

## Selecting Discriminative Binary Patterns for a Local Feature

Yingying Li<sup>1,2</sup>, Jieqing Tan<sup>1,3</sup>, Jinqin Zhong<sup>1</sup>

<sup>1</sup> School of Computer and Information, Hefei University of Technology, Hefei 230009 Anhui, China

<sup>2</sup> School of Electronics and Information, Anhui Jianzhu University, Hefei 230009 Anhui, China

<sup>3</sup> School of Mathematics, Hefei University of Technology, Hefei 230009 Anhui, China

Emails: jieqingtan@hfut.edu.cn    yyhg@ahjzu.edu.cn    jinqinzhong@163.com

**Abstract:** *The local descriptors based on a binary pattern feature have state-of-the-art distinctiveness. However, their high dimensionality resists them from matching faster and being used in a low-end device. In this paper we propose an efficient and feasible learning method to select discriminative binary patterns for constructing a compact local descriptor. In the selection, a searching tree with Branch&Bound is used instead of the exhaustive enumeration, in order to avoid tremendous computation in training. New local descriptors are constructed based on the selected patterns. The efficiency of selecting binary patterns has been confirmed by the evaluation of these new local descriptors' performance in experiments of image matching and object recognition.*

**Keywords:** *Selecting patterns, searching tree, local descriptor, matching, binary pattern.*

### 1. Introduction

Local feature is widely used in many computer vision tasks, such as image retrieval [1, 2], object detection and recognition [3, 4], scene categories [4, 5] and action classification [6, 7]. The local feature is the first stage to get mid-level visual representation or high-level semantic description in most of the vision applications [1-7]. Plentiful applications drive the researches to propose a variety of local feature descriptors.

Among these presented descriptors, CS-LBP [8], MRRID [9], OC-LBP [4] have taken the limelight for their high distinctiveness, robustness to illumination and simple computation. These three local descriptors are all based on the histogram of a binary pattern feature on the given spatial pooling. The local descriptors CS-LBP and MRRID use CS-LBP histogram [8], and the local descriptor OC-LBP uses OC-LBP histogram [4]. These three descriptors have more discrimination than SIFT [10], however the higher dimensionality impedes them from matching faster and being used in a low-end device.

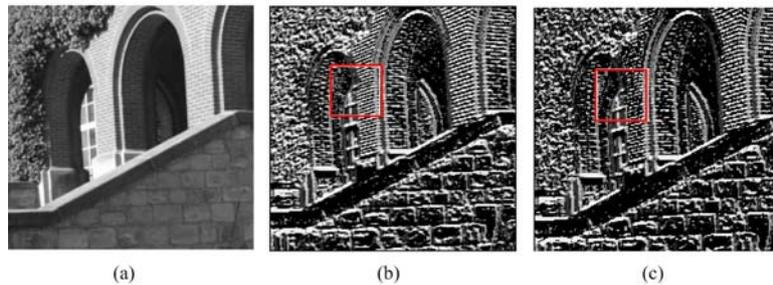


Fig. 1. CS-LBP transformed image: (a) original image, (b) CS-LBP transformed image with 16 patterns, (c) CS-LBP transformed image with 8 selected patterns

In our work we find that some patterns contribute more distinctiveness to the image description, while some patterns contribute less. For example, Fig. 1 shows two CS-LBP transformed images with 16 patterns and 8 patterns having similar distinctiveness in appearance, though (c) has lost the information of 8 patterns. Accordingly, different patterns may contribute differential discrimination to the local descriptor. We aim to select discriminative patterns to construct a more compact local descriptor and discard less informative patterns in order to reduce the dimension of the local descriptors with low loss of distinctiveness.

Many learning methods on local descriptors target at achieving a compact higher level feature, such as middle feature, image feature, leading to a different learned result in different vision scenes and tasks. Few learning is used to construct a local descriptor providing a low level feature. However, recently learning for constructing a local descriptor has received increased interest by the researchers. Ke and Sukthankar [11] use Principal Components Analysis (PCA) to the normalized gradient patch to gain a more compact descriptor. Cai, Mikolajczyk and Matas [12] propose Linear Discriminant Projections (LDP) for reducing the dimensionality of local descriptors. A large data set from multi-image 3D reconstruction is introduced by Brown, Hua and Winder [13], which makes possible the learning result to be suitable for universal application. LDB [14], RFD [15] and discriminative Learning of local image descriptors [13] are all trained on this large data set. The authors of [13] make use of Linear Discriminant Analysis (LDA) and Powell minimization to solve the problem of the parameters for gaining a more discriminative descriptor. LDB, RFD both use greedy learning techniques to select a salient bit for constructing a compact and distinctive binary descriptor. LDB uses a similar AdaBoost-based Bit Selection and RFD employs iterative optimization.

This paper’s contributions lie in: 1) we propose an efficient learning method to select discriminative binary patterns in a non-greed way; 2) we design a Loss function as a selecting criterion to evaluate the discrimination of the selected patterns in constructing a local descriptor. The efficiency of patterns selection is confirmed by evaluating the performance of the generated local descriptors based on the selected patterns in experiments of image matching and objects recognition.

## 2. The method proposed

The aim is to select a generic patterns subset for constructing a compact local descriptor. There are three problems to solve: the first one is to generalize the uniform flow for constructing a local descriptor based on a binary pattern; the second is to define a selecting criterion to decide which is the discriminative pattern for the local descriptor; the third is how to learn feasibly on a large data set to avoid a local optimum.

### 2.1. Uniform flow for constructing a local descriptor

In order to make patterns selection independent on descriptor construction, we generalize the uniform construction flow from CS-LBP, OC-LBP and MRRID, as shown in Fig. 2. Firstly, detect the region. These three descriptors can employ the same region detector. In this paper Hessian-Affine detector is used. Secondly, partition the region into several sub-regions. CS-LBP and OC-LBP use SIFT-like grid, MRRID uses intensity order pooling partition. Thirdly, compute the pattern occurrence histogram on each sub-region. CS-LBP and MRRID both use the center-symmetric local binary pattern, while OC-LBP uses orthogonal combination of a local binary pattern. At the fourth and fifth steps, these three descriptors do the same thing: concatenating the histograms of each sub-region together and generating the final local descriptor.

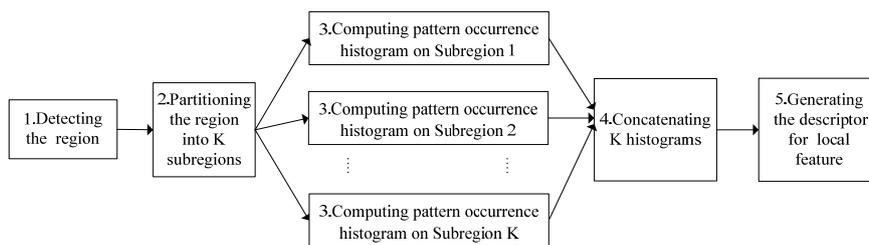


Fig. 2. The uniform flowchart for constructing a local descriptor

### 2.2. Selecting criterion

As well known, the recall versus 1-precision curve [16] is widely accepted as the evaluation criterion by many local descriptors, which assesses the performance of the local descriptor under different thresholds. In order to eliminate the influence of the different thresholds on the learning result, we evaluate the local descriptor using

its ROC area of the recall versus 1-precision curve between 80% and 95% precision, as shown in Fig. 3.

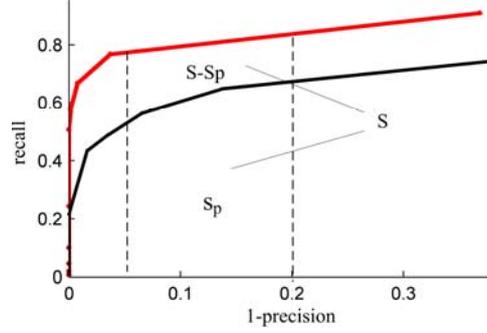


Fig. 3. The recall versus 1-precision curves for two descriptors.  $S_p$  is the ROC area between 80% and 95% precision of the descriptor below.  $S$  is the ROC area of the descriptor above

Based on ROC area, a Loss function is designed to evaluate the discrimination loss of the descriptor based on the selected patterns subset, compared with the original descriptor. It is defined as (1)

$$(1) \quad \text{Loss}(P) = 1 - \frac{S_p}{S} = 1 - \frac{\int_{0.8}^{0.95} F(x, \text{precision}_p^{i-1}, \text{recall}_p^{i-1}, \text{precision}_p^i, \text{recall}_p^i, \text{precision}_p^{i+1}, \text{recall}_p^{i+1}, \text{precision}_p^{i+2}, \text{recall}_p^{i+2}) dx}{\int_{0.8}^{0.95} F(x, \text{precision}^{i-1}, \text{recall}^{i-1}, \text{precision}^i, \text{recall}^i, \text{precision}^{i+1}, \text{recall}^{i+1}, \text{precision}^{i+2}, \text{recall}^{i+2}) dx},$$

where

$$F(x, p_1, r_1, p_2, r_2, p_3, r_3, p_4, r_4) = \left( \left( 1 + 2 \frac{x - (1 - p_2)}{p_2 - p_3} \right) r_2 + \frac{1}{2} (x - (1 - p_2)) \left( \frac{r_2 - r_1}{p_1 - p_2} + \frac{r_3 - r_2}{p_2 - p_3} \right) \right) \left( \frac{x - (1 - p_3)}{p_2 - p_3} \right)^2 + \left( \left( 1 - 2 \frac{x - (1 - p_3)}{p_2 - p_3} \right) r_3 + \frac{1}{2} (x - (1 - p_3)) \left( \frac{r_3 - r_2}{p_2 - p_3} + \frac{r_4 - r_3}{p_3 - p_4} \right) \right) \left( \frac{x - (1 - p_2)}{p_2 - p_3} \right)^2, \quad x \in [1 - p_2, 1 - p_3],$$

where  $S$  is the ROC area of the original local descriptor between 80% and 95% precision (the area below the upper curve), and  $S_p$  is the ROC area of the local descriptor based on the selected pattern subset  $P$  (the area below the lower curve);  $\text{precision}^i$  and  $\text{recall}^i$  ( $i = 0, 1, \dots, n$ ) are multiple precisions and recalls of the local descriptors generated on the candidate patterns set under different thresholds,  $\text{Precision}_p^i$  and  $\text{recall}_p^i$  ( $i = 0, 1, n$ ) are multiple precisions and recalls on the pattern subset  $P$  and  $n$  is the number of thresholds we used ( $n = 8$  in our selecting process). The call and precision are computed in (2) by patches matching in a training set using the corresponding local descriptor,

$$(2) \quad \text{recall} = \frac{\#\text{correct matches}}{\#\text{correspondences}}; \quad 1 - \text{precision} = \frac{\#\text{false matches}}{\#\text{all matches}},$$

where “#correspondences” is the ground truth number of matches in the training set; “#correct matches” is the number of correctly returned matches; “#false matches” represents the number of falsely returned matches; the total of

“#false matches” and “#correct matches” is “#all matches”. The ROC area is computed using the cubic Hermite interpolation of these precisions and recalls in the interval  $[0.8, 0.95]$  that 1-precision belongs to.

The dimensionality of the local descriptor is determined by the number of patterns used in its construction. The most discriminative patterns are selected to construct a local descriptor with low dimension at minimum loss of distinctiveness. A Loss function is used as our selecting criterion. The smaller the Loss function is, the better the corresponding pattern subset is in selecting.

### 2.3. Selecting with a searching tree

The source of our training data comes from publicly available data sets (Liberty, Yosemite, and Notre Dame [13]). Each data set contains over 400K patches, sampled densely from 3D reconstruction scenes. In each data set, the ground truth data indicating the match and mismatch has been given. In preparing the training set, we randomly chose 50K pairs of matching patches and 200K pairs of non-matching patches from Liberty and Notre Dame dataset as the training set. We use Yosemite as a testing data set.

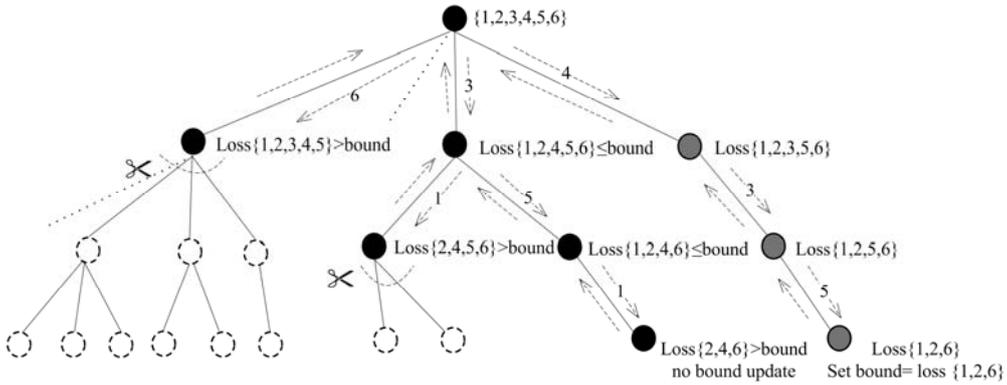


Fig. 4. Selecting with a searching tree

In order to gain a stable and generic result, we use a non-greedy learning to select the optimum. However, the evaluation of the selecting criterion for all candidate subsets in exhaustive enumeration is a huge amount of work. For each candidate pattern subset, we must construct their corresponding descriptors for all the patches in the training data set and compare their Euclidean distance in pairs to decide the match or the mismatch, finally compute its loss function as (1) by the statistics of the match and the mismatch and the given ground truth data. To avoid exhaustive enumeration and local optimum, our selecting employs the selecting tree with Branch&Bound [17], as shown in Fig. 4.

In this searching tree, the root represents the candidate pattern set before selecting and each node is the subset of patterns corresponding to a local descriptor based on this subset. The middle nodes represent the intermediate subset in selecting process and the leaves contain the final selected result. The node except the root is achieved by removing a pattern from its father node. The patterns number in the node is reduced successively with the tree level down. The depth of

the tree relates to the pattern number to be selected, which is determined by the requirement, such as memory store limitation. The rightmost successor has one branch, which is achieved by a greedy way. This means that each node in the rightmost branch (see Fig. 4) has the least loss among the subsets enumeration that are achieved by removing a pattern from its father node respectively.

We construct the asymmetric tree structure with sparser branches in the right, than in the left, as in the method given in [17]. Based on this tree structure, we try to position a good pattern subset node into the right sparse part of the tree. We need to order the nodes among their brothers according to their discrimination loss. In order to avoid a mass of heavy computation on the Loss function evaluation for each node, we utilize a simple prediction mechanism. We predict the Loss function of the subset using the distribution proportion of the corresponding discarded pattern from its father node, since the extensive experiments in [18, 19] have demonstrated that the pattern with a small proportion is inadequate to provide reliable and discriminative information. In this prediction mechanism, the nodes in the right part of the searching tree may likely provide a better selected result than in the left with high probability.

Our searching path is a pre-ordered traversal from right to left, shown as a dashed line in Fig. 4. When the node is visited, we compute its Loss function and compare it with the bound: when the Loss function of the current node is lower than the bound, we continue to visit the next node in the order of the pre-order traversal; otherwise cut off its subtrees. An initial bound is set as the Loss function of the rightmost leaf. In the traversal, the bound would be updated by the lower Loss function of the visited leaf. The bound represents the currently least discrimination loss for the visited leaves, and its corresponding leaf is saved as the currently best pattern subset for the given pattern number to be selected. The bound is also used to cut off the subtrees of the node whose loss function is higher than the bound. We speed up the searching process by cutting off the bad subset. In order to avoid a discarding possible optimum in cutting off the subtrees, the monotonicity condition should be ensured: the loss function of the subset should not be lower than that of the set which contains this subset. In general, without disturbing of noise or blurring, the pattern feature always obeys the monotonicity condition. In our training set, most patches are high-quality, the patches disturbed by the blurring and noises are only minority. So the losses of the nodes satisfy the monotonicity in our selecting.

Suppose that  $n$  is the pattern number to be selected and  $N$  is the total number of the candidate patterns. The above selecting is efficient when  $n$  is not less than  $N/2$ . When  $n$  is less than  $N/2$ , we deal with selecting in dichotomization:

Step 1. Select the  $N/2$  most discriminative patterns from  $N$  candidate patterns with the above selecting method.

Step 2. If  $n$  is not less than  $N/4$ , then select  $n$  patterns from the  $N/2$  most discriminative patterns selected at Step 1, else set the selecting  $n$  patterns from the  $N/2$  most discriminative patterns as a new selecting problem and iterate the whole process.

## 2. 4. Selecting result

Our learning is conducted three times: one is selecting from 16 patterns of CS-LBP based on a Cartesian grid (4×4); the second is selecting from 32 patterns of OC-LBP on a Cartesian grid (4×4); the third is selecting from 16 patterns of CS-LBP on intensity order pooling (4 partitions).

From the first and the second learning based on the same partition, we gain two groups of selected subsets with different sizes. We compare the discrimination of these pattern subsets with different pattern number in serving to construct a local descriptor by evaluating their corresponding generated local descriptors on a test data set Yosemite, as shown in Fig. 5. The evaluation experiment demonstrates that CS-LBP provides better distinctive patterns than OC-LBP in constructing a local descriptor. Considering the balance of performance and dimension, CS-LBP patterns serve better in constructing a local descriptor than OC-LBP.

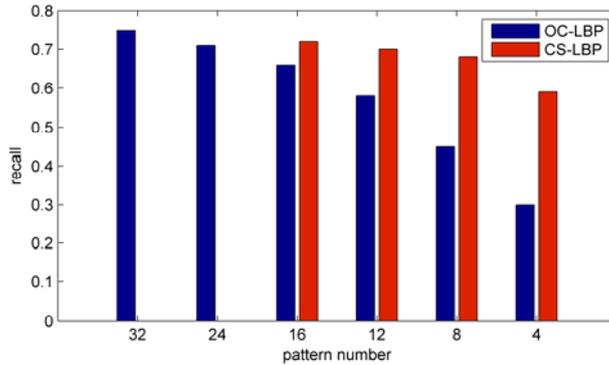


Fig. 5. The call at 95% precision for different descriptors on the selected pattern subsets with different pattern numbers. Train: Liberty and Notre Dame (250k), Test: Yosemite (100k)

For the first and the third selecting from the same pattern set on different partition, we gain the same subset. This demonstrates that our selected result is independent on the partition in local descriptor construction.

## 3. Experiments

In order to demonstrate that the selected result is discriminative and suitable for constructing a powerful local descriptor with lower dimension, we use the selected CS-LBP patterns and MRRID-like construction to generate the new local descriptors P4\_MRRID and P8\_MRRID. P4\_MRRID uses 4 selected binary patterns; P8\_MRRID uses 8 selected binary patterns; P4\_MRRID and P8\_MRRID are based on 4 partition sub-regions and 4 support regions. They both have lower dimension than their original descriptors. Their dimensions are denoted in the legend of Fig. 6, which is labelled behind the notation ‘\_’. We compare the performance of these new local descriptors with the state-of-the-art local descriptors in experiments of image matching and image recognition.

### 3. 1. Matching on the Oxford data set

The new generated descriptors based on the selected patterns are evaluated on a standard Oxford data set (<http://www.robots.ox.ac.uk/~vgg/research/affine/>), which contains images in different transformations: viewpoint change (“graf”, “wall”), scale change and image rotation (“boat”), image blur (“bike”, “tree”), illumination change (“leuven”), and JPEG compression (“ubc”). The evaluation criterion uses the call at 95% precision, where the call and the precision are defined as 2.2 (2).

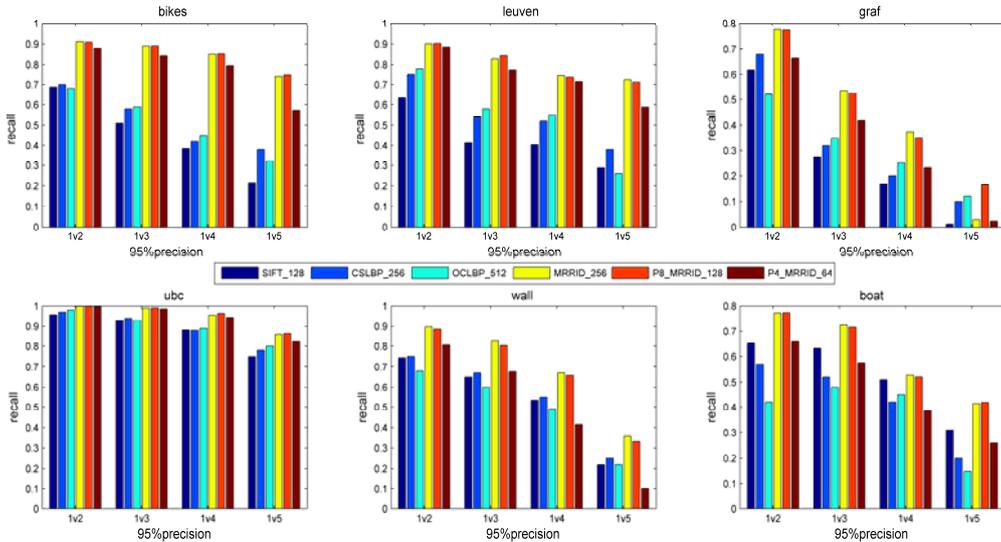


Fig. 6. Experimental results under various image transformations in the Oxford data set for Hessian-Affine Region

We compare the performance of the proposed descriptors with SIFT, original CS-LBP, original MRRID, original OC-LBP with Hessian-Affine regions, as shown in Fig. 6. P8\_MRRID has similar performance to original MRRID, while it only has half of the dimensions of the latter. It surpasses SIFT, original CS-LBP and original OC-LBP with the same dimension or lower dimension. In some cases like “leuven”, “bike”, “ubc”, P4\_MRRID has also shown pretty better distinctiveness than SIFT, original CS-LBP and original OC-LBP, though its dimension is much lower. This is mainly because of two reasons: 1) the generated local descriptor uses better construction than SIFT, CS-LBP and OC-LBP; 2) the selected pattern reserves most information of the original patterns.

### 3. 2. Object recognition on the Kentucky data set

There are 10199 images in the Kentucky data set on the website <http://vis.uky.edu/~stewe/ukbench/data/>, in groups of four that belong to the same object. One image of every group is queried as the test, and the measure is based on whether the other three images of its group are returned in the top three with the largest similarities. The recognition accuracy is defined as the ratio of the number of images returned in the top three correctly to the number of the total returned

images [10]. Suppose that  $I^1$  and  $I^2$  are two images in the data set, and  $\{f_1^1, f_2^1, \dots, f_m^1\}$ ,  $\{f_1^2, f_2^2, \dots, f_n^2\}$  are two sets of local descriptors of the same type extracted from  $I^1$  and  $I^2$ . The similarity of the two images is defined as (3):

$$(3) \quad \text{similarity} = \frac{\sum_{i,j} g(f_i^1, f_j^2)}{m \times n},$$

where

$$g(f_i^1, f_j^2) = \begin{cases} 1 & \text{if } \text{dist}(f_i^1, f_j^2) \leq T \\ 0 & \text{otherwise} \end{cases},$$

and  $T$  is a threshold which is tuned to give the best result for each evaluated descriptor.

The object recognition result is shown in Table 1. P8\_MRRID is approaching the best performance of MRRID with half of its dimensions. This means that binary pattern selecting is efficient and the selected patterns are discriminative for constructing a local descriptor. P8\_MRRID surpasses SIFT, OC-LBP and OC-LBP. P4\_MRRID has comparable performance of SIFT, OC-LBP and OC-LBP with the least dimensions among them.

Table 1. Object recognition results on the Kentucky data set

Descriptor	SIFT (128)	MRRID (256)	CS-LBP (256)	OC-LBP (512)	P8-mrrid (128)	P4-mrrid (64)
Recognition accuracy	48.2%	57.5%	49.1%	48.8%	57.0%	49.0%

#### 4. Conclusion

In the paper we propose a feasible learning method to select a discriminative binary pattern in order to construct a compact local descriptor. We design a Loss function, using the area of the call versus 1-precision curves as the selecting criterion to gain a stable and distinctive result. In selecting, a simple prediction mechanism and Branch&Bound technology are employed to accommodate the tremendous computation in training on a large data set. The experiments in image matching and object recognition demonstrate that the selected pattern is distinctive in serving for constructing a local descriptor; the generated local descriptors are compact and have comparable performance to the original local descriptors, being with lower dimensionality than their competitors.

**Acknowledgements:** This work was supported by the NSFC-Guangdong Joint Foundation Key Project under Grant No U1135003, the National Natural Science Foundation of China under Grant No 61070227 and No 61472466 and the National Science-technology Support Plan Projects under Grant No 2012BAJ08B01.

## References

1. Jégou, H., F. Perronnin, M. Douze, J. Sánchez, P. Pérez. Aggregating Local Image Descriptors into Compact Codes. – IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. **34**, 2012, No 9, pp. 1704-1716.
2. Yang, Y., S. Newsam. Geographic Image Retrieval Using Local Invariant Features. – IEEE Transactions on Geoscience and Remote Sensing, Vol. **51**, 2013, No 2, pp. 818-832.
3. Seidenari, L., G. Serra, A. Bagdanov, S. Lorenzo, S. Giuseppe, D. B. Andrew. Local Pyramidal Descriptors for Image Recognition. – IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. **36**, 2014, No 5, pp. 1033-1040.
4. Zhu, C., C. Bichot, L. Chen. Image Region Description Using Orthogonal Combination of Local Binary Patterns Enhanced with Color Information. – Pattern Recognition, Vol. **46**, 2013, No 7, pp. 1949-1963.
5. Doersch, C., S. Singh, A. Gupta. What Makes Paris Look like Paris?. – ACM Transactions on Graphics, Vol. **31**, 2012, No 4, p. 101.
6. Kaâniche, M., F. Brémont. Recognizing Gestures by Learning Local Motion Signatures of HOG Descriptors. – IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. **34**, 2012, No 11, pp. 2247-2258.
7. Thi, T. H., L. Cheng, J. Zhang, L. Wang, S. Sato. Structured Learning of Local Features for Human Action Classification and Localization. – Image and Vision Computing, Vol. **30**, 2012, No 1, pp. 1-14.
8. Heikkilä, M., M. Pietikäinen, C. Schmid. Description of Interest Regions with Local Binary Patterns. – Pattern Recognition, Vol. **42**, 2009, No 3, pp. 425-436.
9. Fan, B., F. Wu, Z. Hu. Rotationally Invariant Descriptors Using Intensity Order Pooling. – IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. **34**, 2012, No 10, pp. 2031-2045.
10. Lowe, D. G. Distinctive Image Features from Scale-Invariant Keypoints. – International Journal of Computer Vision, Vol. **60**, 2004, No 2, pp. 91-110.
11. Ke, Y., R. Sukthankar. PCA-SIFT: A More Distinctive Representation for Local Image Descriptors. – In: IEEE Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 2004, pp. 511-517.
12. Cai, H., K. Mikolajczyk, J. Matas. Learning Linear Discriminant Projections for Dimensionality Reduction of Image Descriptors. – IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. **33**, 2011, No 2, pp. 338-352.
13. Brown, M., G. Hua, S. Winder. Discriminative Learning of Local Image Descriptors. – IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. **33**, 2011, No 1, pp. 43-57.
14. Xin, Y., C. Tim. Local Difference Binary for Ultrafast and Distinctive Feature Description. – IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. **36**, 2014, No 1, pp. 188-194.
15. Fan, B., Q. Kong, T. Trzcinski, W. Zhiheng. Receptive Fields Selection for Binary Feature Description. – IEEE Transactions on Image Processing, Vol. **23**, 2014, No 6, pp. 2583-2595.
16. Mikolajczyk, K., C. Schmid. A Performance Evaluation of Local Descriptors. – IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. **27**, 2005, No 10, pp. 1615-1630.
17. Somol, P., P. Pudil, J. Kittler. Fast Branch & Bound Algorithms for Optimal Feature Selection. – IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. **26**, 2004, No 7, pp. 900-912.
18. Ojala, T., M. Pietikäinen, T. Mäenpää. Multi Resolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. – IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. **24**, 2002, No 7, pp. 971-987.
19. Bongjin, J., K. Taewan, K. Daijin. A Compact Local Binary Pattern Using Maximization of Mutual Information for Face Analysis. – Pattern Recognition, Vol. **44**, 2011, pp. 532-543.