International Journal of Multimedia and Ubiquitous Engineering
Vol.10, No.12 (2015), pp.109-118
http://dx.doi.org/10.14257/ijmue.2015.10.12.12

Face Recognition based on a Novel Nonlinear Version of LBP

Zhanghongyi and Zhao Feng

School of Optoelectronic & Communication, School of Computer Science and Technology
Engineering, Xiamen University of Technology, Shandong Institute of Business and Technology
Xiamen, China, Yantai, China
Zhanghongyi@xmut.edu.cn, zhaofeng1016@126.com

Abstract

How to extract the robust discrimination features is the key of face recognition (FR). Local binary pattern is one of the most widely used feature extracting method in FR for its comprehensive representation of the visual content of face image. However, the feature vector extracted by LBP is usually very high-dimensional and maybe contains information redundancy. To deal with the drawback of LBP, a novel nonlinear version of LBP is presented. The main idea is firstly all the feature vectors extracted by LBP are mapped into a feature space by a nonlinear mapping, and then the mapped features are expressed using the corresponding projection vectors. Lastly, FR is performed based on the projection vectors. Compared with LBP, the new method has two advantages. Firstly, it can capture the nonlinear information of the feature vector extracted by LBP. Secondly, it avoids the complex expression of the nonlinear mapping. The experimental results on two public standard visual face datasets demonstrate the proposed method is superior to LBP in recognition accuracy while its computational complexity is considerably reduced.

Keywords: *local binary pattern (LBP), Kernel method, Face recognition*

1. Introduction

As an important biometrics technology, face recognition (FR) is characterized with being direct, convenient, and non-contact. Therefore, it will have a wide application in the future in some important fields such as security, customs and finance and in some civil fields such as intelligent entrance guard, intelligent attendance and mobile phone [1-2].

From the 1960s when the research of FR was started, great progress has been made and many effective algorithms have been proposed one after another in face recognition technology. For instance, the subspace method is based on global feature [3-4], which includes principal component analysis (PCA), fisher discriminate analysis, independence component analysis and their corresponding kernel methods(KPCA,KFDA、KICA). Meanwhile, another method is based on local feature [5-6], which includes local binary pattern (LBP) and Gabor feature *et al*.

Compared with the other feature extracting methods, the feature extracted by LBP is better in the comprehensive representation of the visual content of face image and in improving the accuracy of face recognition [5-7]. So LBP becomes one of the most widely used feature extracting method in FR. However, a drawback of LBP is that the feature vector extracted by LBP is usually very high-dimensional and maybe contains information redundancy, which affects the recognition accuracy and the computational complex in FR [8-9].

In order to deal with the disadvantage of LBP, a nonlinear version of LBP (NLBP) is proposed inspired by the idea of kernel method. Compared with the kernel method, its most different is that the final process of FR is performed based on a projection vectors

ISSN: 1975-0080 IJMUE

which are expressed in the explicit. The procedure can be described as follows. Firstly, all the feature vectors extracted by LBP are mapped into the feature space by a nonlinear mapping. Secondly, based on a standard orthogonal basis spanned by all the mapped feature vectors, we get the projection vectors of all the mapped feature vectors, where the projection vectors are expressed in the explicit form. Finally, we perform FR based on the projection vectors in the projection space.

An overview of NLBP can be illustrated in Figure 1. We suppose $x_i(i = 1,2,\cdots,N)$ denotes the feature vectors extracted by LBP. N denotes the number of the feature vectors, which is consistent to the number of face images. d is the dimensionality of x_i. The whole process of NLBP can be described as follows. (1)Inspired by the kernel function, we firstly map all the feature vectors $\{x_i\}_{i=1}^{N}$ in to another feature space by a nonlinear mapping φ. Where $\{\varphi(x_i)\}_{i=1}^{N}$ denotes the mapped dataset and h denotes the dimensionality of the mapped space. Based on the step, we can obtain the nonlinear extension of $x_i(i = 1,2,\cdots,N)$. (2) Based a standard orthogonal basis $\beta_i(i = 1,2,\cdots,r)$ of the subspace which is spanned by all the mapped data $\{\varphi(x_i)\}$, we get the projection vectors $\{y_i\}_{i=1}^{N}$ which can capture the data's nonlinear information as well as $\{\varphi(x_i)\}_{i=1}^{N}$ do. (3) FR is performed based on $\{y_i\}_{i=1}^{N}$ in the projection space. In this step, according to $\{y_i\}_{i=1}^{N}$ is expressed in the explicit form and r usual be much smaller than N, we not only avoid the expression of the complex nonlinear mapping φ but also reduce the computational complexity compared with the kernel method. The contributions of NLBP can be summarized as follows. (1) NLBP can extract the nonlinear feature of FR. (2) Its computational complexity is much low for avoiding the expression of the mapping using the explicit.

$$\{x_i\}_{i=1}^{N} \subset R^d \xrightarrow{\varphi} \{\varphi(x_i)\}_{i=1}^{N} \subset R^h \xrightarrow{\beta_1,\beta_2,\cdots,\beta_r} \{Y_i\}_{i=1}^{N} \subset R^r \to FR$$

Feature space (Explicit form) | Kernel space (Implicit form) | Projection space (Explicit form)

Figure 1. Framework of NLBP

The rest of the paper is organized as follows. In Section 2, we give a brief introduction to LBP. Section 3 elucidates the proposed method--NLBP, in which we will provide in details the concrete steps for obtaining the explicit form of the mapped feature vectors and give some theoretical analysis of NLBP. In Section 4, we illustrate the effectiveness of our algorithm by the experiments. Finally, we give conclusions of our work In Section 5.

2. Description of LBP

This feature extracted by Local binary pattern (LBP) is widely used in texture analysis including image retrieval and face recognition. Its main idea is the local texture feature is described by a binary code which is obtained by the comparison of the center pixel value with its neighborhood. The basic algorithm of LBP is shown in Figure2. Firstly, the value of the center pixel is compared with that of its neighborhood (see Figure2 (a)). If its value is bigger than that of the center pixel, then its value is set to 0; otherwise, the value is 1. By this way, we get a binary pattern of the window (see Figure2 (b)). Secondly, the binary pattern in Figure2 (b) is arranged in clockwise and then a binary string is obtained (see Figure2(c)). Finally, the decimal representation of the binary string is the LBP code of the center pixel (see Figure2(c)). In an image, the LBP code of every pixel can reflect the distribution of its neighborhood gray value. The texture structure of a local region can be described by the histogram of the LBP code in this region.

In the LBP approach for texture classification [1], the occurrences of the LBP codes in an image are collected into a histogram. The classification is then performed by computing simple histogram similarities. However, considering a similar approach for facial image representation results in a loss of spatial information and therefore one should codify the texture information while retaining also their locations. One way to achieve this goal is to use the LBP texture descriptors to build several local descriptions of the face and combine them into a global description. Such local descriptions have been gaining interest lately which is understandable given the limitations of the holistic representations. These local feature based methods are more robust against variations in pose or illumination than holistic methods.

This histogram effectively has a description of the face on three different levels of locality: the LBP labels for the histogram contain information about the patterns on a pixel-level, the labels are summed over a small region to produce information on a regional level and the regional histograms are concatenated to build a global description of the face.

25	25	42
34	35	36
40	56	53

0	0	1
0		1
1	1	1

00111110→62

(a) Rectangular Area (b) Binary Pattern (c) LBP Code

Figure 2. LBP Algorithm

For a face image, the LBP code value of the whole image is firstly calculated and then the statistical histogram is used to describe the appearance frequency of various patterns in an image, such as edge, brightness, smooth area *etc*. In order to express the spatial information of the texture in a image, the idea of dividing the whole image into some blocks is introduced. That is to say, a face image is divided into different blocks, and then the histogram of LBP is extracted in each block. Finally, we obtain the LBP of a face image by the components of the histograms of LBP from all of the block regions. In this paper, a face image is evenly divided into four sections and the histogram of LBP is got in each block. Finally, the global histogram of LBP is obtained by the components in the four block regions, as a feature vector for subsequent processing.

3. The Nonlinear Version of LBP (NLBP)

Based on the algorithm of LBP presented in the second section, we get the feature vectors $\{x_i\}_{i=1}^{N}$. Usually, x_i is very high dimensional and maybe contains information redundancy, which effects the accuracy of FR and the computational complex. To capture the nonlinear information of $\{x_i\}_{i=1}^{N}$ extracted by LBP, we get a nonlinear version of LBP (NLBP) which is inspired by the kernel method.

3.1. The Main Idea of NLBP

The main idea of NLBP is all the feature vectors $\{x_i\}_{i=1}^{N}$ are firstly mapped in another feature space called the kernel space by a nonlinear mapping φ. And then we get a standard orthogonal basis $\{\beta_i\}_{i=1}^{r}$ of the subspace spanned by all the mapped vectors $\{\varphi(x_i)\}_{i=1}^{N}$. Subsequently, we obtain the projection vector y_i of each $\varphi(x_i)$ using Eq. (1). Finally, we perform FR based the data $\{y_i\}_{i=1}^{N}$.

$$y_i = (\beta_1, \beta_2, \cdots, \beta_r)^T \cdot \phi(x_i) \tag{1}$$

In the process of NLBP, we will focus on two things. Firstly, $\{y_i\}_{i=1}^{N}$ and $\{\varphi(x_i)\}_{i=1}^{N}$ is equivalent for Eq. (1) is an orthogonal transformation, which means $\{y_i\}_{i=1}^{N}$ contains the same nonlinear like $\{\varphi(x_i)\}_{i=1}^{N}$ does. Secondly, compared with performing FR based on the kernel space, the computational complexity of NLBP is much lower than that of the kernel method. The reason is that we usually have $r \ll N$, especially when N is very large [10].

3.2. How to Get a Standard Orthogonal Basis

In order to get a standard orthogonal basis $\{\beta_i\}_{i=1}^{r}$ of the subspace spanned by all the mapped samples $\{\varphi(x_i)\}_{i=1}^{N}$, we designed an iterative algorithm to find a basis $\{\phi(x_{bi})\}_{i=1}^{b}$ and then make the orthogonalization of $\{\varphi(x_i)\}_{i=1}^{N}$ for getting a standard orthogonal basis $\{\beta_i\}_{i=1}^{r}$. Firstly, we give the theoretical deduction about the scheme based on the vector correlation principle to select a basis $\{\phi(x_{bi})\}_{i=1}^{b}$ [11].

Theorem 1. Support $\varphi(x_1), \varphi(x_2), \cdots, \varphi(x_l)$ is the mapped samples which are linearly independence, x is any training sample, then $\varphi(x_1), \varphi(x_2), \cdots, \varphi(x_l), \varphi(x)$ is linearly independence if and only if $k_{xx} - k_{lx}^T K_{ll}^{-1} k_{lx} \neq 0$, where $k_{xx} = k(x, x)$, $K_{ll} = (k(x_i, x_j))_{1 \le i, j \le l}$, $k_{lx} = (k(x_1, x), k(x_2, x), \cdots, k(x_l, x))^T$.

Proof: Let K_{l+1} denotes the kernel matrix of $\varphi(x_1), \varphi(x_2), \cdots, \varphi(x_l), \varphi(x)$. Then we have the following equation.

$$\det(K_{l+1}) = \begin{vmatrix} K_{ll} & k_{lx} \\ k_{lx}^T & k_{xx} \end{vmatrix} = \begin{vmatrix} K_{ll} & k_{lx} \\ 0 & k_{xx} - k_{lx}^T K_{ll}^{-1} k_{lx} \end{vmatrix} \tag{2}$$

Since $\varphi(x_1), \varphi(x_2), \cdots, \varphi(x_l)$ is linearly independence, we have $\det(K_{l+1}) \neq 0$, so $\det(K_{l+1}) \neq 0$ if and only if $k_{xx} - k_{lx}^T K_{ll}^{-1} k_{lx} \neq 0$. On the other word, $\varphi(x_1), \varphi(x_2), \cdots, \varphi(x_l), \varphi(x)$ is linearly independence if and only if $k_{xx} - k_{lx}^T K_{ll}^{-1} k_{lx} \neq 0$.

According to Theorem 1, we can get an iterative algorithm to find a basis of $\{\phi(x_{bi})\}_{i=1}^{b}$. The mainly idea can be describe as follows. Support we have finished the train phase using training data $\{\phi(x_i)\}_{i=1}^{t}$ where $t(t \le n)$ denotes the number of the training data and got a basis $\varphi(x_{b1}), \varphi(x_{b2}), \cdots, \varphi(x_{bl})$ of $\{\phi(x_i)\}_{i=1}^{t}$. Let x denotes the next new sample, it is easy to verify whether $\varphi(x_1), \varphi(x_2), \cdots, \varphi(x_l), \varphi(x)$ is linearly independence based on theorem 1. if $\varphi(x_1), \varphi(x_2), \cdots, \varphi(x_l), \varphi(x)$ is linearly independence, then let $\varphi(x_{b(l+1)}) = \varphi(x)$ and we get a new linearly independence set $\varphi(x_1), \varphi(x_2), \cdots, \varphi(x_l), \varphi(x_{b(l+1)})$. When t

travel all the training samples, we get the final linearly in dependence set $\{\phi(x_{bi})\}_{i=1}^{b}$ which a basis is of the subspace spanned by all the samples $\{\varphi(x_i)\}_{i=1}^{N}$.

However, using theorem 1 to verify the linear independent of a vector set, it involves the inverse operation of K_{ll} which is a $l \times l$ dimension matrix. With the increasing of t, l usually be very large, and so the computation complexity of the inverse operation will be very large. Therefore, to tackle this problem, an optimization algorithm is presented in the paper, which can convert the inverse operation into the multiplication for reducing the computation complexity.

Theorem 2. Support $\varphi(x_{b1}), \varphi(x_{b2}), \cdots, \varphi(x_{bl})$ is linearly independence and the corresponding kernel matrix is K_{ll} .The kernel matrix of $\varphi(x_1), \varphi(x_2), \cdots, \varphi(x_l), \varphi(x_{b(l-1)})$ is denoted by $K_{(l-1)(l-1)}$ .Let $\gamma = (k(x_{b1}, x_{bl}), k(x_{b2}, x_{bl}), \cdots, k(x_{b(l-1)}, x_{bl}))^T, \alpha = k(x_{bl}, x_{bl})$ and $D = K_{ll}^{-1}$. We have

$$K_{ll}^{-1} = \frac{1}{\alpha - \gamma^T D\gamma} \begin{bmatrix} (\alpha - \gamma^T D\gamma)D + D\gamma\gamma^T D & -D\gamma \\ -\gamma^T D & 1 \end{bmatrix} \tag{3}$$

It is easy to verify the theorem 2. According to theorem 2, the inverse operation of $l \times l$ matrix is converted into the inverse operation of $(l-1)\times(l-1)$ matrix. Followed by analogy, the inverse operation is converted into the multiplication operation. Combined theorem 1, it is easy to find out a basis $\{\phi(x_{bi})\}_{i=1}^{b}$ of the subspace spanned by all the samples $\{\varphi(x_i)\}_{i=1}^{N}$ .Based on the above analysis, the main procedure of finding a basis of the subspace spanned by all the samples will be described in the following.

Alogrithm1. The iterative algorithm to find a basis

Step1. Initialization. Select randomly a sample x from the set $X = \{x_i\}_{i=1}^{N}$,and $k(x, x) \neq 0$. Where $k(\cdot)$ denotes the kernel function. Let $S = \{x\}$, $D = \{x\}$, $G = 1/k(x, x)$, $t = 1$.

Step2. The condition of ending the procedure. If $t = N$, then put out D and let the procedure end. Otherwise, next step.

Step3. Discrimination function. Select randomly a sample $x^* \in X - S$. let $S = S \cup \{x^*\}$, $t = t + 1$ and verify Eq. (4) is correct or not.

$$k_{tt} - k_{st}^T G k_{st} = 0 \tag{4}$$

Where $k_{tt} = k(x^*, x^*)$, $(k_{st})_i = k(\bar{x}_i, x^*)$, $\bar{x}_i \in D$.

Step4. Update the procedure. If Eq.4) is correct, go back step2. Otherwise, do the following and go to step2. Otherwise, let

$$G = \frac{1}{k_{st} - k_{st}^T G k_{st}} \begin{bmatrix} (k_{tt} - k_{st}^T G k_{st})G + G k_{st} k_{st}^T G & -G k_{st} \\ -k_{st}^T & 1 \end{bmatrix}, D = D \cup \{x^*\}$$

and go to back .

When the above procedures stop, we get a samples set $D = \{x_{bi}\}_{i=1}^{r}$. Its mapped samples $\phi(D) = \{\varphi(x_{bi})\}_{i=1}^{r}$ will be a basis of the subspace spanned by $\{\varphi(x_i)\}_{i=1}^{N}$, which can be easily verified using the linear correlation theories. After that, we can orthogonalize the basis $\phi(D) = \{\varphi(x_{bi})\}_{i=1}^{r}$ using Eq.(5).

$$(\beta_1, \beta_2, \cdots, \beta_r) = (\varphi(x_{b1}), \varphi(x_{b2}), \cdots \varphi(x_{br}))C \tag{5}$$

Where $u_i, \lambda_i (i, j = 1,2 \cdots, r)$ denotes respectively the eigenvector and the corresponding eigenvalue of the kernel matrix $K_{rr} = (k(x_{bj}, x_{bk}))_{1 \le j,k \le r}$, and

$$C = \left(u_1 \Big/ \sqrt{\lambda_1}, u_2 \Big/ \sqrt{\lambda_2}, \cdots, u_r \Big/ \sqrt{\lambda_r} \right). \tag{6}$$

Combined with Eq. (5) and Eq.(6), the projection vector of any mapped sample $\varphi(x_i)$ can obtained by Eq.(6).

$$y_i = C^T (k(x_{b1}, x_i), k(x_{b2}, x_i), \cdots, k(x_{br}, x_i)) \quad (i = 1,2, \cdots, N) \tag{7}$$

At the end of this section, we give the overview about the nonlinear version of LBP (NLBP) for FR.

Algorithm2. The nonlinear version of LBP (NLBP) for FR.

Step1. Based on algorithm 1, we get a basis $\varphi(D) = \{\varphi(x_{bi})\}_{i=1}^{r}$.

Step2.Computing the kernel matrix $K_{rr} = (k(x_{bj}, x_{bk}))_{1 \le j,k \le r}$ of the basis $\varphi(D) = \{\varphi(x_{bi})\}_{i=1}^{r}$ and eigenvalue decomposition of K_{rr} for getting the matrix C (see Eq. (5)).

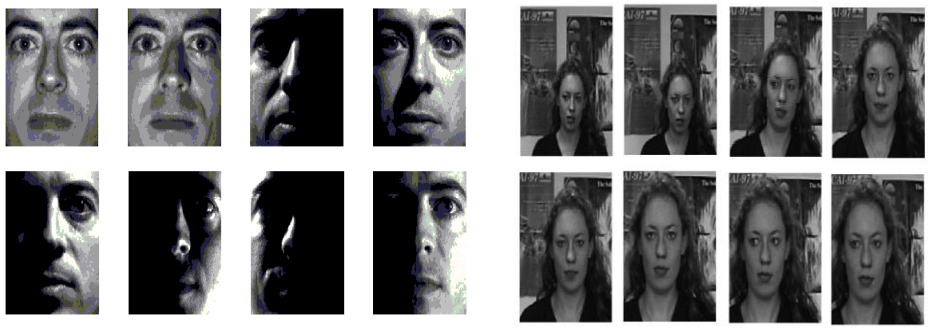
Step3. Using Eq. (6), we get the projection vectors $Y = \{y_i\}_{i=1}^{N}$.

Step4. Performing FR based on $Y = \{y_i\}_{i=1}^{N}$. In this paper, we employ the nearest neighbor classifier.

4. Experiments Analysis

To evaluate the performance of the presented method for FR, we perform extensive experiments and compared out method with LBP based on two well-known face datasets, YaleB database and Face96 database. Both of them are the public standard visual face images for FR. YaleB database consists of 10 subjects and each subject has approximately 64 frontal view images under various lighting conditions. The size of all image data is 168×192 pixels. Figure 3 (a) shows part of the face images of one subject. Face96 is established by Dr. Libor Spacek in University of Essex. Images of each person reflect the change of scale and position caused by the person moving towards the camera. 20 images of each person and 10 persons are chosen at random for the experiment in the database. Each image is of 200×180 pixels and instances of those images are shown in Figure3 (b).

In the experiment, we select 64 frontal view images per person from YaleB database, of which 54 images per person are used for training and the others for the testing. From the Face96 database, 15 images per person are selected for training and the rest 5 images for the testing. During the feature extraction processing, the kernel function is Gauss kernel function $k = \exp\left(- \left| x - y \right|^2 \Big/ d \right)$.



(a) Samples from Extended YaleB. (b) Samples Form Face96

Figure 3. Instance of Face Images

Obviously, as a nonlinear feature extraction method, the performance of KPCA is affected by the kernel parameters. The different parameters will finally lead to differences in the recognition rate. Figure 4 shows the relationship between the kernel parameter and the recognition rate which is based YaleB database.

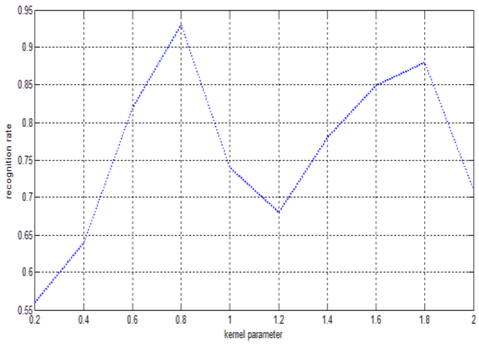


Figure 4. The Variety of Recognition Rate and the Kernel Parameter

Table 1. Average Recognition Rates (%) and Standard Deviations of Compared Methods

Feature / dataset	LBP	NLBP
YaleB	88.2±4.45	95.48±5.46
Face96	87.5±6.45	93.26±7.22

As can be seem form Figure4. The recognition rate is strongly influenced by the kernel parameter. The recognition rate varies with the variation of parameter. The reason is that the choice of kernel function is essentially determined by the selection of the kernel parameters.

In order to comprehensively display the performance of the presented features extraction method, the average recognition rate of 10 random tests and the corresponding variance of using the different features are shown in the Table 1. NLBP, as the nonlinear extraction of LBP, its average recognition rate is much higher than that of LBP. On the other hand, the corresponding standard deviation is little difference. TSable 1 indicate better recognition power of NLBP than LBP, which demonstrates our method can extract the most effective discriminate feature and proves a strong robustness in statistically significance.

5. Conclusion

A novel nonlinear version of LBP was proposed based the kernel theory, which is very effective in capturing the nonlinear feature of data. The experiments over the YaleB and Face96 databases were conducted to verify the proposed method. The comparative results of average recognition rate and standard deviations demonstrated that the proposed method outperforms LBP. Moreover, we also analyze the relationship between the change of parameters and the recognition rate.

However, with the method proposed in this paper, how to optimize the kernel parameter still needs to be deeply research. We think it a good solution to find out the optimal parameter by using genetic algorithm. Studying how to optimize the kernel parameter can be a direction of our future work.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (NSFC) under Grant No. 60970105, 60673153, 60773053; Fujian Province Education Scientific Research Fund(JK) (“mobile logistics chain management system based on the cloud platform implementation") of China.

References

[1] X. Bing X., Gao X. B. and Tao D. C., “Review face recognition in the shape–texture domain”, Pattern Recognition, vol. 46, **(2013)**, pp. 1906-1919.
[2] J. Rabia and R. A. Hamid, “A survey of face recognition techniques”, Journal of information processing systems, vol. 5, no. 2, **(2009)**, pp. 41-68.
[3] W. Chao and C. G. Guo, “Face recognition based on gabor-enhanced manifold learning and SVM”, Advances in neural network, vol. 6064, **(2010)**, pp. 184-191.
[4] H. Zhao, P. C. Yue and T. K. James, “A novel incremental principal component analysis and its application for face recognition”, IEEE transactions on systems, man, and cybernetics-part B: cybernetics, vol. 36, no. 4, **(2006)**, pp. 873-886.
[5] H. L. Tang, Y. F. Sun and B. C. Yin, “Face recognition based on Haar LBP histogram”, ICACTE, vol. 6, **(2010)**, pp. 235-238.
[6] A. B. Ashraf, S. Lucey and T. Chen, “Reinterpreting the application of Gabor filters as a manipulation of the margin in linear support vector machines”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 7, **(2010)**, pp. 1335:1341.
[7] M. Daniel, M. Domingo and S. Alvaro, “Face recognition with local binary patters, Spatial pyramid histograms and naïve bayes nearest neighbor classification”, 2009 International Conference of the Chilean Computer Science Society, **(2009)**, pp. 125-132.
[8] T. Ahonen, A. Hadid and M. Pietikainen, “Face Description with Local Binary Patterns: Application to Face Recognition”, IEEE Transactions Pattern analysis and machine intelligence, vol. 28, no. 12, **(2006)**, pp. 2037-2041.
[9] S. S. Liu, W. Y. Tian and C. Wan, “Facial Expression Recognition Method Based on Gabor Multi-orientation Features Fusion and Block Histogram”, Acta Automatica Sinica (China), vol. 37, no. 2, **(2011)**, pp. 1455-1462.
[10] G. Baudat and F. Anouar, “Kernel-based methods and function approximation”, In Proceeding of International Conference of Neural Networks, Washington, DC, **(2001)**, pp. 1244-1249.
[11] F. Zhao, Y. Liu and Z. Hua, “Simplified Solution for Support Vector Domain Description”, International Journal of Digital Content Technology and its Applications, vol. 5, no. 2, **(2011)**, pp. 292-300.

Authors

Zhang Hongyi, Professor in School of Optoelectronic & Communication Engineering, Xiamen University of Technology. Her research focuses on the field of pattern recognition.

Zhao Feng, Associate Professor, School of Computer Science and Technology, Shandong Institute of Business and Technology. His research focuses on the field of image processing.