Pattern Recognition Letters 51 (2015) 44–49

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Feature coding for image classification combining global saliency and local difference☆

Shuhan Chen[a,*], Weiren Shi[b], Xiao Lv[c]

[a] College of Information Engineering, Yangzhou University, Yangzhou 225127, China
[b] College of Automation, Chongqing University, Chongqing 400044, China
[c] Chongqing Special Equipment Inspection and Research Institute, Chongqing 401121, China

ARTICLE INFO

Article history:
Received 24 September 2013
Available online 16 September 2014

Keywords:
Image classification
Feature coding
Global saliency
Local difference

ABSTRACT

Saliency[1] based coding proposed recently have been proven to perform well in both performance and efficiency for image classification. However, we find that they are sensitive to unusual features, e.g., noisy features, which we call poor robustness. To address this problem, we propose a novel coding scheme by combining global saliency and local difference together, which are applied for improving stability or robustness and exploring the latent structure information of the codebook respectively. Thorough experiments on various datasets show that our coding consistently performs better than local saliency based coding, in terms of both accuracy and computation cost. Furthermore, it is more robust to unusual features than localized soft-assignment coding. In addition, a combination of our global saliency with local saliency based coding can usually improve both.

1. Introduction

As an important and challenging problem in computer vision, image classification has gained more and more attention in recent years. Many good approaches for image classification have been proposed in the literatures. Among them, the bag-of-words (BOW) [1] model and its extensions (such as spatial pyramid matching [2]) achieve the state-of-the-art performance and have been widely used in many applications. They commonly consist of the following five steps: feature extraction, codebook generation, feature coding and pooling, classification. Feature coding means how to express each descriptor by a codebook to obtain an image-level representation, and has significant influence on classification performance.

We group the existing coding approaches into four categories according to their motivations, as shown in Fig. 1. Voting-based methods are the simplest coding in the literature. Hard-assignment [1] represents a local descriptor to the closest codeword and gives one nonzero coefficient. Without considering codeword ambiguity [3], it always introduces large quantization error. To improve it, soft-assignment [4] is proposed by assigning a local descriptor to all the codewords. Reconstruction-based methods choose a group of codewords to reconstruct descriptors via resolving a least square optimization problem with sparse or locality constraints, e.g., sparse coding [5], local coordinate coding [6], locality-constrained linear coding [7]. Compared with voting-based methods, they always achieve better performance. To reduce reconstruction error, Ren et al. [8] proposed local hypersphere coding, which made reconstruction on a local smooth hypersphere and obtained more distinctive representation. High dimensional methods, proposed for large-scale image classification, such as Fisher kernel coding [9], improved Fisher kernel [10], super vector coding [11], achieve impressive performance [12]. However, they require a large quantity of memory [13]. More recently, saliency based methods are developed, whose core idea is that saliency is a fundamental characteristic of feature coding in the framework with max-pooling [5]. They make a good compromise on efficiency and classification performance. The original salient coding (SaC) [14] encodes each descriptor using the closest codeword by the saliency degree. However, this hard assignment strategy is coarse for feature description [15]. Then, group saliency coding (GSC) [13] is proposed to improve it, whose main idea is calculating the saliency response in a group of codewords. It explores more latent structure information, thus, it performs well.

As mentioned in [15], there are four characteristics we should consider in designing coding method: robustness, adaptiveness, accuracy and independency. Among them, robustness plays the most important role. However, saliency based coding are sensitive to unusual features, e.g., noisy features, in other words, they have poor robustness. In this paper, we propose a novel coding method with good

☆ This paper has been recommended for acceptance by J.K. Kämäräinen.

* Corresponding author. Tel.: +86 18662386487.
E-mail addresses: c.shuhan@gmail.com (S. Chen), wrs@cqu.edu.cn (W. Shi), lvxiao87@126.com (X. Lv).

[1] Saliency in this paper denotes descriptor space saliency.

http://dx.doi.org/10.1016/j.patrec.2014.08.008
0167-8655/

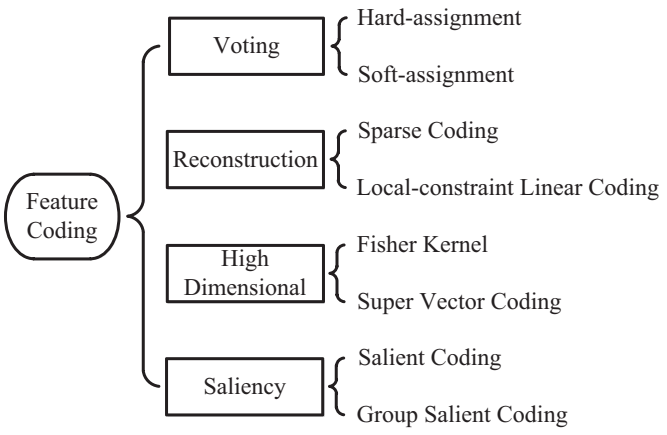


Fig. 1. A taxonomy of coding methods in image classification. Several representatives are listed for each type of coding schemes.

robustness, adaptiveness and independency. Specially, it is achieved by combing global saliency and local difference together, thus, we call it global and local saliency based coding (GLSC). It is noted that they are applied for improving stability or robustness and exploring the latent structure information of the codebook respectively. In addition, our global saliency is complementary to the previous local saliency based coding, thus, a combination can usually improve both.

The remainder of the paper is organized as follows. In Section 2, we briefly review saliency based coding schemes in BOW model. The proposed coding method is presented in Section 3. Section 4 provides experimental results on three datasets: Caltech-101, Scene-15 and UIUC-Sport. Finally, conclusions are drawn in Section 5.

2. Related work

In this section, we mainly concentrate on saliency based coding strategies, introduce their motivations and analyze their limitations. Let x_i ($x_i \in \mathbb{R}^d$) be a d dimensional descriptor, such as scale-invariant feature transform (SIFT) descriptor [16], $B_{d\times M} = (b_1, b_2, \ldots, b_M)$ be a codebook with M cluster centers, and u_i ($u_i \in \mathbb{R}^d$) be the coding coefficient vector of x_i, e.g., u_{ij} be the response of x_i on codeword b_j.

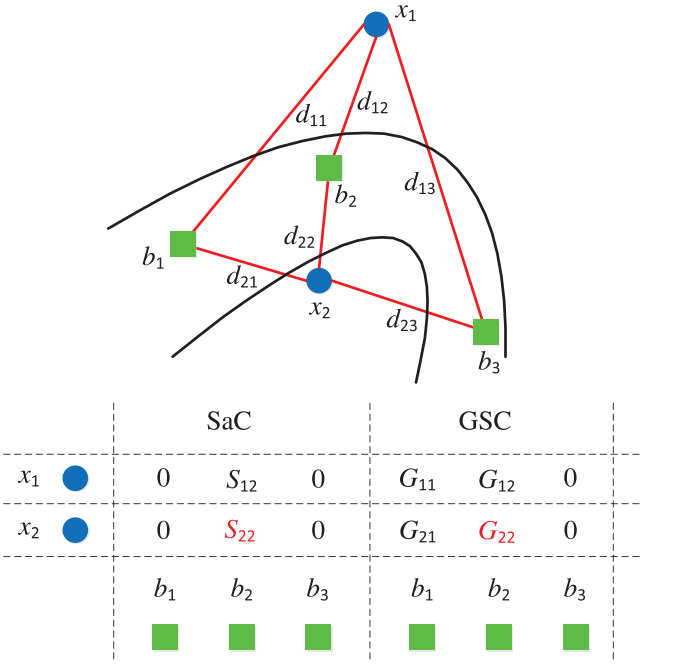
Currently, the framework of using a sparse or local coding scheme combining with max-pooling is regarded as the state-of-the-art. Pooling operation is used to obtain an image-level representation. In the max-pooling, only the strongest response will be preserved. Let p_j be the ith component of image representation p, then max-pooling can be defined as:

$$p_j = \max_i u_{ij} \tag{1}$$

where $i = 1, 2, \ldots, l$, and l is the total number of local features in an image. A detailed analysis of feature pooling was conducted in [22], including average [1], sum [2], max pooling, we only concentrate on max-pooling in this paper.

In saliency based coding, a strong response on a codeword means that this codeword is much closer to a descriptor belonging to it comparing with the other codewords [15]. It indicates the codeword can represent this descriptor independently, which is measured by saliency degree in saliency based coding. In the original salient coding, it is defined by measuring the difference between the closest code and other K-1 codes. In detail, a descriptor is represented as:

$$u_{ij} = \begin{cases} \Psi(x, b_j), & \text{if } j = \arg\min_j \left(\|x - b_j\|^2\right) \\ 0, & \text{otherwise} \end{cases}$$
$$\Psi(x_i, \tilde{b}_j) = 1 - \frac{\|x_i - \tilde{b}_j\|_2}{[1/(K-1)]\sum_{k\neq j}^{K} \|x_i - \tilde{b}_k\|_2} \tag{2}$$



$$S_{12} = 1 - \frac{d_{12}}{d_{11}} > S_{22} = 1 - \frac{d_{22}}{d_{21}}$$
$$G_{12} = (d_{11} - d_{12}) + (d_{13} - d_{12}) > G_{22} = (d_{23} - d_{22}) + (d_{21} - d_{22})$$

Fig. 2. Illustration of saliency based coding. The blue balls are local descriptors and the green rectangles are codewords. The red lines denote the Euclidean distance between them in descriptor space. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

where Ψ denotes the saliency degree, and is the set of K closest codewords to descriptor x.

Although performs well in both effectiveness and efficiency, there still exists a limitation caused by the coarse hard assignment strategy. Only considering the closest codeword, the representations of some descriptors may be suppressed in the subsequent max-pooling. Take Fig. 2 for example, wherein x_1, x_2, x_3 and b_1, b_2, b_3 denote local descriptors and codewords respectively, S_{ij} and G_{ij} denote the response of x_i to b_j in SaC and GSC respectively. As described in Fig. 2, S_{22} is suppressed by S_{12} (since $S_{22} < S_{12}$), thus, it will lose the representation of descriptor x_2.

To improve it, Wu et al. [13] proposed GSC method by introducing group coding. Its main idea is to compute the saliency response in a group of codewords with different group code sizes, and the maximum of all responses is preserved in the final coding result. Let v^g denote the coding response with group code size g, then the GSC representation can be described as:

$$u_{ij} = \max\{v_{ij}^g\}, g = 1, \ldots, G$$
$$v_{i,j}^g = \begin{cases} \Phi^g(x_i), & \text{if } b_j \in g(x_i, g) \\ 0, & \text{otherwise} \end{cases} \tag{3}$$
$$\Phi^g(x_i) = \sum_{t=1}^{G+1-g} \left(\|x_i - \tilde{b}_{g+t}\|_2 - \|x_i - \tilde{b}_g\|_2\right)$$

where in Φ^g denotes group saliency degree, $g(x_i, g)$ denotes the set of the g closest codewords of descriptor x_i, and G is the maximum group code size.

GSC not only preserves the good properties of effectiveness and efficiency in SaC with the help of group coding, but also performs more stably and robustly than SaC. Consider the example in Fig. 2 ($G = 2$), although the response G_{22} is also suppressed by G_{12} (since $G_{22} < G_{12}$), we can still find the representation of descriptor x_2 on

b_1 (G_{21}). However, consider the special case, if the response G_{21} is still suppressed by G_{11} ($G_{21} < G_{11}$), could it be about to happen again? The answer is true, and we will discuss it in the next section.

3. The proposed method

3.1. Global saliency

Both of SaC and GSC calculate saliency response on the k closest codewords, which indicates that they only reflect the local saliency characteristic, thus, we call them local saliency based coding in this work. As mentioned before, the response of the neighbor codeword still will be suppressed in GSC. We analyze that it is caused by disregarding the global saliency information. We also find that preserving more representations of various descriptors would improve the stability and robustness of coding, nor more representations of only some salient descriptors. Specifically, we will illustrate it though the following two cases.

Case 1: A descriptor x_1 is far away from all the codewords. In this case, any single code cannot well represent the descriptor. However, it may produce high response on the K closest codewords in local saliency based coding, in other words, it could produce high local saliency response which is not stable and robust. Table 1(a) is an example in this case.

Case 2: A descriptor x_2 is close to all the codewords, especially to these K closest codewords. In this case, all these codes should be used to describe the descriptor. However, it may produce weak response relative to case 1, in other words, it could produce low local saliency, thus, it may be suppressed by the response of x_1, which will cause the loss of the x_2's representation. We show an example of this case in Table 1(b).

As can be found in Table 2 that the response u_{21} will be suppressed by u_{11} (since $u_{21} < u_{11}$) in SaC, while in GSC, all the responses of x_2 will be suppressed by x_1, thus, we will lose the representation of x_2 in both SaC and GSC. To solve it, we first propose a global saliency based coding method. In common sense, saliency indicates the most noticeable property, which is described as the closest codeword is much closer to a descriptor than the other K-1 closest codewords in [14]. Different with [14], we extend the local saliency to global saliency to improve its stability and robustness. In detail, it is defined as the ratio of the difference between the closest code and all the other codes.

$$\Psi_G(x_i, \tilde{b}_j) = \Phi\left(\frac{\|x_i - \tilde{b}_j\|_2}{[1/(n-1)]\sum_{k\neq j}^{n}\|x_i - \tilde{b}_k\|_2}\right) \tag{4}$$

where Φ is a monotonically decreasing function which denotes the global saliency degree, here, we define it as:

$$\Phi(a) = 1 - a \tag{5}$$

Rigidly calculating the response of each descriptor on all the codewords nor the closest ones has been proven to degrade the classification performance. Because it fails to capture the underlying manifold structure in codewords [17]. Locality constraint which has been proven to outperform the sparse constraint is usually applied to reduce computation cost. Thus, we also make locality constraint on our global saliency, that is only computing response on the K closest codewords:

$$u_{i,j} = \begin{cases} \Psi'_G(x, b_j), & \text{if } b_j \in \mathrm{N}_K(x_i) \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

where $\mathrm{N}_K(x_i)$ denotes the set of K closest codewords to descriptor x_i. For convenience of computation, we further make an approximation as below:

$$\Psi'_G(x_i, \tilde{b}_j) = \Phi\left(\frac{\|x_i - \tilde{b}_j\|_2}{(1/n)\sum_{k=1}^{n}\|x_i - \tilde{b}_k\|_2}\right) \tag{7}$$

Consider the above two examples again, coding responses produced by our global saliency are shown in Table 1. We can find that the response u_{11} is suppressed by u_{21} (since $u_{11} < u_{21}$) in global saliency coding, which is opposite to the local saliency based coding. Although obtained better performance than the local saliency based coding (can be seen in Section 4), the problem mentioned above still exists. We will solve it by introducing local difference, which will be discussed in the following subsection. In addition, a combination of our global saliency with previous methods lacking global saliency information can usually improve both.

Table 1
Two examples showing global saliency.

	(a)				
	$\|x_1 - b_1\|^2$ 0.4406	$\|x_1 - b_2\|^2$ 0.5873	$\|x_1 - b_3\|^2$ 0.5918	$\|x_1 - b_4\|^2$ 0.6459	$\|x_1 - b_5\|^2$ 0.6597
	u_{11}	u_{12}	u_{13}	u_{14}	u_{15}
SaC	0.2907	0	0	0	0
GSC	0.9448	0.2113	0.1933	0.0310	0.0104
Global	0.5594	0.4127	0.4082	0.3541	0.3403
GLSC	0.7918	0.1908	0.1724	−0.0493	−0.1058
	(b)				
	$\|x_2 - b_1\|^2$ 0.2368	$\|x_2 - b_2\|^2$ 0.2815	$\|x_2 - b_3\|^2$ 0.2991	$\|x_2 - b_4\|^2$ 0.3355	$\|x_2 - b_5\|^2$ 0.3395
	u_{21}	u_{22}	u_{23}	u_{24}	u_{25}
SaC	0.2457	0	0	0	0
GSC	0.4215	0.1980	0.1276	0.0184	0.0034
Global	0.7632	0.7185	0.7009	0.6645	0.6605
GLSC	0.3495	0.2411	0.1985	0.1103	0.1006

Table 2
Max-pooling results of Table 1. The italic ones denote the responses of descriptor x_2, and the others denote the responses of descriptor x_1.

	b_1	b_2	b_3	b_4	b_5
SaC	0.2907	0	0	0	0
GSC	0.9448	0.2113	0.1933	0.0310	0.0104
GLSC	0.7918	*0.2411*	*0.1985*	*0.1103*	*0.1006*

3.2. Local difference

Research shows that local saliency is a fundamental characteristic in feature coding, these codes with high local saliency could independently describe the descriptors [14]. In SaC, local saliency is defined as the ratio of the difference between the closest code and other K-1 codes, the bigger the ratio, the higher the local saliency. In GSC, local saliency is defined as the sum of the difference between the closest code and other K-1 codes in a group code size, the larger the difference, the higher the local saliency. In this work, we define local difference to reflect local saliency, the higher the local difference, the higher the local saliency, which is defined as:

$$\Psi_L(x_i, \tilde{b}_j) = \sum_{k=1}^{K}\|x_i - \tilde{b}_k\|_2 - K\|x_i - \tilde{b}_j\|_2 \tag{8}$$

In our local difference coding, large non-negative value reflects high local saliency, while small negative value reflects low local saliency. However, negative response is meaningless, because it will

be suppressed by zero in the subsequent max-pooling. Thus, we simply set them zero:

$$\Psi'_L = \begin{cases} \Psi_L, & \text{if } \Psi_L > 0 \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

Finally, our local difference coding can be written as:

$$u_{i,j} = \begin{cases} \Psi'_L(x, b_j), & \text{if } b_j \in \mathrm{N}_K(x_i) \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

3.3. GLSC coding

We think that a salient code should both have high global and local saliency. Therefore, we combine global saliency and local difference together to reflect saliency degree. Here, we simply combine them by linear summation, in which local difference is used to revise global saliency. In detail, it is defined as:

$$\Psi = \Psi'_G + q\Psi_L \tag{11}$$

where q is a weighted factor.

In our coding scheme, high salient coding response is only produced when with high global saliency and high local difference simultaneously. It is noted that the local difference with negative response is used to revise global saliency. Our coding can perform more stably and robustly than local saliency based coding by considering global saliency and local difference together. Come back to the above two examples, we can find that although the response u_{21} is suppressed by u_{11} (since $u_{21} < u_{11}$), the representation of x_2 can still be found on $u_{22}, u_{23}, u_{24}, u_{25}$ simultaneously. Our GLSC coding not only preserves the effectiveness and efficiency of local saliency based coding, but also shows better stability and robustness.

4. Experiments

4.1. Datasets and experimental settings

The following three datasets are used for test in our experiments:

Caltech-101[18]: It contains 9144 images in 102 classes (including a background class) including animals, vehicles, flowers, etc., with high intra-class appearance shape variability. The number of images per category ranges from 31 to 800. Most of the images are medium resolution, i.e. 300 × 300 pixels.

Scene-15 [19]: It contains 4485 images spread over 15 categories, including outdoor and indoor scenes, e.g., mountains, forest, living room and kitchen. There are 200–400 images per category with average image resolution of 300 × 250 pixels.

UIUC-Sport[20]: It is a sport event dataset which contains 1792 images in eight categories including badminton, bocce, croquet, polo, rock climbing, rowing, sailing and snowboarding. The number of images varies from 137 to 250 in each category.

We choose the SaC [14], GSC [13] and localized soft-assignment coding (LSC) [17] for comparison. Note that all of them are efficient coding techniques, and LSC achieves top performance except IFK and SVC in state-of-the-art methods. The source code of LSC is available on the author's project site. For fair comparison, we integrate all the other methods (including our GLSC) into the LSC framework. Thus, we can guarantee that all the configurations other than the coding part is the same. In our experiments, only a single descriptor is used, the SIFT descriptor, which is densely extracted from images on a grid with step size of six pixels under one scale 16×16 pixel patches. Codebook is generated by the standard K-means clustering algorithm. The codebook size and other parameters in our method will be discussed in the next subsection. Max-pooling is used in all our experiments. SPM kernel with three levels of 1×1, 2×2 and 4×4 is adopted. Lib-linear SVM is used for classification wherein the penalty coefficient is set to 1. All experiments are repeated 10 times with different random selected training and testing images in a PC with an Intel Core 2 Duo CPU (2.26 GHz) and 4 GB RAM, and results are shown with average accuracy and the standard deviation.

4.2. Experimental results and analysis

Caltech-101: On this dataset, we use 30 images per class for training while leaving the rest for test. We first study the influence of K in Eq. (6), (10) and q in Eq. (11) to our algorithm, and then test whether our global saliency is complementary to previous local saliency based coding methods, finally, compare the classification performance of various coding approaches with different size of codebook. We also make an additional experiment to study the robustness of different approaches, in which random noises in different proportions per image are added to replace the original SIFT descriptors as did in [15].

Fig. 3(a) shows the performance when $K = 2, 5, 10, 20, 40$ with 1024 codebook. From the experimental results, we can see that $K \leq 10$ leads to a good performance, while $K \geq 20$, the performance degrades quickly. We fix K to 5 in the rest of tests. Fig. 3(b) shows the results when $q = 0.5, 1.0, 1.5, 2.0, 2.5$ with 1024 codebook on three datasets. As seen, $q = 1.5$ obtains best performance except on UIUC-Sport ($q = 2.0$) dataset. For simplicity, we set $q = 1.5$ in the following experiments. We further test the effectiveness of our global saliency, the results are shown in Fig. 4, in which the global saliency and local difference are calculated by Eqs. (6) and (10) respectively. The global saliency coding outperforms the local saliency based coding. Furthermore, all the tested methods (SaC, GSC) perform over a wide range after combining our global saliency due to its complementary property. The results of various coding strategies under different size of

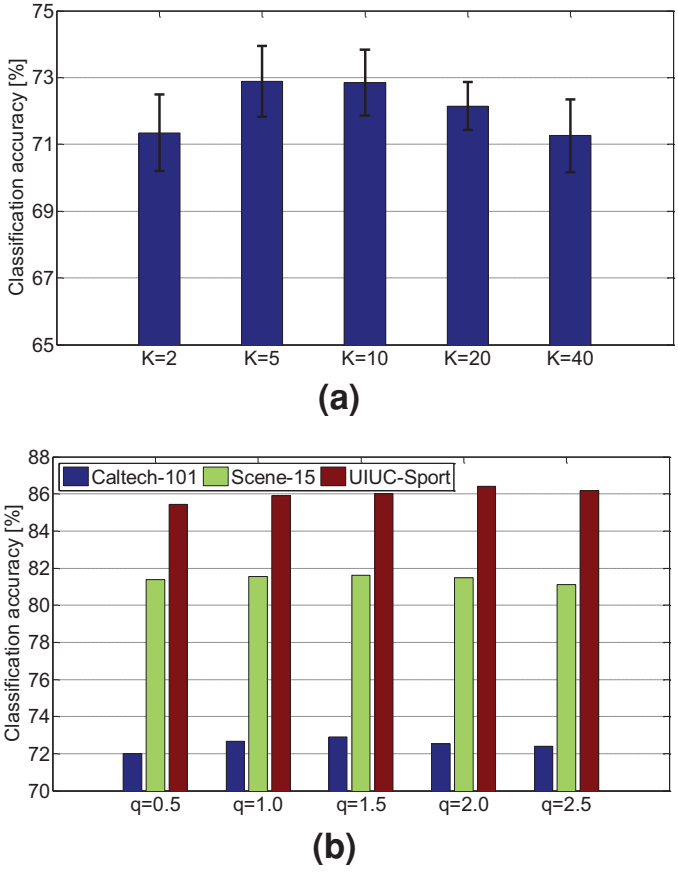


Fig. 3. Performance of our method under different K, q.

Table 3
Performance comparison of various coding strategies under different sizes of codebook.

Codebook size	SaC	GSC	LSC	GLSC
		(a) Caltech-101		
256	65.20 ± 0.57	66.73 ± 0.38	68.95 ± 0.84	**69.13 ± 1.03**
512	66.66 ± 0.46	69.87 ± 1.22	**71.39 ± 1.08**	71.31 ± 0.63
1024	66.21 ± 1.30	70.76 ± 0.87	72.57 ± 1.06	**72.89 ± 1.06**
2048	65.67 ± 0.96	71.32 ± 1.21	**74.21 ± 0.72**	74.02 ± 1.38
		(b) Scene-15		
256	76.63 ± 0.50	76.60 ± 0.52	78.08 ± 0.61	**78.28 ± 0.56**
512	78.15 ± 0.37	78.28 ± 0.65	**80.05 ± 0.57**	79.92 ± 0.78
1024	77.68 ± 0.62	79.27 ± 0.41	81.55 ± 0.35	**81.62 ± 0.56**
2048	78.01 ± 0.64	80.03 ± 0.75	82.30 ± 0.77	**82.51 ± 0.64**
		(c) UIUC-Sport		
256	**83.64 ± 1.05**	82.30 ± 1.29	83.28 ± 0.90	83.29 ± 1.01
512	84.33 ± 0.59	84.64 ± 0.72	84.61 ± 0.99	**84.72 ± 1.01**
1024	85.08 ± 0.72	85.79 ± 0.79	85.65 ± 0.85	**86.01 ± 0.71**
2048	85.51 ± 1.24	86.59 ± 0.75	86.15 ± 1.31	**86.75 ± 1.27**

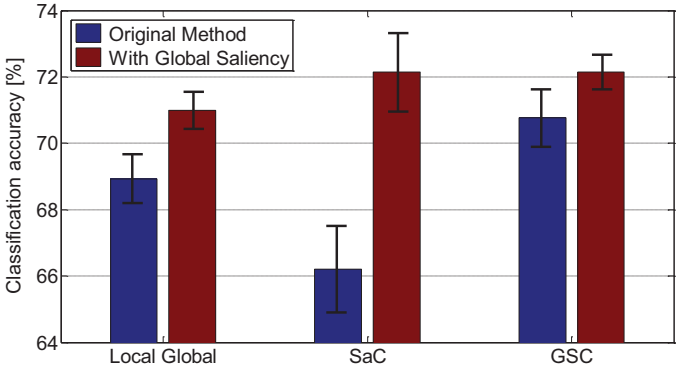


Fig. 4. Effectiveness of our global saliency.

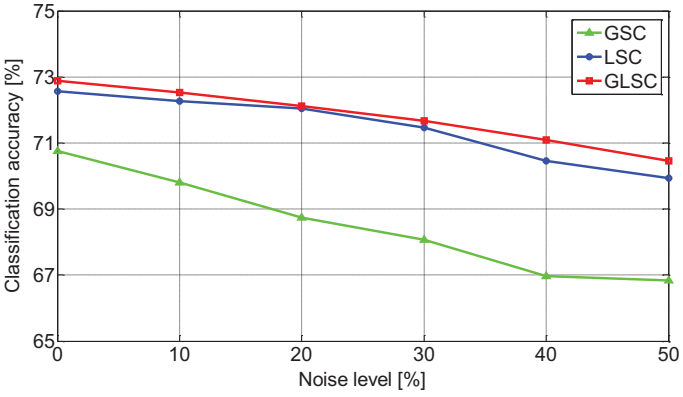


Fig. 5. The influence of random noises on coding algorithms.

codebook are shown in Table 3(a). It is obvious that our method and LSC outperform the others a large margin. With the increase of the codebook size, our method performs better and better. But for SaC, the performance decreases when the codebook size becomes large. This conclusion is the same as drawn in [13]. From Fig. 5, we can see clearly that our coding method performs best under different noise level, which indicates its good robustness to noises. Thus, we can say our GLSC coding is better than LSC. It is noted that the SaC in our implementation is slightly worse than reported in the original paper, we analyze that it may be caused by the different engineering details, e.g., normalization of descriptors, scale of densely grid in SIFT (only one scale is used in our framework).

Scene-15: Following the experimental setup of Lazebnik et al. [2], we randomly pick out 100 images from each category for training and the remainder for testing. As illustrated in Table 3(b), our proposed method and LSC still perform better than the others.

UIUC-Sport: We follow the common experimental setup as did in [17,21], and randomly selected 70 training images from each category and test on the rest images. Table 3(c) gives the performance comparison of the various methods. Again, our GLSC shows better performance than the other coding methods.

5. Conclusion

In this paper, we first analyzed various coding strategies in BOW model, and then deeply discussed the advantages and limitations of local saliency based coding. We have demonstrated that the global saliency is an important characteristic of coding, which has not been fully considered in the literatures. Based on this analysis, we proposed a novel and efficient coding method by combining global saliency and local difference together, called GLSC. The experimental results on different databases (Caltech-101, Scene-15 and UIUC-Sport) have shown its superiority (stability and robustness) to local saliency based coding and also performed as well as LSC which achieves top classification performance in state-of-the-art methods. Furthermore, compared with LSC, our coding approach is more robust when dealing with noisy features. It is also worth noting that global saliency coding can also be cooperated with other coding strategies such as SaC and GSC to improve both. In future, we will further study the influence of different combination strategies of global saliency and local difference, and experimentally analyze on more challenging datasets.

Acknowledgments

The authors would like to express their sincere thanks to the anonymous reviewers for their invaluable suggestions and comments to improve this paper.

References

[1] G. Csurka, C. Bray, C. Dance, L. Fan, Visual categorization with bags of keypoints, in: Workshop on Statistical Learning in Computer Vision, ECCV, 2004, pp. 1–22.
[2] S. Lazebnik, C. Schmid, J. Ponce, Beyond bags of features: spatial pyramid matching for recognizing natural scene categories, vol. 2, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2006, pp. 2169–2178.
[3] J.C. Van Gemert, C.J. Veenman, A.W.M. Smeulders, J.-M. Geusebroek, Visual word ambiguity, IEEE Trans. Pattern Anal. Mach. Intell. 32 (7) (2010) 1271–1283.
[4] J.C. Van Gemert, J.-M. Geusebroek, C.J. Veenman, A.W.M. Smeulders, Kernel codebooks for scene categorization, in: European Conference on Computer Vision (ECCV) 2008, pp. 696–709.

[5] J. Yang, K. Yu, Y. Gong, T. Huang, Linear spatial pyramid matching using sparse coding for image classification, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2009, pp. 1794–1801.
[6] K. Yu, T. Zhang, Y. Gong, Nonlinear learning using local coordinate coding, in: Conference on Neural Information Processing Systems (NIPS) 2009, pp. 2223–2231.
[7] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, Y. Gong, Locality-constrained linear coding for image classification, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2010, pp. 3360–3367.
[8] W. Ren, Y. Huang, X. Zhao, K. Huang, T. Tan, Local hypersphere coding based on edges between visual words, in: Asian Conference on Computer Vision (ACCV) 2013, pp. 190–203.
[9] F. Perronnin, C. Dance, Fisher kernels on visual vocabularies for image categorization, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2007, pp. 1–8.
[10] F. Perronnin, J. Sánchez, T. Mensink, Improving the fisher kernel for large-scale image classification, in: European Conference on Computer Vision (ECCV) 2010, pp. 143–156.
[11] X. Zhou, K. Yu, T. Zhang, T.S. Huang, Image classification using super-vector coding of local image descriptors, in: European Conference on Computer Vision (ECCV) 2010, pp. 141–154.
[12] K. Chatfield, V. Lempitsky, A. Vedaldi, A. Zisserman, The devil is in the details: an evaluation of recent feature encoding methods, in: British Machine Vision Conference (BMVC) 2011.
[13] Z. Wu, Y. Huang, L. Wang, T. Tan, Group encoding of local features in image classification, in: International Conference on Pattern Recognition (ICPR) 2012, pp. 1505–1508.
[14] Y. Huang, K. Huang, Y. Yu, T. Tan, Salient coding for image classification, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2011, pp. 1753–1760.
[15] Y. Huang, Z. Wu, L. Wang, T. Tan, Feature coding in image classification: a comprehensive study, IEEE Trans. Pattern Anal. Mach. Intell. 36 (3) (2014) 493–506.
[16] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.
[17] L. Liu, L. Wang, X. Liu, In defense of soft-assignment coding, in: International Conference on Computer Vision (ICCV) 2011, pp. 2486–2493.
[18] L. Fei-Fei, R. Fergus, P. Perona, Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories, in: IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPRW) 2004.
[19] L. Fei-Fei, P. Perona, A bayesian hierarchical model for learning natural scene categories, vol. 2, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2005, pp. 524–531.
[20] L.-J. Li, L. Fei-Fei, What, where and who? Classifying events by scene and object recognition, in: International Conference on Computer Vision (ICCV) 2007, pp. 1–8.
[21] 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: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2011, pp. 1673–1680.
[22] Y. Boureau, J. Ponce, Y. LeCun, A theoretical analysis of feature pooling in vision algorithms, in: International Conference on Machine Learning (ICML) 2010.