ICIC Express Letters
Volume **10**, Number **11**, November **2016**

 ISSN 1881-803X
pp. **2669–2675**

EFFICIENT LOCALITY AND SALIENCY SIMILARITY CONSTRAINED CODING FOR IMAGE CLASSIFICATION

Shengsheng Wang, Rongchuan Cao and Dong Liu

College of Computer Science and Technology
Jilin University
No. 2699, Qianjin Street, Changchun 130012, P. R. China
wss@jlu.edu.cn; caorc14@mails.jlu.edu.cn

Received April 2016; accepted July 2016

Abstract. *Locality-constrained linear coding (LLC) has gained remarkable success in image classification. In this paper, we put forward an efficient feature encoding method for image classification by combining the saliency similarity constraint with LLC method. Different from the previous saliency coding methods that merge the visual saliency information to the coding stage, we consider the saliency similarity relationship and constitute a novel coding method using it. Our method consists of two parts: saliency similarity kNN (k-Nearest Neighbor) search and saliency similarity constrained coding. In the first part, we combine the saliency similarity relationship among descriptors with kNN method to search the local base. In the second part, we calculate the saliency similarity relationship between the codewords and descriptors, which is used as saliency similarity constraint in coding stage. Experiments on the Scene-15, Caltech-101 and Caltech-256 datasets show that our approach achieves better performance.*
Keywords: Locality-constrained linear coding (LLC), Saliency similarity relationship, Feature coding, Image classification

1. **Introduction.** Recently, the bag of words (BOW) model [1] has been widely used and has gained remarkable success. The BOW framework mainly consists of the three parts: extracting scale-invariant feature transform (SIFT) feature as local descriptors, obtaining a vector as representation of an image via several coding and pooling schemes, and finally putting the vector to a classifier. The coding scheme is the core process during the BOW model and a number of coding methods have been proposed in these years. The original vector quantization (VQ) method assigned each local feature to its closest codeword in the coding layer. To reduce the quantization error, Yang et al. [2] proposed the ScSPM method, in which the sparse coding (SC) [3] instead of VQ has been employed to encode the image feature. Later, Wang et al. [4] proposed the locality-constrained linear coding (LLC) method using the local constraint. However, LLC model ignores the saliency information of image, and has poor interpretation of saliency relationship of features for image classification.

To address this issue, many works [5-7] have been proposed to obtain essential representation of image by incorporating saliency information, such as saliency map. However, they ignored the saliency similarity relationship, including the saliency similarity among descriptors and the saliency similarity between descriptors and codewords. The saliency similarity relation provides a new relation of descriptors that the more saliency similar of the descriptors, the more common information they have. In this paper, we propose to merge the saliency similarity relation into LLC method. More specifically, we first calculate the saliency similarity relationship among the descriptors during the process of kNN search. We call this novel search method as saliency similarity kNN search (SS-kNN). Instead of searching nearest codewords only based on locality in LLC, SS-kNN method makes saliency similar descriptors share their neighboring codewords, and then searches

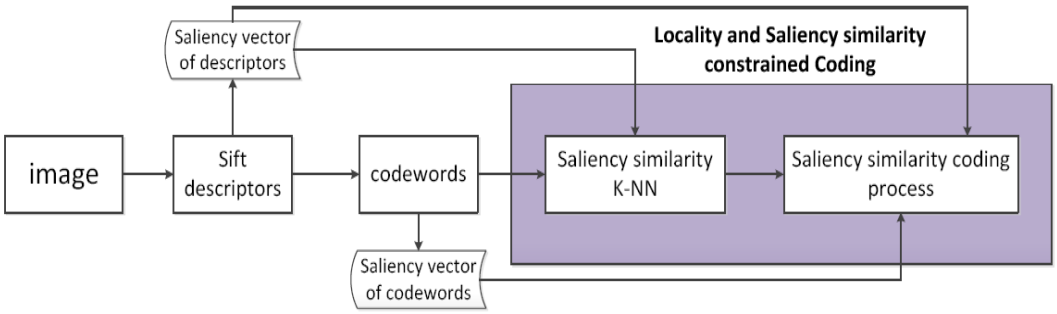

FIGURE 1. The flowchart of proposed framework

the nearest codewords with both locality and saliency similarity. Finally, we calculate the saliency similarity relation between descriptor and codewords, which was applied in the process of coding as the saliency similarity constraint. This method is called as saliency similarity-constrained coding (SSC). In order to gain the saliency similarity relation, we need to calculate the saliency information of descriptors and codewords. In contrast to previous methods, we use the saliency vector to represent the saliency information instead of only a saliency value. The framework of our method is shown in Figure 1.

Finally, we summarize our contribution is threefold:

1). Propose the saliency similarity relationship;

2). Construct the SS-kNN method to search the more suitable local base than LLC;

3). Propose the SSC method. We highlight the saliency similarity constraint in coding step and achieve better classification accuracy.

The rest of the paper is organized as follows. Section 2 briefly reports the LLC model. Section 3 computes the saliency vectors of descriptors and codewords. Section 4 presents our approach to merge the saliency similarity relation into the LLC method. The experimental results are shown in Section 5. We conclude our work in Section 6.

2. **Locality-Constrained Linear Coding.** As suggested in [8], locality is more essential than sparsity. Therefore, many coding methods [9,10] are proposed relying on the locality property. We first briefly describe the LLC [4] algorithm. Let $X = [x_1, \ldots, x_N] \in R^{D\times N}$ denote a matrix with N descriptors. $B = [b_1, \ldots, b_M] \in R^{D\times M}$ shows the codebook entries. $Q = [q_1, \ldots, q_N]$ is the set of codes for X. The LLC algorithm uses the locality criteria to obtain a representation of the sample:

$$\min_{Q} \sum_{i=1}^{N} ||x_i - Bq_i||^2 + \lambda ||d_i \odot q_i||^2 \quad (1)$$
$$\text{s.t. } 1^T q_i = 1, \ \forall i$$

where $d_i = \exp\left(\frac{dist(x_i,B)}{\sigma}\right)$, $dist(x_i, B) = [dist(x_i, b_1), \ldots, dist(x_i, b_M)]^T$ is the Euclidean distance between x_i and the basis vectors; and the parameter σ controls the weight decay speed for the locality; $\lambda \in R$ is the local coefficient; $\odot$ denotes the element-wise multiplication. They also provide an approximated LLC method for fast encoding. The kNN search strategy was applied to search the local bases for each descriptor to form a local coordinate system. It can be calculated as:

$$\min_{\tilde{Q}} \sum_{i=1}^{N} ||x_i - \tilde{q}_i B_i||^2 \quad (2)$$

where $B_i \in R^{D\times K}$ is a local base containing the k nearest codewords of x_i and could replace the elements of d_i. The approximated LLC method realizes the locality constraint and also reduces the computational complexity.

3. **Saliency Vector Computation.** In this paper, we use the phase spectrum of quaternion fourier transform (PQFT) [11] approach to calculate saliency map of image. In order to represent an image by BOW model, local patches with equal space are extracted from an image and every patch is described by SIFT descriptors. After that, we get the saliency values of pixels using PQFT method in each patch and then use them to compute the saliency vector of SIFT descriptor. For the patch i, saliency vector of the SIFT descriptor extracted from it can be derived as:

$$m_i = (m_{i1}, \ldots, m_{ij}, \ldots, m_{iS})^T \tag{3}$$

where S is the number of the pixels in patch i. The m_{ij} is the saliency value of the jth pixel in the patch. In the following step, we calculate the saliency vectors of codewords. First, we use k-means [12] method to cluster the descriptors into m centers and propose a_i is the cluster label of x_i. Then, we denote $H = \{h_1, \ldots, h_j, \ldots, h_m\}$ as the matrix that consists of the saliency vectors of codewords. It can be calculated as:

$$h_j = \frac{\sum_{i=1}^{n} \vartheta(a_i = j) m_i}{\sum_{i=1}^{n} \vartheta(a_i = j)}, \quad \vartheta(x) = \begin{cases} 1 & \text{if } x \text{ is true} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where m_i is the saliency vector of SIFT descriptor. h_j is the saliency vector of the jth codeword, $j \in \{1, 2, \ldots, m\}$. Through the above method, the saliency vectors of descriptors and codewords have been calculated.

4. **Locality and Saliency Similarity Constrained Coding.** Considering the locality constraint alone might not be sufficient to ensure the final representation has discriminative power, we propose the locality and saliency similarity constrained coding. In our method, coefficient vector is decided by two constraints in our method: the locality constraint and the saliency similarity constraint. In detail, we first merge the saliency similarity into original kNN method to search the local base, and finally we apply the saliency similarity constraint in the coding stage.

4.1. **Saliency similarity kNN search algorithm (SS-kNN).** Taking advantage of the saliency similarity constraint, we propose a novel approach to search the local base. Suppose there is a SIFT descriptor x_n, the original kNN search method searches the K-Nearest neighbors of x_n as the local base $B = [b_1, \ldots, b_{K-1}, b_K]$. In our method, we search the local base of x_n not only taking account of its K nearest codewords but also using the codewords of the descriptors that have the most saliency similar to it. More specifically, we first compute the saliency similarity between x_n and other descriptors in turn and search two descriptors which are most saliency similar to the x_n. The saliency similarity among descriptors was calculated through measuring the distance among the saliency vectors of them. Finally, we suppose x_i and x_j are the two SIFT descriptors which are most saliency similar to x_n. We search the nearest codeword C_i for x_i and C_j for x_j, and the local base B of the descriptor x_n is updated through the use of C_i and C_j.

In the LLC method, the elements in the local base B are gradually far from X_n and the locality descends orderly. Therefore, we consider that C_i and C_j are more suitable as local base than the last elements of B, like b_K and b_{K-1}. To improve accuracy, we update the set B using C_i and C_j through the following three situations.

First, the set B contains C_i and C_j. In this situation, the set B does not need to change,

$$B' = [b_1, \ldots, b_{K-1}, b_K] \quad B \cap C_i \neq \emptyset \;\&\; B \cap C_j \neq \emptyset \tag{5}$$

Second, only one of C_i and C_j is in the set B. We assume that set B only includes C_j. In this situation, we only update the set B using the C_i. If C_j is not the last codeword

of B, we add the C_i in the set B and delete the B_K. Else we add the C_i and delete the b_{K-1},

$$B' = \begin{cases} [b_1, \ldots, b_{K-1}, C_i] & B \cap C_i = \emptyset \ \& \ B \cap C_j \neq \emptyset \ \& \ C_j \neq b_K \\ [b_1, \ldots, b_{K-2}, C_i, b_K] & B \cap C_i = \emptyset \ \& \ B \cap C_j \neq \emptyset \ \& \ C_j = b_K \end{cases} \tag{6}$$

Third, both C_i and C_j are all not in the set B. In this instance, we add the C_i and C_j into the set of B. If $C_i \neq C_j$, we use the C_i and C_j to replace the b_K and b_{K-1}. If $C_i = C_j$, we add the C_i into the set of B and just need to delete b_K,

$$B' = \begin{cases} [b_1, \ldots, b_{K-2}, C_i, C_j] & B \cap (C_i \cup C_j) = \emptyset \ \& \ C_i \neq C_j \\ [b_1, \ldots, b_{K-1}, C_i] & B \cap (C_i \cup C_j) = \emptyset \ \& \ C_i = C_j \end{cases} \tag{7}$$

So far, a new local base B' is generated by the SS-kNN search method. It contains the properties not only the locality but the saliency similarity. Later, it will be applied to encode the descriptors in the encoding process.

4.2. **Saliency similarity-constrained coding (SSC).** In this phase, we consider the saliency similarity between the codeword and SIFT descriptor. In order to reduce the reconstruction error during the coding stage, descriptor should be encoded by the codeword which is more saliency similar to it. Therefore, a new coding scheme called saliency similarity-constrained coding (SSC) method is proposed. It adds a saliency similarity constraint to the objective function of approximated LLC. The SSC method enhances the codeword's weight which is more saliency similar to descriptor. And the rough procedures are as follows.

First, we select the K codewords as the local base of x_i through SS-kNN algorithm. Next, we encode the SIFT descriptors using the approximated LLC as (2). Then, the Euclidean distance of saliency vector was used to measure the saliency similarity between x_i and the elements of the set B_i. After that, we can obtain a weighted vector $D = \{d_1, \ldots, d_K\}$ with K columns and each column d_m, $m \in \{1, \ldots, K\}$ represents saliency similarity weight of the codeword b_m,

$$d_m = D(m_i, h_m) = \sum_{j=1}^{S} (m_{ij} - h_{mj})^2 \tag{8}$$

where m_i is the saliency vector of x_i and h_m is the saliency vector of codeword b_m. Then, we add the weighted vector to the basis of $\tilde{q}_i$ as Equation (9).

$$q_i = \frac{\tilde{q}_i + \lambda(1 + D)^{-1}}{\sum\limits_{\tilde{q}_i} \tilde{q}_i + \lambda(1 + D)^{-1}} \tag{9}$$

where q_i is the new coding vector for x_i. Through the SSC method, the codeword that is more similar of the SIFT will be assigned larger weight.

5. **Experimental Results.** In this section, we verify the effectiveness of our method on three widely used datasets: Scene-15, Caltech-101, and Caltech-256. We extract dense SIFT features for all images with a step width of 8 pixels, and the descriptor is extracted at 16×16 pixels. During the SS-kNN search processing, the number of neighbors is set to 5. After all features are encoded, the SPM with levels of $[1 \times 1, 2 \times 2, 4 \times 4]$ is performed. We use max pooling method to normalize the codewords as the final image feature representation. Our method contains two sub methods, which can not only improve the LLC respectively, but also combine together to further improve the LLC. To evaluate the validity of our three strategies, we compared with LLC as well as several state-of-the-art methods for LLC improvement [13,14].

TABLE 1. Classification accuracies on Scene-15 dataset

Training images	20	40	60	80	100
LLC	71.41	76.45	78.86	80.43	81.47
Yang et al. [13]	–	–	–	–	83.32
LLC + SS-kNN	73.22	78.09	80.17	81.91	82.69
LLC + SSC	73.60	78.41	80.52	82.14	83.06
LLC + SS-kNN + SSC	**74.18**	**79.03**	**81.04**	**82.85**	**83.88**

5.1. **Scene-15 dataset.** The Scene-15 dataset contains 4485 images of fifteen scene categories. We trained a codebook with 1024 bases and the results compared with previous methods are shown in Table 1.

5.2. **Caltech-101 dataset.** The Caltech-101 dataset contains 9144 images in 101 different classes. These categories contain from 31 to 800 different numbers of images. We randomly partition the whole dataset into 10, 15, ..., 30 training images per class. And the codebook we trained has 2048 bases. The results are listed in Table 2. To further prove our method, we compare the performance of the four methods when codebooks with different sizes are used. Figure 2 shows the classification accuracy on Caltech-101 dataset in the case of the training images being set to 30.

5.3. **Caltech-256 dataset.** The Caltech-256 dataset consists of 257 classes with a minimum of 80 images per class and a total number of images equal to 30607. We train a codebook of 4096 bases and follow the common experimental setup as the Scene-15 dataset. We train our algorithm on 15, ..., 60 images per class respectively.

TABLE 2. Classification accuracies on Caltech-101 dataset

Training images	10	15	20	25	30
LLC	59.77	65.43	67.74	70.16	73.44
Yang et al. [13]	62.73	67.25	70.37	72.36	74.09
Min et al. [14]	62.90	67.50	69.20	71.60	74.00
LLC + SS-kNN	62.12	66.97	69.42	71.95	75.21
LLC + SSC	62.69	67.48	69.98	72.41	75.67
LLC + SS-kNN + SSC	**63.34**	**68.06**	**70.75**	**73.23**	**76.55**

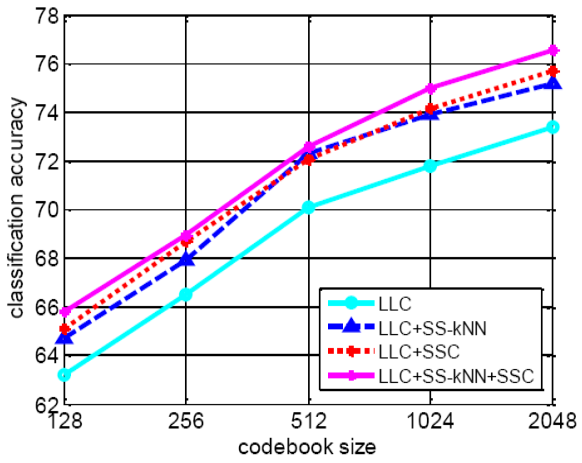

FIGURE 2. Performance comparison with different codebook sizes on Caltech-101 dataset

Table 3. Classification accuracies on Caltech-256 dataset

Training images	15	30	45	60
LLC	34.36	41.49	45.31	47.68
Yang et al. [13]	–	–	–	39.73
LLC + SS-kNN	35.94	42.96	46.62	48.84
LLC + SSC	36.25	43.19	47.03	49.25
LLC + SS-kNN + SSC	**36.90**	**43.83**	**47.70**	**49.87**

5.4. **Discussion.** From the experiments on above three datasets, we can see that the two sub methods of our method, SS-kNN method and SSC method, either of them could improve the LLC respectively. And combining with the two sub methods, our method can achieve significant better result than some of the best coding methods. No matter how many the training images, the accuracy is improved after using our method. Figure 2 shows the experimental result in different sizes of the codebook. It proves that our methods outperform the LLC method under different codebook sizes and the classification accuracies of our methods increase as the size of codebook enhances. It can be seen that the saliency similarity could make the relationship between features more compact and stable, and let the coding results more accurate.

6. **Conclusion and Future Work.** In this paper, we propose locality and saliency similarity constrained coding method. This method integrates saliency similarity to the commonly used approximated LLC method perfectly. It consists of two parts. First, we applied an SS-kNN algorithm to searching the more suitable bases. Second, we proposed SSC method, which employs the saliency similarity as constraint to reduce information loss in coding stage. Experimental results show our method can achieve better performance in image classification. In future work, we plan to combine spatial information with our method and take our method to the video field.

Acknowledgment. This work is supported by the National Natural Science Foundation of China (61472161, 61133011, 61402195, 61502198, 61303132, 61202308), Science & Technology Development Project of Jilin Province (20140101201JC).

REFERENCES

[1] H. S. Yue, W. H. Chen and X. M. Wu, Visualizing bag-of-words for high-resolution remote sensing image classification, *Journal of Applied Remote Sensing*, vol.10, no.1, 2016.

[2] J. Yang, K. Yu and Y. Gong, Linear spatial pyramid matching using sparse coding for image classification, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp.1794-1801, 2009.

[3] B. B. Ni and P. Moulin, Order preserving sparse coding, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.37, no.8, pp.1615-1628, 2015.

[4] J. Wang, J. Yang and K. Yu, Locality-constrained linear coding for image classification, *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.3360-3367, 2010.

[5] W. B. Zou and K. Nikos, HARF: Hierarchy-associated rich features for salient object detection, *IEEE International Conference on Computer Vision*, 2015.

[6] S. D. Jain and K. Grauman, Supervoxel-consistent foreground propagation in video, *The 13th European Conference on Computer Vision*, 2014.

[7] Z. Yang and H. L. Xiong, Image classification based on saliency coding with category-specific codebooks, *Neurocomputing*, 2015.

[8] S. Y. Lu and Z. Y. Wang, A bag-of-importance model with locality-constrained coding based feature learning for video summarization, *IEEE Trans. Multimedia*, vol.16, no.6, 2014.

[9] J. B. Pang, L. Qin and C. J. Zhang, Local laplacian coding from theoretical analysis of local coding schemes for locally linear classification, *IEEE Trans. Cybernetics*, vol.45, no.12, pp.2937-2947, 2015.

[10] G. F. Wang, X. Y. Qin and F. Zhong, Visual tracking via sparse and local linear coding, *IEEE Trans. Image Processing*, vol.24, no.11, pp.3796-3809, 2015.

[11] C. L. Guo and Q. Ma, Spatio-temporal saliency detection using phase spectrum of quaternion Fourier transform, *The 26th IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[12] Z. Sobia, G. M. Ali and K. Asra, Novel centroid selection approaches for KMeans-clustering based recommender systems, *Information Sciences*, vol.320, no.156-189, 2015.

[13] Y. B. Yang, Q. H. Zhu and X. J. Mao, Visual feature coding for image classification integrating dictionary structure, *Pattern Recognition*, vol.48, no.10, pp.3067-3075, 2015.

[14] H. Q. Min, M. J. Liang and R. H. Luo, Laplacian regularized locality-constrained coding for image classification, *Neurocomputing*, vol.171, pp.1486-1495, 2016.