ICIC Express Letters
Part B: Applications
Volume **7**, Number **10**, October **2016**

ICIC International ©2016 ISSN 2185-2766
pp. **2189–2196**

ONE NOVEL FACE RECOGNITION APPROACH USING HISTOGRAMS OF ORIENTED GRADIENTS DERIVING FROM MULTI-LAYER PYRAMID FEATURE BLOCKS

Hao Zhang, Zhezhou Yu, Xiangchun Yu
Shuchao Pang and Jiachen Zhang*

College of Computer Science and Technology
Jilin University
No. 2699, Qianjin Street, Changchun 130012, P. R. China
*Corresponding author: zhangjc@jlu.edu.cn

Received April 2016; accepted July 2016

Abstract. *Face recognition has been one of the challenging tasks and a long-standing problem which has been applied in more and more fields. In actual environment, however, face images are possibly obtained in complex environments inundated with illumination changes, expression variations and occlusions. In this paper, by virtue of the unique advantages of HOG features, a novel face recognition approach named Multi-layer Pyramid Feature Blocks (MPFB) was proposed. At each layer, we conduct a number of combinations that are related to the numbers, sizes, positions of the feature blocks and take the best combination to extract features with the purpose of getting more characteristic and distinguishing information from a processed standard face image. Then, we merge each feature block descriptor of a face image together so as to form more representative and integral descriptors. Besides, we employ multi-layer pyramid and construct adjacency graph for Locality Preserving Projections algorithm to reduce the dimensionality of descriptors and make the classification process less prone to over-fitting. Finally, experimental results on well-known face databases illustrate the preponderance and robustness of our approach.*
Keywords: Face recognition, Multi-layer pyramid feature blocks, Histograms of Oriented Gradients

1. **Introduction.** Face recognition technology has a variety of potential applications such as information security, surveillance, and law enforcement. However, in practice, face images are possibly obtained in complex environments inundated with illumination changes, expression variations and occlusions. Our new approach has good properties against these adverse conditions for face recognition.

Histograms of Oriented Gradients (HOG) [1], actually inherited from Scale Invariant Feature Transform (SIFT) [2], were initially proposed to detect pedestrians. The basic idea is that local appearance and structure information can often be characterized rather well by the distributions of local edge gradients or intensity gradients. Recently, Tan et al. [3,4] successfully applied Histograms of Oriented Gradients for face recognition and showed good performance. However, they extracted HOG features from the whole face images and led to capture of some less important, less distinguishing or fundamentally useless information.

Obviously, facial landmark localization can be smoothly applied for face recognition [5,6]. However, the performances of facial landmark algorithms are not 100% perfect even many algorithms [7-9] were put forward. In this case, we propose to extract features from feature blocks which will be detailedly described in Section 3.

As we can figure out, the dimensionality of HOG features is too high. So dimensionality reduction operation is required and understandable. We make use of the multi-layer pyramid, calculate neighborhoods on every layer and construct adjacency graph for LPP algorithm.

The rest of this paper is organized as follows. In Section 2, we will give relevant theories and related works including Histograms of Oriented Gradients and Locality Preserving Projections. In Section 3, we will describe our proposed method in detail. In Section 4, the benefits and robustness of our approach will be evaluated, followed by conclusions in Section 5.

2. Relevant Theories and Related Works.

2.1. Histograms of Oriented Gradients (HOG).

HOG features are based on evaluating well-normalized local histograms of image gradient orientations in a dense grid [1]. The features possess good invariance to ambient lighting changes or geometric transformations and have continued to provide one of the more robust features even in recent analyses [10-12]. Besides, some researchers have done further study [13,14] and showed HOG features are provided with good potentials. The main extraction procedures in our paper are as follows.

Step 1: Gradient, magnitude of gradient and gradient orientation computation. Considering a feature block, the grids, which are shown in Figure 1(a), and gradient estimation filters $h_x = [-1 \quad 0 \quad 1]$, $h_y = [-1 \quad 0 \quad 1]^{\mathrm{T}}$, we compute gradients of feature blocks $G_x(x, y), G_y(x, y)$ by the following formula, and $I(x, y)$ represents pixel gray value.

$$G_x(x, y) = I(x+1, y) - I(x-1, y) = I(x, y) * [-1 \quad 0 \quad 1] \tag{1}$$

$$G_y(x, y) = I(x, y+1) - I(x, y-1) = I(x, y) * [-1 \quad 0 \quad 1]^{\mathrm{T}} \tag{2}$$

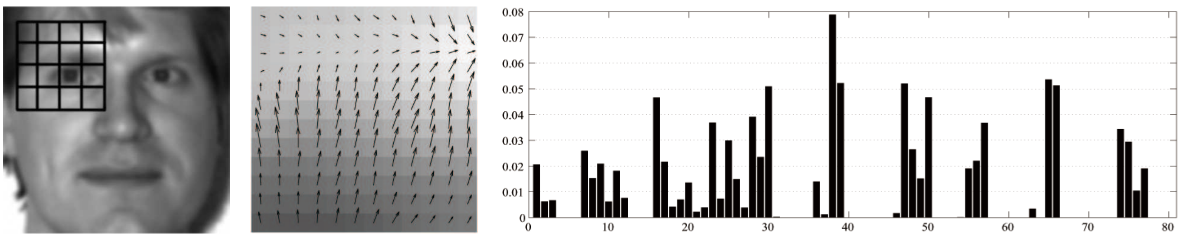
where $*$ represents convolution operation. The gradient magnitude $G(x, y)$ and orientation $\alpha(x, y)$ can be calculated by:

$$G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \tag{3}$$

$$\alpha(x, y) = \arctan \frac{G_y(x, y)}{G_x(x, y)} \tag{4}$$

Step 2: Creating orientation histograms. There are two important units named cell and block. In our paper, cell is a square region constituted by 16 (4×4) pixels. A block is composed by 9 (3×3) cells which is shown in Figure 1(b). We collect statistical information of each cell into histograms including 9 bins. And in our paper, the range of gradient direction is 360° which contains signed information.

Step 3: Cell information grouping. We group cell information into block. In the meantime, a Gaussian spatial window is applied to reducing the influencing of illumination



(a) (b) (c)

FIGURE 1. Figure 1(a) shows one of the feature blocks marked on a 100% scale face image. Figure 1(b) displays one of the blocks with composition of nine cells, magnitude of gradient and orientation of each pixel. Figure 1(c) exhibits gradient information statistical results of Figure 1(b).

changes and the weight near edges. In our paper, we do not overlap blocks since the sampling density is relatively high which is 16 (4×4) pixels per cell.

Step 4: Block normalization and descriptor vector formation. In our paper, L1-norm method is adopted while proceeding block normalization and performs better than other normalization schemes. Figure 1(c) shows the gradient information statistical results of Figure 1(b). Finally, the descriptor vector is composed with cell orientation histograms of all blocks.

2.2. **Locality Preserving Projections (LPP).** LPP [15] is a typical linear graph-based dimensionality reduction algorithm which has been successfully applied in many practical problems [16-18]. It is a linear approximation of the Laplacian eigenmap, and so much of the performance depends on the adjacency graph constructing method. The purpose is to find a transformation matrix to project the input data $X = \{X_1, X_2, \ldots, X_N\}$ with N sample images into a low-dimensional subspace Y. Given the objective function of LPP:

$$\frac{1}{2}\sum_{mn}(Y_m - Y_n)^2 S_{mn} \tag{5}$$

where $S_{mn} = e^{-\frac{\|X_m - X_n\|^2}{t}}$ when sample m is among the k nearest neighbors of sample n or sample n is among the k nearest neighbors of sample m; otherwise $S_{mn} = 0$. Besides, parameter t is adjusted according to [19].

3. Multi-Layer Pyramid Feature Blocks Using HOG Features.

3.1. **Features extraction from multi-layer pyramid feature blocks.** We, human beings, recognize individuals by multi-information such as faces, voices, and habitual behaviors. However, the face plays a more important part. Unfortunately, precisely locating local and more useful information such as eyes, nose, mouth positions and exploring their complicated correlative distributions is definitely not an easy work for computer, let alone suffering occlusions, illuminations or expression variations, even though it will help us to get less computational work and better recognition rates. Therefore, after a host of experiments and comparison, we propose to divide five regions from face image at each scale named feature blocks which possess good properties and capture the majority of important and distinctive information. More specifically, different individuals have distinctive looks, facial details and facial local structures while our feature blocks, gained from processed standard face images like AR and Yale faces, capture plenty of detailed, local structural and distinguishing information from human faces. Moreover, less useful or interferential information will be excluded.

We conduct a number of combinations that are related to the numbers, sizes and positions of feature blocks and select the best combination. The feature blocks are shown in Figure 2 as A, B, C, D, E. Besides, HOG features, insensitive to local optical deformation and geometric variations, are introduced to promote the robustness and performance. All

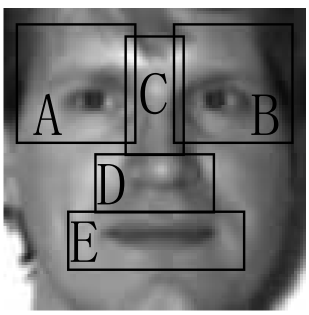


FIGURE 2. Five feature blocks of a face image

the parameters of HOG algorithm are obtained by our experiments. Considering one of training samples, the main steps of HOG features extraction from Multi-layer Pyramid Feature Blocks are as follows.

Step 1: Color, gamma normalization and multi-layer pyramid establishment. The original input image is resized to 128×128 at scale 100%. In order to get more detailed and integral information, we resize the input to 192×192 at scale 150%, 64×64 at scale 50% and construct multi-layer pyramid.

Step 2: Dividing feature blocks from face image. At scale 100%, we set 48×48, 48×48, 48×24, 24×48, 24×72 corresponding to feature block A, B, C, D, E. As for positions, at scale 100%, lengths to the upper edge for A, B, C, D, E are 4, 4, 14, 62, 86; A and B are 4 pixels away from the left and right side; C, D and E are in the middle of two vertical sides. Concerning scale 50% and scale 150%, we set 0.5 and 1.5 times value of each parameter including feature block sizes and lengths to image boundaries. In this step, we capture 15 feature blocks from three-layer pyramid. The multi-layer pyramid of first individual on Yale database is shown in Figure 3.

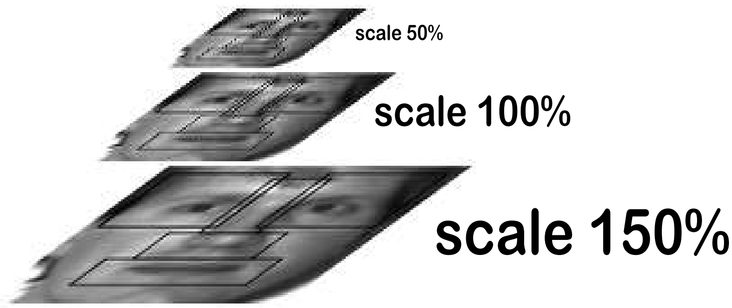


Figure 3. Multi-layer pyramid of first individual on Yale database

Step 3: Employing the method explained in Step 1 of Section 2.1. We get gradient, magnitude of gradient and gradient orientation of each feature block.

Step 4: Voting and histogram constructing. Each feature block is divided into a number of cells which are 4×4 square regions. Through vote for each bin by the magnitude of gradient and dominant orientation determination, we construct histograms of gradient for each cell. The range of gradient direction, in this paper, is $360°$ which includes signed information and performs better than $180°$. The statistical gradient information is gathered into histograms with 9 bins and we just take gradient magnitude as the weighted value into histograms.

Step 5: Collecting cell information into block. We gather cell information into block with no overlap. And the block width, in our paper, is 12 pixels. In this process, a Gaussian window is applied with $\sigma = block\_width/2$ for each block to reduce the weight of pixels far from the block center. Besides, trilinear interpolation operation is adopted to amend votes to histograms and reduce aliasing. Then, for better invariance to illumination, we perform block normalization using L1-norm method which performs best in our experiments in comparison with L1-sqrt, L2 and L2-hys.

Step 6: Final descriptor formation. Till now, we have gained the HOG descriptors of each feature block at three scales. Then, we fuse feature block descriptors with the sequence of A, B, C, D, and E to make up the final descriptor.

Finally, at each scale, a face image is represented by a fusion descriptor. Scale 100% with 60 blocks, scale 50% with 15 blocks, scale 150% with 135 blocks and each block contains 1×81 dimensionality information.

All of above-mentioned steps are executed on training samples. However, for testing samples, we just extract HOG descriptors from feature blocks at scale 100% for matching procedure.

3.2. **Dimensionality reduction using improved LPP algorithm and classification process.** In order to remove noises and make the classification process less prone to over-fitting, dimensionality reduction operation is required. We make use of multi-layer pyramid and construct a neighbor graph for Locality Preserving Projections. The main steps are as follows.

Step 1: Supposing $\{S_1, S_2, S_3\}$ denotes the HOG descriptors of training samples. $S_1 = \{P_i | i = 1, 2, 3, \ldots, r\}$, $S_2 = \{Q_j | j = 1, 2, 3, \ldots, r\}$, $S_3 = \{R_k | k = 1, 2, 3, \ldots, r\}$, where r denotes the number of images. G denotes the neighbor graph which is an $r \times r$ square matrix, and N denotes an $r \times r$ square matrix. We initialize N to zeros.

Step 2: Computing the k nearest neighbors of each row. For more explicit explanation, supposing the mth row of S_1 is P_m, we find the second nearest neighbor (i.e., second nearest row) P_n, and add one to N_{mn}. Then we find the third, fourth, $\cdots$, kth and do the same operation. For S_1, S_2, S_3, we loop each row from 1 to r. In this case, for matrix N, the summation of each row is $3 \times (k-1)$. For example, $k = 5$, the 5 nearest neighbors of the first face image on AR database are shown in Figure 4.

Figure 4. Five nearest neighbors of the first face image on AR database at multi-layer pyramid

Step 3: For matrix N, we make summation of every element to variable SUM and count the number of nonzero elements to variable NUM. Then we compute a new matrix S. As for the weight plan, we adopt $\{0, 1\}$, given:

$$S_{mn} = \begin{cases} 1, & N_{mn} \geq \frac{SUM}{NUM} \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

each row of N_{mn} denotes the similarity between the mth descriptor and the rest, so S_{mn} is equal to G_{mn}. The final neighbor graph $G = \max\{G, G^{\mathrm{T}}\}$.

Objective function of LPP in this paper is as follows:

$$\begin{aligned} \frac{1}{2}\sum_{mn}(Y_m - Y_n)^2 S_{mn} &= \frac{1}{2}\sum_{mn}\left(W^{\mathrm{T}}X_m - W^{\mathrm{T}}X_n\right)^2 S_{mn} \\ &= \sum_{m} W^{\mathrm{T}}X_m D_{mm} X_m^{\mathrm{T}} - W^{\mathrm{T}}XS_{mn}X^{\mathrm{T}}W \\ &= W^{\mathrm{T}}XDX^{\mathrm{T}}W - W^{\mathrm{T}}XS_{mn}X^{\mathrm{T}}W \\ &= W^{\mathrm{T}}X(D - S_{mn})X^{\mathrm{T}}W = W^{\mathrm{T}}XLX^{\mathrm{T}}W \end{aligned} \tag{7}$$

where $X = \{X_1, X_2, \ldots, X_N\}$ implies the sample set with N samples. D is a diagonal matrix whose entries are the sums of row or column of weighted matrix S_{mn}. $D_{mm} = \sum_m S_{nm}$, $L = D - S$, where L denotes the Laplacian matrix. W implies the transformation vector. By the following formula, we compute W and real number λ:

$$XLX^{\mathrm{T}}W = \lambda XDX^{\mathrm{T}}W \tag{8}$$

Eventually, we employ nearest neighbor classifier. We employ training features of scale 100% and testing features of scale 100% to process matching procedure. The test sample would be classified into the same class with training sample which has the nearest distance to the test sample.

4. **Experiments Setup and Results.**

4.1. **Robust tests on AR database.** In this subsection, we conduct experiments on AR database to show the robustness against occlusions, expression variations and illumination changes of our approach. The AR database consists of over 4,000 frontal face images of 126 individuals and captures different expressions. Some of images are occluded by glasses or scarves and suffer illumination changes. In our experiment, we select the first 13 images of first 8 males and first 8 females, totally 208 face images, to construct a sub-database of AR. The first person of sub-AR database is shown in Figure 5.

In order to show the advantage of our approach, on the same sub-AR database, we also conduct experiments including PCA, LPP, and LPP using HOG features and Single-layer Feature Blocks. We make 7 groups: $\{G_i|i = 2, 3, 4, \ldots, 8\}$, $\{P_j|j = 11, 10, 9, \ldots, 5\}$. G_i and P_j denote training and testing samples, thereinto $i + j = 13$. Finally over 50 times of iterating, the average of experimental results on sub-AR is shown in Figure 6.

From recognition rates, we can draw the conclusion. On the one hand, our proposed approach performs well against occlusions, expression variations, and illumination changes, and gains a better recognition rates than some classical or relevant algorithms on sub-AR database. On the other hand, experimental results of single-layer feature blocks at scale 100% are not as good as our proposed method and demonstrate the necessity and advantage of extracting HOG features at multi-layer pyramid.

4.2. **Robust tests on Yale database.** In this subsection, we perform robust tests against illumination changes, expression variations and slight decoration. The Yale database contains 165 face images of 15 individuals, gained at different illumination intensity and expression variations. Some of face images are captured when the individual wears

Figure 5. The 13 face images of the first individual on sub-AR database

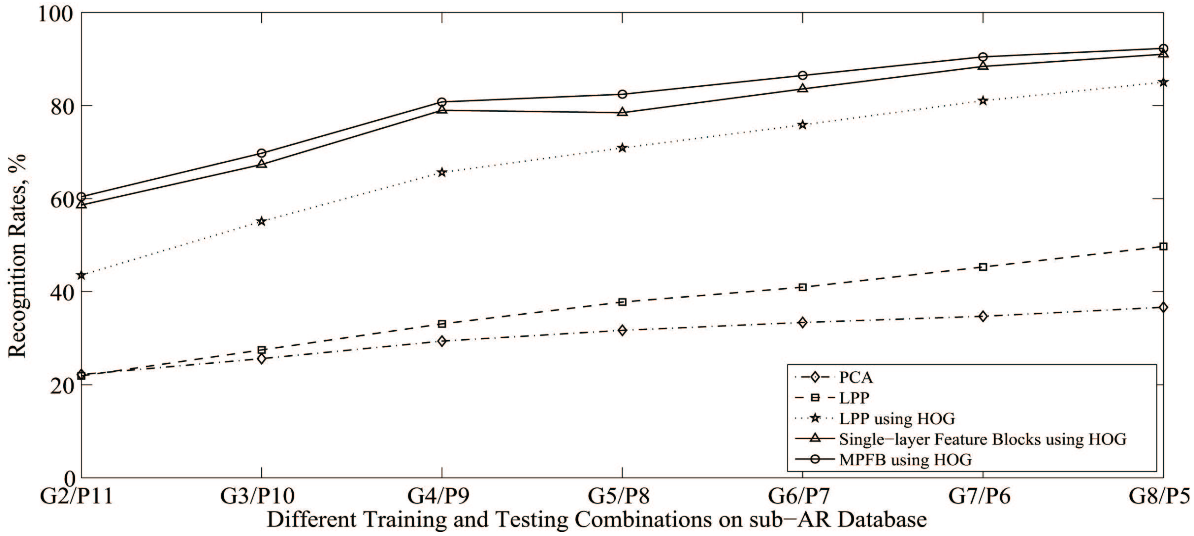


Figure 6. Recognition rates of different algorithms on sub-AR database

Figure 7. The 11 face images of the first individual on Yale database

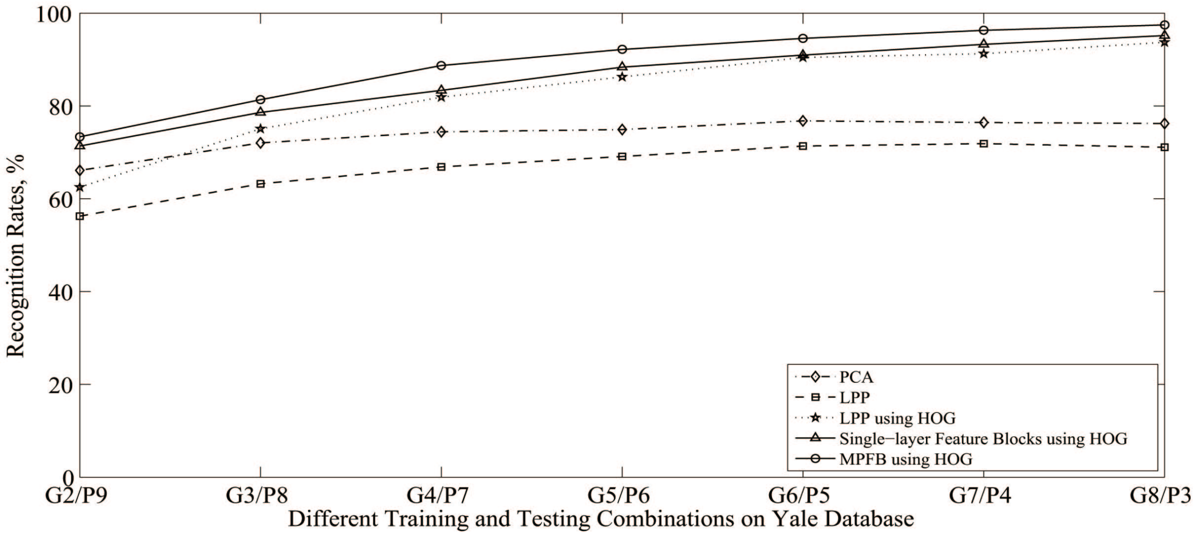


Figure 8. Recognition rates of different algorithms on Yale database

different styles of glasses. Figure 7 shows the first individual's 11 face images of Yale database. In order to minimize occasionality, we also iterate 50 times and get average recognition rates as the final result. Figure 8 shows the results of our approach and some other algorithms.

As we can see, our approach gives a better result than some classical or relevant recognition algorithms and shows a good robustness against illumination changes, expression variations and slight decoration on Yale database.

5. **Conclusions.** In this paper, we propose a novel face recognition approach to extract HOG features from Multi-layer Pyramid Feature Blocks. We create a multi-layer pyramid of three scales and extract features from specific regions named feature blocks which possess more representative and discriminative information for face recognition. Then, at each scale, we fuse feature block descriptor to bring the syncretic descriptor into being. In order to increase recognition efficiency and accuracy, we construct a neighbor graph for LPP to perform dimensionality reduction operation. Experimental results on sub-AR and Yale database show good robustness against occlusions, expression variations and illumination changes.

In future work, we plan to do further study on weight, shape and more possible elements of feature blocks. In addition, we propose to combine our proposed algorithm with remarkable face preprocessing, face detection methods and make up a precise and real-time video surveillance system for authentication, security, admission control, and so on.

Acknowledgments. This research is supported by (1) the Specialized Research Fund for the Doctoral Program of Higher Education of China (Grant No. 20120061110045) and (2) the Project of Science and Technology Development Plan of Jilin Province, China (Grant No. 20150204007GX).

REFERENCES

[1] N. Dalal and B. Triggs, Histograms of oriented gradients for human detection, *IEEE Conference on Computer Vision & Pattern Recognition*, pp.886-893, 2005.

[2] D. G. Lowe, Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision*, pp.91-110, 2004.

[3] H. Tan, B. Yang and Z. Ma, Face recognition based on the fusion of global and local HOG features of face images, *IET Computer Vision*, vol.8, no.8, pp.224-234, 2014.

[4] X. Wei, G. Guo, H. Wang and H. Wan, A multiscale method for HOG-based face recognition, *Intelligent Robotics and Applications*, pp.535-545, 2015.

[5] X. Yu, *Unconstrained Face Landmark Localization: Algorithms and Applications*, Ph.D. Theses, Rutgers, The State University of New Jersey – New Brunswick, 2015.

[6] B. M. Smith, *Exemplar-Based Face Image Parsing with Landmark-and Segment-Based Representations*, University of Wisconsin-Madison, 2014.

[7] X. Xiong and F. D. L. Torre, Superviseddescent method and its applications to face alignment, *IEEE Conference on Computer Vision & Pattern Recognition*, pp.532-539, 2013.

[8] Z. Zhang, W. Zhang, H. Ding et al., Hierarchical facial landmark localization via cascaded random binary patterns, *Pattern Recognition*, vol.48, no.4, pp.1277-1288, 2015.

[9] B. R. Campomanes-Álvarez et al., Dispersion assessment in the location of facial landmarks on photographs, *International Journal of Legal Medicine*, vol.129, no.1, pp.227-236, 2015.

[10] U. Mlakar and B. Potočnik, Automated facial expression recognition based on histograms of oriented gradient feature vector differences, *Signal, Image and Video Processing*, vol.9, no.1, pp.245-253, 2015.

[11] Z. Chen, C. Wang, C. Wen et al., Vehicle detection in high-resolution aerial images via sparse representation and superpixels, *IEEE Trans. Geoscience & Remote Sensing*, vol.54, no.1, pp.1-14, 2016.

[12] J. C. Lee, C. P. Chang and W. K. Chen, Detection of copy-move image forgery using histogram of orientated gradients, *Information Sciences*, no.321, pp.250-262, 2015.

[13] B. Li and G. Huo, Face recognition using locality sensitive histograms of oriented gradients, *Optik – International Journal for Light and Electron Optics*, 2015.

[14] H. Li et al., A novel traffic sign detection method via color segmentation and robust shape matching, *Neurocomputing*, 2015.

[15] X. F. He and P. Niyogi, Locality preserving projections, *Proc. of the Neural Information Processing Systems*, vol.6, pp.1059-1071, 2004.

[16] W. Chao, J. Ding and J. Liu, Facial expression recognition based on improved local binary pattern and class-regularized locality preserving projection, *Signal Processing*, 2015.

[17] J. Soldera, C. A. R. Behaine and J. Scharcanski, Customized orthogonal locality preserving projections with soft-margin maximization for face recognition, *IEEE Trans. Instrumentation and Measurement*, vol.64, no.9, pp.2417-2426, 2015.

[18] P. Huang and G. Gao, Local similarity preserving projections for face recognition, *AEU – International Journal of Electronics and Communications*, vol.69, no.11, pp.1724-1732, 2015.

[19] M. Belkin and P. Niyogi, Laplacian eigenmaps and spectral techniques for embedding and clustering, *NIPS*, no.14, pp.585-591, 2001.