IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 7, 2017, accepted January 6, 2018, date of publication January 17, 2018, date of current version April 23, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2794525

A Dual Locality-Constrained Linear Coding Algorithm for Image Classification

XIANCHEN WANG[1], YANSHAN LI[1,2], AND WEIXIN XIE[1]

[1]ATR National Key Laboratory of Defense Technology, Shenzhen University, Shenzhen 518060, China
[2]Guangdong Key Laboratory of Intelligent Information Processing, College of information Engineering, Shenzhen University, Shenzhen 518060, China

Corresponding author: Yanshan Li (lys@szu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61771319, in part by the Natural Science Foundation of Guangdong Province under Grant 2017A030313343, and in part by the Shenzhen Science and Technology Project under Grant JCYJ20160520173822387 and Grant JCYJ20160307143441261.

ABSTRACT The locality-constrained linear coding (LLC) algorithm has been developed as an effective means for image classification. This algorithm uses locality linear constraints to encode feature points of the image, achieving higher classification accuracy than that of Spatial Pyramid Matching (SPM), spatial pyramid matching using sparse coding, and other traditional algorithms. However, the LLC algorithm uses only the locality information for the visual words in the dictionary while rarely using mutual information between the neighbouring feature points, causing severe ambiguities. Based on the principle of the locality correlation of images, we propose a new algorithm, named the dual locality-constrained linear coding algorithm (DLLC), which applies the locality information of the feature points and visual words and uses the discriminant information provided by the nearest neighbouring feature points. The experimental results demonstrate that the accuracy of the DLLC algorithm is higher than that of the LLC algorithm, particularly when the image category is large, but the set of dictionary training data is small.

INDEX TERMS Dual locality constraints, image classification, LLC.

I. INTRODUCTION

With the arrival of the era of big data, Image processing and image classification[1]–[18], [24]–[32], [34]–[40], have become a research focus in recent years. Among the state-of-the-art image classification technology, BOW [1]–[18] is widely used due to its simplicity and efficiency in the representation of images. Based on BOW, Wang *et al.* [6] proposed the locality-constrained linear coding (LLC) algorithm. The LLC algorithm encodes the extracted locality invariant features (e.g., SIFT) with a visual dictionary and locality linear coding, which provides excellent performance in image classification. However, the LLC algorithm uses only the locality information of the visual words in the dictionary while disregarding mutual information between the neighbouring feature points. Because of the complexity and diversity of images in the real world, the features contain uncertainty. This condition has an effect on the assignment of visual words to a feature in the coding of the LLC algorithm, including the following aspects. The features of different kinds of images may show strong similarities, and the features of the same types of images often show differences. This condition may reduce the accuracy of the classification of the LLC. Here, we propose a dual locality-constrained linear coding algorithm (DLLC). This algorithm makes use of the principle of locality correlation of images to optimize the coding of the feature points to the visual words. The proposed algorithm shows promising classification accuracy relative to that of other state-of-the-art methods.

The rest of the paper is organized as follows. Related works on image classification are introduced in Section II. In Section III, we briefly describe the motivation and contribution of our work. Section IV gives the details of the proposed DLLC algorithm. An experimental evaluation is conducted in Sections V and VI. Finally, we conclude the paper in Section VII.

II. RELATED WORK

Csurka *et al.* [19] and Sivic and Zisserman [20] applied the BOW to image classification and proposed the vector quantization (VQ) method. VQ is a basic method of local feature coding. First, this method extracts the local feature points from the image and represents them with a nearest-neighbour visual word. It then quantifies the entire image according to the statistical data of the visual words. This method involves hard-coding of image features. Van Gemert *et al.* [8] noted that the hard-coding method of image features embodies the rationality and uncertainty of a dictionary. The rationality of the dictionary denotes that the

local feature point of the image is assigned to one of the most similar visual words. The uncertainty of the dictionary denotes that the image features are assigned only to one of the most strongly matching visual words, which ignores the similarity with other visual words. Zhou *et al.* [21] proposed a super-vector coding method to overcome the low accuracy of the VQ. Van Gemert *et al.* [22] proposed a kernel dictionary coding method and used the uncertainty of the codebook as the criterion for the coding weight allocation. Jaakkola and Haussler [23] combined generative probability models and discriminative methods to implement the Fisher kernel (FK) in image classification. Perronnin and Dance [24] applied the FK to image categorization in which the input signals were images, and the underlying generative model was a visual vocabulary. This Gaussian mixture model approximates the distribution of low-level features in images. Jianchao *et al.* [7] proposed the spatial pyramid matching using sparse coding (ScSPM) linear coding method using sparsity [25]–[29] to code local image features that can be assigned to only a few visual words. Zhu *et al.* [30] proposed a novel two-layer model, in which dictionaries are learned through three different stages, and the locality-constrained sparse representation is improved. Yu *et al.* [31] found that in sparse coding, a local non-zero-weight coefficient is typically assigned to the nearest-neighbour visual words of local feature points; they subsequently proposed the locality coordinate coding (LCC) method. The LCC explicitly notes that the coding of local features of the image is local and considers that the properties in the image's local feature space are more essential than the sparsity. However, the LCC is a solution to the optimization problem of the L1 model and requires substantial computation. Wang *et al.* [6] proposed LLC, which is considered a fast implementation of the LCC method. On the basis of local features and morphological transformation, Zhu *et al.* [32] proposed an ensemble method based on supervised learning for segmenting the retinal vessels in colour fundus images. Yuan *et al.* [14] combined this method with the idea of fuzzy geometry and proposed the FG-LLC image classification method. Gao *et al.* [25] improved the LLC and applied it to the classification of human organs in CT images. Based on temporal and morphological features, He *et al.* [33] presented an online unsupervised learning classification of pedestrians and vehicles for video surveillance. Huang *et al.* [34], [35] used ultrasound image classification for the diagnosis of breast lesions.

The process of image classification based on the above local feature algorithms is shown in Fig. 1. This algorithm can be divided into the following steps.

Step 1. Extract feature points (such as the SIFT feature) from the image and describe them.

Step 2. Use K-means or other methods to train the feature points into a visual dictionary.

Step 3. Represent the unclassified image features points with visual words [6], [7], [13], [18] in the dictionary using the image feature coding method.

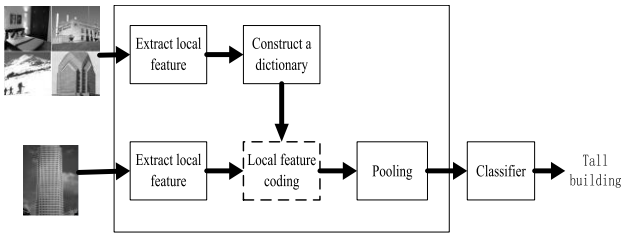


FIGURE 1. Framework of the image classification model based on the local feature algorithms.

Step 4. Form the image descriptors using the pooling algorithm [37]–[39].

Step 5. Classify the image descriptors using support vector machines (SVMs) or other classification algorithms.

Image feature coding is a process that quantifies the image feature points into visual words. Coding error is the predominant factor that affects the accuracy of image classification.

III. MOTIVATION AND CONTRIBUTION

The LLC algorithm takes the feature point x_i and visual word b_i as two vectors. The algorithm calculates the weight coefficient z_i by calculating the Euclidean distance of two vectors x_i and b_i. The LLC algorithm assigns visual words to a feature according to the similarity. However, there substantial uncertainties exist in image coding based on a visual dictionary, mainly from the following aspects:

A. AMBIGUITY OF THE IMAGE

Different images are obtained from the same object due to the shooting angle and light, and other factors. Ambiguity of the image results in local feature uncertainty and variations in plausibility.

B. UNCERTAINTIES IN FEATURE EXTRACTION AND DESCRIPTION

Local features extracted from an image are imperfect information since they cover only partial intrinsic variations in visual appearance. When these features are used to learn the codebook, the uncertainty translates to the codebook as well.

C. UNCERTAINTIES IN CREATION OF VISUAL DICTIONARIES

In the state-of-the-art approaches, the visual words of the codebook are generated by cluster algorithms in which the centres of the clusters are viewed as visual words. This process also introduces ambiguity to the mapping between the visual words and the features, as the cluster is an approximate method for partitioning data, and the centres themselves lose a large amount of information on the features.

D. UNCERTAINTIES IN THE MAPPING OF VISUAL WORDS AND IMAGE FEATURE POINTS

Traditional approaches assign visual words to a feature according to similarity. Unfortunately, the similarity measure

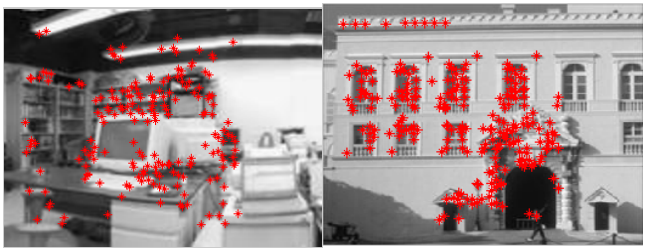


FIGURE 2. Distributions of the extracted SIFT features in two images.

and assignment both introduce ambiguity to the codebook approach for image classification.

The LLC uses soft assignment to perform feature coding to reduce these uncertainties. However, it is still impossible to completely eliminate them using vector metrics alone. As shown in Fig. 2, the two images in the Scene15 dataset have completely different contents. If the same visual dictionary is used to encode their features, then more than 42% of them will have the same visual words.

Therefore, we present a useful image classification algorithm termed DLLC. By taking advantage of the relativity of nearest neighbours to optimize the coding of the feature points and visual words, the DLLC algorithm reduces the uncertainty that occurs when image feature points and visual words are mapped. The algorithm improves the recognition of similar images, thus constituting an efficient and promising scheme for big data consisting of images.

IV. CODING ALGORITHM BASED ON DUAL LOCALITY CONSTRAINTS

A. LOCALITY CONSTRAINTS BETWEEN ADJACENT FEATURE POINTS

It is well known that an image has locality correlation on the grey scale [2]. In addition, the image local invariant feature is the reflection of the grey level distribution of image pixels and others around it. Thus, local features for adjacent images have relevance. In addition, the general probability of adjacent features belonging to the same semantic is much higher than that of distant features, and image classification is a type of semantic recognition to a certain extent. Therefore, the spatial information of the feature plays an important role in feature coding and image classification and will help reduce the influence of uncertainty on image classification. To verify that semantic correlations exist between adjacent local features, we perform the following analysis.

Suppose that the image SIFT feature set is $X = [x_1, x_2, \cdots, x_N] \in R^{D\times N}$. Two feature points $x_i, x_j \in X$ exist, and their spatial distance in the image isd_{ij} (in pixels). The number of dictionary words shared by the two feature points is S. The number of SIFT feature points extracted from the image is N. The number of feature points sharing S. words is M_S. In this study, the Scene15 dataset is analysed, and the ratio of feature points of sharing words is quantified by the following correlation formula:

$$\psi = {}^{M_S}/_{N} \tag{1}$$

The statistical results are shown in Table 1.

TABLE 1. Statistical table of visual words shared by the feature points.

	S =3	S =2	S =1
d_{ij} =20	0.33	0.62	0.83
d_{ij} =30	0.20	0.54	0.80

As shown in Table 1, when the distance between the two feature points is shorter, the possibility of sharing visual words is greater. The correlation between local feature points is strong. This condition leads to a high probability that the nearest-neighbour features correspond to the same visual word during dictionary learning. Thus, it is most possible that the same visual words are assigned to adjacent feature points. The locality constraints between adjacent feature points are important in image classification.

B. ALGORITHM

Based on the LLC algorithm and the above correlation measure, we propose the DLLC algorithm. The DLLC algorithm uses the following criteria:

$$\begin{aligned} \underset{z}{\operatorname{argmin}} & \sum_{i=1}^{N} \|x_i - Bz_i\|^2 + \lambda_1 \|d_i \odot z_i\|^2 \\ & + \lambda_2 \sum_{1\le i<j\le N} d_{ij} I \left\|z_j - z_i\right\|^2 \\ s.t.\ & 1^T z_i = 1, \quad \forall i, \end{aligned} \tag{2}$$

where

$$I = \begin{cases} 1 & d_{ij} \le H \\ 0 & d_{ij} > H \end{cases} \tag{3}$$

$$d_{ij} = \left\|x_j - x_i\right\|^2 \tag{4}$$

$$d_i = \exp(\frac{dist(x_i, B)}{\sigma}) \tag{5}$$

$X = [x_1, x_2, \ldots, x_N] \in R^{D\times N}$ is the image feature point set, and $B = [b_1, b_2, \ldots, b_K] \in R^{D\times K}$ is the visual dictionary. $Z = [z_1, z_2, \ldots, z_N] \in R^{D\times N}$ is the coding coefficient of image feature points X and the visual words B solved by the DLLC algorithm. λ_1 and λ_2 are regularization parameters that establish the relative importance of the reconstruction error $\|x_i - Bz_i\|^2$ with respect to the regularization terms $\|d_i \odot z_i\|^2$ and $d_{ij} I \left\|z_j - z_i\right\|^2$. H is the threshold for calculating the nearest neighbour's feature points.

Where $\text{dist}(x_i, B) = [\text{dist}(x_i, b_1), \ldots, \text{dist}(x_i, b_n)]^T$. $\text{dist}(x_i, b_i)$ is the Euclidean distance between and codebook. σ is used to adjust the weight decay speed for the locality adaptor. The constraint $1^T z_i = 1$ in Eq. (2) follows the sift-invariant requirement of the DLLC algorithm.

The first factor of Eq. (2) ($\|x_i - Bz_i\|^2$) is the signal fidelity, which ensures that the classification signal energy is not lost. The second factor ($\|d_i \odot z_i\|^2$) is the constrained term of the coefficient Z by the neighbour words of the feature points, which ensures that the local feature points are mapped to the nearest-neighbour words. The third factor ($d_{ij}I \left\|z_j - z_i\right\|^2$) is the constrained term of the coefficient Z by the local feature points to reduce the fuzziness and uncertainty of the image classification caused by external factors, such as the change of light.

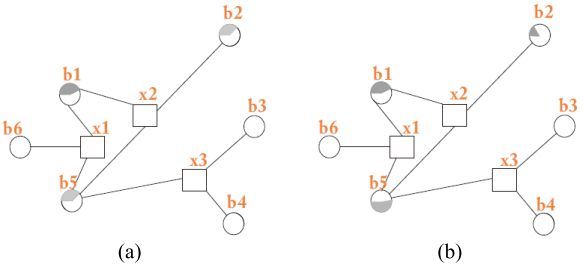


FIGURE 3. Comparison of the LLC and DLLC algorithms. (a) LLC algorithm. (b) DLLC algorithm.

In Fig. 3, to further illustrate the LLC and DLLC algorithms, we use an example to illustrate how the algorithms work. The square represents the image feature points. The circle represents the visual words in the dictionary. The shadow size of the circle represents the weight coefficient of the visual words mapped to the feature point x_2. In Fig. 3(a), only the locality constraints of the feature points and the nearest-neighbour words in the visual dictionary are considered in the LLC algorithm. Therefore, when the distances between the feature points and a few visual words are the same, the coding coefficient is similar. As shown in Fig. 3(a), the shadow sizes of circle b_2 and circle b_5 are the same. In Fig. 3(b), the nearest three visual words are mapped to the feature point b_2. Since the characteristics of x_2 are affected by the feature points x_3 and x_1, the shadow sizes of circle b_2 and circle $b_5$have changed. Because both the nearest neighbours x_1 and x_3 of the image feature point x_2 have a mapping relationship with the visual word b_5, the coding coefficient of the visual words b_2 and b_5 are adjusted, and the visual word b_5 coding coefficient is greater than that of the visual word b_2.

C. APPLICATION IN IMAGE CLASSIFICATION

The image classification model based on the proposed DLLC algorithm first uses the KNN algorithm to find the nearest-neighbour words of feature points in the dictionary. The closer words are to the feature points, the greater the quantization coefficient. Furthermore, to apply the constraints of the nearest-neighbour feature points, the coefficients of the nearest-neighbour feature points mapped to a visual word also affect the coefficients of the feature points mapped to the words in the dictionary. When the coefficients of the nearest-neighbour feature points mapped to a word are larger, the coefficients of the feature points mapped to the words in the dictionary are also larger, and vice versa.

The algorithm reduces ambiguous assignments between features points and visual words caused by an illumination change, viewpoint change, or scale change in image classification. The reasons for the reduction of the matching ambiguity lie in the following aspects. First, the DLLC algorithm uses the locality constraint of the feature points and visual words. The nearest words are assigned to a feature using spatial distance information. Second, based on the principles of locality correlation of the image, the DLLC algorithm revises the coefficient in Eq. (2) by introducing the discriminant information provided by the nearest-neighbour feature points.

Let input $X = [x_1, x_2, \ldots, x_N] \in R^{D\times N}$ be the image feature points set, $B = [b_1, b_2, \ldots, b_K] \in R^{D\times K}$ be the visual dictionary, and the output $Z \in R^{D\times N}$ be the coding coefficients of the feature points and the words in the dictionary solved by the DLLC algorithm. Therefore, the steps of the image classification algorithm based on the DLLC are as follows.

Step 1. Extract local features from training images.

Step 2. Learn the codebook using the K-means algorithm on the features extracted.

Step 3. Find the nearest-neighbour words of a feature from the codebook.

Step 4. Find the nearest feature points for the feature points.

Step 5. Encode features extracted from the test image by solving the constrained least-squares fitting problem.

Step 6. Pool the features using multiscale spatial max pooling.

Step 7. Output the image descriptor for image classification.

V. EXPERIMENTAL SETTING

In the experiments, we implemented and evaluated five classes of local feature classification algorithms on Scene15ĄCCaltech-101 and Caltech-256 datasets. The five methods are:

1) ScSPM algorithm [7]: the ScSPM linear coding method using sparsity to code local image features that can be assigned to only a few visual words.;

2) Traditional LLC algorithm [6]: the Locality-constrained Linear Coding algorithm that uses nearest-neighbor words for the spatial domain to represent a local feature;

3) FG-SPM algorithm [24]: the algorithm that replaces the code book of SPM by FG-Codebook, and measure the similarity between features and words via fuzzy set.

4) FG-LLC algorithm [24]: the algorithm that replaces the code book of LLC by FG-Codebook, and measure the similarity between features and words via fuzzy set.

5) Our algorithm: the framework that uses nearest-neighbor words for the spatial-spectral domain to represent a local feature and spatial-spectral Pyramid Matching mode.

In this paper, we use three widely used image databases, the Scene15, Caltech-101 and Caltech-256 datasets. In the

experiment, the test image is randomly selected from the dataset, the feature points are extracted, and the image dictionary is constructed. The image features used are 128-dimensional dense SIFT features. The number of nearest-neighbour feature points for the feature point to be classified is 3. The classifier used is the linear SVM. The experimental platform is a Supermicro server with an Intel Xeon CPU E7-4880 V2 2.5 GHz processor and 256 GB internal storage.

The classification experiment for each size operates ten rounds with the same parameters.

FIGURE 4. Scene15 dataset image sample.

The Scene15 dataset contains *bedroom*, *kitchen* and 13 other categories. Each sub image set contains 200-400 images, as shown in Fig. 4, yielding 4485 images. In the experiment, we randomly selected 100 images from each class as the training set for the classifier SVM, and the remainder of the images are used as the test set.

The Caltech-101 dataset contains 102 image categories, such as mobile phones, cameras, cars and so on. Each category contains 31 to 800 images, yielding a total of 9144 images, as shown in Fig. 5. In the experiment, we randomly selected 30 images from each class as the training set for the classifier SVM, and the remainder of the images are used as the test set.

FIGURE 5. Caltech-101 dataset image sample.

The Caltech-256 dataset is a challenging set of 256 object categories containing a total of 30607 images. with much higher intra-class variability and higher object location variability compared with Caltech-101. It can be obtained from http://www-cvr.ai.uiuc.edu/ponce_grp/data. There are at least 80 images in each category, and each image is of about 300_250 pixels in size. Fig. 6 lists several images selected from the Caltech-256 dataset.

FIGURE 6. Caltech-256 dataset image sample.

TABLE 2. Image classification accuracy with various values of λ_1 and λ_2.

λ_1 \ λ_2	0.001	0.005	0.01	0.05	0.1
0.001	0.775	0.788	0.772	0.745	0.632
0.005	0.805	0.832	0.804	0.794	0.646
0.01	0.774	0.786	0.778	0.732	0.607
0.05	0.665	0.702	0.693	0.679	0.555
0.1	0.632	0.669	0.656	0.648	0.539

VI. EXPERIMENTAL RESULTS AND ANALYSIS

A. EXPERIMENTAL RESULTS AND ANALYSIS OF THE SCENE15 DATASET

λ_1 and λ_2 in Eq. (2) are adjusting factors and determine the weight of the nearest-neighbour visual words and feature points. Table 2 presents the image classification accuracy of the DLLC when λ_1 and λ_2 are 0.001, 0.005, 0.01, 0.05 and 0.1. The classification accuracy increases with the values of $\lambda_1$until λ_1 reaches 0.005, at which point λ_2 equals 0.005. Then, when λ_1 increases further, the image classification accuracy decreases. When λ_1 is small, the locality information of the feature points and visual words plays a minor role in the representation of features. When λ_1 is large, the locality information of the feature points and visual words plays a major role in the representation of features. Similar to λ_1, λ_2 determines the locality information of the nearest-neighbour feature points. The classification accuracy increases with the values of λ_2 until λ_2 reaches 0.005, at which point λ_1 equals 0.005. However, as λ_2 further increases, the image classification accuracy decreases. Considering special cases in which the value of λ_2 is set to 0, the classification accuracy of the proposed algorithm will be the same as that of traditional LLC.

TABLE 3. Classification results for the scene15 dataset (average of 10 rounds).

CODEBOOK SIZE	SPM	FG-SPM	LLC	FG-LLC	DLLC
256	0.6351	0.6518	0.7743	0.7809	0.7951
1024	0.6515	0.6639	0.8134	0.8198	0.8245
2048	0.6626	0.6642	0.8256	0.8348	0.8401
4096	0.6646	0.6689	0.8308	0.8452	0.8522

Table 3 shows the experimental results on Scene15 dataset with the DLLC algorithm and four existing approaches. The experimental data indicate that the classification accuracy of the DLLC algorithm is better because of locality constraints of the nearest-neighbour feature points. The DLLC algorithm uses dual locality constraints of the visual words in the

dictionary and the nearest-neighbour feature points. This approach reduces ambiguous assignment, e.g., polysemy, which causes low classification accuracy. Specifically, when the codebook size is small, the advantages of our algorithm are more prominent.

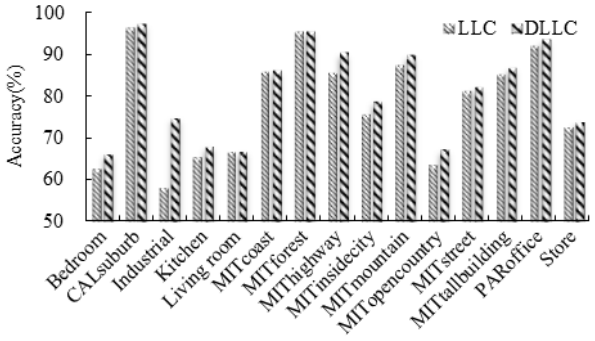


FIGURE 7. Classification accuracy of each image category on Scene15 dataset.

Fig. 7 shows the classification accuracy on all categories of the Scene15 dataset when the codebook size is 256. The DLLC algorithm results in improved accuracy for all categories over the traditional LLC algorithm. With *bedroom*, *industrial*, and *kitchen*, for example, the advantages of our algorithm are more prominent.

The DLLC algorithm brings improved accuracy for all categories over the traditional LLC algorithm, particularly on *MITopencountry*, *bedroom*, *industrial* and *livingroom*. Among all of the categories, the accuracies on the CAL suburb are the best. When the size of the codebook is larger than 256, the accuracy with various sizes of the codebook are all over 95%, and the accuracy of the LLC and DLLC are similar on this category. Because the accuracy on several categories, such as *MITopencountry*, *bedroom*, *industrial* and *livingroom*, is relatively low when using the LLC, the DLLC improves the accuracy more significantly on these categories.

The reasons for the accuracy improvement of the DLLC algorithm lie in the following aspects. First, the DLLC algorithm uses the locality information of the feature points and visual words. The nearest words are assigned to a feature using spatial distance information. Second, based on the principles of locality correlation of the image, the DLLC algorithm revises the coefficient Z in Eq. (2) by introducing the discriminant information provided by the nearest-neighbour feature points. The algorithm reduces ambiguous assignments between features points and visual words caused by an illumination change, viewpoint change, or scale change in image classification.

The DLLC algorithm improves classification accuracy but increases computation. The main reason is that there are two additional sources of consumption. One is the determination of the K nearest-neighbour feature points, and the other is the optimal processing of the coding coefficients of the nearest-neighbour feature points. In this experiment, 20 images are

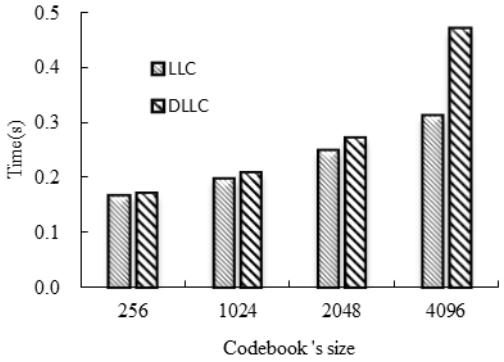


FIGURE 8. Comparison of computational cost between DLLC and LLC algorithms.

randomly selected, and both the DLLC and LLC algorithms are applied for image classification. The average coding time overhead of the 20 images is shown in Fig. 8. The DLLC and LLC algorithms have little time difference when the codebook size is relatively small. Furthermore, the classification accuracy of the DLLC algorithm with a low codebook size is similar to the classification accuracy of the LLC algorithm with a higher codebook size. When the size of the codebook is high, the time consumed between the DLLC and LLC algorithms differs. For example, when the codebook size is 4096, the average computational time of the DLLC algorithm is 1.5 times greater than that of the LLC algorithm.

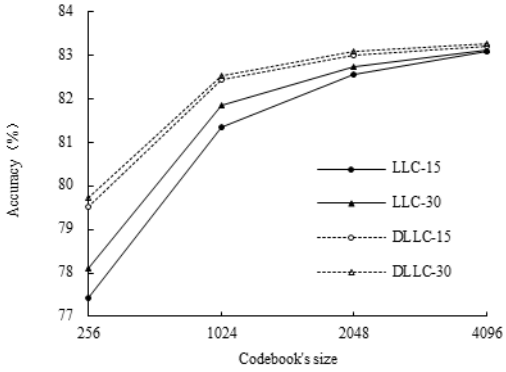


FIGURE 9. Experimental results of different dictionary scales on Scene15 dataset (average of 10 rounds).

To test the influence of different numbers of images used to construct a dictionary on the DLLC and LLC algorithms, in the experiment, 15 and 30 images from each class are selected to construct dictionaries. The experiment results are shown in Fig. 9. When the size of codebook is small, the size of dictionary has a strong influence on the LLC algorithm and less influence on the DLLC algorithm.

B. EXPERIMENTAL RESULTS AND ANALYSIS OF THE CALTECH-101 DATASET

We compared our result with several existing approaches on Caltech-101. Detailed results are shown in Table 4, and it

TABLE 4. Experimental results for the caltech-101 dataset.

CODEBOOK SIZE	SPM	FG-SPM	LLC	FG-LLC	DLLC
256	0.5038	0.5455	0.7174	0.7519	0.7613
1024	0.6227	0.6629	0.7633	0.7681	0.7795
4096	0.6385	0.6541	0.7925	0.7927	0.7931

can be seen that in most cases, our proposed DLLC method leads the performance. The classification accuracy of the five algorithms on Caltech-101 dataset is lower than that of the Scene15 dataset. The main reason is that the image class of 102 in the Caltech-101 dataset is higher than the image class of 15 in the Scene15 dataset. In the Caltech-101 experiment, when training a dictionary, 30 images from each class are selected as the training set of the classifier SVM, which is less than that in the Scene15 dataset. Furthermore, on Caltech-101 dataset, particularly when the codebook size is low, the advantage of the DLLC algorithm is more significant than that using the Scene15 dataset.

TABLE 5. Experimental results for the Caltech-256 dataset.

TRAINING IMAGES	SPM	FG-SPM	LLC	FG-LLC	DLLC
30	0.3402	0.3623	0.4119	0.4456	0.4728
45	0.3746	0.3892	0.4531	0.4626	0.4839
60	0.4014	0.4151	0.4768	0.4792	0.4923

C. EXPERIMENTAL RESULTS AND ANALYSIS OF THE CALTECH-256 DATASET

We followed the common setup during experiment on Caltech-256. We trained a codebook with 4096 bases. In every experiment, we randomly choose 30,45 and 60 images from each class as the training dataset for the SVM classifier and the rest as testing dataset. The classification results are shown in Table 5. Similarly, the proposed algorithm's performance is better than the original SPM, LLC FG-SPM and FG-LLC algorithms under the same experimental conditions. It further demonstrates the advantage of the proposed algorithm. The DLLC with small number of training images can obtain a better result than other algorithms do with larger number. The classification accuracy of the five algorithms in the Caltech-256 dataset is lower than that of other datasets. It is mainly due to different datasets and experimental setups. Caltech-256 dataset has more categories than Scene15 and Caltech-101 dataset. The image classification performance decreases as the number of image category increases.

D. DISCUSSION

These experimental results show that the accuracy of the DLLC algorithm is higher than that of other algorithms in the experiments. The advantages of the DLLC algorithm are more prominent when the image category is higher and when the training data for the dictionary is lower. Therefore, the DLLC algorithm is more suitable for image classification for a large volume of data. Meanwhile, the size of the codebook has weaker effects on the classification accuracy of the DLLC algorithm. With the increase in the number of images used in the construction of the visual dictionary, the classification accuracy of the DLLC and LLC algorithms is improved to a certain extent. When the codebook size is low, the number of images used in the construction of dictionaries has fewer effects on the classification accuracy of the DLLC than on the classification accuracy of other algorithms.

As the nearest-neighbour feature points increase, the classification accuracy of the DLLC is further improved. In this study, to facilitate the experiment, the number of nearest-neighbour feature points is fixed. Considering the computational complexity, the focus of our future research will be how to determine the number of the nearest feature points, i.e., to determine the value of the threshold H in Eq. (2).

VII. CONCLUSION

We propose a linear coding method called DLLC based on the dual locality constraints of visual words and nearest-neighbour points that fully utilizes the locality similarity of the image. The DLLC algorithm enhances the discriminative power of the image classification and reduces ambiguous assignments between feature points and visual words. Experiments on the Scene15, Caltech-101 and Caltech-256 datasets are conducted in this study. The experimental results show that the DLLC algorithm is an effective classification method for image classification in the context of big data.

REFERENCES

[1] C. Grana, R. Cucchiara, D. Borghesani, and M. Manfredi, "A fast approach for integrating ORB descriptors in the bag of words model," *Proc. SPIE*, vol. 8667, no. 15, pp. 2273–2276, Feb. 2013.

[2] M. M. Farhangi, M. Soryani, and M. Fathy, "Improvement the bag of words image representation using spatial information," in *Advances in Computing and Information Technology*. Berlin, Germany: Springer, 2013, pp. 681–690.

[3] R. Ji *et al.*, "Location discriminative vocabulary coding for mobile landmark search," *Int. J. Comput. Vis.*, vol. 96, no. 3, pp. 290–314, Feb. 2012.

[4] R. Ji, H. Yao, W. Liu, X. Sun, and Q. Tian, "Task-dependent visual-codebook compression," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2282–2293, Apr. 2012.

[5] S. Baek, C. D. Yoo, and S. Yun, "Learning a discriminative visual codebook using homonym scheme," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Prague, Czech Republic, May 2011, pp. 2252–2255.

[6] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, San Francisco, CA, USA, Jun. 2010, pp. 3360–3367.

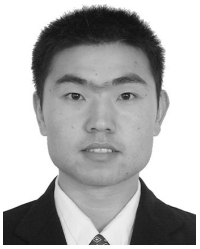
[7] Y. Jianchao, Y. Kai, G. Yihong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Miami, FL, USA, Jun. 2009, pp. 1794–1801.

[8] J. C. van Gemert, C. J. Veenman, A. W. M. Smeulders, and J.-M. Geusebroek, "Visual word ambiguity," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 7, pp. 1271–1283, Jul. 2010.

[9] Y. Li, W. Liu, X. Li, Q. Huang, and X. Li, "GA-SIFT: A new scale invariant feature transform for multispectral image using geometric algebra," *Inf. Sci.*, vol. 281, pp. 559–572, Oct. 2014.

[10] Y. Li, W. Liu, Q. Huang, and X. Li, "Fuzzy bag of words for social image description," *Multimedia Tools Appl.*, vol. 75, no. 3, pp. 1371–1390, Feb. 2016.

[11] Y. Li, W. Liu, and Q. Huang, "Traffic anomaly detection based on image descriptor in videos," *Multimedia Tools Appl.*, vol. 75, no. 5, pp. 2487–2505, Mar. 2015.
[12] Y. Li, W. Xie, Z. Gao, Q. Huang, and Y. Cao, "A new bag of words model based on fuzzy membership for image description," in *Proc. 12th Int. Conf. Signal Process. (ICSP)*, Hangzhou, China, Oct. 2014, pp. 972–976.
[13] Y. Li, Q. Huang, W. Xie, and X. Li, "A novel visual codebook model based on fuzzy geometry for large-scale image classification," *Pattern Recognit.*, vol. 48, no. 10, pp. 3125–3134, Oct. 2015.
[14] Y. Yuan, A. Hoogi, C. F. Beaulieu, M. Q.-H. Meng, and D. L. Rubin, "Weighted locality-constrained linear coding for lesion classification in CT images," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Milan, Italy, Aug. 2015, pp. 6362–6365.
[15] H. Cheng, Z. Liu, L. Hou, and J. Yang, "Sparsity-induced similarity measure and its applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 4, pp. 613–626, Apr. 2016.
[16] L. Liangyue, L. Sheng, and F. Yun, "Discriminative dictionary learning with low-rank regularization for face recognition," in *Proc. IEEE Int. Conf. Workshops Automat. Face Gesture Recognit.*, Shanghai, China, Apr. 2013, pp. 1–6.
[17] Q. Qiu, V. M. Patel, and R. Chellappa, "Information-theoretic dictionary learning for image classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 11, pp. 2173–2184, Nov. 2014.
[18] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. Int. Conf. Comput. Vis.*, Washington, DC, USA, 1999, pp. 1150–1157.
[19] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *Proc. Workshop Statist. Learn. Comput. Vis. (ECCV)*, 2004, pp. 1–22.
[20] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nice, France, 2003, pp. 1470–1477.
[21] X. Zhou, K. Yu, T. Zhang, and T. S. Huang, "Image classification using super-vector coding of local image descriptors," in *Proc. 11th Eur. Conf. Comput. Vis.*, Heraklion, Greece, 2010, pp. 141–154.
[22] J. C. Van Gemert, J.-M. Geusebroek, C. J. Veenman, and A. W. M. Smeulders, "Kernel codebooks for scene categorization," in *Proc. 10th Eur. Conf. Comput. Vis.*, Marseille, France, 2008, pp. 696–709.
[23] T. S. Jaakkola and D. Haussler, "Exploiting generative models in discriminative classifiers," in *Proc. Adv. Neural Inf. Process. Syst.*, Nov. 1998, vol. 11. no. 11, pp. 487–493.
[24] F. Perronnin and C. Dance, "Fisher kernels on visual vocabularies for image categorization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Minneapolis, MN, USA, Jun. 2007, pp. 1–8.
[25] S. Gao, I. W. H. Tsang, L. T. Chia, and P. Zhao, "Local features are not lonely—Laplacian sparse coding for image classification," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, San Francisco, CA, USA, Jun. 2010, pp. 3555–3561.
[26] J. Yang, K. Yu, and T. Huang, "Efficient highly over-complete sparse coding using a mixture model," in *Proc. 11th Eur. Conf. Comput. Vis.*, Heraklion, Greece, 2010, pp. 113–126.
[27] N. Kulkarni and B. Li, "Discriminative affine sparse codes for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, vol. 32. no. 14, pp. 1609–1616.
[28] C. Zhang, J. Liu, Q. Tian, C. Xu, H. Lu, and S. Ma, "Image classification by non-negative sparse coding, low-rank and sparse decomposition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Providence, RI, USA, Jun. 2011, pp. 1673–1680.
[29] S. Gao, L.-T. Chia, and I. W.-H. Tsang, "Multi-layer group sparse—For concurrent image classification and annotation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Washington, DC, USA, Jun. 2011, pp. 2809–2816.
[30] S. Zhu, J. Du, N. Ren, and M. Liang, "Hierarchical-based object detection with improved locality sparse coding," *Chin. J. Electron.*, vol. 25, no. 2, pp. 290–295, Mar. 2016.
[31] K. Yu, T. Zhang, and Y. Gong, "Nonlinear learning using local coordinate coding," in *Proc. 22nd Int. Conf. Neural Inform. Process. Syst.*, Vancouver, BC, Canada, 2009, pp. 2223–2231.
[32] C. Zhu, B. Zou, Y. Xiang, J. Cui, and H. Wu, "An ensemble retinal vessel segmentation based on supervised learning in fundus images," *Chin. J. Electron.*, vol. 25, no. 3, pp. 503–511, May 2016.
[33] Y. He, N. Sang, C. Gao, and J. Han, "Online unsupervised learning classification of pedestrian and vehicle for video surveillance," *Chin. J. Electron.*, vol. 26, no. 1, pp. 145–151, Jan. 2017.
[34] Q. Huang, X. Huang, L. Liu, Y. Lin, X. Long, and L. Li, "A case-oriented Web-based training system for breast cancer diagnosis," *Comput. Methods Programs Biomed.*, vol. 156, pp. 73–83, Mar. 2018.
[35] Q. Huang, F. Yang, L. Liu, and X. Li, "Automatic segmentation of breast lesions for interaction in ultrasonic computer-aided diagnosis," *Inf. Sci.*, vol. 314, pp. 293–310, Sep. 2015.
[36] X. Feng, X. Guo, and Q. Huang, "Systematic evaluation on speckle suppression methods in examination of ultrasound breast images," *Appl. Sci.*, vol. 7, no. 1, p. 37, 2017.
[37] L. Cao, R. Ji, Y. Gao, Y. Yang, and Q. Tian, "Weakly supervised sparse coding with geometric consistency pooling," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Providence, RI, USA, Jun. 2012, pp. 3578–3585.
[38] J. Feng, B. Ni, Q. Tian, and S. Yan, "Geometric ℓp-norm feature pooling for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Providence, RI, USA, Jun. 2011, pp. 2609–2704.
[39] Y. Jia, C. Huang, and T. Darrell, "Beyond spatial pyramids: Receptive field learning for pooled image features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Providence, RI, USA, Jun. 2012, pp. 3370–3377.
[40] F. Yang, Q. Huang, L. Jin, and A. W.-C. Liew, "Segmentation and recognition of multi-model photo event," *Neurocomputing*, vol. 172, pp. 159–167, Jan. 2016.



XIANCHEN WANG received the M.Sc. degree in electronic circuits and systems from the Beijing Jiaotong University, Beijing, China, in 2009. He is currently pursuing the Ph.D. degree in intelligent information processing with the ATR National Key Laboratory of Defense Technology, Shenzhen University. His research interests include intelligent information processing, image analysis, and pattern recognition.

YANSHAN LI received the M.Sc. degree from the Zhejiang University of Technology in 2005 and the Ph.D. degree from the South China University of Technology, China. He is currently an Associate Professor with the ATR National Key Laboratory of Defense Technology, Shenzhen University, China. His research interests include computer vision, machine learning, and image analysis.

WEIXIN XIE was born in 1941. He received the degree from Xidian University, Xi'an, China. He was a Professor, a Doctoral Supervisor and the Vice-President of Xidian University. From 1981 to 1983, he was a Visiting Scholar with the University of Pennsylvania, Philadelphia, PA, USA. In 1989, he was with the University of Pennsylvania, as a Visiting Professor. He is currently a Professor with Shenzhen University, where he is also the Director of the National Natural Science Foundation of China with electronic discipline and the Academic Committee. His current research interests include intelligent information processing, fuzzy information processing, image processing, and pattern recognition.