Hindawi
Wireless Communications and Mobile Computing
Volume 2020, Article ID 8838454, 9 pages
https://doi.org/10.1155/2020/8838454

Research Article

Annular Spatial Pyramid Mapping and Feature Fusion-Based Image Coding Representation and Classification

Mengxi Xu,[1] **Yingshu Lu**,[2] **and Xiaobin Wu**[1]

[1]*School of Computer Engineering, Nanjing Institute of Technology, Nanjing 211167, China*
[2]*Huawei Technologies Co., Ltd., Nanjing 210000, China*

Correspondence should be addressed to Mengxi Xu; mxxu26@126.com

Received 15 July 2020; Revised 13 August 2020; Accepted 22 August 2020; Published 11 September 2020

Academic Editor: Hongju Cheng

Conventional image classification models commonly adopt a single feature vector to represent informative contents. However, a single image feature system can hardly extract the entirety of the information contained in images, and traditional encoding methods have a large loss of feature information. Aiming to solve this problem, this paper proposes a feature fusion-based image classification model. This model combines the principal component analysis (PCA) algorithm, processed scale invariant feature transform (P-SIFT) and color naming (CN) features to generate mutually independent image representation factors. At the encoding stage of the scale-invariant feature transform (SIFT) feature, the bag-of-visual-word model (BOVW) is used for feature reconstruction. Simultaneously, in order to introduce the spatial information to our extracted features, the rotation invariant spatial pyramid mapping method is introduced for the P-SIFT and CN feature division and representation. At the stage of feature fusion, we adopt a support vector machine with two kernels (SVM-2K) algorithm, which divides the training process into two stages and finally learns the knowledge from the corresponding kernel matrix for the classification performance improvement. The experiments show that the proposed method can effectively improve the accuracy of image description and the precision of image classification.

1. Introduction

Image classification is a major topic in the field of image processing and pattern recognition [1, 2]. The conventional image classification methods [3, 4] focus on some specific targets which extract effective features to represent the informative contents of images. However, this kind of method has obvious drawbacks. For example, some specific image features cannot be generalized to strange objects. In addition, image information is likely lost during the coding phase of the method.

Recently, image color features have been widely considered. Traditional color features include the color histogram, color moments, color sets, the color coherence vector, and the color correlogram. These color features are combined with contour-based features (i.e., Hu invariant moments and histogram of gradient) in the field of image classification and have achieved excellent performance results [5, 6]. A major research focus on human linguistics in the color feature representation is color names (CN). In the computer vision, the color properties contain human linguistic labels for pixels in an image. Berlin and Kay [7] presented that languages include eleven universal basic color terms. According to their analysis, Khan et al. [8] proposed to exploit the conventional statistical model to learn the knowledge of the naming color by the human's brain, and Khan et al. [6] adopted this color feature combined with the histogram of gradient for image features fusion, which were successfully applied in target recognition. Therefore, this paper utilizes CN and dimension-reduced scale invariant feature transformation as the complementary fused features. In order to further investigate and optimize image information, a modified spatial pyramid matching method (SPM) is proposed to add the features' spatial location information. In this method, an image would be split into patches whose features are extracted and encoded so that a uniform vector with a spatial

Figure 1: CN label sketch map. The first row is the original images, and the second row is the corresponding labeled images by CN.

location information can be formed. However, the simple division in SPM cannot maintain the vector representation after the rotation of the image.

Aiming to solve the problems above, we propose an image classification model based on an annular spatial pyramid matching and multifeature fusion. Considering the success of the SIFT feature, we adopt the dimension-reduced SIFT feature to speed up the encoding process. A large part of image information is contained in color features, while SIFT never takes this point into account. Thus, if color information can be added into our feature fusion model, the proposed image classification system can be improved significantly. Additionally, an annular spatial pyramid matching is applied to map the extracted features for the rotation invariance spatial vector representation. Due to various P-SIFT features extracted from different images, the sparse coding algorithm is adopted for the generation of the uniform vector. In the phase of feature fusion and the learning parameters of the classification algorithm, support vector machine with two kernels algorithm (SVM-2K) [9, 10] is utilized in our proposed model for training the labeled data set and predicting the unlabeled images. The SVM-2K algorithm is the primary issue when combining the kernel canonical correlation analysis algorithm (KCCA) [11, 12] and SVM classification algorithm in which KCCA is adopted for preprocessing the features (P-SIFT and CN), and two independent SVM models of two features are trained together. The novelty of our method lies in our novel feature fusion-based image classification method that obtains satisfying results in contrast to existing methods, while P-SIFT and CN features can be used as complementary descriptions of the image.

2. Features Extraction and Spatial Mapping

2.1. Color Naming Feature. Recently, the research of image features has focused on the target shape-based local feature instead of the information-rich color features. Compared with the traditional color descriptors including HUE, OPP, and color moments, the term-based histogram statistical feature, color names, is used as the complementary to P-SIFT. Through mapping RGB color channels, the extraction algorithm of CN labels in each pixel of an image is using one of 11 color names: black, blue, brown, grey, green, orange, pink, purple, red, white and yellow. The algorithm also employs the histogram and normalization to get the vector representation. CN labels are different forms of the same color to a specific color name so that the CN feature is equipped with an illumination invariance. The CN algorithm is to predict the color description by humans for a specific color in essence. The experiments in paper [5] shows that, compared with HUE and OPP color feature, the CN descriptor is more selective and has been successfully applied in the field of image recognition. Hence, this paper utilizes the CN feature to mitigate the vacancy of color in our image classification model. The comparison of images with and without color names is shown in Figure 1. For a given pixel X, the CN description of this point should be defined as the probability of the pixel belonging to one of the 11 color names.

$$\mathrm{CN} = [P(\mathrm{cn}_1 \mid X), P(\mathrm{cn}_2 \mid X), \cdots, P(\mathrm{cn}_{11} \mid X)], \quad (1)$$

where $P(\mathrm{cn}_i \mid X)$ represents the probability of the pixel assigned to the ith label, and the probability mapping matrix is determined from a large dataset. This paper utilizes a mapping matrix, which statistically computes the image dataset of 11 color names collected by Google. The max probability of color names is in equation (1). Following the above flows, all pixels in an image would be assigned a specific color name; after that, a 1×11 histogram vector can be formed.

2.2. Dimension-Reduced SIFT Feature. SIFT originally proposed by Lowe [13] is a local gradient histogram used to locate the target shape. The algorithm identifies even the extreme points in a multiscale image space and has been widely applied in the field of image classification and pattern recognition. In recent years, SIFT has had various modifications, including SURF, PCA-SIFT, and HSV-SIFT. However, SIFT is a 1×128 vector and can be utilized to detect image features even when the image properties, including scale

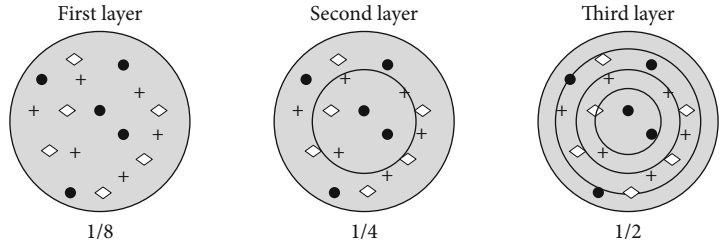

Figure 2: The frame of the rotational spatial pyramid matching model. The different toys in circles represent different features, and the weights of the first to third layer in R-SPM are 1/8, 1/4, 1/2, respectively.

and image noise, are changed. However, the 1×128 SIFT descriptor has a negative impact on the computation and performance of a reconstruction method. In addition, SIFT is typically taken for its accuracy in reconstruction, and the task of image classification in this paper does not need the same level of accuracy. Thus, a dimension-reduced SIFT feature generated from SIFT using a principal component analysis algorithm (PCA) [14] is proposed to extract the main information of the image feature. Due to the various number of SIFT in different image patches, the extracted SIFT cannot be used as the final representation. Thus, we have included the sparse representation in our model to encode the SIFT features of an image.

2.3. Rotational Annular Spatial Pyramid Matching. SIFT features contain corresponding location information. Lazebnik et. al [2] introduced the spatial pyramid matching model (SPM) in which images would be divided into multiscale patches so that SIFT can be a more valuable mapped for encoding. In this SPM model, SIFT features are first extracted from all images and clustered to get a visual dictionary which includes K visual words. Each image is then divided into three layers (1×1, 2×2, 4×4), and 1/4, 1/4, and 1/2 are independently given to the corresponding layer. Finally, the SIFT features in their patches are encoded, and the series is connected as a $21 \times K$ vector. Unfortunately, SPM has no capability of dealing with violent rotation images, so it is unable to maintain the spatial location information. While targets in the image recognition and classification are normally accompanied by large position changes, the vectors of the model would become the original SIFT vectors or CN features if SPM is integrated into our model for mapping.

To make full use of the features' location information, we suggest an annular division based on a rotational spatial pyramid matching model (R-SPM) to map SIFT and CN features. In this model, as shown in SPM, each image will be annularly divided into three layers (1×1, 1×2, and 1×4), which are attached to weights 1/8, 1/4, and 1/2, respectively (Figure 2). Because of the relatively small image patches in R-SPM, the location information of features is more effective. For the larger patches of images, this effect is suppressed, so less weight is given to the larger layers. Finally, all features in annular circle patches will be encoded, and the series is connected as a $9 \times K$ vector.

In this way, the location information in image patches would be attached to the specific location of the vector representation. No matter how the targets rotate, the vectors of the circle patches never change so that the mapped vector by R-SPM owns the rotational spatial locative information.

3. The Theory of the Sparse Representation

In the traditional bag-of-visual-words model (BOVW), the k-means algorithm [15] is used to cluster the terms for a visual dictionary. The vector quantification method (VQ) is then adopted for encoding features by computing Euclidean distances between features and visual words to get a histogram representation. However, VQ fails to consider that the Euclidean distance is unsuitable for histogram-based features (i.e., SIFT and HOG). For the reduction of the feature information in the encoding process, instead of VQ, ScSPM [2] is proposed to utilize the sparse representation algorithm and generate more sparse and selective encoding vectors. The sparse representation [16] is a kind of soft encoding algorithm which can be seen as the extension of the k-means.

3.1. Sparse Representation. The goal of the sparse representation (SR) is to learn an ultracomplete dictionary and use rare atoms to reconstruct original signals for the successfully extraction of the embedded image information. The sparse representation has a wide range of research and practical applications, especially for the collection, compression, and representation of high dimension vectors. For instance, in face recognition, image features are first extracted from a training set and then acquired into an ultracomplete dictionary, generating sparse representations of that training set. The feature extraction and sparse representation processes for the test image set are also then performed, and the test and training samples are compared. The nature of the sparse representation is to solve the convex problem, but some specific problems, such as the low convex degree of sparse matrices, cannot be efficiently solved by a traditional convex optimization algorithm. The sparse dictionary is developed by a K-SVD algorithm [17] which iteratively and simultaneously regenerates atoms and sparse coefficients. However, the time required to run an optimization algorithm is excessive, which has a significant negative effect on the classification performance. Therefore, this paper used the K-means++ algorithm [18] to get a stable ultracomplete dictionary, and orthogonal matching pursuit (OMP) was performed to get reconstruction coefficients. The core problem of the sparse representation is to solve equation (2).

$$\min_V \frac{1}{2}\|x - vU\|_2 + \lambda\|v\|_1, \quad (2)$$

where x represents the feature to be reconstructed, U is the fixed dictionary, and v is the sparse coefficients. The former term is the reconstruction error, and the latter term is to take control of the sparsity.

3.2. Orthogonal Matching Pursuit. An orthogonal matching pursuit algorithm (OMP) [19, 20] is a modified iterative version of a matching pursuit algorithm (MP). Due to the nonorthogonality of the selected atoms by MP, the sparse coefficients are normally local peaks. In the process of the atom selection, OMP follows the laws of MP, but converts the selected atoms into orthogonal, which decreases the iterative number of optimal convergences. At the same time, the OMP algorithm has set a maximum iterative number, so that when the number has achieved the set value, the OMP algorithm is forcibly stopped.

Ideally, each image would get a coefficient matrix when all SIFT are to be encoded. In addition, a pooling algorithm should be used to pool the matrices for uniform vectors. Experiments [21, 22] show that, when compared with other pooling methods, the maximum pooling is more effective for the generation of the sparse vector representation. Thus, we have adopted a maximum pooling algorithm in this paper. If we set that a given image has M features (K dimensions) after the mapping process of the R-SPM model, the coefficient matrix is $V = (v_1, v_2, v_3, \cdots, v_M) \in R^{M\times(9K)}$. The method of the maximum pooling can be defined in equation (3).

$$r = \max(|v_1|, |v_1|, |v_2|, |v_3|, \cdots, |v_M|), \quad (3)$$

where $r \in R^{1\times(9K)}$ is the final vector representation of this image.

4. Support Vectors Machine with Double Fusion Kernels Algorithm

In the field of image recognition, the selection of a specific image feature is determined by the characteristics of the targets to be predicted. There is no effective standard to assess the feature selection. In order to mitigate any influence of feature selection on the performance of image recognition and classification, the learning of the corresponding knowledge from the training image set through machine learning algorithms has been widely researched. This feature fusion learning algorithm mainly focuses on two requirements. The first is that the classifier directly learns the series of connected vectors to achieve the performance of feature fusion. The second requirement is that each image feature is considered an individual unit to train their own models and give each model different weight.

Following the above method, researchers [23, 24] combined the preserved invariant feature with SIFT to design an image classification model that based on feature fusion and got a more brilliant performance than a single feature based model. Therefore, we proposed the support vector machine with a two fusion kernels algorithm (SVM-2K) to complete the task of the image feature fusion and classification. SVM-2K combines the preprocess of the kernel canonical correlation analysis (KCCA) and the parameters of an SVM algorithm to make two features that independently and complementarily describe the images. In the SVM-2K algorithm, a similarity constraint between two hyper plane mappings for the organic combination of preprocess and parameter learning is introduced.

If there are two completely different features (set as A and B) for the same data set after their individual kernel mapping, the mapped features can be set as $\varnothing_A$ and $\varnothing_B$. Then, a specific image can be described in equation (4).

$$S = \{(\varnothing_A(x_1), \varnothing_B(x_1)), (\varnothing_A(x_2), \varnothing_B(x_2)), (\varnothing_A(x_3), \varnothing_B(x_3)), \cdots, (\varnothing_A(x_l), \varnothing_B(x_l))\}. \quad (4)$$

The similarity constraint is defined as follows.

$$|\langle w_A, \varnothing_A(x_i)\rangle + b_A - \langle w_B, \varnothing_B(x_i)\rangle - b_B| \le \gamma_i + \epsilon, \quad (5)$$

where $(w_A, b_A), (w_B, b_B)$ is the weights and thresholds of the SVM model. This constraint is introduced into the SVM function for further optimization.

$$\min L = \frac{1}{2}\|w_A\|^2 + \frac{1}{2}\|w_B\|^2 + C^A\sum_{i=1}^{l}\varepsilon_i^A + C^B\sum_{i=1}^{l}\varepsilon_i^B + \mathrm{D}\sum_{i=1}^{l}\gamma_i. \quad (6)$$

The decisive function of SVM-2K can be expressed as follows.

$$h(\mathrm{x}) = \mathrm{sign}\left(0.5\left(\langle \widehat{w}_A, \varnothing_A(x)\rangle + \widehat{b}_A - \langle \widehat{w}_B, \varnothing_B(x)\rangle - \widehat{b}_B\right)\right). \quad (7)$$

In this paper, our classification model adopts the SVM-2K algorithm to perform feature fusion and classification learning on P-SIFT and CN for better performance than the single feature-based SVM model. Because the SVM-2K algorithm is a binary classifier, however, we follow the implementation of LibSVM [25] to utilize the "one VS one" method to extend SVM-2K to multiclass classification tasks. The "one VS one" method is to train the $\mathrm{N}\times(\mathrm{N}-1)/2$ binary SVM-2K model to vote for the final predicted results of the test image. Figure 3 shows the flow chart of our feature fusion-based image classification model.

5. Experiment Results and Analysis

The experimental datasets used in this paper are Caltech-256 [26] and PASCAL VOC 2011 [27]. Caltech-256 is a traditional dataset in the field of computer vision. It includes 256 category image sets and each set has different number of image patches (from 31 to 800). PASCAL VOC 2011 is a benchmark test set for the detection of visual object classification which provides standard images for testing algorithms and learning performance. We randomly selected 9

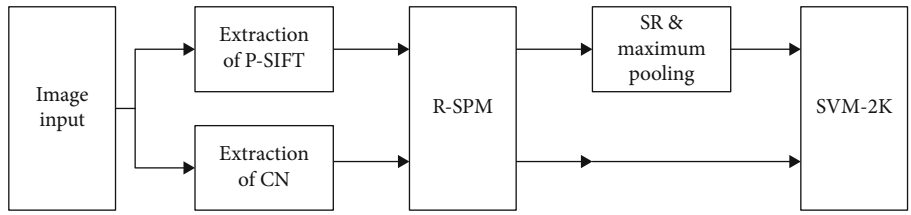

Figure 3: The flow chart of the proposed feature fusion-based image classification model.

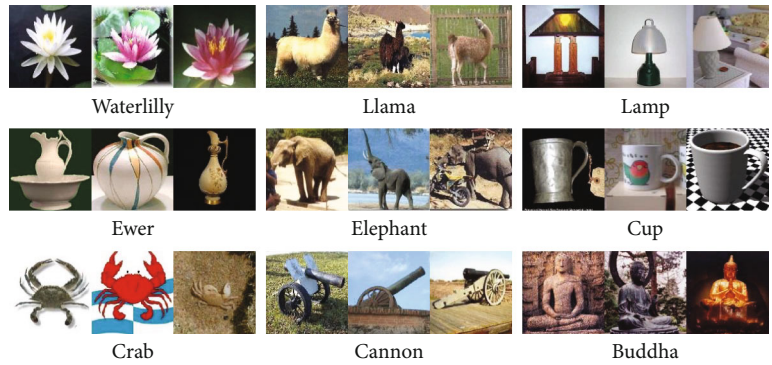

Figure 4: Nine categories image dataset selected from Caltech-256.

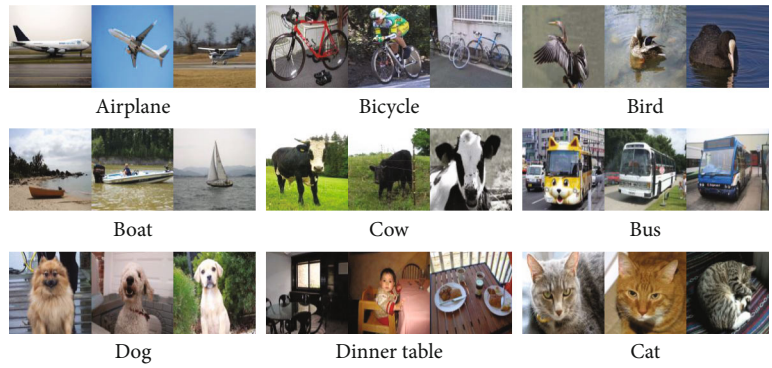

Figure 5: Nine categories image dataset selected from PASCAL VOC 2011.

categories from each dataset and divided them into training and test datasets, which are shown as Figures 4 and 5. The training dataset contained 50 images, the rest of the images in each category were set as the test dataset.

The original SIFT features are extracted by the vl_feat function library [28]. All experiments in this paper were implemented on the MATLAB 2013b platform, and Average Precision (AP) [29, 30] is introduced to assess the performance of the image classification models.

5.1. The Performance Analysis of P-SIFT. P-SIFT was generated by implementing a PCA algorithm on SIFT for dimension reduction. In order to assess the influence of dimension-reduced percentages on the speed of feature reconstruction and the performance of image classification, we collected the mean time (MT) spent by sparse coding and the AP performance of the proposed model. Caltech-256 is utilized as the baseline dataset in this section.

Table 1: The influence of the dimension-reduced percent.

Dimension-reduction	0.1	0.3	0.5	0.7	0.9	1
MT(s)	0.0521	0.0931	0.1240	0.1734	0.2325	0.2571
AP (%)	66.29	75.42	87.42	89.24	89.34	88.57

Table 1 shows that when the dimension-reduction percentage increases, and the mean time spent by feature reconstruction shortens, but the corresponding AP performance is extremely slow. When the percentage is larger than 0.7, the mean time is relatively short, and the performance of classification model tends to be stable. Thus, the experiments all adopt 0.7 as the dimension-reduced percentage.

5.2. The Performance Analysis of R-SPM. For the sake of elevating the rotational performance of R-SPM, this section utilizes an aircraft image and implements 6 kinds of rotation

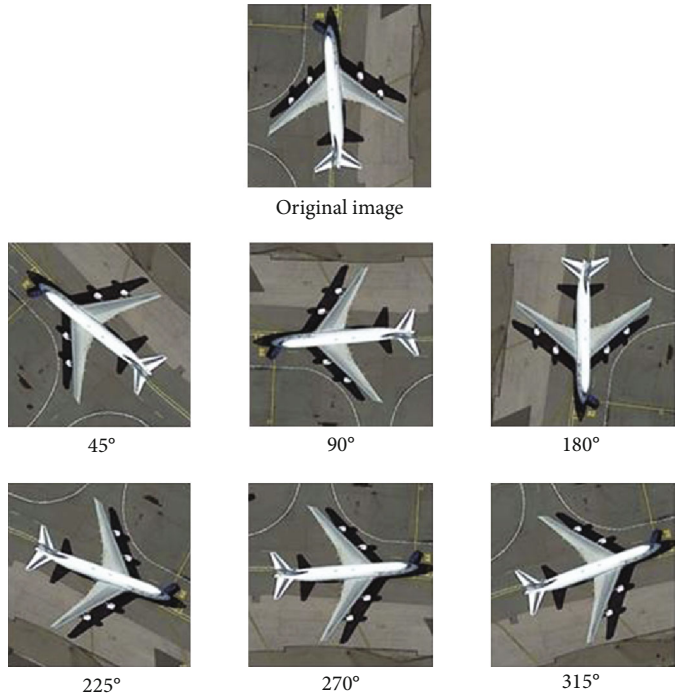

Figure 6: Six kinds of rotation transform on the aircraft image.

transformations on it as shown in Figure 6. The SPM and R-SPM models, respectively, are used to map the extracted features for further sparse representation. The length of dictionary of the sparse representation is set to be 300. After each transformed image in Figure 6 gets its vector representation, their difference degree (Diff) values are defined in equation (8), and the results are shown in Table 2.

$$\text{Diff} = \frac{\sqrt{\left(V - V'\right)\left(V - V'\right)^{T}}}{\sqrt{VV'}}, \tag{8}$$

where V is the vector of the original image, and V' represents the vector of the transformed vector.

As shown in Table 2, with the rotation transformation of the image, the different degree values of SPM become much larger. However, the proposed R-SPM model does not have changes in degree and always keeps a relatively low level which supports the idea that R-SPM has a strong adaptability for the rotation transformation of images.

In order to further compare R-SPM with SPM in terms of performance, we designed experiments to utilize SPM and R-SPM models to map SIFT features and use SVM and AdaBoost algorithms to recognize the images selected from Caltech-256. The kernel of SVM is set as a histogram intersection kernel. The combination of SPM and SVM is set as SSVM, while the combination of SPM and AdaBoost is set as SABT. R-SPM and SVM are set as RSVM, and R-SPM and AdaBoost are set as RABT. The length of reference is set as 400, and other parameters are selected as the above experimental setups.

Table 2: The difference degree values of transformed images.

Diff	45°	90°	180°	225°	270°	315°
SPM (%)	12.15	22.42	28.95	34.26	32.41	36.50
R-SPM (%)	2.33	1.94	1.61	1.84	2.19	1.42

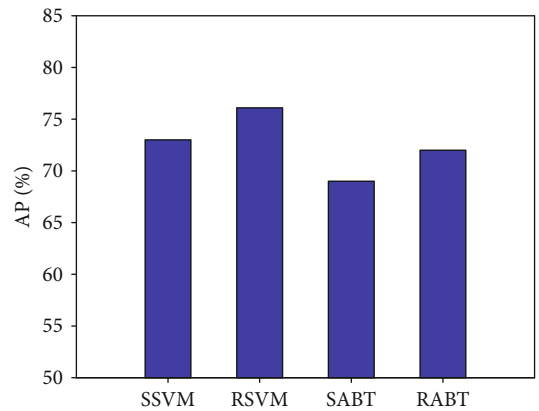

Figure 7: The AP performance of SPM and R-SPM.

It is clearly evident from Figures 7 and 8 that the average accuracy of R-SPM is higher than the 2-3% SPM in terms of the classification performance. The performance of RSVM in the four models is also optimal. At the same time, the image recognition accuracy of each category using R-SPM is higher than SPM, which confirms that the proposed R-SPM model can better obtain the spatial characteristics of images and makes the image vector representation be more selective and robust.

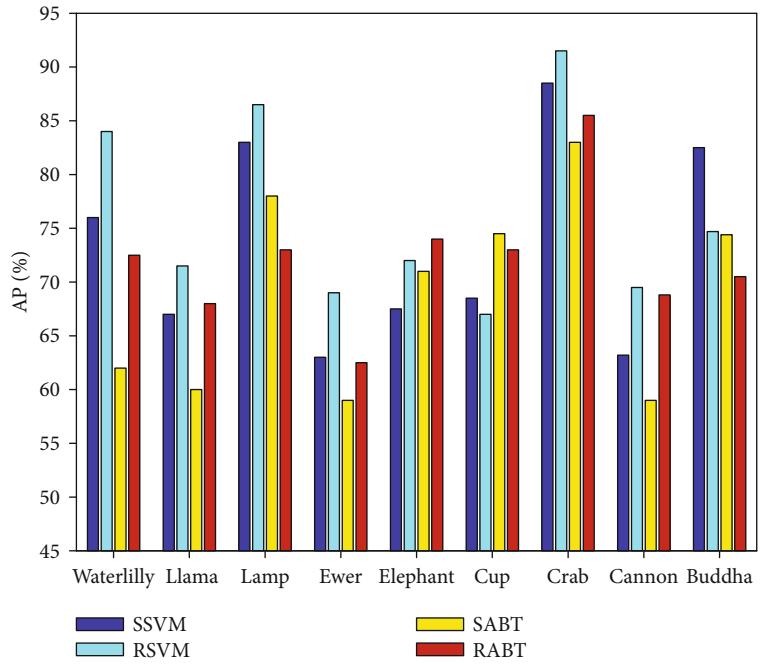

Figure 8: The precision values of each category using SPM and R-SPM.

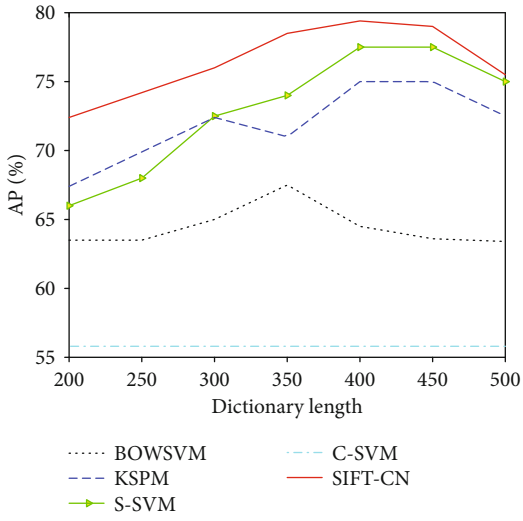

Figure 9: The performance of each model implemented on Caltech-256 with different lengths of the sparse dictionary.

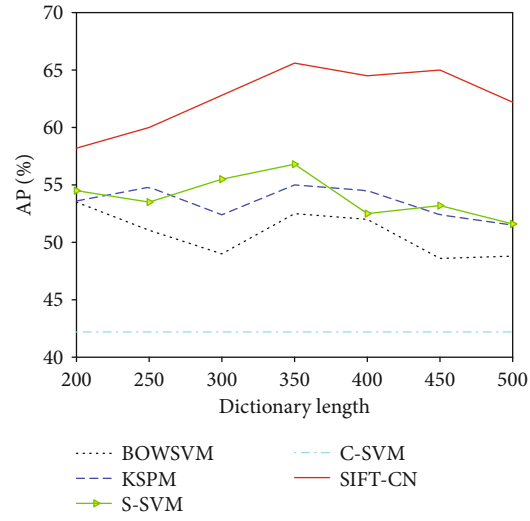

Figure 10: The performance of each model implemented on Caltech-256 with different lengths of the sparse dictionary.

5.3. The Performance of SVM-2K Algorithm. The experiment in this section also adopts BOWSVM [31, 32] and KSPM [33, 34] which are implemented on the Caltech-256 and PASCAL VOC 2011 datasets for the comparison with the proposed model. The feature fusion model in this thesis is set as SIFT-CN; the model using P-SIFT and SVM is set as S-SVM; and the term using CN and SVM is set as C-SVM. In the SVM-2K algorithm of the proposed method, HIK is utilized as the kernel of P-SIFT, and the lineal kernel works as the kernel of CN. The length of the dictionary in the sparse representation varies from 200 to 500 to obtain the optimal value. The performance of the mentioned models is shown in Figures 9 and 10.

As shown in Figures 9 and 10, with the same length of the dictionary, the proposed model using P-SIFT and CN for feature fusion achieved the best performance of all classification models. Compared with other models, the AP results of the fusion model increased by 5-10%. It is clear that the image classification model proposed in this paper has a better recognition effect on the image types. When the sparse dictionary length reaches 400, the performance of SIFT-CN is optimal. Therefore, the length of the dictionary is fixed at the same time for BOWSVM, KSPM, and SIFT-CN image classification models implemented on the rest of the experiments.

Table 3: The Kappa performance results of each model when the dictionary is set as 400.

	BOWSVM	KSPM	S-SVM	C-SVM	SIFT-CN
Caltech-256 (%)	51	58	62	49	68
VOC 2011 (%)	48	53	59	43	63

To better illustrate the performance of the image classification model, we follow recent literature [35, 36] that uses the kappa coefficient (equation (9)) as a measure of standards. The essence of the kappa coefficient is to measure the agreement between the interpretations of different observers. When the Kappa coefficient is -1, a negative correlation is expressed, while when the Kappa coefficient is 1, it indicates that the classification result is in complete agreement.

$$\text{Kappa} = \frac{po - pe}{1 - pe}. \tag{9}$$

Table 3 gives the Kappa performance of each image description model with a 400 length of the dictionary. Taken the Caltech-256 dataset for example, the Kappa performance result of the SIFT-CN model is 0.68, which is nearest to 1 among models listed in Table 3. It can be concluded that the SIFT-CN model is better than the other classification models. It shows that the results of the feature fusion model have a high reliability, in other words, the corresponding classification performance is the most ideal. And in the interior of the SIFT-CN model, Table 3 shows that the feature fusion method is superior to the single feature performance which further verifies the excellent performance of the proposed model.

Through a series of experiments, we can see that the P-SIFT feature modified by this paper has the ability to increase the speed of feature coding through the dimension reduction processing. At the same time, compared with the traditional SPM model, the proposed R-SPM model maps P-SIFT and CN features, which have the ideal rotation invariant phase position information. Finally, experimental comparisons show that the use of a SVM-2K feature fusion classification algorithm is significantly better than using a single feature classification model.

6. Conclusion

In this paper, we propose a method of the rotation invariant spatial pyramid mapping and a feature fusion-based image classification model. In feature extraction and representation, this paper proposes a novel P-SIFT feature, which is the dimension-reduced SIFT feature vector that extracts the principal information of SIFT features to speed up the sparse representation. At the same time, CN and P-SIFT features are fused to describe the image. In order to explore the location information of the feature, an annular division-based spatial pyramid model is proposed, which makes the space vector representation of the features invariant in rotation. At the stage of feature fusion, the SVM-2K algorithm is used to train two independent SVM models, and the final result is obtained by a weighted voting method. In the experimental evaluations, the Caltech-256 and VOC PASCAL image databases are selected. The experimental results show that, compared with other single feature image classification model, our feature fusion-based image description model can extract the characteristic information of the image, so that P-SIFT and CN characteristic can be used as complementary descriptions of the image and finally achieve the satisfying performance improvement. However, we acknowledge that beside the features that we investigated in this paper, there are many other advanced features which are required to be studied in our future work, while the robust should be concerned which means different kinds of data could be tested.

Data Availability

The image data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is supported partly by the University-Level Research Fund Project of Nanjing Institute of Technology (No. ZKJ201907).

References

[1] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2169–2178, New York, NY, USA, 2006.

[2] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1794–1801, Miami, FL, USA, June 2009.

[3] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 3360–3367, San Francisco, CA, USA, June 2010.

[4] H. Wang, Z. Chen, X. Wang, and Y. Ma, "Random finite sets based UPF-CPHD multi-object tracking," *Journal on Communications*, vol. 33, no. 12, pp. 147–153, 2012.

[5] F. S. Khan, J. van de Weijer, and M. Vanrell, "Modulating shape features by color attention for object recognition," *International Journal of Computer Vision*, vol. 98, no. 1, pp. 49–64, 2012.

[6] F. S. Khan, J. Weijer, A. D. Bagdanov, and M. Vanrell, "Portmanteau vocabularies for multi-cue image representation," in *Advances in Neural Information Processing Systems*, pp. 1323–1331, Granada, Spain, 2011.

[7] Y. Wang, J. Liu, J. Wang, Y. Li, and H. Lu, "Color names learning using convolutional neural networks," in *2015 IEEE International Conference on Image Processing (ICIP)*, pp. 217–221, Quebec City, QC, September 2015.

[8] F. S. Khan, R. M. Anwer, J. Weijer, A. D. Bagdanov, M. Vanrell, and A. M. Lopez, "Color attributes for object

detection," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3306–3313, Providence, RI, USA, June 2012.

[9] J. D. Farquhar, D. R. Hardoon, H. Meng, J. Shawe-Taylor, and S. Szedmak, "Two view learning: SVM-2K, theory and practice," in *Advances in Neural Information Processing Systems*, pp. 355–362, Vancouver, BC, Canada, 2005.

[10] Y. Lu, *Research on targets detection method in high resolution remote sensing image*, Hohai University, 2016.

[11] W. Zheng, X. Zhou, C. Zou, and L. Zhao, "Facial expression recognition using kernel canonical correlation analysis (KCCA)," *Neural Networks*, vol. 17, no. 1, pp. 233–238, 2006.

[12] X. Wang, X. Yan, G. Lv, and T. Fan, "Balloon-borne spectrum–polarization imaging for river surface velocimetry under extreme conditions," *Infrared Physics & Technology*, vol. 58, pp. 5–11, 2013.

[13] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[14] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 4, pp. 433–459, 2010.

[15] S. E. Hickman, N. D. Kingery, T. K. Ohsumi et al., "The microglial sensome revealed by direct RNA sequencing," *Nature Neuroscience*, vol. 16, no. 12, pp. 1896–1905, 2013.

[16] H. Lee, A. Battle, R. Raina, and A. Y. Ng, "Efficient sparse coding algorithms," in *Advances in Neural Information Processing Systems*, pp. 801–808, Vancouver, BC, Canada, 2006.

[17] Z. Jiang, Z. Lin, and L. S. Davis, "Label consistent K-SVD: learning a discriminative dictionary for recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2651–2664, 2013.

[18] D. Arthur and S. Vassilvitskii, "K-means++: the advantages of careful seeding," in *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms. Society for Industrial and Applied Mathematics*, pp. 1027–1035, Philadelphia, PA, USA, 2007.

[19] M. Tan, I. W. Tsang, and L. Wang, "Matching pursuit lasso part I: sparse recovery over big dictionary," *IEEE Transactions on Signal Processing*, vol. 63, no. 3, pp. 727–741, 2013.

[20] W. Guo, N. Xiong, A. V. Vasilakos, G. Chen, and C. Yu, "Distributed k–connected fault–tolerant topology control algorithms with PSO in future autonomic sensor systems," *International Journal of Sensor Networks*, vol. 12, no. 1, pp. 53–62, 2012.

[21] J. X. Wu and J. M. Rehg, "Beyond the euclidean distance: creating effective visual codebooks using the histogram intersection kernel," in *2009 IEEE 12th International Conference on Computer Vision*, pp. 630–637, Kyoto, Japan, September 2009.

[22] H. Liang, J. Zou, Z. Li, M. J. Khan, and Y. Lu, "Dynamic evaluation of drilling leakage risk based on fuzzy theory and PSO-SVR algorithm," *Future Generation Computer Systems*, vol. 95, pp. 454–466, 2019.

[23] Q. Wang, X. Zhang, M. Li, X. Dong, Q. Zhou, and Y. Yin, "Adaboost and multi-orientation 2D Gabor-based noisy iris recognition," *Pattern Recognition Letters*, vol. 33, no. 8, pp. 978–983, 2012.

[24] J. Li, N. Xiong, J. H. Park, C. Liu, S. MA, and S. E. Cho, "Intelligent model design of cluster supply chain with horizontal cooperation," *Journal of Intelligent Manufacturing*, vol. 23, no. 4, pp. 917–931, 2012.

[25] C. C. Chang and C. J. Lin, *LIBSVM: a library for support vector machines*, 2001, Software, http://www.csie.ntu.edu.tw/cjlin/libsvm.

[26] G. Griffin, A. Holub, and P. Perona, *Caltech-256 Object Category Dataset*, California Institute of Technology, 2007.

[27] K. Chatfield, V. Lempitsky, A. Vedaldi, and A. Zisserman, "The devil is in the details: an evaluation of recent feature encoding methods," *BMVC*, vol. 2, no. 4, p. 8, 2011.

[28] J. Li, X. Li, B. Yang, and X. Sun, "Segmentation-based image copy-move forgery detection scheme," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 3, pp. 507–518, 2015.

[29] Y. Yue, T. W. Finley, F. Radlinski, and T. Joachims, "A support vector method for optimizing average precision," in *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '07*, pp. 271–278, New York, NY, USA, July 2007.

[30] Y. Liu, M. Ma, X. Liu, N. Xiong, A. Liu, and Y. Zhu, "Design and analysis of probing route to defense sink-hole attacks for Internet of Things security," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 1, pp. 356–372, 2020.

[31] A. P. López-Monroy, M. Montes-y-Gómez, H. J. Escalante, A. Cruz-Roa, and F. A. González, "Improving the BoVW via discriminative visual n-grams and MKL strategies," *Neurocomputing*, vol. 175, pp. 768–781, 2016.

[32] M. Xu, Q. Sun, Y. Lu, and C. Shen, "Nearest-neighbors based weighted method for the BOVW applied to image classification," *Journal of Electrical Engineering & Technology*, vol. 10, no. 4, pp. 1877–1885, 2015.

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.

[34] H. Liang, J. Zou, K. Zuo, and M. J. Khan, "An improved genetic algorithm optimization fuzzy controller applied to the wellhead back pressure control system," *Mechanical Systems and Signal Processing*, vol. 142, article 106708, 2020.

[35] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2094–2107, 2014.

[36] F. Long, N. Xiong, A. V. Vasilakos, L. T. Yang, and F. Sun, "A sustainable heuristic QoS routing algorithm for pervasive multi-layered satellite wireless networks," *Wireless Networks*, vol. 16, no. 6, pp. 1657–1673, 2010.