Visual categorization with bags of keypoints"*, ECCV Workshop on Statistical Learning for Computer Vision, 2004.

[2] F. Moosman, E. Nowak, and F. Jurie, *"Randomized clustering forests for image classification"*, IEEE Transactions on PAMI, vol. 30, pp. 1632-1646, 2008.

[3] F. Jurie and B. Triggs, *"Creating efficient codebooks for visual recognition"*, ICCV, 2005.

[4] G. Lowe, *"Distinctive image features from scale-invariant keypoints"*, International Journal of Computer Vision, vol. 60, 2004.

[5] Nowak, F. Jurie, and B. Triggs, *"Sampling strategies for bag-of-features image classification"*, ECCV, 2006.

[6] T. Leung, J. Malik, *"Representing and recognizing the visual appearance of materials using three-dimensional textons"*, International Journal of Computer Vision, vol. 43, pp. 29-44, 2001.

[7] K. Barnard, P. Duygulu, R. Guru, P. Gabbur, and D. Forsyth, *"The effects of segmentation and feature choice in a translation model of object recognition"*, CVPR, 2003.

[8] P. Koniusz and K. Mikolajczyk, *"On a quest for image descriptors based on unsupervised segmentation maps"*, International Conference on Pattern Recognition, 2010.

[9] J. Sivic, B. C. Russell, A. A. Efros, A. Zisserman, and W. T. Freeman, *"Discovering object categories in image collections*", CVPR, 2005.

[10] Ursani, K. Kpalma, and J. Ronsin, *"Texture features based on Fourier transform and Gabor filters: an empirical comparison"*, International Conference on Machine Vision, 2007.

[11] M. Heikkila, M.Pietikainen, and C. Schmid, *"Description of interest regions with local binary patterns"*, Pattern Recognition, vol. 42, pp. 425-436, 2009.

[12] Zhou, J.-F. Feng, and Q.-Y. Shi, *"Texture feature based on local Fourier transform"*, International Conference on Image Processing, 2001.

[13] Ursani, K. Kpalma, and J. Ronsin, *"Texture features based on local Fourier histogram: self-compensation against rotation"*, Journal of Electronic Imaging, 2008.

[14] T. Ahonen, J. Matas, C. He, and M. Pietikainen, *"Rotation invariant image description with local binary pattern histogram Fourier features"*, SCIA '09 Proceedings of the 16th Scandinavian Conference on Image Analysis, 2009.

[15] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, *"Surf: Speeded up robust features"*, Computer Vision and Image Understanding, vol. 110, pp. 346-359, 2008.

[16] S. Belongie, J. Malik, and J. Puzicha, *"Shape matching and object recognition using shape contexts"*, IEEE Transactions on PAMI, vol. 24, pp. 509-521, 2002.

[17] N. Dalal and B. Triggs, *"Histograms of oriented gradients for human detection"*, CVPR, 2005.

[18] X. Tan and B. Triggs, *"Enhanced local texture feature sets for face recognition under difficult lighting conditions"*, IEEE Transactions on Image Processing, vol. 19, pp. 1635-1650, 2010.

[19] H. Harzallah, F. Jurie, and C. Schmid, *"Combining efficient object localization and image classification"*, ICCV, 2009.

[20] R. Fergus, L. Fei-Fei, P. Perona, and A. Zisserman, *"Learning object categories from Google's image search"*, ICCV, 2005.

[21] Nister and H. Stewenius, *"Scalable recognition with a vocabulary tree"*, CVPR, 2006.

[22] Available at *http://www.vision.caltech.edu/html-files/archive.html.*

[23] Available at *http://www.cs.columbia.edu/CAVE/software/softlib/coil-100.php*