

Sparse Kernel Principal Components Analysis for Face Recognition in RGB Spaces

Minghai Xin^{1,2}, Yang Liu¹ and Jingjie Yan¹

¹*Research Centre for Learning Science, Southeast University,
Nanjing, China, 210096*

²*School of Computer Science and Technology, Huaqiao University, Xiamen 361021
xin_minghai@163.com, liuyang@seu.edu.cn, yanjingjie1212@163.com*

Abstract

This paper presents a kinds of information fusion algorithm based on multi-channel color image. The color face image is first separated into three pseudo grayscale images: R, G, and B, then the partial characteristics of face is extracted by use of Gabor wavelet transform from each component to be eigenvector in series connection, which will be through dimensionality reduction by sparse kernel principal components to be recognized and classified by the nearest classifier. In order to testify the validity, we make experiment by use of XM2VTS color face dataset and the experimental result supports the proposed method.

Keywords: *Sparse Kernel Principal Components Analysis, face recognition, RGB spaces*

1. Introduction

Face recognition is a research focus in the field of pattern recognition and artificial intelligence with a history of nearly 30 years. Due to its easy access and handling characteristics, the majority of face recognition methods are based on grayscale face image, but the real face image is in color with richer information. With the development of information technology as well as hardware update, the color information plays a more important role in the image processing. A grayscale image contains a maximum of 256 levels, while a color image may contain tens of thousands. Therefore, the color space greatly increases the amount of information contained with higher image recognition accuracy than grayscale space. It has been found that the use of color information can enhance the recognition rate of face image with respect to grayscale image only [1]. Most research methods of face color image recognition do not directly process the face color image but transform the nonlinear color space to make color components completely independent of luminance components, thereby eliminating the impact of the brightness variation caused by the light [2]. Luo *et al.* [3] has proposed to convert the color image RGB into HSV space and then transform it through quaternion; Li *et al.* [4] has suggested the color image fusion technology based on Contourlet transform to convert the color visible image from RGB space into HIS space. However, the nonlinear transformation of color space will lose the information of original image and the converted image is very sensitive to noise. To this end, the paper presents an algorithm to separate the color image into three pseudo grayscale images - R, G, and B - in the RGB space and extract the partial characteristics of face from each component respectively by use of Gabor wavelet transform. Compared with grayscale feature, the new feature is characterized with the following:

1. Higher data dimensions. For example, the SIFT feature dimensions extracted from an image is 3840 in the grayscale space and is 11520 extracted from the same image in the color space.

2. A lot of information contained but with possible data redundancy. Because of the correlation between pixels, the feature data extracted may have redundancy. In addition, the combination of different channels in color space may also cause data redundancy.

3. Effect of noise. Since noise is always included in the original image, we extract the feature in the color space, and also extended the noise to the color space simultaneously.

Feature extraction tend to affect the final image recognition result. Many effective feature extraction methods have been proposed so far, mainly divided into two categories: one is a linear feature extraction method, and the other is non-linear feature extraction method. Principal Component Analysis (PCA) [6] is a classical dimension-reduction method for feature extraction in the linear domain, PCA finds a linear projection of high-dimensional data into a low-dimensional data, and largely retained the properties of the original data. Sparse principal component analysis [7] is built on the basis of the principal component analysis, principal component analysis rewritten as a form of regression to solve optimization problems, and in its regression criterion function added penalty term, resulting in the main component of the sparse coefficients. But the principal component analysis methods are essentially based on linear transformation, are unable to extract high-level information on the image. To avoid these drawbacks, some researchers have improved by the PCA process, a new non-linear approach. A new Kernel principal component analysis (KPCA) [8] is proposed for nonlinear feature extraction method. Studies show that KPCA method is superior to PCA [9-15].

In this paper, the sparse kernel principal components analysis (SKPCA) of elastic net is used for analyzing and handling color characteristics. The sparse kernel principal components analysis based on elastic net is to combine two methods – ridge regression and LASSO. By introduction of L1 and L2 penalty term, these features with little contribution to recognition will be ignored in order to achieve the goal.

2. KPCA Algorithm Thought

The PCA (Principal Component Analysis) has been introduced into face recognition by M. Turk and A. Pentland. The main principle is to project a high-dimensional vector to a low-dimensional vector space through a special feature vector matrix, which will be characterized by a low-dimensional vector without loss of any information. The main steps consist of two: (1) in the training phase, the system will produce a number of characteristic face images according to the trained face with consistent shape and size; (2) in the recognition phase, the new face image will be expressed as the coefficients of these characteristic faces. Thus, the face image can be represented by a one-dimensional vector to achieve the purpose of dimensionality reduction. The kernel principal components analysis (KPCA) is a nonlinear expansion of the traditional PCA algorithm. The basic idea is to use the kernel method to project the input data into feature space F where the kernel principal components analysis will be analyzed for effective extraction of non-linear feature.

With observation sample set $x_k, k = 1, 2, \dots, M, x_k \in R^N$, we assume that it has been centralized, namely: $\sum_{k=1}^M x_k = 0$. Then, the covariance matrix of sample is:

$$C = \frac{1}{M} \sum_{i=1}^M x_i x_i^T \quad (1)$$

The characteristic equation of C is

$$\lambda v = Cv \quad (2)$$

Where, $\lambda \geq 0$ is the eigenvalue and $v \in R^N$ is its corresponding eigenvector. Combine equation (1) and (2) and get:

$$\lambda v = Cv = \left(\frac{1}{M} \sum_{k=1}^M x_k x_k^T \right) \bullet v = \frac{1}{M} \sum_{k=1}^M (x_k \bullet v) x_k \quad (3)$$

As it can be seen that all eigenvectors are located in the space spanned by the training samples, if we map all the training samples from the input space to a high-dimensional space by the introduction of a nonlinear transform $\Phi: R^N \longrightarrow F$.

At this point, the corresponding eigenvectors in F are also located in the space spanned by $\{\Phi(x_1), \Phi(x_2), \dots, \Phi(x_M)\}$, i.e., a set of one-dimensional array $\alpha_i (i = 1, 2, \dots, M)$ existed, makes:

$$V = \sum_{i=1}^M \alpha_i \Phi(x_i) \quad (4)$$

While equation (3) is equivalent to:

$$\lambda(\Phi(x_k) \bullet V) = (\Phi(x_k) \bullet \bar{C}V), \quad k = 1, 2, \dots, M \quad (5)$$

From equation (4) and (5):

$$\lambda \sum_{i=1}^M \alpha_i (\Phi(x_k) \bullet \Phi(x_i)) = \frac{1}{M} \sum_{i=1}^M \alpha_i \left(\Phi(x_k) \bullet \sum_{j=1}^M \Phi(x_j) \right) (\Phi(x_j) \bullet \Phi(x_i)) \quad (6)$$

Define $M \times M$ order matrix K and each element in K is

$$K_{i,j} = k(x_i, x_j) = (\Phi(x_i) \bullet \Phi(x_j)) \quad (7)$$

$$\text{Then equation (6) becomes } M\lambda K\alpha = K^2\alpha \quad (8)$$

Where, α is the column vector: $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_M)^T$

$$\text{And because the equation 1.8 is equivalent to } M\lambda\alpha = K\alpha \quad (9)$$

Set $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_M$ to represent the eigenvalue of matrix K and the corresponding eigenvector is expressed as $\alpha^1, \alpha^2, \dots, \alpha^M$

λ^p presents the first non-zero eigenvalue. Standardize the corresponding eigenvector $\alpha^p, \alpha^{p+1}, \dots, \alpha^M$:

$$(V^k \bullet V^k) = 1, k = p, \dots, M \quad (10)$$

The above equation contains no significant eigenvector $\alpha^p, \alpha^{p+1}, \dots, \alpha^M$, which can also be expressed as:

$$1 = \sum_{i,j=1}^M \alpha_i^k \alpha_j^k (\Phi(x_i) \bullet \Phi(x_j)) = \sum_{i,j=1}^M \alpha_i^k \alpha_j^k K_{i,j} = (\alpha^k \bullet K\alpha^k) = \lambda_k (\alpha^k \bullet \alpha^k) \quad (11)$$

So as long as $\alpha^p, \alpha^{p+1}, \dots, \alpha^M$ satisfies (11), a set of orthonormal eigenvector V^k of feature space F can be obtained.

For feature extraction, calculate the projection from input sample to the eigenvector only. Assuming the input sample as x and mapping $\Phi(x)$ into feature space F through mapping $\Phi(x)$, the projection of its eigenvector set V^k will be

$$(V^k \bullet \Phi(x)) = \sum_{i=1}^M \alpha_i^k \Phi^T(x_i) \Phi(x) = \sum_{i=1}^M \alpha_i^k k(x_i \bullet x) \quad (12)$$

As the KPCA analysis is based on the assumption of $\frac{1}{M} \sum_{i=1}^M \Phi(x_k) = 0$, we must centralize the mapping $\Phi(x_i)$ of each training sample in the feature space for removal of assumption.

$$\bar{\Phi}(x_i) = \Phi(x_i) - \frac{1}{M} \sum_{i=1}^M \Phi(x_i) \quad (13)$$

At this point, the equation (11) can still be met. With similar analysis, we get:

$$\bar{K} = \bar{k}(x_i, x_j) = (\bar{\Phi}(x_i) \bullet \bar{\Phi}(x_j)) = \bar{\Phi}(x_i)^T \bar{\Phi}(x_j) \quad (14)$$

$$\bar{\lambda} \alpha = \bar{K} \alpha \quad (15)$$

$$\bar{V} = \sum_{i=1}^M \alpha_i \bar{\Phi}(x_i) \quad (16)$$

However, since the feature space is very high-dimensional and even infinite dimensional, we cannot centralize it according to (14). Therefore, \bar{K} cannot be obtained from direct calculation but usually be calculated through the corresponding K of decentralized data:

$$\bar{K}_{ij} = K_{ij} - \frac{1}{M} \sum_{n=1}^M 1_{in} K_{mj} - \frac{1}{M} \sum_{n=1}^M K_{in} 1_{nj} + \frac{1}{M^2} 1_{in} K_{mn} 1_{nj} \quad (17)$$

Where, $1_{ij} = 1(i = 1, \dots, M; j = 1, \dots, M)$.

Summarize the basic principal of KPCA algorithm and use it to extract the face image characteristics with the following process:

Expand the matrix corresponding to face image into a column vector by columns; select appropriate kernel function and its parameters for training sample. The kernel matrix K is obtained from equation (7) and the kernel matrix \bar{K}

Obtain the eigenvalue $\bar{\lambda}$ and eigenvector $\bar{\alpha}$ of kernel matrix from (15); arrange the eigenvalue in ascending order to find the on-zero one and its corresponding eigenvector; standardize the vector $\bar{\alpha}$ of feature by referring to (11);

Select the eigenvector $\bar{\alpha}^k$ corresponding to the top k eigenvalues and get the eigenvector V^k of principal components in the high-dimensional space; where, k is the smallest integer satisfying the equation below and d is the threshold value:

$$\frac{\sum_{i=1}^k \bar{\lambda}_i}{\sum_{i=1}^{M-p} \bar{\lambda}_i} \geq d \quad (18)$$

Calculate the projection of training sample in the feature space V^k according to (2), normalize the projection coefficient matrix, and train the selected classifier:

In the recognition phase, project the face to be recognized Y to the feature V^k according to (12) and obtain its projection coefficient α^γ , which will be recognized by the selected classifier after it being normalized.

3. Sparse Kernel Principal Components Analysis

Sparse Kernel Principal Components Analysis, which is on the basis of PCA, is an optimization problem of solving regression form improved from PCA, adding one penalty term L1 into regression criterion function, and producing the coefficient of sparse principal component.

Using SKPCA to process the feature extracted. It can not only reduce the dimension of data feature to a large extent, but also select some useful features and shrink any other useless feature to zero by sparse feature after analysis of the validity of data features.

For the purpose of obtaining the sparse principal component, first we should express the principal component as the regression form including penalty term of elastic net [16].

Provided both $\alpha_{p \times k} = (\alpha_1, \dots, \alpha_k)$ and $\beta_{p \times k} = (\beta_1, \dots, \beta_k)$ are matrix of $p \times k$, to random $\lambda > 0$, provided

$$\begin{aligned} (\hat{\alpha}, \hat{\beta}) = \operatorname{argmin}_{\alpha, \beta} & \left\| \mathbf{X} - \mathbf{X}\beta\beta^T \right\|^2 + \sum_{j=1}^k \lambda_{1,j} \|\beta_j\|_1 + \lambda_2 \sum_{j=1}^k \|\beta_j\|^2 \\ \text{s.t. } & \mathbf{a}^T \mathbf{a} = I_{k \times k} \end{aligned} \quad (18)$$

Hence, $\hat{\beta}_j \propto V_j, \quad j = 1, \dots, k$

As to a criterion function including two unknown variables, we first fix one variable α , and compute the estimate of another variable β . Then fixes one variable β , and computes the estimate of another variable α .

Provided α_{\perp} stands for almost orthogonal matrix, and then $[\alpha; \alpha_{\perp}]$ is that of order p. Hence,

$$\left\| \mathbf{X} - \mathbf{X}\beta\beta^T \right\|^2 = \left\| \mathbf{X}\alpha_{\perp} \right\|^2 + \left\| \mathbf{X}\alpha - \mathbf{X}\beta \right\|^2 = \left\| \mathbf{X}\alpha_{\perp} \right\|^2 + \sum_{j=1}^k \left\| \mathbf{X}\alpha_j - \mathbf{X}\beta_j \right\|^2$$

Thereinto, when α is fixed, equality (18) can be changed into minimum:

$$\operatorname{argmin}_{\beta} \sum_{j=1}^k \left\{ \left\| \mathbf{X}\alpha_j - \mathbf{X}\beta_j \right\|^2 + \lambda_{1,j} \|\beta_j\|_1 + \lambda_2 \|\beta_j\|^2 \right\} \quad (19)$$

Provided $Y_j^* = \mathbf{X}\alpha_j$, then $\hat{\beta}_j$ can be regarded as estimate of elastic net:

$$\hat{\beta}_j = \operatorname{argmin}_{\beta_j} \left\| Y_j^* - \mathbf{X}\beta_j \right\|^2 + \lambda_{1,j} \|\beta_j\|_1 + \lambda_2 \|\beta_j\|^2 \quad (20)$$

There into,

$$\left\| Y_j^* - \mathbf{X}\beta_j \right\|^2 + \lambda_{1,j} \|\beta_j\|_1 + \lambda_2 \|\beta_j\|^2 = (\alpha_j - \beta_j)^T \mathbf{X}^T \mathbf{X} (\alpha_j - \beta_j) + \lambda_{1,j} \|\beta_j\|_1 + \lambda_2 \|\beta_j\|^2 \quad (21)$$

Then, β is fixed:

Let β is fixed, then our only concern is the 1st term of (18), namely:

$$\hat{\alpha} = \operatorname{argmin}_{\alpha} \left\| \mathbf{X} - \mathbf{X}\beta\beta^T \right\|^2, \text{ s.t. } \mathbf{a}^T \mathbf{a} = I_{k \times k} \quad (22)$$

The minimum problem in such form can be resolved by SVD of matrix decomposition. Then provided

$$\mathbf{X}^T \mathbf{X} \beta = \mathbf{U} \mathbf{D} \mathbf{V}^T \quad (23)$$

Hence $\hat{\alpha} = \mathbf{U} \mathbf{V}^T$.

4. Sparse Kernel Principal Components Analysis for Color Face Recognition

For the purpose of effective dimension reduction of high dimensional feature extracted before, SKPCA which is based on elastic net will have application in recognition process. The first step is to select a data base to be used in test, and extract the color feature. The process of feature extraction has been described in detail in Chapter Three, therefore there are no repeating details. It needs to mix all features of the three color channels in series. As to high-dimensional characteristic data, we use SKPCA for dimension reduction and feature selection. The procedure of algorithm of SKPCA lists as below :

First, obtain the principal component of the characteristic data \mathbf{X} , and use \mathbf{V} to denote the direction of the principal component, and $\alpha = \mathbf{V}(:, 1:k)$ denote the direction of previous k principal components.

Second, fix α to solve the problem of the elastic net

$$\beta_j = \underset{\beta}{\operatorname{argmin}} (\alpha_j - \beta)^T \mathbf{X}^T \mathbf{X} (\alpha_j - \beta) + \lambda_{1,j} \|\beta\|_1 + \lambda_2 \|\beta\|^2, \text{ where } j = 1, \dots, k.$$

Third, based on $\beta = [\beta_1, \dots, \beta_k]$ obtained in the second step, obtain the matrix $\mathbf{M} = \mathbf{X}^T \mathbf{X} \beta$, and then decompose \mathbf{M} as a singular value, $\mathbf{M} = \mathbf{U} \mathbf{D} \mathbf{V}^T$, hence obtain \mathbf{U} and \mathbf{V} . Where $\alpha = \mathbf{U} \mathbf{V}^T$.

Fourth, repeat the second and third steps till the convergence of the estimated constant β .

Fifth, make $\hat{v}_j = \frac{\beta_j}{\|\beta_j\|}$, where $j = 1, \dots, k$.

It is necessary to explain the fourth step: because during the process of iteration, the estimated constant β is not always convergent. The method adopted here is to set an upper limit for the cycle, and if the number of iterations exceeds the upper limit, iteration is over.

Now we reduce the original high dimensional data of dimensional p to dimensional k . In addition, when the coefficient matrix contains many zero elements, this plays an important role in the characteristics selection of data \mathbf{X} .

The last part of the flow chart is an identification of the characteristics after the dimension reduction. In order to be compared with the previous experimental processes, this experiment adopts the classification method of Nearest Neighbors to classify the characteristic data.

5. Experimental

As we know, the selection of the parameters is essential for the experimental results. Using the SKPCA to handle the data, this experiment takes into consideration three parameters: s , k and λ . The parameter s denotes the number of the non-zero variables reserved finally, k denotes the number of the reserved sparse principal elements, namely the reduced dimension, and λ denotes the shrinkage parameter. For example, provided, parameter $s = 3$ and $k = 10$, after the SKPCA on the original data, it can be obtained that the sparse projection matrix contains 10 column vectors, and every column vector contains 3 non-zero numbers.

During the analysis process of principal components, the reduced dimension of data is determined by the data themselves. It is usual to select the number of the principal elements when the cumulative contribution of characteristic values reaches the pre-set threshold value. During the analysis process of the sparse principal components, it is necessary to first set the reduced dimension. Therefore, we first estimate the dimensional parameters. Table 1 and Figure 1 denote the recognition rates when the characteristic dimensions retain 50, 100, 150 and 200 relatively:

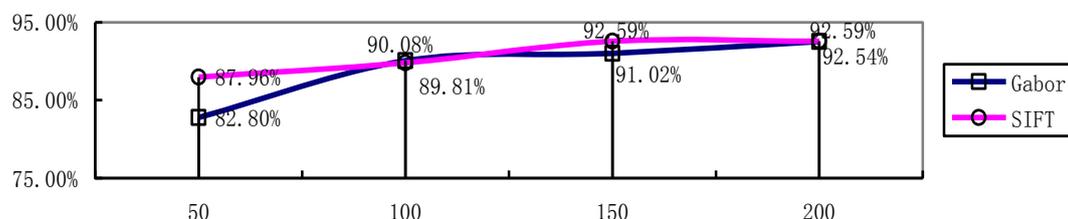


Figure 1. Recognition Rate Changes with Dimensions

Thereinto, the horizontal coordinate denotes the reduced dimension, and vertical coordinate denotes the recognition rate; the blue line denotes the Gabor face features extracted from the color XM2VTS facial database, and the red line denotes the SIFT

expression characteristics extracted from the color child images. At this moment, set the parameter $s = 20$, and the shrinkage parameter $\lambda_1 = \lambda_2$. It is known based on the equation $\lambda = \frac{\lambda_2}{\lambda_1 + \lambda_2} = 0.5$; set the maximum value of the regression frequency to be 200.

Table 1. Recognition Rate Changes with Dimensions

	50	100	150	200
XM2VTS Facial Database Gabor Characteristics	82.80%	90.08%	91.02%	92.54%

It can be seen from the Table that with the increase in the reserved characteristic dimension, the recognition rate increases. However, because the increase of the dimensions will cause the increase of the calculated amount, and between the calculated amount and the dimension increase there is non-linear relationship, when the dimension increases to 200, the coefficient matrix of resolve characteristics needs more time. For example, handling the Gabor characteristic of XM2VTS facial database under the same condition, it will take 10 hours to reduce the dimension to 200; while it only takes 1 hour for the dimension to reduce to 100. It can be seen from the cognition rate that the recognition rate for dimension 100 is 90.08% and for 200 is 92.54%. Therefore, there are no significant differences between cognition rates. Similarly, when handling the SIFT characteristics of the child images, it both takes a long time for reducing dimension to 150 and 200, and it is rare to increase the recognition rates.

Therefore, during the experimental process, we make a compromise, both taking into consideration the operation time and the effectiveness of the characteristics, and finally obtain $k = 100$ as the reduction dimension parameter.

Based on the above experimental results, there is monotone increasing function relationship between the recognition rate increase and reduction dimension parameter k . Due to the strong association between the value of parameter k and the operation time, here we set the value of dimension parameter k to be 50 in order to improve the efficiency.

During the experiment, set $s = 20$, and the maximum of the regression frequency is 200.. The changes of the recognition rate with those of the shrinkage parameter λ are shown in Figure 2. The horizontal coordinate denotes the value of parameter λ , and the vertical coordinate denotes the recognition rate. The selection range of the shrinkage parameter λ covers from 0.2 to 0.8. Figure 2 denotes that the Gabor characteristics extracted from XM2VTS facial database. It can be seen from the figures when the shrinkage parameter $\lambda = 0.4$, there comes the highest recognition rate.

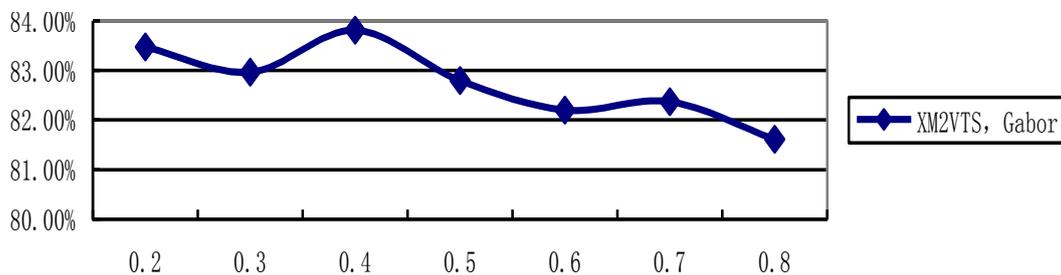


Figure 2. Change of XM2VTS Facial Recognition Rate with

Therefore, during the following experimental process, we select 0.4 for those from XM2VTS facial images.

Table 2 denotes the recognition rate of using XM2VTS color facial images to carry out facial recognition. The data in the first column denotes the recognition rate of using the pure grey-scale feature for recognition, the second denotes the recognition rate after reducing the

dimension by directly using standard PCA, and the third denotes the recognition rate of re-recognition after the characteristic selection by using the SKPCA.

Table 2. XM2VTS Color Adult Facial Recognition Result Contrast

	Grey-scale Feature	Color (RGB) Feature (PCA)	Color (RGB) Feature (SKPCA)
SIFT	80.34%	78.73%	85.00%
Gabor	83.56%	85.00%	88.39%

During the experiment, the characteristic dimension of color space feature reduces to 100 after the SKPCA selection. Considering from the perspective of contrast experimental results, the dimension reduces to 100 by means of the standard PCA.

According to the data in the first line of Table 2, the SIFT characteristic dimension of grey-scale feature is 3840, that of the RGB color space is 11520. After the SKPCA analysis characteristic selection, the dimension reduces to 100. It can be seen that the recognition increases to 85.00%.

According to the data in the second line, the characteristic dimension of grey-scale Gabor feature is 2560, that of the RGB color space is 7680, and the recognition rate increases to 88.39% after SKPCA analysis characteristic selection.

It can be seen that, during the process of facial expression recognition, the characteristic dimension reduces significantly while the recognition rate increases to certain degree by using the SKPCA to analyze the color characteristics.

6. Summary

This paper comes to a detailed description on the SKPCA method based on the elastic net, and applies the SKPCA method to the characteristic selection of face color image recognition. In the experiment, we first uses the data to determine the appropriate parameters, and then puts forward specific experimental procedures and results. It is shown from the result that, comparing with the standard PCA, the recognition rate of face image color characteristics selected by using SKPCA method increases significantly. Therefore, this further proves the effectiveness of color characteristics.

Acknowledgements

This work was supported by National Nature Science Foundation of China (No. 60972023), National Science and Technology Important Special Project (2011ZX03003-002 & 2011ZX03003-004).

References

- [1] C. D. Scott and R. E. Smalley, J. Nanosci, Nanotechnol., vol. 3, (2003), pp. 75.
- [2] L. Orres, J. Reutter and L. Lorente, "The importance of the color information in face recognition", International Conference on Image Processing, (1999), pp. 627- 631, Kobe, Japan.
- [3] S. -J. Wang, H. Sun and H. -Y. Mo, "A new feature space transformation algorithm in color image and its application", Acta Electronica Sinica, vol. 35, (2007), pp. 193-196.
- [4] B. J. Chen, X. M. Sun, D. C. Wang, *et al.*, "Color Face Recognition Using Quaternion Representation of Color Image", vol. 11, (2012), pp. 1817.
- [5] G. -X. Li and K. Wang, "Color image fusion algorithm using the Contourlet transform", Acta Electronica Sinica, vol. 25, (2007), pp. 111-117.
- [6] H. Zou, T. Hastie and R. Tibshirani, "Sparse principal component analysis", Journal of computational and graphical statistics, (2006), pp. 265-286.
- [7] S. Wold, K. Esbensen and P. Geladi, "Principal component analysis", Chemometrics and intelligent laboratory systems, vol. 2, (1987), pp. 37-52.
- [8] H. Shen and J. Z. Huang, "Sparse principal component analysis via regularized low rank matrix approximation", Journal of multivariate analysis, vol. 99, (2008), pp. 1015-1034.

- [9] B. Schölkopf, A. Smola and K. R. Müller, “Kernel principal component analysis”, Artificial Neural Networks—ICANN ’97, (1997), Springer Berlin Heidelberg.
- [10] K. I. Kim, K. Jung and H. J. Kim, “Face recognition using kernel principal component analysis”, Signal Processing Letters, IEEE, vol. 9, (2002), pp. 40-42.
- [11] P. Honeine, “Online kernel principal component analysis: a reduced-order model”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, (2012), pp. 1814-1826.
- [12] D. Widjaja, C. Varon, A. Dorado, *et al.*, “Application of Kernel Principal Component Analysis for Single-Lead-ECG-Derived Respiration”, IEEE Transactions on Biomedical Engineering, vol. 59, (2012), pp. 1169-1176.
- [13] P. M. Rasmussen, T. J. Abrahamsen, K. H. Madsen, *et al.*, “Nonlinear denoising and analysis of neuroimages with kernel principal component analysis and pre-image estimation”, NeuroImage, vol. 60, (2012), pp. 1807-1818.
- [14] F. Mathieu, C. Jocelyn and B. Jón Atli, “Kernel principal component analysis for the classification of hyperspectral remote sensing data over urban areas”, EURASIP Journal on Advances in Signal Processing, (2009).
- [15] T. Hastie, R. Tibshirani and J. Friedman, “The Elements of Statistical Learnin”, (2009).
- [16] R. Jenssen, “Kernel entropy component analysis”, Pattern Analysis and Machine Intelligence, IEEE Transactions, vol. 32, (2010), pp. 847-860
- [17] H. Zou, T. Hastie and R. Tibshirani, “Sparse principal component analysis”, Journal of computational and graphical statistics, vol. 15, (2006), pp. 265-286.

Authors



Minghai Xin

He received the M.S. degrees from Huaqiao University, Xiamen, China. Now, he is currently working toward the Ph.D. degree in the Research Centre for Learning Science, Southeast University, Nanjing, China. His research interests are in pattern recognition.



Yang Liu

She received the M.S. degree in the Research Centre for Learning Science, Southeast University, Nanjing, China 2011. Her current research interests include affective computing and pattern recognition.



Jingjie Yan

He received the M.S. degree in signal processing from China University of Mining and Technology, Xuzhou, China, 2009. Currently, he is a PH.D student with the School of Information Science and Engineering, Southeast University, Nanjing. His current research interests include affective computing and pattern recognition.

