

Support Vector Machine with Mixture of Kernels for Image Classification

Dongping Tian^{1,2}, Xiaofei Zhao¹, and Zhongzhi Shi¹

¹ Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100190, China

² Graduate University of the Chinese Academy of Sciences, Beijing, 100049, China
{tiandp, zhaoxf, shizz}@ics.ict.ac.cn

Abstract. Image classification is a challenging problem in computer vision. Its performance heavily depends on image features extracted and classifiers to be constructed. In this paper, we present a new support vector machine with mixture of kernels (SVM-MK) for image classification. On the one hand, the combined global and local block-based image features are extracted in order to reflect the intrinsic content of images as complete as possible. SVM-MK, on the other hand, is constructed to shoot for better classification performance. Experimental results on the Berg dataset show that the proposed image feature representation method together with the constructed image classifier, SVM-MK, can achieve higher classification accuracy than conventional SVM with any single kernels as well as compare favorably with several state-of-the-art approaches.

Keywords. Image classification, SVM, Kernel function, PCA.

1 Introduction

Image classification is a challenging problem in computer vision. With the rapid explosion of images available from various multimedia platforms, effective technologies for organizing, searching and browsing these images are urgently required by common users. Fortunately, image classification can give much help to image indexing and retrieval. In recent years, many methods have been developed for image classification. As the representative work using SVM as classifiers, Chapelle et al. [1] use multiclass classifier framework to train 14 SVM classifiers for 14 image level concepts based on 4096 dimensional HSV histograms. Goh et al. [2] study several ensemble SVM binary classifiers, including one per class (OPC), pairwise coupling (PWC) and error-correction output coding (ECOC) through margin boosting and noise reduction methods to enhance the classification accuracy. Autio and Elomaa [3] focus on indoor image classification by using SVM as the classifier. Yang and Dong [4] propose ASVM-MIL to extend the conventional support vector machine through multiple-instance learning (MIL) so as to train the SVM in a new space. Recently, Anthony et al [5] employ SVM to handle the multiple classification tasks common in remote sensing studies for land cover mapping. And they find that

whereas the one-against-all technique is more predisposed to yielding unclassified and mixed pixels, the resulting classification accuracy is not significantly different from one-against-one approach. In addition, Gao and Lin [6] present CGSVM for semantic image retrieval, which utilizes clustering result to select the most informative image samples to be labeled and optimize the penalty coefficient. As a result, CGSVM can get higher search accuracy than conventional SVM for semantic image retrieval. In more recent years, Jiang, He and Guo [7] propose a novel method which adopts learning vector quantization (LVQ) technique to optimize low-level features extracted from given images, and then some representative vectors are selected with LVQ to train support vector machine classifier instead of using all feature data. Agrawal et al. [8] use SVM to implement image classification based on the feature of color image histograms, which benefits from the insensitivity of translation and rotation.

In general, to construct an image classifier includes two important stages: (1) features extracted from a set of training images are used to train the classifier. (2) features extracted from a set of testing images are fed into the built classifier so as to implement the image classification. However, previously related work only focuses on using either global or local (region) image features as the input to train the classifier. In literature [9], Tsai and Lin compare various combinations of feature representation methods including the global and local block-based and region-based features for image database categorization. Then the significant conclusion, i.e. the combined global and block-based feature representation performs best, is drawn in the end. Chow and Rahman [10] present an image classification approach through a tree-structured feature set, in which the image content is organized in a two-level tree, and the root node denotes the whole image features while the child nodes represent the local region-based features. Subsequently, the tree-structured image features are processed by a two-level self-organizing map (SOM) to implement the classification of images. In addition, Lu and Zhang [11] propose a block-based image feature representation in order to reflect the spatial features for a specific concept. For an image with $N \times N$ blocks, partition it by rows with blocks and connect into a block-line. The block lies at the more left side of the block-line represents upper and/or left position in the image. Shi et al. [12] present a square symmetrical local binary pattern texture descriptor, which is a compact symmetrical-invariant variation of local binary pattern, to capture the low-dimensional optimal discriminative features of images, etc. Most of these approaches can achieve state-of-the-art performance and motivate us to explore image classification with the help of their excellent experiences and knowledge. So in this paper, we present a new support vector machine with mixture of kernels (SVM-MK) for image classification. On the one hand, the combined global and local block-based image features are extracted so as to reflect the intrinsic content of images as complete as possible. SVM-MK, on the other hand, is constructed to shoot for better classification performance. Finally, the approach is experimentally tested on the Berg dataset and the precision we obtain is higher than conventional SVM with any single kernels as well as comparative to the state-of-the-art methods.

The rest of the paper is organized as follows. Section 2 elaborates the image classification method proposed in this paper, including the image feature

representation and the support vector machine with mixture of kernels (SVM-MK). Section 3 reports the experimental results on the Berg dataset. Finally, some important conclusions and future work are summarized in Section 4.

2 Classification Method

Motivated by the references aforementioned, global and local block-based image feature representation is proposed in this paper. Meanwhile, support vector machine with mixture of kernels (SVM-MK) is constructed as the image classifier. To our best knowledge, this is the first study to apply SVM with mixture of kernels in image classification. Details of them will be described in the following subsections, respectively.

2.1 Visual Feature Representation

(1) Global Feature Representation

To begin with, the image in RGB color space is transformed into HSV (hue, saturation and value) color space since it is more perceptually uniform than RGB. Without loss of generality, color histogram is used to represent the global feature of images in this paper. In details, the histogram of each channel is computed as follows: $h_G = n_G / n_T$, $G=1,2,\dots,q$, where G denotes a quantized level of an HSV color channel, n_G is the total number of pixels in that level, n_T is the total number of pixels, and q is the number of quantized levels. The complete histogram vector can be represented as $H_q = [h_{H1}, \dots, h_{Hq}, h_{S1}, \dots, h_{Sq}, h_{V1}, \dots, h_{Vq}]$.

(2) Local Feature Representation

The first-order, second-order and third-order color moments of image blocks are calculated for each channel as follows:

$$\mu_i = \frac{1}{N} \sum_{j=1}^N f_{ij} \quad (1)$$

$$\sigma_i = \left(\frac{1}{N} \sum_{j=1}^N (f_{ij} - \mu_i)^2 \right)^{\frac{1}{2}} \quad (2)$$

$$\gamma_i = \left(\frac{1}{N} \sum_{j=1}^N (f_{ij} - \mu_i)^3 \right)^{\frac{1}{3}} \quad (3)$$

where f_{ij} is the color value of the i -th color component of the j -th image pixel and N is the total number of pixels in the image. Hence, the color moments can be represented as $CM = [\mu_{c1}, \mu_{c2}, \mu_{c3}, \sigma_{c1}, \sigma_{c2}, \sigma_{c3}, \lambda_{c1}, \lambda_{c2}, \lambda_{c3}]$, where $\mu_{ci}, \sigma_{ci}, \lambda_{ci}$ ($i=1,2,3$)

denote the mean, variance and skewness of each channel of an image block, respectively.

Texture is a very useful characterization for a wide range of image. It is generally believed that human visual system uses textures for recognition and interpretation. The common methods for texture feature analysis include Gabor filter, wavelet transform and gray co-occurrence matrices, etc. Among which Gabor filter is the most commonly used method in extracting texture features. Given an image $I(x,y)$, Gabor wavelet transform $G(x,y)$ convolves $I(x,y)$ with a set of Gabor filters of different spatial frequencies and orientations. Assuming the wavelet transform is at u -th scale and v -th orientation, then means and standard deviations of $G_{uv}(x,y)$ can be calculated to represent the Gabor features:

$$\mu_{uv} = \frac{1}{PQ} \sum_x \sum_y |G_{uv}(x, y)| \tag{4}$$

$$\sigma_{uv} = \frac{1}{PQ} \sqrt{\sum_x \sum_y |G_{uv}(x, y)| - \mu_{uv}} \tag{5}$$

where $P \times Q$ is the size of image. The complete Gabor feature is then represented as $G_{uv}=[\mu_{00}, \sigma_{00}, \mu_{01}, \sigma_{01}, \dots, \mu_{0(v-1)}, \sigma_{0(v-1)}, \dots, \mu_{(u-1)(v-1)}, \sigma_{(u-1)(v-1)}]$.

In addition, shape feature is also considered, whose purpose is to encode simple geometrical forms such as straight lines in different directions. Since Sobel operator is insensitive to noise than other edge detectors. Here, the shape features can be extracted by the convolution of 3×3 masks with the image in 4 different directions (horizontal, 45° , vertical and 135°), and the corresponding features can be represented as $E=[e_0, e_{45}, e_{90}, e_{135}]$. Fig.1 shows several masks used here to compute four specific shapes.

Up to now, the approach for local feature extraction is presented above. The next step is how to partition the image and how to organize the block-based image features. The idea in reference [11] is adopted here, to start with, the image is

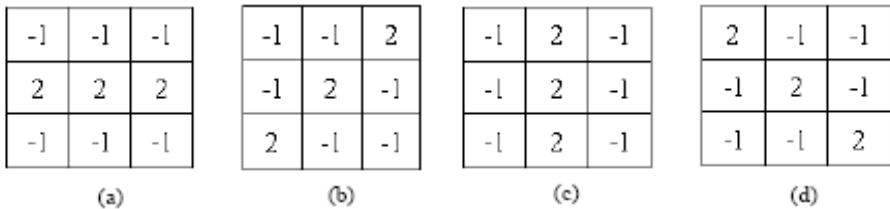


Fig. 1. Mask to detect shape features (a) Horizontal lines. (b) 45° slanted lines. (c) Vertical lines. (d) 135° slanted lines.

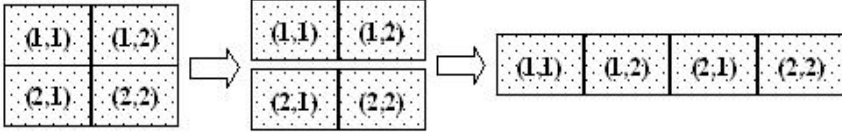


Fig. 2. Partition an image into 2*2 blocks and connect into a block-line

partitioned into regular grids (blocks) and then connect the blocks by left-right and top-down orders. In short, partition the image by rows with blocks and connect into a block-line. For the sake of simplicity, the image is partitioned into 2x2 blocks here, and the basic procedure is illustrated in Fig.2.

2.2 Kernel Function Constructing

Support vector machine (SVM) is a supervised classifier proposed by Vapnik which is based on statistical learning theory [13]. SVM works by mapping the training data into a high dimensional feature space. Then it separates the two classes of data with a hyperplane, and maximizes the distance which is called the margin. By introducing Kernels into the algorithm, it is possible to maximize the margin in the feature space, which is equivalent to nonlinear decision boundaries in the original input space. The algorithm comes with all the theoretical guarantees given by VC theory as well as the convergence properties studied in the statistical mechanics literature. In general case, assume that the labeled training examples $(x_1, y_1), \dots, (x_n, y_n)$, where each $x_k \in R^n$ is the k -th input sample and $y_k \in R$ is the k -th output pattern. Suppose that each example is assigned a binary label $y_k \in \{+1, -1\}$. In their simplest form, SVM could find out the hyperplanes that separate the training data by a maximal margin. All the vectors lying on one side of the hyperplane are labeled as +1, and all the vectors lying on the other side are labeled as -1. The training instances that lie closest to the hyperplane are called support vectors. As we known, the challenging problem in the case of SVM is the choice of the kernel function which is actually a measure of similarity between two vectors. Different choices of kernel functions have been proposed and extensively used in the last decades [14]. The most popular kernel functions (RBF & polynomial) are given as follows:

$$K(x_k, x)_{RBF} = \exp(-\rho \|x_k - x\|^2), \rho > 0 \quad (6)$$

$$K(x_k, x)_{Poly} = [(x_k * x) + c]^q \quad (7)$$

It is reported that kernels used by SVM can be divided into two classes: global and local kernels. In global kernels, points far away from the test point have a great effect on kernel values. While in local kernels, only those close to the test point have a great effect on kernel values. The RBF and polynomial kernel functions are two typical

local and global kernels, respectively. Therefore, a kind of SVM modeling method based on mixture of kernels can be constructed as follows:

$$K(x_k, x_l)_{Mix} = \lambda K(x_k, x_l)_{RBF} + (1 - \lambda) K(x_k, x_l)_{Poly} \quad (8)$$

where $K(x_k, x_l)_{RBF}$ denotes the RBF kernel function and $K(x_k, x_l)_{Poly}$ denotes polynomial kernel function, $\lambda(0 \leq \lambda \leq 1)$ is the mixed coefficient used to adjust the weight of the two kinds of kernel functions.

3 Experiments and Analysis

For the sake of comparison, we test the proposed image classification scheme on the ten-animal-class dataset of Berg available from [15]. Fig.3 shows some images of Berg dataset¹. Each image is represented with 668-dimensional visual features since the image representation adopted in this paper consists of global and local block-based segmentation features reduced by principal component analysis (PCA). SVM-MK is implemented by adapting the source code of Libsvm [16] so as to construct the mixture of polynomial kernel and RBF. To determine the optimal classification parameters C and λ , 10-fold cross validation is conducted on the training image dataset and the best parameters are used in testing. Here, the corresponding parameters are predefined as $C=1000$, $c=1$, $q=1$ and the best representative λ value is set to 0.05 by trial and error.



Fig. 3. Example images from Berg dataset

To show the effectiveness of the SVM-MK for image classification, we conduct two experiments in this paper. The first one is the image classification comparison between SVM-MK and the SVM with any single kernels (RBF & polynomial) under usual image feature representation and the image feature representation proposed in

¹ Download from <http://tamaraberg.com/animalDataset/index.html>

Table 1. Comparison among SVM-Poly., SVM-RBF and SVM-MK under different image feature representations

Classifier	SVM-Poly.	SVM-RBF	SVM-MK
Avg. Prec.(usual feature)	43.6	45.4	46.1
Avg. Prec.(presented feature)	46.8	48.5	50.3

this paper, respectively. The average classification precision over different kinds of images is used to measure the overall quality of SVM-MK.

From table 1, it easily can be seen that the performance of all classifiers with the image feature representation proposed in this paper is better than the corresponding ones with the usual image features. On the other hand, the classification accuracy of SVM-MK is higher than that of SVM-Poly. and SVM-RBF under the same image feature representation, respectively. It is also noticeable that SVM-MK with the proposed image feature representation performs the best, which further demonstrates that the mixture of global and local kernel functions as well as the block-line image feature representation plays a very important role in SVM classifiers. In addition, we also make a comparison between SVM-MK and other state-of-the-art approaches. Here we create two SVM classifiers (SVM-EQ, SVM-TW) with different settings for five categories. SVM-EQ is the SVM created using the same positive and negative data as in SVM-MK; SVM-TW is the SVM created using the ground truth data twice the number of the labeled data in SVM-MK. According to table 2, SVM-MK performs significantly better than the SVM-EQ model. Particularly, the precision of “ant” has been increased from 30 to 56.3. But it is partially outperformed by SVM-TW for the classification of “alligator”, “bear” and “leopard”.

Table 2. Comparison among SVMs: Precision at top 100 images

	alligator	ant	bear	leopard	penguin
SVM-EQ	58.0	30.0	38.0	52.0	48.0
SVM-TW	77.0	52.0	61.0	67.0	62.0
SVM-MK	73.5	56.3	55.0	61.8	63.6

To further illustrate the effect of SVM-MK proposed in this paper for image classification, examples for the precision of the top 100 images on the Berg dataset by Berg [15] and Schroff [17], together with ours are depicted in Fig.4. From Fig.4, we can clearly see that our approach gives superior precision to both Berg and Schroff over five categories and outperforms [15] on eight categories except “monkey” and “ant”. In summary, the performance of our approach is holistically superior to other methods.

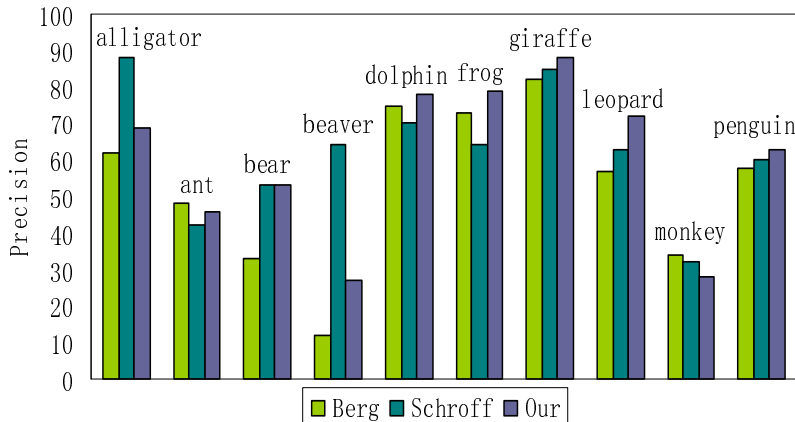


Fig. 4. Comparison with Berg and Schroff

4 Conclusions and Future Work

In this paper, a novel support vector machine with mixture of kernels (SVM-MK) for image classification has been proposed. On the one hand, the combined global and local block-based image features are extracted in order to reflect the intrinsic content of images. SVM-MK, on the other hand, is constructed in this paper as classifiers. Experimental results on Berg dataset show that the proposed image feature representation method together with the constructed image classifier, SVM-MK, can achieve higher classification accuracy than conventional SVM with any single kernels as well as comparable performance with several state-of-the-art approaches.

For the future work, we plan to introduce semi-supervised learning into our approach under the conditions that there are only a few labeled but a large amount of unlabeled images to implement image classification. In addition, we intend to apply other more complicated image datasets to further evaluate the performance of SVM-MK proposed in this paper comprehensively.

Acknowledgements. This work is supported by the National Natural Science Foundation of China (No.61035003, No. 60933004, No. 60903141, No.61072085), the National Program on Key Basic Research Project (973 Program) (No.2007CB311004), the National High-tech R&D Program of China (863 Program) (No.2012AA011003) and the National Science and Technology Support Program (2012BA107B02).

References

1. Chapelle, O., Haffner, P., Vapnik, V.: Support Vector Machines for Histogram-Based Image Classification. *IEEE Transactions on Neural Networks* 10(5), 1055–1064 (1999)
2. Goh, K.S., Chang, E., Cheng, K.T.: SVM Binary Classifier Ensembles for Image Classification. In: *CIKM 2001*, pp. 395–402 (2001)
3. Autio, I., Elomaa, T.: Flexible view recognition for indoor navigation based on Gabor filters and support vector machines. *Pattern Recognition* 36(12), 2769–2779 (2003)
4. Yang, C.B., Dong, M.: Region-based image annotation using asymmetrical support vector machine-based multiple instance learning. In: *CVPR 2006*, pp. 2057–2063 (2006)
5. Anthony, G., Gregg, H., Tshilidzi, M.: Image Classification Using SVMs: one-against- one Vs one-against-all. In: *ACRS (2007)*
6. Gao, K., Lin, S.X., Zhang, Y.D., et al.: Clustering guided SVM for semantic image retrieval. In: *ICPCA 2007*, pp. 199–203 (2007)
7. Jiang, Z.H., He, J., Guo, P.: Feature data optimization with LVQ technique in semantic image annotation. In: *ISDA 2010*, pp. 906–911 (2010)
8. Agrawal, S., Verma, N.K., Tamrakar, P., et al.: Content Based Color Image Classification using SVM. In: *ITNG 2011*, pp. 1090–1094 (2011)
9. Tsai, C.F., Lin, W.C.: A Comparative Study of Global and Local Feature Representations in Image Database Categorization. In: *NCM 2009*, pp. 1563–1566 (2009)
10. Chow, T.W.S., Rahman, M.K.M.: A new image classification technique using tree-structured regional features. *Neurocomputing* 70(4-6), 1040–1050 (2007)
11. Lu, H., Zheng, Y.B., Xue, X.Y., Zhang, Y.J.: Content and Context-Based Multi-Label Image Annotation. In: *CVPRW 2009*, pp. 61–68 (2009)
12. Shi, Z.P., Liu, X., Li, Q.Y., He, Q., Shi, Z.Z.: Extracting discriminative features for CBIR. *Multimedia Tools and Applications* (2011)
13. Vapnik, V.N.: *The Nature of Statistical Learning Theory*. Springer, New York (1995)
14. Zhu, Y.F., Tian, L.F., Mao, Z.Y., Wei, L.: *Mixtures of Kernels for SVM Modeling*. Springer, Heidelberg (2005)
15. Berg, T.L., Forsyth, D.A.: Animals on the web. In: *CVPR 2006*, pp. 1463–1470 (2006)
16. Chang, C.C., Lin, C.J.: *Libsvm: a library for support vector machines* (2001)
17. Schroff, F., Criminisi, A., Zisserman, A.: Harvesting image databases from the web. In: *ICCV 2007*, pp. 1–8 (2007)