



# Evaluation of Face Recognition Techniques Based on Symlet 2 Wavelet and Support Vector Machine

Zhipeng Li, Xuesong Jiang<sup>(✉)</sup>, and Yewen Pang

School of Computer Science and Technology, Qilu University of Technology  
(Shandong Academy of Sciences), Jinan 250353, China  
lizhipengqilu@gmail.com, jxs@qlu.edu.cn

**Abstract.** In this study, we analyze and study the feature extraction method based on wavelet transform and Principal Component Analysis (PCA), and propose a method combining symlet 2 wavelet (sym2) and Support Vector Machine (SVM) for face recognition. Firstly, the sym2 is used to wavelet decomposition of the original image, and the decomposed low-frequency information is selected for image reconstruction. Then, the reconstructed data is multi-classified using support vector machine. Experiments were carried out using the internationally published ORL face recognition database, and compared with some classical methods used in ORL face recognition. The experimental results show that the recognition success rate increases with the increase in the training set. When 7 poses are selected for each face (A total of 280 photos are used as training sets) in the data set, the recognition success rate reaches 100%. Therefore, the method proposed in this paper is better than the previous method.

**Keywords:** Symlet 2 wavelet · Support Vector Machine · ORL face recognition, feature extraction · Principal component analysis

## 1 Introduction

With the rise of artificial intelligence and big data technology, the research and application of face recognition technology has been promoted. In recent years, domestic and foreign scholars have proposed some feasible algorithms for face recognition technology. Such as eigenface method, neural network, wavelet transform, support vector machine, hidden Markov and other methods [1]. Most algorithms currently require hardware and long training time. Generally, the face recognition method commonly uses wavelet transform for feature extraction and dimensionality reduction or PCA for feature dimensionality reduction and for performing classification operations using a neural network. In this paper, we use wavelet transform and PCA for image feature extraction and dimensionality reduction. Wavelet transform is a major breakthrough after Fourier analysis in recent years. The multi-resolution analysis of wavelet has good localization characteristics of space and frequency, and the detailed analysis of the object can be realized by adopting gradually detailed step size for the high-frequency part of the signal. Therefore, wavelet transform is especially suitable

for the processing of non-stationary signals such as images [2]. PCA is a commonly used method for feature dimension reduction. The basic idea is to extract the main information of the data, and then discard the redundant information to achieve the purpose of compression. SVM is a common classification algorithm in machine learning. For multi-classification problems, SVM maps data from low-dimensional to high-dimensional space by introducing kernel functions, so as to achieve linear divisibility and achieve multi-classification goals. In this paper, we use the international standard universal face recognition library which is ORL face recognition data set. The ORL face database consists of a series of face images, with 40 objects of different ages, genders and races. Each person's 10 images consists of a total of 400 grayscale images, the image size is  $92 \times 112$ , and the image background is black, as shown in Fig. 1. In this paper, we introduce the feature extraction methods using PCA and wavelet transform, respectively, and use SVM for classification and recognition. The structure of this paper is as follows. Section 2 introduces the feature extraction method wavelet transform and PCA. In Sect. 3, we introduced the classification algorithm SVM. In Sect. 4, we detail the experimental process and experimental results. Finally, In Sect. 5, we draw conclusions.

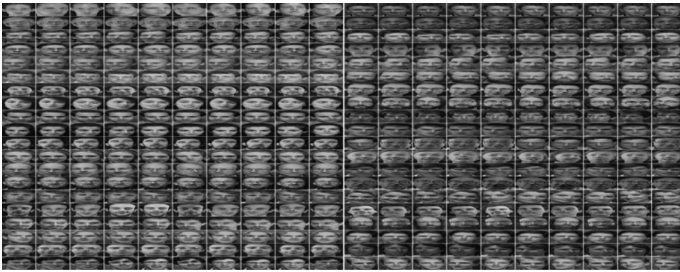


Fig. 1. ORL face recognition data set

## 2 Feature Extraction Method

### 2.1 Wavelet Transform

Wavelet transform is applied to face recognition, mainly because the resolution of the subgraph is reduced in different directions after the wavelet image is decomposed by wavelet, and the computational complexity is reduced accordingly [3]. At the same time it provides good local information in both the airspace and the frequency domain. The low frequency portion of the face image information describes the overall shape of the image. The high frequency part describes the details of the image. The face low frequency information obtained by wavelet transform can better describe the face features useful for classification [4].

In our study, we used a multi-layer (4-layer) SWT (stationary wavelet transform) wavelet decomposition on the image set. SWT consists of two opposite processes of decomposition and reconstruction, where the decomposition is the time series data

y (recorded as  $a^0$ ) through J iterations. Transform into a set of wavelet coefficients distributed in J + 1 wavelet metrics

$$c_y = [a^J, b^J, b^{J-1}, \dots, b^1] \tag{1}$$

$$\begin{cases} a^j = H^{[j-1]}a^{[j-1]}, b^j = G^{[j-1]}a^{[j-1]} \\ H^{[j]} = U_oH^{[j-1]}, G^{[j]} = U_oG^{[j-1]} \end{cases} \tag{2}$$

In the above formula, decomposition series  $j = 1, 2, \dots, J$ ;  $H^{[0]}$  and  $G^{[0]}$  are wavelet low pass, high pass decomposition filters.  $U_o$  indicates that zeros are inserted after each coefficient of the filter to double the filter length.  $a^j$  and  $b^j$  are called the j level low frequency scale and the high frequency scale, respectively. Through wavelet decomposition, different frequency band components in y are separated into different scales in  $c_y$ . The above formula is therefore called multi-scale analysis.

Unlike DWT, each wavelet scale of SWT is the same length as y, so SWT is a redundant transform [5].  $H^{[j-1]}, G^{[j-1]}$  decomposing  $a^{j-1}$  into  $(a^j, b^j)$  is non-orthogonal. Its inverse transformation is not unique. But if defined  $D_0$  and  $D_1$  perform a downsampling operator for the second choice of keeping the even and odd terms. Then transforms  $(D_oH^{[j-1]}, D_oG^{[j-1]})$  and  $(D_1H^{[j-1]}, D_1G^{[j-1]})$  are orthogonal transforms, and gives the even and odd terms of  $a^j$  and  $b^j$  respectively. Therefore, if we remember that the inverse transformation of the two is  $R_0^{[j-1]}$  and  $R_1^{[j-1]}$ . Then refactored SWT is

$$a^{j-1} = \frac{1}{2} (R_0^{[j-1]} + R_1^{[j-1]}) (a^j, b^j) \tag{3}$$

If  $j = 1, 2, \dots, J$ , then the sequence data y can be obtained from  $c_y$ . If the coefficient of the partial scale in  $c_y$  is kept unchanged, and the coefficient of the remaining scale. The discarding scale is set to zero, it is a partial scale reconstruction.

Because it is a redundant transform, SWT retains more information than DWT and has translation invariance, which is more conducive to time series analysis [6]. However, if the length of y is  $2^J$ , the complexity of SWT decomposition J is  $O(J2^J)$ , which is larger than  $O(2^J)$  of DWT.

For simple calculation, use W to represent J-level SWT decomposition, which is  $c_y = W_y$ . Use  $W^-$  to indicate the corresponding refactoring, which is  $y = W^-c_y$ . The use of  $S_0$  for partial scale reconstruction adds zero to the coefficient of the discarding scale.

Replace DWT with SWT and use wavelet sym2 to study wavelet extraction [7]. Sym2 is an approximate symmetric wavelet function with better symmetry, which can reduce the phase distortion when analyzing and reconstructing signals to a certain extent.

### 2.2 Principal Component Analysis

Principal Component Analysis (PCA) is the most commonly used linear mapping method in pattern recognition analysis [8]. It is based on the position distribution of sample points in multi-mode space, and the maximum direction of sample points in space. The direction with the largest variance is used as the discriminant vector to achieve feature extraction [9].

Face frequency map as raw data of PCA [10]. Let a known low-frequency sub-graph form a column vector whose size  $D = M \times N$  is a dimension. Let  $n$  be the number of training samples, and  $X_i$  denote the face vector formed by the  $i$ -th face low-frequency sub-graph, then the covariance matrix of the required samples is

$$S_r = \sum_{i=1}^n (X_i - \mu)(X_i - \mu)^T \tag{4}$$

Where  $\mu$  is the average image vector of the training sample

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i \tag{5}$$

Let  $A = [x_1 - \mu, x_2 - \mu, \dots, x_n - \mu]$ , there is  $S_r = AA^T$ , and its dimension is  $D \times D$ . According to the K-L transform principle [11], the new coordinate system to be required is composed of the feature vectors corresponding to the non-zero eigenvalues of the matrix  $AA^T$ .

The calculation of direct calculation is relatively large, so the eigenvalues and eigenvectors of  $AA^T$  are obtained by solving the eigenvalues and eigenvectors of  $A^T A$  using the SVD (singular value decomposition) theorem.

According to the SVD theorem, let  $\lambda_i = (i = 1, 2, \dots, r)$  be the  $r$  non-zero eigenvalues of matrix  $A^T A$ .  $v_i$  is the eigenvector of  $A^T A$  corresponding to  $\lambda_i$ , then the orthogonal normalized eigenvector  $\mu_i$  of  $AA^T$  is

$$\mu_i = \frac{1}{\sqrt{\lambda_i}} \tag{6}$$

Then the feature face subspace is  $\omega = (\mu_1, \mu_2, \dots, \mu_r)$ . The training sample is projected into the feature face subspace, and a set of projection vectors  $W = w^T \mu$  is obtained to form a database for face recognition [12]. When identifying, each image of the face to be recognized is first projected into the feature face subspace, and is identified as input data of the SVM classifier.

### 3 Classification and Recognition Based on SVM

Support Vector Machine is a learning method based on statistical learning theory developed [13] in the 1990s. The generalization ability of machine learning is improved by seeking the minimum of structured sorting, and the empirical risk and

confidence range are minimized. In the case of a small amount of statistical samples, good statistical laws can also be obtained.

In the training data, each data has  $n$  attributes and a second class of category markers, we can think of these data in an  $n$ -dimensional space. Our goal is to find an  $n-1$  dimensional hyperplane. This hyperplane divides the data into two parts, each of which belongs to the same category. In fact, there are many such hyperplanes, we have to find the best one [14]. Therefore, a constraint has been added, the distance from this hyperplane to the nearest data point of each class is the largest.

A linearly separable data set  $\{(\vec{x}_1, y_1), (\vec{x}_2, y_2), \dots, (\vec{x}_N, y_N)\}$ , ample feature vector  $\vec{x} \in D^T$ , that is,  $\vec{x}$  is a vector in the  $D$ -dimensional real space. Class label  $y \in \{-1, +1\}$ , that is, there are only two types of samples. Usually, a sample with a class label of  $+1$  is a positive example, and a sample with a class label of  $-1$  is a counterexample. Now classify these two types of samples [15]. The goal is to find the optimal segmentation hyperplane, i.e. the segmentation hyperplane of the largest classification interval determined from the training samples. We set the equation for the optimal hyperplane as  $\frac{\vec{w}}{w} \cdot \frac{\vec{x}}{x} + b = 0$ . According to the point-to-plane distance formula,

the distance between sample  $\frac{\vec{x}}{x}$  and the best hyperplane  $\left(\frac{\vec{w}}{w}, b\right)$  is  $\frac{\frac{\vec{w}}{w} \cdot \frac{\vec{x}}{x} + b}{\left\| \frac{\vec{w}}{w} \right\|}$ . By

scaling the vector  $\frac{\vec{w}}{w}$  and the deviation term  $b$  proportionally, there are many solutions to the optimal hyperplane. The hyperplane is normalized to select the sample  $\vec{x}_k$  that is closest to the hyperplane to satisfy  $\frac{\vec{w}}{w}$  and  $b$  of  $\left| \frac{\vec{w}}{w} \cdot \frac{\vec{x}_k}{x_k} + b = 0 \right| = 1$ . We can get normalized sample hyperplane. The distance from the nearest sample to the edge is

$$\frac{\frac{\vec{w}}{w} \cdot \frac{\vec{x}}{x} + b}{\left\| \frac{\vec{w}}{w} \right\|} = \frac{1}{\left\| \frac{\vec{w}}{w} \right\|} \tag{7}$$

And the classification interval becomes

$$m = \frac{2}{\left\| \frac{\vec{w}}{w} \right\|} \tag{8}$$

In fact, it can be seen from the above formula that the most critical point for finding the support vector is to find the upper method to maximize the normal vector  $\frac{\vec{w}}{w}$ . Put  $\frac{\vec{w}}{w}$  with a relationship  $\frac{\vec{w}}{w} \cdot \frac{\vec{x}}{x} + b = 1$  to get  $b$ . The key to the entire support vector is to find the following objective function:

$$\arg \max_{\vec{w}, b} \left\{ \frac{1}{\left\| \frac{\vec{w}}{w} \right\|} \min_i \left[ y_i \left( \frac{\vec{w}}{w} \cdot \frac{\vec{x}_i}{x_i} + b \right) \right] \right\} \tag{9}$$

Find the maximum spacing of the nearest support vector points for the two types of data, and satisfy the conditions  $\vec{w}$  and  $b$ . At the same time, there is a constraint on this formula:

$$y_i \left( \vec{w}^T \vec{x}_i + b \right) \geq 1 \tag{10}$$

At this point, the problem of finding support vector points is transformed into an extremum problem of a function with constraints. For the problem of this kind, we can solve it by Lagrangian multiplier method to get the following formula:

$$L(\vec{w}, b, \alpha) = \frac{1}{2} \|\vec{w}\|^2 - \sum_{i=1}^N \alpha_i \left[ y_i \left( \vec{w}^T \vec{x}_i + b \right) - 1 \right], \alpha_i > 0 \tag{11}$$

Find the partial conductance of L on  $\vec{w}$  and  $b$  make it equal to zero.

$$\frac{\partial(L, \vec{w}, b, \alpha)}{\partial \vec{w}} = 0 \Rightarrow \vec{w} = \sum_{i=1}^N \alpha_i y_i \vec{x}_i \tag{12}$$

$$\frac{\partial(L, \vec{w}, b, \alpha)}{\partial b} = 0 \Rightarrow \sum_{i=1}^N \alpha_i y_i = 0 \tag{13}$$

Bring the above two styles into the following formula.

$$L(\vec{w}, b, \alpha) = \frac{1}{2} \vec{w}^T \vec{w} - \sum_{i=1}^N \alpha_i y_i \vec{w}^T \vec{x}_i - b \sum_{i=1}^N \alpha_i y_i + \sum_{i=1}^N \alpha_i \tag{14}$$

Find the support vectors of the two types of data sets by finding  $\vec{w}$  and  $b$  that satisfy the maximum value of the above formula.

The commonly used kernel functions are linear kernel function, polynomial kernel function, Gaussian kernel function (RBF kernel function) and sigmoid kernel function [16]. The choice of kernel function has a great influence on the classifier. We compare and analyze the following kernel functions [17].

The mapping function corresponding to the RBF kernel function projects the sample into an infinite dimensional space, and performs a polynomial expansion on the RBF kernel function to obtain the result. Second, after mapping to the new space, all sample points are distributed over a 1/4 sphere with a radius of 1 at the origin as:

$$K(x_i, x_j) = \exp \left( - \frac{\|x_i - x_j\|^2}{2\sigma^2} \right) \tag{15}$$

Linear kernel are mainly used for linear separability. We can see that the dimension of the feature space to the input space is the same. Its parameters are less fast, and for linearly separable data, the classification effect is very good, so we try to use linear kernel function to do classification as:

$$K(x, x_i) = x \cdot x_i \quad (16)$$

Polynomial kernel functions can map low-dimensional input spaces to high-dimensional feature spaces, but the polynomial kernel function has many parameters. When the order of the polynomial is relatively high, the element values of the kernel matrix will tend to infinity or infinitesimal. The computational complexity is too large to calculate as:

$$K(x, x_i) = ((x \cdot x_i) + 1)^d \quad (17)$$



**Fig. 2.** (a) The original image; (b) Decompose the original image to the image of the fourth layer based on sym2 wavelet;

## 4 Experimental Results and Analysis

We first used wavelet transform to extract features. Wavelet transform can reduce dimension while extracting features, and the extracted feature data was classified by SVM. In this experiment, we use python language for code writing, wavelet transform USES the third-party package pywt in python, and SVM USES the popular third-party machine learning extension library scikit-learn in python. The experimental hardware environment we used was inter(R) Core(TM) i7-8700k CPU @3.70 ghz processor, 32G RAM and 1080TI GPU.

In this study, we used the standard face recognition data set ORL. In this data set, each person has 10 different pose faces and 40 people have 400 photos. During the experiment, we selected one, two, three, four, five, six and seven face poses of each person as the training set, and the rest as the test set.

In the experimental process, we use the sym2 wavelet base in the pywt library. The Symlets wavelet system is a finitely tightly supported orthogonal wavelet, which has strong localization ability in the time domain and frequency domain, especially in the wavelet decomposition process of the signal. Actually more specific digital filters, so the sym2 wavelet base in the Symlets wavelet system was chosen to perform wavelet transform on image features. In the experiment, we use wavelet to decompose the original image into the fourth layer which as shown in Fig. 2. The feature data after wavelet decomposition is used. The image before and after decomposition is shown in Fig. 3. For the feature data after wavelet transform, we use SVM for multi-classification. The prediction and real results of SVM classification test set are shown in Fig. 4. In SVM, we set the value of parameter  $c$  to 1000.  $C$  is the error item. Penalty factor. The larger  $C$ , the greater the degree of punishment for the fault-divided sample, so the higher the accuracy in the training sample, but the lower the generalization ability, that is, the classification accuracy of the test data is reduced. Conversely, if  $C$  is reduced, there are some misclassification errors in the training samples, and the generalization ability is strong. For the case where the training sample is noisy, the latter is generally used, and the sample that is misclassified in the training sample set is used as noise. The kernel functions we can use are 'linear', 'poly', 'rbf', etc. We set the kernel function coefficient  $\gamma$  to 0.001.

All figures and tables should be cited in the main text as Fig. 1, Table 1, etc.

In the experiment, we also tried to use PCA as image feature extraction. We use the PCA in scikit-learn for dimension reduction. The feature data before and after dimension reduction is shown in Fig. 5. We set the value of `n_components` in sklearn.decomposition.PCA to be equal to 0.9, which is to retain 90% of the principal components. Then we use SVM to complete the classification identification. We also use wavelet transform for feature extraction, using the PCA for dimension reduction (the dimensionality reduction method is the same as above), and then using SVM to complete the classification and recognition. After three different methods of testing, we obtained the experimental results shown in Table 1.

From Table 1, we can see that the four-layer wavelet decomposition and SVM face recognition using Sym2 are better than the other two methods. With the increase of training samples, when the sample pose is increased to 7, the recognition rate classified by the method proposed in this paper can reach 100%.

Consider previous studies [18]. Comparing the proposed method with the previous method, we plot the recognition rate of two different methods using different kernel functions, as shown in Fig. 6 (RBF kernel), Fig. 7 (LINEAR kernel) and Fig. 8 (POLY kernel). When the face number of each training set is less than 2 (80 in total), the support vector machine with three different kernel functions is higher than the previous method. When the training set selects less than 4 faces per face (160 total), the previous method is superior to this method. When the training set is greater than 160, the recognition rate of this method is significantly higher than the previous method.





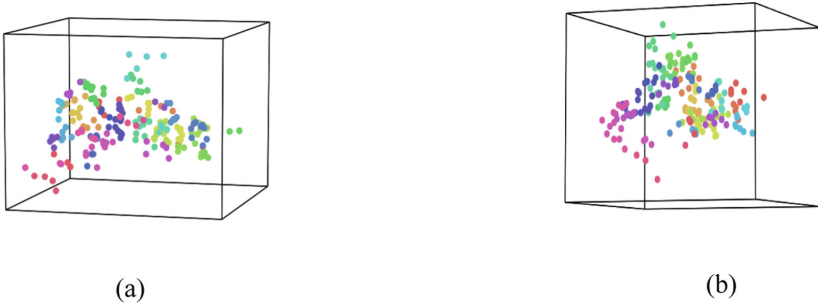
**Fig. 3.** (a) The distribution of the features of the first primitive face (the first one on the left in Fig. 2) in space; (b) The distribution of the features of the first original face after four layers of wavelet transform (the graph on the right in Fig. 2)



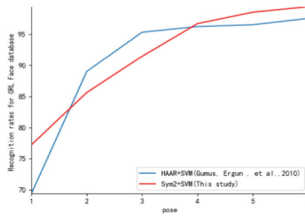
**Fig. 4.** (a) Training sample 240, test sample 160, when inputting test sample predicted classification result; (b) Training sample 240, test sample 160, when entering the correct classification result of the test sample

**Table 1.** Recognition rates of applied techniques according to increasing pose count

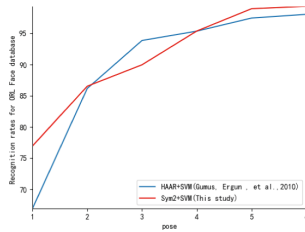
Technique	Pose count per individual in training (number of test images)						
	1/360	2/320	3/280	4/240	5/200	6/160	7/160
	%	%	%	%	%	%	%
Sym2_SVM(rbf)	77.22	85.62	91.43	96.67	98.50	99.38	100.0
Sym2_SVM(linear)	77.22	85.62	91.43	95.83	99.00	99.38	100.0
Sym2_SVM(poly)	76.94	86.56	90.00	95.42	99.00	99.38	100.0
PCA_SVM(rbf)	69.17	86.56	89.64	95.00	95.50	96.88	98.33
PCA_SVM(linear)	69.17	86.56	89.64	94.58	96.00	96.88	98.33
PCA_SVM(poly)	34.44	52.19	58.21	76.25	83.00	86.88	89.17
Sym2_PCA_SVM(rbf)	72.22	82.81	88.93	96.67	97.50	98.75	99.17
Sym2_PCA_SVM(linear)	72.22	82.81	88.57	96.25	97.50	98.12	98.33
Sym2_PCA_SVM(poly)	60.00	71.25	81.43	92.08	94.00	94.38	96.67



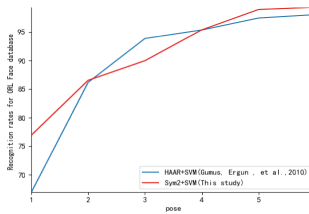
**Fig. 5.** (a) The spatial distribution of the training set before dimension reduction; (b) After the dimension reduction, the training set is spatially distributed.



**Fig. 6.** Comparison of two methods based on RBF kernel function recognition rate



**Fig. 7.** Comparison of two methods based on LINEAR kernel function recognition rate



**Fig. 8.** Comparison of two methods based on POLY kernel function recognition rate

## 5 Conclusions

In this paper, we propose a method based on Sym2 based wavelet transform and face recognition. This method is validated on the ORL data set. We select the feature data of the face through the wavelet transform of the 4-layer sym2 wavelet, and then use svm to complete the classification and recognition. We use three different kernel functions and compare them with the other two methods (One is using pca and svm, another is using wavelet transform, pca and svm). The experimental results show that the method is better than due to two other methods. This paper also compares this method with the previous method. With the increase of training samples, the recognition rate of this method is higher than the previous method. When the training sample reaches 280, the recognition rate of this method reaches 100%.

**Acknowledgments.** This work was supported by Key Research and Development Plan Project of Shandong Province, China (No. 2017CXGC0614).

## References

1. Chen, X., Wu, W., Fan, J.: Overview of face recognition technology. In: China Academic Association of Instrumentation Youth Conference (2010)
2. Mi, C., Wei, F.: Application of Haar wavelet transform in image processing. *Ningxia Eng. Technol.* **2**, 70–72 (2003)
3. Eslami, R., Radha, H.: Wavelet-based contourlet transform and its application to image coding. In: International Conference on Image Processing, Singapore, p. 3189. IEEE (2004)
4. Wang, Y., Chen, J., Wu, Q., Castleman, K.R.: Fast frequency estimation by zero crossings of differential spline wavelet transform. *EURASIP J. Appl. Signal Process.* **2005**, 1251 (2005)
5. Alickovic, E., Kevric, J., Subasi, A.: Performance evaluation of empirical mode decomposition, discrete wavelet transform, and wavelet packed decomposition for automated epileptic seizure detection and prediction. *Biomed. Signal Process. Control* **39**, 94 (2018)
6. Ma, L., Stückler, J., Wu, T., Cremers, D.: Detailed Dense Inference with Convolutional Neural Networks via Discrete Wavelet Transform (2018)
7. Bhattacharyya, A., Sharma, M., Pachori, R.B., Sircar, P., Acharya, U.R.: A novel approach for automated detection of focal EEG signals using empirical wavelet transform. *Neural Comput. Appl.* **29**, 47 (2018)
8. Mavroudis, D., Vazirgiannis, M.: Stability based sparse LSI/PCA: incorporating feature selection in LSI and PCA. In: Kok, J.N., Koronacki, J., Mantaras, R., Matwin, S., Mladenič, D., Skowron, A. (eds.) *ECML 2007. LNCS (LNAI)*, vol. 4701, pp. 226–237. Springer, Heidelberg (2007). [https://doi.org/10.1007/978-3-540-74958-5\\_23](https://doi.org/10.1007/978-3-540-74958-5_23)
9. Lee, M.-S., Chen, M.-Y., Lin, F.-S.: Face recognition under variant illumination using PCA and wavelets. In: Salberg, A.-B., Hardeberg, J.Y., Jensen, R. (eds.) *SCIA 2009. LNCS*, vol. 5575, pp. 341–350. Springer, Heidelberg (2009). [https://doi.org/10.1007/978-3-642-02230-2\\_35](https://doi.org/10.1007/978-3-642-02230-2_35)
10. Poon, B., Amin, M.A., Yan, H.: Performance evaluation and comparison of PCA Based human face recognition methods for distorted images. *Int. J. Mach. Learn. Cybern.* **2**, 245 (2011)

11. Caballero-Morales, S.-O.: Noise-removal markers to improve PCA-based face recognition. In: Martínez-Trinidad, J.F., Carrasco-Ochoa, J.A., Olvera-Lopez, J.A., Salas-Rodríguez, J., Suen, C.Y. (eds.) MCPR 2014. LNCS, vol. 8495, pp. 192–200. Springer, Cham (2014). [https://doi.org/10.1007/978-3-319-07491-7\\_20](https://doi.org/10.1007/978-3-319-07491-7_20)
12. Biswas, S., Sil, J., Maity, S.P.: PCA based face recognition on curvelet compressive measurements. In: Mandal, J.K., Dutta, P., Mukhopadhyay, S. (eds.) CICBA 2017. CCIS, vol. 775, pp. 217–229. Springer, Singapore (2017). [https://doi.org/10.1007/978-981-10-6427-2\\_18](https://doi.org/10.1007/978-981-10-6427-2_18)
13. Schuldt, C., Laptev, I., Caputo, B.: Recognizing human actions: a local SVM approach. In: Proceedings - International Conference on Pattern Recognition, vol. 3, p. 32. Institute of Electrical and Electronics Engineers Inc., Cambridge (2004)
14. Sadeghi, M.T., Khoshrou, S., Kittler, J.: SVM-based selection of colour space experts for face authentication. In: Lee, S.-W., Li, S.Z. (eds.) ICB 2007. LNCS, vol. 4642, pp. 907–916. Springer, Heidelberg (2007). [https://doi.org/10.1007/978-3-540-74549-5\\_95](https://doi.org/10.1007/978-3-540-74549-5_95)
15. Kar, N.B., Babu, K.S., Sangaiah, A.K., Bakshi, S.: Face expression recognition system based on ripplelet transform type II and least square SVM. *Multimedia Tools Appl.* **78**, 4789 (2019)
16. Manolova, A., Neshov, N., Panev, S., Tonchev, K.: Facial expression classification using supervised descent method combined with PCA and SVM. In: Cantoni, V., Dimov, D., Tistarelli, M. (eds.) Biometric Authentication. LNCS, vol. 8897, pp. 165–175. Springer, Cham (2014). [https://doi.org/10.1007/978-3-319-13386-7\\_13](https://doi.org/10.1007/978-3-319-13386-7_13)
17. Li, M., Yu, X., Ryu, K.H., Lee, S., Theera-Umpon, N.: Face recognition technology development with Gabor, PCA and SVM methodology under illumination normalization condition. *Cluster Comput.* **21**(1), 1117–1126 (2017)
18. Gumus, E., Kilic, N., Sertbas, A., Ucan, O.N.: Evaluation of face recognition techniques using PCA, wavelets and SVM. *Expert Syst. Appl.* **37**, 6404 (2010)