Comparison of Reference-set Selection Methods for Reference-based Scheme

Qun Li, Ding Xu, Le An, Xiaochuan Sun, and Chi Zhang,

***Abstract*—Reference-based representation scheme achieves significant reduced dimensionality and better discriminability compared to using original image features. More specifically, a greater degree of diversity among the reference-set increases the probability that the descriptor of different images will be unique and well-spread in the data space while maintaining low redundancy. To ensure good diversity of reference-set, we study and compare five different methods to automatically construct the reference-set, including max-variation scheme, max-mean scheme, max-correlation scheme, min-correlation scheme, and affinity propagation algorithm to the task of scene categorization on some benchmark datasets. Extensive experimental results demonstrate that the proposed reference-set selection methods improves reference-based scene categorization method and outperforms *state-of-the-art* methods.**

***Index Terms*—Scene image categorization, reference-set selection, min-correlation, affinity propagation algorithm (APA).**

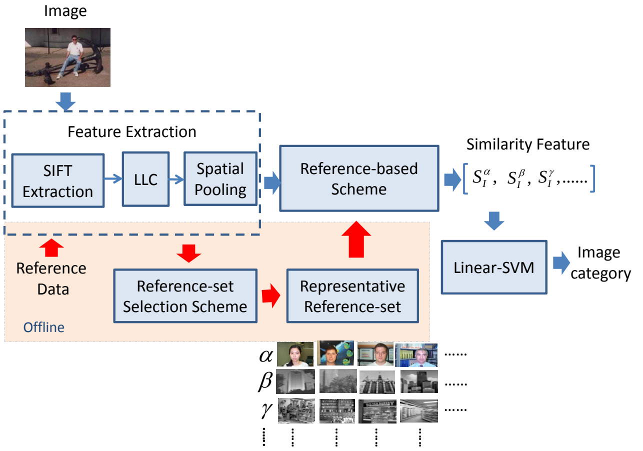


Fig. 1. Illustration of the reference-based scheme with a reference-set selection process. The reference-based scene image categorization method utilizes a reference-set composed of images from different categories, and the representation of a given image is generated by computing the similarities between this image and the images in the reference-set. The image similarity is computed using the extracted image features. To achieve better performance, before the reference-based scheme, the reference-set selection scheme is processed which constructs a representative reference-set.

I. Introduction

Image categorization is a fundamental problem in computer vision, which offers noteworthy assistance to other computer vision problems, such as image retrieval, image completion, face recognition [1], human activity analysis, object recognition [2] [3], so it has attracted a lot of interest and efforts in recent years. A well designed Bag-of-words (BoW) combined with spatial pyramid matching (SPM) model [4] trained on a single feature, such as the sparse coding (SC) method [5] and locality-constrained linear coding (LLC) [6], achieves *state-of-the-art* performance. Sadeghi and Tappen [7] design a representation based on discriminative scene regions. Zhang *et al.* [8] learn structured low-rank representations for image classification. A probabilistic label tree model is introduced by Liu *et al.* [9] for efficient large scale image classification with significantly improved recognition accuracy.

Although the problem of scene categorization has been approached in a variety of different angles, it remains very challenging mainly due to the complexity and diversity of the scene image itself. A recent method proposes to replace the traditional feature-based representation of individual images by a similarity representation that could be applied to structural models [10]. Reference-based scheme is an image-to-class measure, instead of an image-to-image measure widely used in previous work. It appears that a similarity-based representation defined on top of a traditional feature-based representation possesses unique advantages in classification. Specifically, classification in reference space is more robust and less affected by intra-class variations in the original image feature space [11]. However, for the reference-based scheme [10], a reference-set is selected randomly to form a set of basis, in which the image is to be represented. To achieve better performance a key issue is how to construct a representative reference-set space [11].

In this work, we show that a more descriptive and discriminative reference-set which is generated automatically and well spans the data space while maintaining low redundancy can further improve the classification accuracy.

II. Selecting Reference-set for Reference-based Scene Image Categorization

The reference-based scene image categorization method utilizes a reference-set composed of images from different categories, and the representation of a given image is generated by computing the similarities between this image and the images in the reference-set. The image similarity is computed using the extracted image features. In the original method [10], a reference-set is a set of randomly generated images. To achieve better performance, one important issue is about how to construct a representative reference-set space. Fig. 1 illustrates the proposed method by incorporating reference-set selection process into the reference-based scheme. Before executing the reference-based scheme, we

Manuscript received March 24, 2016; revised July 18, 2016. This work was partially supported by Nanjing University of Posts and Telecommunications Program under Grant No. NY213071, No. NY214062 and NY213086, National Natural Science Foundation of China under Grant No.61401218, 61571238 and 61302158.

Qun Li, Ding Xu and Chi Zhang are with the Nanjing University of Posts and Telecommunications, Nanjing, China. e-mail: liqun@njupt.edu.cn and xuding@njupt.edu.cn.

Le An is with the University of California, Riverside.

Xiaochuan Sun is with the Hebei United University.

Algorithm 1 The Max-variation Algorithm

```
Require: the reference data.
Ensure: the reference-set.
1: For each image $f_i$ in subclass of reference data, calculate
   $m_i = \frac{\sum_{j=1,j\neq i}^{Q} s(f_i,f_j)}{Q-1}$,
   $v_i = \frac{\sum_{j=1,j\neq i}^{Q} (s(f_i,f_j)-m_i)^2}{Q-1}$.
2: Sort $v_i$ in descending order.
3: Select the top $n$ images as reference-subset. Repeat for all subclasses of reference-data.
4: Concatenate all reference-subsets to generate the reference-set.
```

Algorithm 2 The Min-correlation Algorithm

```
Require: the reference data.
Ensure: the reference-set.
1: For each subclass of reference data, remove the image whose average correlation to other images in the reference-subset is the highest. If $C$ is the covariance matrix, $C = COV(X)$, then correlation coefficients $C_{COEF}(X)$ is the matrix whose $(i,j)$'th element is
   $C_{COEF}(i,j) = \frac{C(i,j)}{sqrt(C(i,i)*C(j,j))}$,
   $sqrt$ is square root of codistributed array.
2: Repeat the first process until $n$ reference images is obtained.
3: Repeat the above two steps for all subclasses of reference-data.
4: Concatenate all selected reference-subsets to generate the reference-set.
```

perform reference-set selection using reference data (candidates of reference-set) based on LLC features. Note that the reference-set selection can be done offline and the efficiency for online image categorization will not be affected.

A greater degree of diversity in the reference-set increases the probability that the descriptor of images from different categories will be distinct. We consider five different selection rules including *max-variation scheme, max-mean scheme, max-correlation, min-correlation scheme and APA*.

A. Max-variation and Max-mean Schemes

The max-variation scheme selects reference-set with the largest variances. Let $F = \{f_1, f_2, \ldots, f_Q\}$ be the reference data which is the candidate pool of reference-set, and $s(x, y)$ be the similarity between images x and y. Algorithm 1 shows the key steps of the max-variation algorithm.

To avoid selecting reference-set resulting in sparse index codes (i.e., index codes that contain many zeros), the operator in the max-variation algorithm is replaced by the sample mean operator, which is named max-mean algorithm. The max-mean rule selects images have a large mean value. The program of this rule is the same with the max-variation algorithm except v_i is replaced by m_i as shown in algorithm 1.

B. Min-correlation and Max-correlation Schemes

The above two selection schemes do not consider similarities among images in the reference-set. Thus, some of images in the reference-set may have very similar characteristics and result in redundant entries. Aiming to overcome this drawback, we present the min-correlation scheme to reduce the pairwise correlation among the reference images.

The min-correlation scheme selects an optimal reference-set in the algorithm 2. For comparison, the max-correlation scheme is also presented in this paper. It is same to the min-correlation scheme, except that the first step is changed to 1 : For each subclass of reference data, remove the image whose average correlation to other images in the reference-subset is the lowest.

C. Affinity Propagation Algorithm

An idea is reference-set should contain exemplar images from all classes while maintaining compactness. APA outperforms other clustering methods and is able to select exemplar data samples automatically without a predefined n. Inspired by its merits, we adapt APA [12] to automatically identify the exemplar images in the reference candidate pool.

Originally designed for data clustering, affinity propagation selects exemplars by message passing between data points. Two kinds of message, "responsibility" $r(i;k)$ and "availability" $a(i;k)$, are exchanged between data points i and k. $r(i;k)$ reflects the accumulated evidence for how well suited point k is to serve as the exemplar for point i, and $a(i;k)$ reflects the accumulated evidence for how appropriate it would be for point i to choose point k as its exemplar. After initialization, such as $a(i;k) = 0$, the responsibility and availability are computed in iteration as follows:

$$r(i,k) \longleftarrow s(i,k) - \max_{k' \neq k}\{a(i,k') + s(i,k')\}, \tag{1}$$

$$a(i,k) \longleftarrow \min\left\{0, r(k,k) + \sum_{i' \notin \{i,k\}} \max\{0, r(i',k)\}\right\}, \tag{2}$$

where $s(i;k)$ denotes the similarity between point i and point k. When the goal is to minimize squared error, each similarity is set to a negative squared error (Euclidean distance): $s(i,k) = -||x_i - x_k||^2$. The algorithm stops once the exemplar decisions remain unchanged or the maximum allowed iteration number is achieved. In our problem, we define the similarity as the negative chi-square distance [12].

III. Experimental Results

A. Image Datasets and Parameters

1) Image Datasets: The proposed reference-set selection schemes are tested on four commonly used image databases: fifteen scene categories [13], Caltech-101 [14], UIUC sports event dataset [12] and Pascal VOC2007 [15].

The fifteen scene categories is one of the most complete scene category dataset used in the literature. The number of images per category varies from 200 to 400, and the average image size is 300 × 250 pixels. The major sources of the pictures in the dataset include the COREL collection, personal photographs, and Google image search.

The Caltech-101 dataset contains 9144 images of both natural and man-made objects belonging to 101 categories.

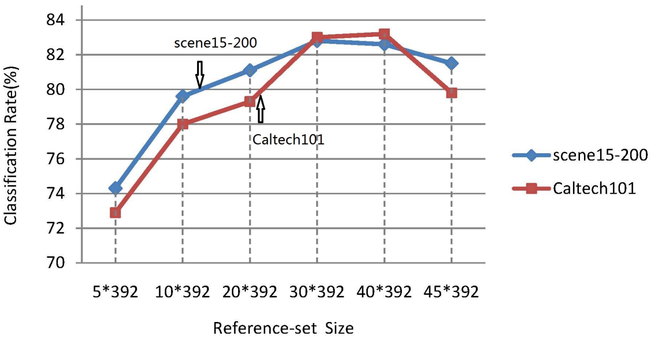


Fig. 2. The classification rate with different sizes of reference-subset on fifteen scene categories(Scene-15) and Caltech-101. We saw that we can achieve best results when the size is 30 of reference-subset.

Each category has 31 to 800 images, and most images are of medium resolution, i.e., about 300 × 300 pixels.

The PASCAL VOC2007 dataset is a most challenging dataset which holds of 9,963 images in 20 classes. All the images in this dataset are daily pictures got from Flicker where the size, viewing angle, illumination, appearances of objects and their poses vary greatly, with frequent occlusions.

UIUC Sports Event dataset contains 8 categories including badminton, bocce, croquet, polo, rock climbing, rowing, sailing and snow boarding. The number of images within each category varies from 137 to 250, and 1579 images in total.

2) Parameter Settings: In this paper, our methods are compared with each other and several *state-of-the-art* methods including the original reference-based method [10]. Throughout all the experiments, we use only the SIFT descriptors of 16×16 pixel patches computed over a grid with a spacing of 8 pixels, and 4×4, 2×2, 1×1 sub-regions for LLC. Dictionary sizes for fifteen scene categories are 200 and 400, and for Caltech-101, UIUC sports event dataset and PASCAL VOC2007, the size is 1024. We partition the whole dataset of Scene 15 into 100 training images per class and the rest for testing images, and up to 30 training images per class for the Caltech-101, and 70 images per class as training data for the UIUC sports event dataset. For PASCAL VOC2007 dataset, the training images and testing images are the same with used by PASCAL challenge.

The reference data which is the candidate pool of reference-set is collected from images in 392 different classes of fifteen scene categories, Caltech-101, Caltech-256 [16] and Pascal VOC2007. So the dimension of the final image feature is reduced significantly to 392. We repeat the experiments 10 times with different random splits of the training and testing images to obtain reliable results and the final classification rates are reported as the average of all runs.

Table I lists the size of reference-set of our proposed different methods, the first row corresponds to fifteen scene categories with 200 bases, and the next is with 400 bases. In our experiments, we stop APA once the iteration number achieves 100. From Table I, we can see that the size of the reference-set is reduced significantly by APA compared to other selecting schemes. To further validate that 30 is the proper number for reference-subset, specifically, we repeat the classification procedure using the reference-based scheme on scene-15 dataset with 200 bases and clatech-101 dataset with 1024 bases while with different reference-subset sizes as shown in Fig. 2. Obviously, the classification rate achieves the best when the size is 30 of reference-subset.

TABLE I
THE SIZE OF REFERENCE-SET WITH DIFFERENT REFERENCE-SET SELECTION SCHEMES. OBVIOUSLY, THE APA METHOD REDUCED THE REFERENCE-SET SIZE COMPARED WITH THE OTHER METHODS.

Dataset	Original	Max- variation/mean	Max/min- correlation	APA
Scene15-200	30× 392	30× 392	30× 392	18× 392
Scene15-400	30× 392	30× 392	30× 392	20× 392
Caltech101	30× 392	30× 392	30× 392	19× 392
VOC2007	30× 392	30× 392	30× 392	20× 392

TABLE II
IMAGE CLASSIFICATION RESULTS ON SCENE-15 DATABASE WITH 200 BASES AND 400 BASES. THE BEST RESULTS ARE SHOWN IN BOLD.

Classification	Accuracy(%)		Classification	Accuracy(%)	
Method	200	400	Method	200	400
Lazebnik [4]	74.5	74.8	Yang [5]	-	80.28
Gemert [17]	74.3	76.67	Wang [6]	78.5	80.2
Reference-based[10]	82.8	83.2	Max-variation	78.5	79.9
Max-mean	78.5	79.8	Max-correlation	77.6	78.9
APA	**83.8**	84.4	Min-correlation	83.5	**84.7**

B. Performance of the Proposed Methods

1) Scene Category Recognition: Results of five different reference-set selection methods are compared with each other and with several *state-of-the-art* approaches [4], [5], [6], [17] including the original reference-based scheme [10] noted as “Reference-based” in Table II. Table II shows that min-correlation yields the best results with 200 bases and APA yields the best results with 400 bases, achieving 1% and 1.5% absolute increase in terms of accuracy over the original reference-based method. Max-variation, max-mean, and max-correlation schemes do not consider or neglect similarities among images in the reference-set. Thus, some of images in the selected reference-set have very similar characteristics and result in redundant entries, further degrading the classification performance over the original reference-based method. The min-correlation scheme and APA overcome the drawback, by reducing the pairwise correlation among the reference images.

Fig.3 gives the confusion table with 400 bases of min-correlation scheme, and Fig. 4 lists the classification accuracy of each class, which reveals that some classes such as suburb, forest, street can be accurately classified by the proposed method with over 95% in terms of accuracy, and the highest block of errors occurs among the four categories: industrial, kitchen, livingroom, and bedroom.

2) Caltech-101: We randomly partition the whole dataset of Caltech-101 into 5, 10, 15, 20, 25, 30 training images per class and the rest for testing images respectively.

As can be seen from Table III, compared with several *state-of-the- art* approaches [4], [17], [5], [6], [18] including the original reference-based scheme [10], min-correlation achieves the best result with a 6.3% increase in terms of accuracy over the original reference-based method with 30 training images per class, while remarkably outperforming all the competing approaches with nearly 16% relative accuracy increase compared to the next best result. Moreover,

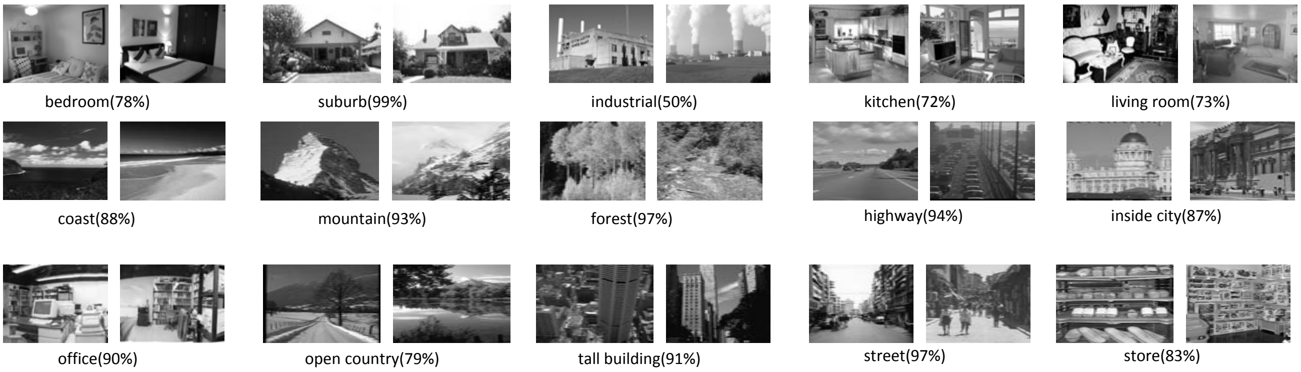
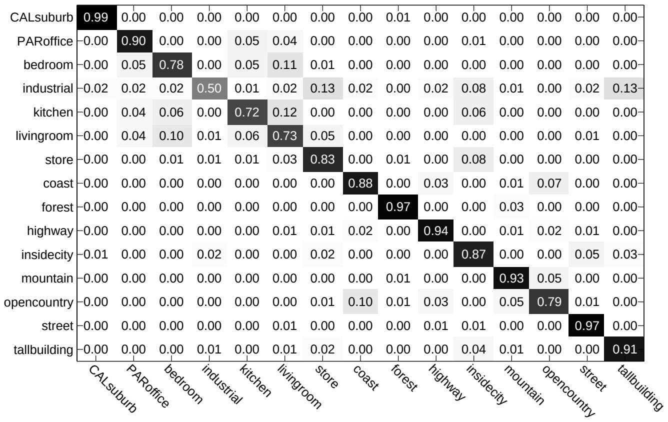


Fig. 4. Example images from Scene-15 dataset with their classification accuracy respectively using 400 bases of min-correlation scheme.



	CALsuburb	PARoffice	bedroom	industrial	kitchen	livingroom	store	coast	forest	highway	insidecity	mountain	opencountry	street	tallbuilding
CALsuburb	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
PARoffice	0.00	0.90	0.00	0.00	0.05	0.04	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
bedroom	0.00	0.05	0.78	0.00	0.05	0.11	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
industrial	0.02	0.02	0.02	0.50	0.01	0.02	0.13	0.02	0.00	0.02	0.08	0.01	0.00	0.02	0.13
kitchen	0.00	0.04	0.06	0.00	0.72	0.12	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00
livingroom	0.00	0.04	0.10	0.01	0.06	0.73	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
store	0.00	0.00	0.01	0.01	0.01	0.03	0.83	0.00	0.01	0.00	0.08	0.00	0.00	0.00	0.00
coast	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.88	0.00	0.03	0.00	0.01	0.07	0.00	0.00
forest	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.97	0.00	0.00	0.03	0.00	0.00	0.00
highway	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.02	0.00	0.94	0.00	0.01	0.02	0.01	0.00
insidecity	0.01	0.00	0.00	0.02	0.00	0.00	0.02	0.00	0.00	0.00	0.87	0.00	0.00	0.05	0.03
mountain	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.93	0.05	0.00	0.00
opencountry	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.10	0.01	0.03	0.00	0.05	0.79	0.01	0.00
street	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.97	0.00
tallbuilding	0.00	0.00	0.00	0.01	0.00	0.01	0.02	0.00	0.00	0.00	0.04	0.01	0.00	0.00	0.91

Fig. 3. Confusion table of Scene-15 dataset using 400 dictionary, the grid detector and patch based representation. The average performance is 84.7%.

TABLE III
IMAGE CLASSIFICATION RESULTS ON CALTECH-101 DATABASE. THE BEST RESULTS ARE SHOWN IN BOLD.

Classification Method	5	10	15	20	25	30
	Classification Accuracy(%)					
Lazebnik [4]	-	-	56.4	-	-	64.6
Gemert [17]	-	-	-	-	-	64.16
Yang [5]	-	-	67.0	-	-	73.2
Wang [6]	51.15	59.77	65.43	67.74	70.16	73.44
K-SVD [18]	49.8	59.8	65.2	68.7	71.0	73.2
Reference-based[10]	72.5	77.9	79.7	81.4	82.3	83.0
Max-variation	69.3	73.3	75.4	76.5	77.2	78.1
Max-mean	68.4	72.1	74.4	76.0	76.8	77.4
Max-correlation	76.1	81.7	84.6	86.0	87.3	87.5
APA	72.8	78.4	80.7	82.5	83.3	84.6
Min-correlation	**79.4**	**84.6**	**86.5**	**87.5**	**88.3**	**89.3**

the classification accuracy of min-correlation with 5 training images per class is still nearly 6% higher than the other methods. APA method achieves 1.6% increase with the most compact reference-set.

Fig. 5 presents the classification accuracy of various coding methods including LLC, k-SVD, reference-based and our five schemes with 5, 10, 15, 20, 25, 30 training images per class on the Caltech-101 dataset. Min-correlation scheme uniformly achieved lower error. At the same time, this comparison proves that reference-based scheme has better robustness to the number of training data than the other coding methods.

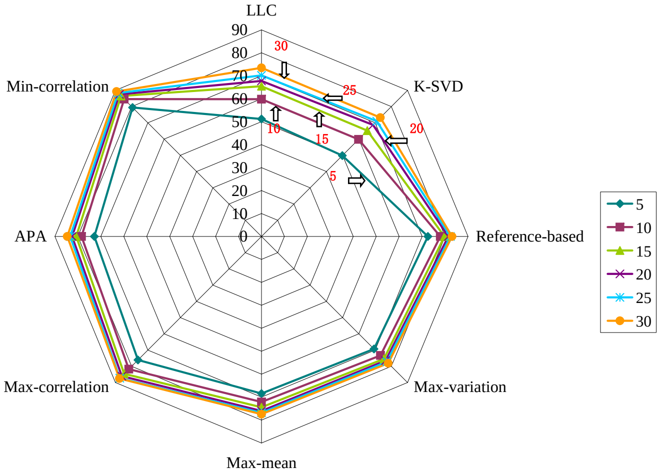


Fig. 5. Classification rate(%) comparison of various coding methods with 5, 10, 15, 20, 25, 30 training images per class on the Caltech-101 dataset. Min-correlation scheme uniformly achieved lower error. At the same time, this comparison proves that reference-based scheme has better robustness to the number of training data than the other coding methods.

3) UIUC sports event dataset: We test our algorithms on UIUC sports event dataset and compare the results with other three coding methods and the original reference-based scheme when the dictionary size is 1024. The results present in Fig.6 and Table IV. As Table IV shown, the most remarkable classification accuracy is 86.6% achieved by min-correlation scheme, with 1.7% accuracy increase compared to the next best result. We can draw the conclusion that the best classification accuracy of our methods is better than that of the other three methods, and our max-correlation, APA and min-correlation schemes promote the original reference-based scheme on the UIUC sports event dataset. Fig.6 lists the classification accuracy of each category by using min-correlation scheme. From the Fig. 6, we can see that the classification results of bocce class and croquet class are in low performance while the classification rates are much higher in the badminton class, rockclimbing class and sailing class. The backgrounds, such as person, hill and water help our schemes to improve classification rate.

4) Pascal VOC2007: For the Pascal VOC2007 dataset, the classification evaluation criterion is the standard metric used by PASCAL challenge. It computes the area under the Precision /Recall curve, and the higher the score, the better the performance.

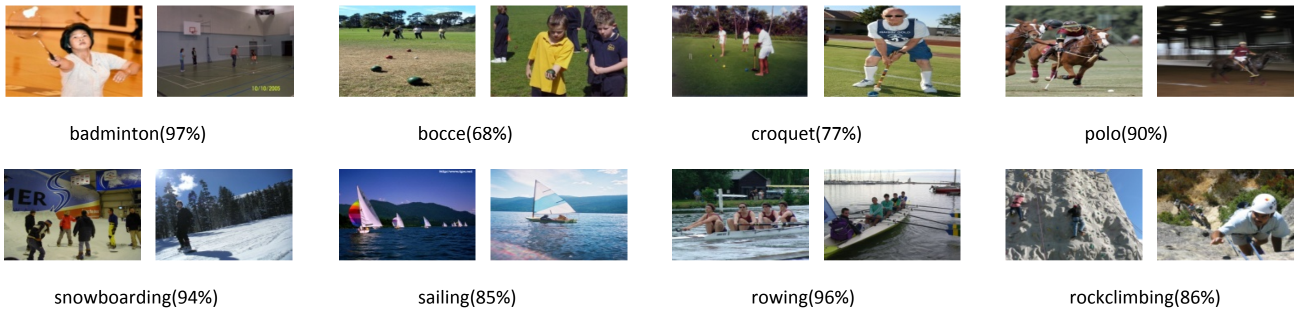


Fig. 6. Example images from UIUC sports event dataset with their average classification accuracy respectively using our min-correlation method.

TABLE V
IMAGE CLASSIFICATION RESULTS ON PASCAL VOC2007 DATABASE. THE BEST RESULTS ARE SHOWN IN BOLD FOR EACH CLASS AND AVERAGE ONE.

Object Class	aero	bicyc	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	**Average**
LLC [6]	74.8	65.2	50.7	70.9	28.7	68.8	78.5	61.7	54.3	48.6	51.8	44.1	76.6	66.9	83.5	30.8	44.6	53.4	78.2	53.5	59.3
Best PASCAL'07 [15]	77.5	63.6	56.1	71.9	**33.1**	60.6	78.0	58.8	53.5	42.6	54.9	45.8	**77.5**	64.0	85.9	36.3	44.7	50.9	**79.2**	53.2	59.4
Reference-based[10]	79.0	72.8	57.9	72.6	29.9	71.8	81.9	65.1	61.6	53.5	64.6	44.8	71.4	69.7	88.8	38.9	45.3	52.9	78.4	59.3	63.0
APA	**79.7**	73.3	58.2	72.5	33.0	**72.8**	82.1	67.3	**61.9**	55.4	64.9	44.7	76.9	**72.2**	89.8	**40.9**	**46.5**	54.9	78.4	**60.3**	64.3
Min-correlation	79.5	**73.8**	**59.8**	**73.6**	33.0	72.3	**83.5**	**68.2**	61.8	**56.7**	**66.0**	**45.9**	77.4	71.9	**90.5**	39.6	45.8	**55.5**	79.0	59.7	**64.7**

TABLE IV
IMAGE CLASSIFICATION RESULTS ON UIUC SPORTS EVENT DATASET. THE BEST RESULTS ARE SHOWN IN BOLD.

Classification Method	Classification Accuracy(%)
ScSPM [5]	81.5
LLC [6]	83.4
Salient Coding[19]	81.8
Reference-based[10]	84.9
Max-variation	83.3
Max-mean	83.2
Max-correlation	85.5
APA	86.0
Min-correlation	**86.6**

Table V lists our scores for all 20 classes by min-correlation and APA schemes in comparison with the LLC method [6], the best performance of the 2007 challenge [15], as well as the reference-based method. As seen from Table V, our min-correlation and APA schemes achieve the best performance in most classes, and the min-correlation method gets the best average score with 64.7%.

IV. CONCLUSIONS

Reference-set selection is an important topic for reference-based scene image categorization scheme. By using a few, but well chosen reference-sets, it is possible to achieve a better classification performance in both accuracy and efficiency than by using randomly selected ones. In this paper, we studied and analyzed five different approaches for reference-set selection and evaluated them on four widely used image datasets. Overall, a systematic selection of reference-set is better than a random selection. Given an appropriate reference-set size, the min-correlation scheme performs well and achieves satisfactory performance, in general. Meanwhile, the affinity propagation algorithm is able to automatically identify exemplar reference images from a large candidate pool. Experimental results showed that a more descriptive and discriminative reference-set with a greater degree of diversity increases the accuracy of classification while retaining computational efficiency compared to the *state-of-the-art* methods.

REFERENCES

[1] Z. Sufyanu, F. S. Mohamad, A. A. Yusuf, and M. B. Mamat, "Enhanced Face Recognition Using Discrete Cosine Transform," *Engineering Letters*, vol. 24, no. 1, pp. 52–61, 2016.

[2] J. T. Lalis, "A New Multiclass Classification Method for Objects with Geometric Attributes Using Simple Linear Regression," *IAENG International Journal of Computer Science*, vol. 43, no. 2, pp. 198–203, 2016.

[3] R. Maca, M. Benes, and J. Tintera, "Segmentation of MRI Images by Adaptive Degenerate Diffusion," *IAENG International Journal of Applied Mathematics*, vol. 45, no. 3, pp. 208–217, 2015.

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

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

[6] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 3360–3367.

[7] F. Sadeghi and M. F. Tappen, "Latent pyramidal regions for recognizing scenes," in *Proc. of European Conference on Computer Vision*, 2012, pp. 228–241.

[8] Y. Zhang, Z. Jiang, and L. S. Davis, "Learning structured low-rank representations for image classification," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 676–683.

[9] B. Liu, F. Sadeghi, M. Tappen, O. Shamir, and C. Liu, “Probabilistic label trees for efficient large scale image classification,” in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 843–850.

[10] Q. Li, H. Zhang, J. Guo, B. Bhanu, and L. An, “Reference-based scheme combined with k-svd for scene image categorization,” vol. 20, no. 1, pp. 67–70, 2013.

[11] L. An, M. Kafai, S. Yang, and B. Bhanu, “Reference-based person re-identification,” in *Proc. of IEEE International Conference on Advanced Video and Signal-Based Surveillance*, 2013, pp. 244–249.

[12] Q. Li, D. Xu, and L. An, “Discriminative reference-based scene image categorization,” *IEICE transactions on Information and Systems*, vol. E97-D, no. 10, pp. 2823–2826, 2014.

[13] F.-F. Li and P. Perona, “A Bayesian hierarchical model for learning natural scene categories,” in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, 2005, pp. 524–531.

[14] F.-F. Li, R. Fergus, and P. Perona, “Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories,” in *IEEE Conference on Computer Vision and Pattern Recognition Workshop on Generative-Model Based Vision*, 2004, p. 178.

[15] M. Everingham, L. Gool, C. Williams, J. Winn, and A. Zisserman, *The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results*.

[16] G. Griffin, A. Holub, and P. Perona, “Caltech-256 object category dataset,” California Institute of Technology, Tech. Rep. 7694, 2007.

[17] J. C. van Gemert, J. M. Geusebroek, C. J. Veenman, and A. W. M. Smeulders, “Kernel codebooks for scene categorization,” in *Proc. of European Conference on Computer Vision*, 2008, pp. 696–709.

[18] M. Aharon, M. Elad, and A. Bruckstein, “K-SVD: Design of dictionaries for sparse representation,” in *Proc. of Signal Processing with Adaptive Sparse Structured Representations*, 2005, pp. 9–12.

[19] Y. Huang, K. Huang, Y. Yu, and T. Tan, “Salient coding for image classification.” in *IEEE International Conference on Computer Vision*. IEEE Computer Society, 2011, pp. 1753–1760.