Signal Processing 93 (2013) 2111–2118

Contents lists available at SciVerse ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Image classification using Harr-like transformation of local features with coding residuals

Chunjie Zhang [a], Jing Liu [b,*], Chao Liang [c], Qingming Huang [a], Qi Tian [d]

[a] School of Computer and Control, University of Chinese Academy of Sciences, 100049 Beijing, China
[b] National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, P.O. Box 2728, Beijing, China
[c] National Engineering Research Center for Multimedia Software, Wuhan University, 430072 Wuhan, China
[d] Department of Computer Sciences, University of Texas at San Antonio, Tx 78249, USA

ARTICLE INFO

Article history:
Received 16 November 2011
Received in revised form
5 September 2012
Accepted 9 September 2012
Available online 18 September 2012

Keywords:
Image classification
Harr-like transformation
Coding residuals

ABSTRACT

Recently, the bag-of-visual-words (BoW) model has been proven very effective for image classification. However, most researchers used local features directly while neglecting their spatial information and correlations. Besides, the encoding of local features causes some information loss which also hinders the final image classification performance. To tackle these problems, in this paper, we proposed a novel image classification method using Harr-like transformation of local features with additional consideration of coding residuals. We apply Harr-like transformation on local features to combine the spatial information as well as the correlations of local features. These Harr-like transformed local features are then encoded using non-negative sparse coding. We jointly consider the coding parameters and the coding residuals as the local representation in order to reduce the information loss during the local feature encoding process. Experiments on several public datasets demonstrate the effectiveness of the proposed method.

1. Introduction

In recent years, the bag-of-visual-words (BoW) model becomes popular and has been proven very effective for image classification. Inspired by the bag-of-words approach to text categorization [1], the BoW model first extracts local features from image patches and quantizes them into "visual words". Each image is then represented by the histogram of visual words occurrences. Finally, SVM classifiers are trained to predict the categories of images.

Although proven very effective for visual applications, the BoW model still has two drawbacks. First, it ignores the spatial information as well as correlations among local features. Second, the *k*-means clustering based codebook generation and nearest neighbor assignment result in information loss which hinders the final performance. To attack these problems, researchers have proposed various models which greatly improve the image classification performance.

To combine the spatial information and correlations of local features, a lot of studies [2–9] have been done. Local features are firstly quantized into visual words. The quantization of local features helps to reduce computational cost and makes it possible to represent images using histograms which can then be combined with the state-of-the-art models for image classification. However, some useful information is lost during local feature encoding process which may causes it unable to handle with subtle differences, especially when we are dealing with large-scale images. To distinguish different classes of similar images, it would be more effective to directly

* Corresponding author.
E-mail addresses: cjzhang@jdl.ac.cn (C. Zhang), jliu@nlpr.ia.ac.cn (J. Liu), liangchao827@gmail.com (C. Liang), qmhuang@jdl.ac.cn (Q. Huang), qitian@cs.utsa.edu (Q. Tian).

0165-1684/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.sigpro.2012.09.007

model the spatial information and correlations at the local feature level instead of using quantized visual words.

In order to reduce the quantization loss during nearest neighbor assignment process, the soft-assignment technique is widely used [10–18]. These soft assignment strategies preserve more information than nearest neighbor assignment, hence can improve the image classification performance. However, these soft-assignment based methods also suffer from heavy computation for large-scale applications, especially when the number of local features is large which has been proven effective for improving image classification performance. Another way to reduce information loss during the local feature encoding process is to use sparse coding [14–18] which can be viewed as a special case of soft assignment. Instead of using only one visual word to encode each local feature, the sparse coding method tries to reconstruct local features using all the visual words with sparsity constraints on the coding parameters. This helps to reduce the information loss. Max pooling is then used to get the feature representation of images. However, there is still some information loss during the sparse coding process. The performance can be further improved when we take these lost information into consideration.

In this paper, we propose a novel image classification method using Harr-like transformation of local features with coding residuals. We first extract local features from images and do Harr-like transform to these local features in order to combine the spatial as well as the correlations of local features. Non-negative sparse coding with residuals are then used to encode local features and we use max pooling for final image representation. Classifiers are then trained to predict the categories of images. Fig. 1 shows the flowchart of the proposed method.

The proposed framework consists of three contributions.

1. First,we model the spatial information and correlation of local features by doing Harr-like transformation on local features. These Harr-like transformed local features are more discriminative and robust than modeling at the visual word level.
2. Second, we encode the Harr-like transformed local features by non-negative sparse coding and use the coding parameters as well as the coding residuals for image representation, in order to reduce the information loss during local feature encoding process.
3. Third, the proposed method can be efficiently computed and easily extended to adapt to the task of large-scale image classification.

The rest of the paper is organized as follows. We give the related work in Section 2. Section 3 presents the proposed Harr-like transformation of local features. Details of non-negative sparse coding with residuals for image representation are given in Section 4. Experimental results are given in Section 5. Finally we conclude in Section 6.

2. Related work

The use of the bag-of-visual words (BoW) model [1] has been proven very useful and effective for image classification. There are mainly three parts of the BoW model: the extraction of local features (SIFT feature or its variants are often extracted), the encoding or quantization of local features and BoW representation of images, the classification model training for image class prediction. The BoW model is able to cope with some scale and rotation variations both for the use of local features and the histogram representation of images. However, the ignorance of spatial and correlations as well as the quantization loss of local features hinder the final performance. Over the past few years, many works have been done to improve the performance of the BoW model.

To take advantages of local feature's spatial information, motivated by the pyramid matching in feature space proposed by Grauman and Darrell [2], Lazebnik et al. [3] proposed the spatial pyramid matching (SPM) method and was widely used since its introduction. Wu et al. [4] buddled features together for web image near-duplicate detection. In order to combine the contextual information, Feature context [5] is also proposed for better image classification. Zhang et al. [6] proposed using components to incorporate spatial contextual information into histogram based image representation. Yao et al. [7] used mutual context to model objects and human poses for action classification. Lee and Grauman [8] constructed object-graphs to automatically discover object categories by modeling object relationships with graphs. Belongie et al. [9] proposed a semi local shape descriptor, called Shape Context. The Shape Context represents a binary shape as a discrete set of points sampled from its contour. These points are then mapped into a log-polar coordinate system centered at a reference point. Each bin of the log-polar space is determined by the distance and angle intervals.

To reduce the information quantization loss during traditional nearest neighbor based assignment strategy, Gemert et al. [10] proposed a kernel codebook method which tried to encode local features in a kernel space by soft assignment. Each local feature is encoded with all the visual words and the weights of visual words are based on their distances with this local feature. Wang et al. [11] proposed a radial basis coding method. A local feature is encoded according to its activations of neurons placed at the cluster centers which helps to reduce the encoding information loss. Huang et al. [12] used a salient coding method which gives strong response to visual words that are closer to a local feature

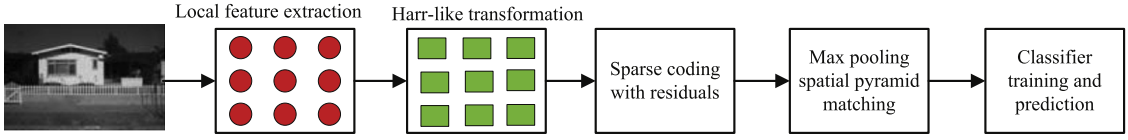

Fig. 1. Flowchart of the proposed image classification method using Harr-like transformation of local features.

than other visual words hence improved image classification performance. Liu et al. [13] studied the mechanism of soft assignment coding and interpreted soft assignment in a probabilistic way.

Inspired by a biological model [14] for object recognition, Yang et al. [15] proposed sparse coding and max pooling to represent images and achieved the state-of-the-art performance for image classification. Spatial pyramid matching is then used to combine the spatial information of local features. Motivated by [15], Wang et al. [16] found that locality is more important than sparsity and proposed the locality-constrained linear coding method (LLC) to speed up computation and improve the sparse coding quality. To ensure coding consistency, Gao et al. [17] proposed a Laplacian sparse coding algorithm by adding a Laplacian constraint term into the sparse coding process and solved it in a simplified way. Since the information of negative sparse coding parameters was lost during max pooling process, Zhang et al. [18] proposed to use non-negative sparse coding in order to alleviate the information loss of sparse coding with max pooling strategy. Cheng et al. [19] proposed a non-negative class-specific entropy component analysis method with adaptive step search.

3. Harr-like transformation of local features

The use of Harr-like features has been proven very effective and efficient for visual applications [6,20], especially for face recognition [20]. This simple transformation of image pixels makes it very robust and effective than directly using image pixels. A simple rectangular Harr-like feature can be calculated as the difference of the sum of pixels of areas inside the rectangle. This modified feature set is called 2-rectangle feature. The 3-rectangle features and 4-rectangle features can be defined in a similar way. These values indicate certain characteristics of a particular area of image and may indicate the existence of certain characteristics in the image. For example, a 2-rectangle feature can indicate where the border of a dark region and a light region lies. This Harr-like transformation is able to combine the spatial information and correlations of near-by pixels. Besides, it is also very easy to compute. If we can transform local features in a similar way, we will be able to improve the performance of image classification.

We propose to use Harr-like transformation of local features to combine the spatial information as well as the correlations among local features. We choose to use SIFT feature or its variants [21] as the local feature representation in this paper. Note that other local features such as histograms of oriented gradients (HOG) [22] can also be used. We use SIFT feature or its variants both for their good performance and for fair comparison with other methods. The Harr-like transformation of local features is defined as the difference of the sum of the local features inside the rectangle. In this paper, we view each local feature as a rectangle and apply Harr-like transformation on these local features. The 14 types of Harr-like transformation are shown in Fig. 2. We use the Harr-like transformation in a general form. Besides using the difference of local features, we also apply sum and concatenation to these local features. Note that more types of transformation can also be used to further improve the image classification performance. For each dimension of local features within the same image, the "Integral image" technique in [20] can be used to speed up computation.

4. Image classification using non-negative sparse coding with residuals

After extracting the Harr-like transformed local features, we can encode these transformed local features to

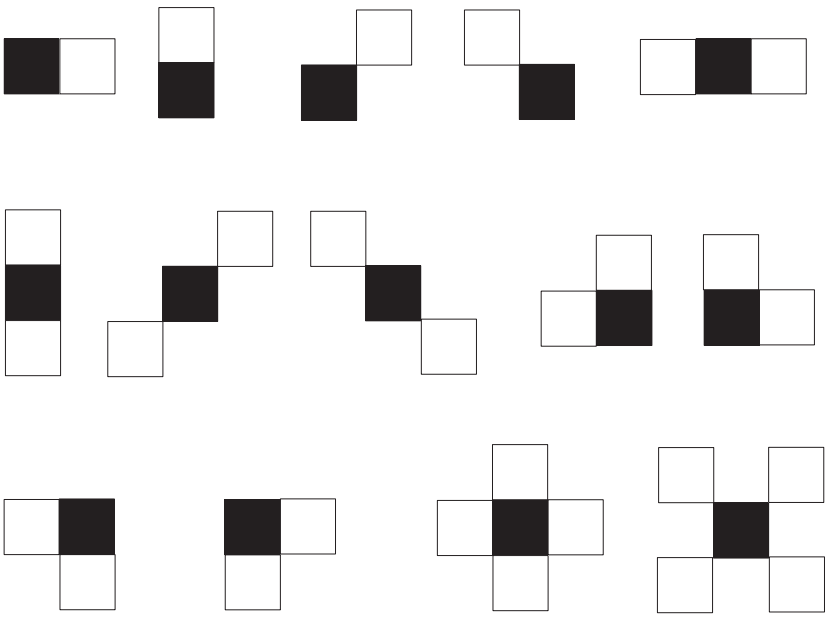

Fig. 2. The Harr-like transformations for local features.

represent images. Sivic and Zisserman [1] proposed to use the k-means clustering algorithm and viewed the cluster centers as visual words. Each local feature is then assigned to the nearest visual word by calculating the Euclidean distances between local features with all the visual words.

Formally, let $X=[x_1,\ldots,x_N]\in\mathbb{R}^{D\times N}$ be the set of D-dimensional local features with the number of local features is N, where $x_i\in\mathbb{R}^{D\times 1}$, $i=1,\ldots,N$. The codebook B has M visual words, where $B=[b_1,\ldots,b_M]\in\mathbb{R}^{D\times M}$. The k-means clustering in [1] tries to find the optimal codebook B and assignment C by solving the following problem as

$$[B,C]=\arg\min_{B,C}\sum_{i=1}^{N}\|x_i-Bc_i\|^2$$
$$\text{s.t.}\quad \|c_i\|_{\ell^0}=1,\quad \|c_i\|_{\ell^1}=1,\quad c_{ij}\geq 0,\ \forall i,j \tag{1}$$

where $C=[c_1,\ldots,c_N]$ is the codes for X and $c_j\in\mathbb{R}^{M\times 1}$, $j=1,\ldots,M$ is the jth element of C. This means each local feature can only be assigned to one visual word. The constraint in problem (1) is too restrictive and bound to bring information loss. To alleviate this problem, Yang et al. [15] proposed a sparse coding solution, in which the optimization problem is defined as:

$$[B,C]=\arg\min_{B,C}\sum_{i=1}^{N}\|x_i-Bc_i\|^2$$
$$\text{s.t.}\quad \|c_i\|<\alpha,\quad \forall c_i \tag{2}$$

where α is the regularization parameter which controls the sparsity of C. Each local feature is then encoded by solving problem (2) with codebook B fixed. Compared with nearest neighbor assignment, the sparse coding can preserve more information which benefits to image classification. However, the sparse coding strategy is sub-optimal when used along with max pooling. To alleviate this problem, Zhang et al. [18] proposed to use non-negative sparse coding along with max pooling instead. This is achieved by forcing the coding parameters C in problem (2) to be non-negative. Besides, we also use locality constraints to speed up computation as Wang et al. [16] did and try to solve the optimization problem as:

$$[B,C]=\arg\min_{B,C}\sum_{i=1}^{N}\|x_i-Bc_i\|^2+\alpha\|d_i\odot c_i\|^2$$
$$\text{s.t.}\quad \|c_i\|_1<\alpha,\quad c_i\geq 0,\quad 1^Tc_i=1,\ \forall c_i \tag{3}$$

where $\odot$ is the element-wise multiplication. And $d_i=\exp(dist(h_i,B)/\sigma)$ with $dist(h_i,B)=[dist(h_i,b_i),\ldots,dist(h_i,b_M)]$.

However, there is still some information loss in non-negative sparse coding which can be used to improve image classification performance. The encoding loss of non-negative sparse coding for local feature x_i is $\text{loss}_i=\|x_i-B\times c_i\|$, where $\|\cdot\|$ denotes the Euclidean distance. In this paper, we represent the encoding loss in another way and try to measure the encoding loss of using one visual word each time for the non-negative sparse coding. Formally, let $e_{i,j}=\|x_i-B\times c_{i,j}\|_2$, where $c_{i,j}\in\mathbb{R}^{M\times 1}$. The jth element of $c_{i,j}$ is the same as the j-th element of c_i while all the other elements of $c_{i,j}$ are set to zero, $j=1,2,\ldots,M$. Let $e_i=[e_{i,1},\ldots,e_{i,M}]\in\mathbb{R}^{M\times 1}$, then

$$\text{loss}_i=\sqrt{\sum_{j=1}^{M}e_{i,j}^2} \tag{4}$$

To represent local feature x_i, we use the non-negative sparse coding parameter c_i and the coding residuals $\overrightarrow{e_i}$ jointly by concatenating the two vectors to form the final local feature encoding vector $\overrightarrow{c_i}\in\mathbb{R}^{2M\times 1}$ as

$$\overrightarrow{c_i}=[c_i;e_i] \tag{5}$$

The proposed non-negative sparse coding with residuals image representation combines the encoding information as well as the information loss during non-negative sparse coding, hence can preserve more information for classification tasks. Max pooling with spatial pyramid matching is then used to extract information for image representation. We can then train classifiers to predict the categories of images. The multiple kernel learning technique [23] is used to combine the discriminative power of these Harr-like transformed local features with different feature types.

Table 1
Performance comparison on the Oxford Flower 17 dataset.

Methods	Classification rate
Nilsback and Zisserman [24]	71.76 ± 1.76
Varma and Ray [27]	82.55 ± 0.34
χ^2 [28]	87.45 ± 1.13
LP-B [29]	85.40 ± 2.40
KMTJSRC-CG [30]	88.90 ± 2.30
Harr-SIFT	$\mathbf{90.15\pm 1.39}$
Harr-SIFT-SC$^+$R	$\mathbf{91.87\pm 1.43}$

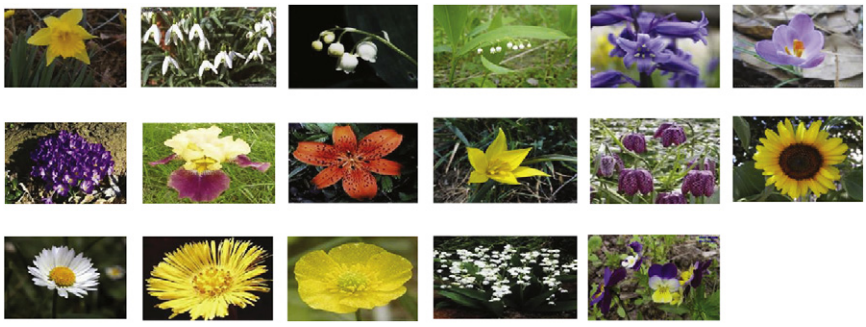

Fig. 3. Example images for the Oxford Flower 17 dataset.

5. Experiments

We evaluate the proposed Harr-like transformed local features with coding residuals for image classification (abbreviated as Harr-SIFT-SC^{+}R) on three public datasets: the Oxford Flower 17 dataset [24], the Oxford Flower 102 dataset [25] and the Caltech-256 dataset [26].We use two types of local patch selection method: Harris–Laplace and dense sampling. For the Oxford Flower datasets, as color plays an important role for flower classification, we use the color SIFT proposed by Sande et al. [21] to represent local patches. For the Caltech-256 dataset, we process images in gray scale and extract SIFT features for local patch representation. The pyramids of $2^l \times 2^l$ with $l = 0,1,2$ are used to take advantage of local feature's spatial information. The codebook size is set to 1000 for the three datasets. The main computational cost lies in the BoW representation of the transformed Harr-SIFT features. Since we used fourteen types of transformation as in Fig. 2, the computational cost is fourteen times of the traditional BoW model. However, since the fourteen types of Harr-like transformation are independent with each other, the computational time can be saved by calculating in parallel. In our experiments, it takes about 10 h to construct the codebook and encode these transformed local features using a HP Z800 Xeon workstation with twelve 2.13 GHz cores and 48 GB memory.

5.1. Oxford Flower 17 dataset

The 17 category dataset has 1360 images with 17 classes of flowers (buttercup, colts' foot, daffodil, daisy, dandelion, fritillary, iris, pansy, sunflower, windflower, snowdrop

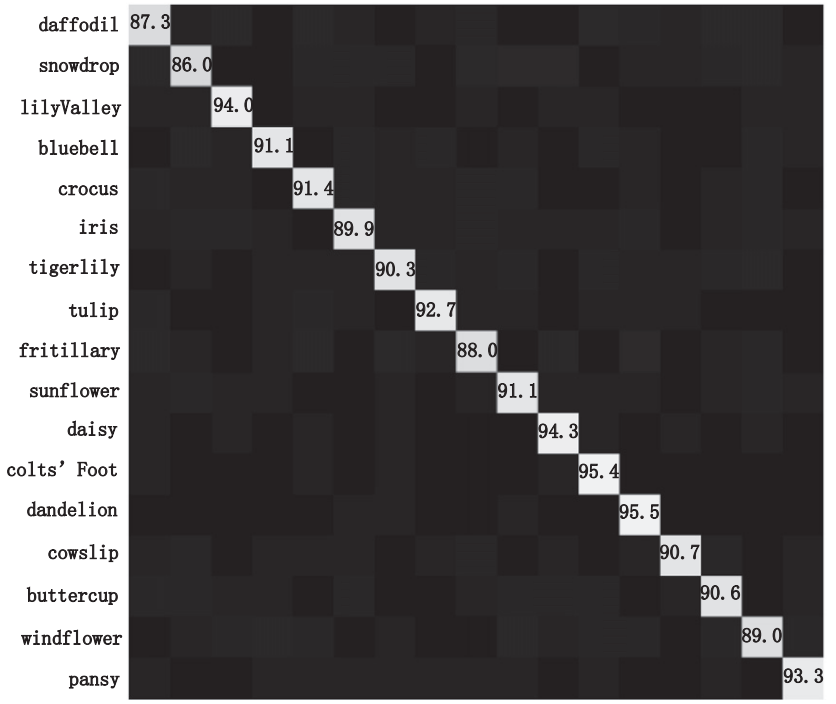

Fig. 4. Confusion matrix for the Oxford Flower 17 dataset.

Fig. 5. Example images for the Oxford Flower 102 dataset.

lilyvalley, bluebell, crocus, tigerlily, tulip and cowslip). Each class has 80 images. For each class, 40 and 20 images are used for training and validation with the rest images are used for testing. We use the three splits provided by Nilsback and Zisserman [24] for fair comparison. Fig. 3 shows some example images of the Oxford Flower 17 dataset.

Table 1 gives the performance comparison on the Oxford Flower 17 dataset. Nilsback and Zisserman [24] tried to learn a proper visual vocabulary for classification while Varma and Ray [27] tried to learn the most discriminative features. Xie et al. [28] used the bin-ratio information for image distance calculation. Gehler and Nowozin [29] combined different types of features using the boosting procedure. Yuan and Yan [30] used a multi-task joint sparse representation method to combine different types of features for flower classification. We can see from Table 1 that the proposed method achieves good performance which is a clear demonstration of the effectiveness of the proposed method. By doing Harr-like transformation on SIFT features or its variants, we can make better use of local features hence improve the classification rate. Besides, by representing images with sparse coding parameters as well as coding residuals, we can preserve more information which further improves the image classification performance.

For detailed comparison, we also give the confusion matrix of the proposed method on the Oxford Flower 17 dataset in Fig. 4. There are flower species that have very similar appearance like dandelions and colts' feet, or similar shapes such as buttercup, daffodil and the wind-flower. This cannot be distinguished effectively by only using local features [24]. It is more effective to consider the correlations of local features. The proposed Harr-like transform of local features takes this advantage by make Harr-like transformation on nearby local features hence is able to cope with this problem more effectively.

Table 2
Performance comparison on the Oxford Flower 102 dataset.

Methods	Classification rate
Nilsback and Zisserman [25]	72.8
KMTJSRC-CG [30]	74.1
Harr-SIFT	**75.6**
Harr-SIFT-SC^{+}R	**76.9**

5.2. Oxford Flower 102 dataset

The Oxford Flower 102 dataset has 8189 images of 102 classes of flowers. There are 40–250 images per class. This dataset is more challenging than the Oxford Flower 17 dataset with more images and classes. We follow the same experimental setup as in [25,29] and use 10 images per class for training, 10 images per class for validation and the rest of images for testing. Fig. 5 gives some example images of the Oxford Flower 102 dataset.

We give the performance comparison in Table 2. We can have similar conclusions as on the Oxford Flower 17 dataset from Table 2. The proposed method makes use of the spatial

Fig. 6. Example image classes with good (red: passion flower, californian poppy and wallflower) and bad (green: anthurium, petunia and alpine sea holly) performances of the proposed method. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

relationship and correlations among local features as well as the coding information loss, hence is more discriminative than other methods which either viewed local features individually or only used the encoding parameters without considering the coding residuals. On analyzing the per class classification rate, the proposed Harr-SIFT-SC$^+$R performs relatively poor on flower classes with larger inter class variation (*e.g.* ball moss and red ginger) than flower species (*e.g.* geranium and gazania) with smaller inter class variation. This is consistent with [25] that image classes with relatively larger inter class variations are more difficult that image classes smaller inter class variations. Fig. 6 shows some example flower classes with good and bad performances of the proposed method, respectively. Besides, as the number of flower classes increases, it is relatively more difficult to separate flower species with small intra class variation (*e.g.* sunflower and dandelion) compared with the Oxford Flower 17 dataset. This is because it is relatively difficult to learn a SVM classifier with proper separating plane for a larger dataset.

5.3. Caltech-256 dataset

The Caltech-256 dataset has 256 categories of 29,780 images with high intra-class variability and object location variability. There are at least 80 images in each class of the Caltech-256 dataset. We randomly choose 15, 30 images per class to train classifiers and use the rest for testing, as [3,10] did.

Table 3
Performance comparison on the Caltech-256 dataset.

Methods	15 training	30 training
KCSPM [10]	–	27.17 ± 0.46
SPM [3]	–	34.10
SPM [15]	23.34 ± 0.42	29.51 ± 0.52
SCSPM [15]	27.73 ± 0.51	34.02 ± 0.35
LLC [16]	34.36	41.19
Harr-SIFT	$\mathbf{36.87 \pm 0.53}$	$\mathbf{42.79 \pm 0.41}$
Harr-SIFT-SC$^+$R	$\mathbf{38.15 \pm 0.47}$	$\mathbf{43.93 \pm 0.54}$

Table 3 shows the performance comparison of the proposed Harr-SIFT-SC$^+$R with methods in [3,10,15,16]. Gemert et al. [10] used kernel codebook for soft assignment of local features and considered the spatial information of local features with spatial pyramid matching. Yang et al. [15] re-implemented the SPM algorithm for fair comparison with the sparse coding spatial pyramid matching method. Wang et al. [16] used locality constraints for efficient sparse coding and speeding up computation. We can see from Table 3 that the proposed method outperforms LLC which used sparse coding with locality constraints. Compared with nearest neighbor assignment based local feature quantization [3], our method can reduce the information loss during local feature encoding process. Besides, the use of Harr-like transformation of local features is more discriminative and robust than using local features individually [10,15,16]. We also give the per class classification rates of the proposed Harr-SIFT-SC$^+$R method in Fig. 7 on the Caltech-256 dataset with 30 training images per class in descending order. We can have similar conclusions as on the Oxford Flower datasets that the proposed method performs well on image classes with small inter class variation and classes that can be distinguished by local feature combination. This also demonstrates the efficiency of the proposed method.

6. Conclusions

This paper proposed a novel image representation for classification using Harr-like transformation of local features with coding residuals. The Harr-like transformed local features can take the spatial information as well as correlations among local features into consideration, hence is more discriminative and robust than using local features individually. Besides, to reduce the information loss during local feature encoding process, we proposed to use the non-negative sparse coding parameters and residuals to jointly represent images. Experimental results on the Oxford Flower 17 dataset, the Oxford Flower 102 dataset and the Caltech-256 dataset demonstrate the effectiveness of the proposed method. Our future work will concentrate on how to encode local features more

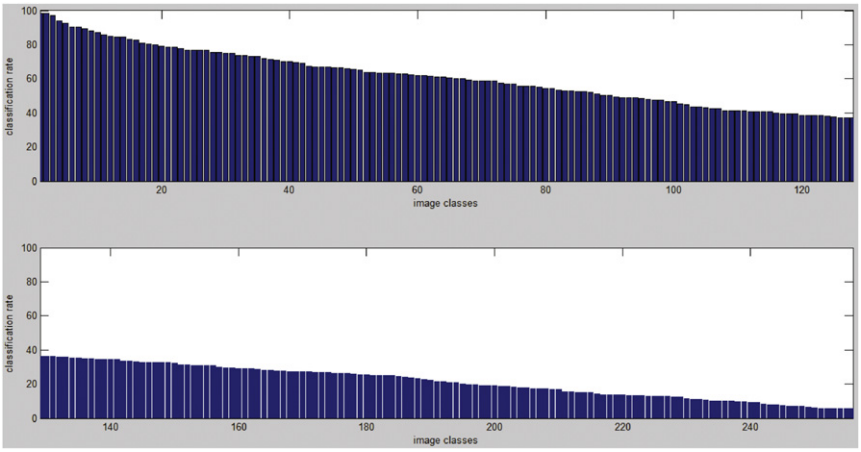

Fig. 7. Per class classification rate (descending order) of the proposed method on the Caltech-256 dataset (30 training images).

effectively and extend to the task of large-scale image classification. Besides, how to separate images with small intra class variations will also be studied.

Acknowledgments

This work is supported in part by National Basic Research Program of China (973 Program): 2012CB316400 and National Natural Science Foundation of China: 61025011 and 60833006.

References

[1] J. Sivic, A. Zisserman, Video Google: a text retrieval approach to object matching in videos, in: Proceedings of the International Conference on Computer Vision, Nice, France, 14–17 October, 2003, pp. 1470–1477.
[2] K. Grauman, T. Darrell, The pyramid match kernel: discriminative classification with sets of image features, in: Proceedings of the International Conference on Computer Vision, 2005.
[3] S. Lazebnik, C. Schmid, J. Ponce, Beyond bags of features: spatial pyramid matching for recognizing natural scene categories, in: Proceedings of the Computer Vision and Pattern Recognition, New York, USA, 17–22 June, 2006, pp. 2169–2178.
[4] Z. Wu, Q. Ke, J. Sun, Bundling features for large scale partial-duplicate web image search, in: Proceedings of the Computer Vision and Pattern Recognition, 2009.
[5] X. Wang, X. Bai, W. Liu, L. Latecki, Feature context for image classification and object detection, in: Proceedings of the Computer Vision and Pattern Recognition, 2011.
[6] C. Zhang, J. Liu, Q. Tian, Y. Han, H. Lu, S. Ma, A boosting sparsity-constrained bilinear model for object recognition, IEEE Multimedia 19 (2) (2012) 58–68.
[7] B. Yao, A. Khosla, L. Fei-Fei, Classifying actions and measuring action similarity by modeling the mutual context of objects and human poses, in: Proceedings of the International Conference on Machine Learning, 2009.
[8] Y. Lee, K. Grauman, Object-graphs for context-aware category discovery, in: Proceedings of the Computer Vision and Pattern Recognition, 2010.
[9] S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape contexts, in: IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002.
[10] J. Gemert, C. Veenman, A. Smedulders, J. Geusebroek, Visual word ambiguity, in: IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010.
[11] X. Wang, X. Bai, W. Liu, L. Latecki, Feature context for image classification and object detection, in: Proceedings of the Computer Vision and Pattern Recognition, 2011.
[12] Y. Huang, K. Huang, Y. Yu, T. Tan, Salient coding for image classification, in: Proceedings of the Computer Vision and Pattern Recognition, 2011.
[13] L. Liu, L. Wang, X. Liu, In defense of soft-assignment coding, in: Proceedings of the Computer Vision and Pattern Recognition, 2011.
[14] T. Serre, L. Wolf, T. Poggio, Object recognition with features inspired by visual cortex, in: Proceedings of the Computer Vision and Pattern Recognition, 2005.
[15] J. Yang, K. Yu, Y. Gong, T. Huang, Linear spatial pyramid matching using sparse coding for image classification, in: Proceedings of the Computer Vision and Pattern Recognition, 2009.
[16] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, Y. Gong, Locality-constrained linear coding for image classification, in: Proceedings of the Computer Vision and Pattern Recognition, 2010.
[17] S. Gao, I. Tsang, L. Chia, P. Zhao. Local features are not lonely-Laplacian sparse coding for image classification, in: Proceedings of the Computer Vision and Pattern Recognition, 2010.
[18] C. Zhang, J. Liu, Q. Tian, C. Xu, H. Lu, S. Ma, Image classification by non-negative sparse coding, low-rank and sparse decomposition, in: Proceedings of the Computer Vision and Pattern Recognition, 2011.
[19] M. Cheng, C. Pun, Y.Y. Tang, Nonnegative class-specific entropy component analysis with adaptive step search criterion, Pattern Analysis & Applications (2011) 1–15.
[20] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in: Proceedings of the Computer Vision and Pattern Recognition, 2001.
[21] K. Sande, T. Gevers, C. Snoek, Evaluating color descriptors for object and scene recognition, in: IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010.
[22] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: Proceedings of the Computer Vision and Pattern Recognition, 2005.
[23] F. Bach, G. Lanckriet, M. Jordan. Multiple kernel learning, conic duality, and the SMO algorithm, in: Proceedings of the International Conference on Machine Learning, 2004.
[24] M. Nilsback, A. Zisserman, A visual vocabulary for flower classification, in: Proceedings of the Computer Vision and Pattern Recognition, 2006.
[25] M. Nilsback, A. Zisserman, Automated flower classification over a large number of classes, in: Proceedings of the ICCVGIP, 2008.
[26] G. Griffin, A. Holub, P. Perona, Caltech-256 Object Category Dataset, Technical Report, CalTech, 2007.
[27] M. Varma, D. Ray, Learning the discriminative power invariance trade-off, in: Proceedings of the International Conference on Computer Vision, 2007.
[28] N. Xie, H. Ling, W. Hu, X. Zhang, Use bin-ratio information for category and scene classification, in: Proceedings of the Computer Vision and Pattern Recognition, 2010.
[29] P. Gehler, S. Nowozin, On feature combination for multiclass object classification, in: Proceedings of the International Conference on Computer Vision, 2009.
[30] X. Yuan, S. Yan, Visual classification with multi-task joint sparse representation, in: Proceedings of the Computer Vision and Pattern Recognition, 2010.