FACIAL EXPRESSION RECOGNITION USING STATISTICAL SUBSPACE

Dang-Khoa Tan Le, Hung Phuoc Truong, Thai Hoang Le

Faculty of Information Technology, Ho Chi Minh City University of Science, VNU-HCM
District 5, Ho Chi Minh City, Vietnam
Email: 1112144@student.hcmus.edu.vn, {tphung, lhthai}@fit.hcmus.edu.vn

ABSTRACT

Face recognition is one of the main areas of research in computer vision. Although many studies address to, there are many challenges in this subject such as accuracy, performance, real-time applications, etc. We propose a novel model based on bilateral 2-dimensional fractional principle component analysis and examine 2-dimensional characteristic of image to retain information structure. After that, we apply statistical features to facial expression recognition problem in order to evaluate the efficiency of feature descriptor with facial images. Our proposed method is named the statistical subspace. For experiments, Cohn-Kanade dataset is used to compare the proposed model with previous methods. The empirical results show that our model is stable and efficient.

***Index Terms*—** bilateral 2D principal analysis, statistical texture feature, fractional variance matrix, face recognition

1. INTRODUCTION

Facial expression recognition is an appealing research discipline in computer vision. There are many prospective applications of this field in living and science: human-computer interaction, telecommunication, behavioral science, video games, etc. Above all, there are many challenges for researchers about performance and accuracy.

There are two main approaches in face recognition problem: holistic method and component method. Holistic method uses a whole face image as an input. For instance, there are many holistic methods such as Principal Component Analysis (PCA), Linear Discrimination Analysis (LDA), etc. Component based, also known as feature-based method, uses local facial features to recognition. A good example of this is Gabor Wavelets [1].

PCA is the holistic approach [2], which is popular and wide used. PCA represents an image from high-dimensional space to lower dimensional subspace. Normally in PCA, the 2D image matrices are transformed to 1D image vectors by vectorization. The major disadvantage of this algorithm is the curse of dimensionality. It means that when the dimensionality increases, the volume of the space increases so fast (because of the size of image) that the data become sparse. As a result, it is difficult that the algorithm recognizes images precisely. To get over the limitation of traditional PCA, Two-Dimensional Principal Component Analysis (2DPCA) was proposed by Yang *et al*. (2004) [3]. Unlike PCA, 2DPCA calculates the covariance matrix directly on the image matrix without vectorization. Therefore, the dimension of covariance matrix is equal to width or height of image.

There are two approaches in 2DPCA: row-based 2DPCA and column-based 2DPCA. However, two kinds of 2DPCA have a drawback: they ignore the spatial information for other side. B2DPCA is proposed by H Kong *et al* [4] to overcome this problem: it uses the row and column vector to create two different subspaces.

In the other hand, Chaobang *et al* [5] obtain two new techniques called fractional principal component analysis (FPCA) and two-dimensional fractional principal component analysis (2D-FPCA). Based on a statistical of fractional moments – that is turned out effective in detection. FPCA aims to produce dense pixel value of covariance matrix of image dataset. According to [5], 2D-FPCA has higher performance and accuracy than 2D-PCA methods.

Motivated by the previous work about B2D-FPCA [6], we develop the statistical subspace model. That is a combination of B2D-FPCA and statistical texture features. The model creates a covariance matrix having dense value of elements and aims to perform dimensionality reduction into space as small as possible. Grayscale image is split to smaller blocks. All blocks of training dataset are projected into the subspace using B2D-FPCA. Then, the bases in new subspace are used to be patterns for statistical texture features (STF) [7]. Finally, the SVM technique classifies features which are extracted from patterns.

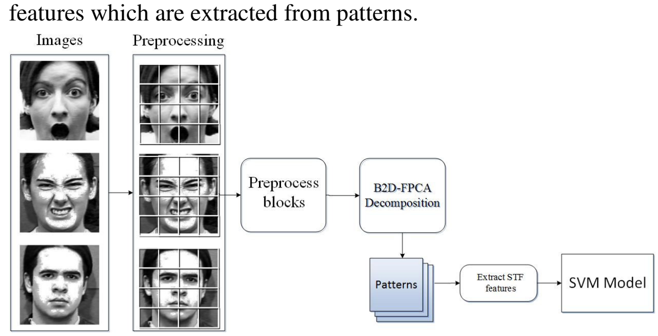

Figure 1. The proposed model

The rest of paper is organized as follow. The diagram of Fig. 1 describes our proposed model. In the next section, we describe the detail of the new model. We explain the reasons why we use 2D-FPCA and B2D-PCA in order to apply this model in Section 3. The experiments about this model are presented in Section 4. Finally, the conclusions and our future works are given in Section 5.

2. PROPOSED MODEL

There are four stages in our model. Firstly, we detect only face in image dataset using Haar-like feature, and then enhance the contrast quality of them. Subsequently, we compute eigenvectors to create patterns by B2D-FPCA algorithm. Thirdly, from these patterns, we extract features using statistical texture feature. Finally, we use SVM model for classification. The description of model is defined as follow.

2.1. Image preprocessing

Supposing that there be N images in training phase, which denoted by $(m \times n)$ matrix $I_i (i = 1, 2, ..., N)$, the first stage is to detect face by using Haar-like features. With aim of contrast adjustment, the image is enhanced by histogram equalization technique. After that, each image I_i is split into $p \times p = K$ blocks by using Eq. (1). Therefore, each of the blocks is a $(\frac{m}{p} \times \frac{n}{p})$ matrix. The Figures 2 demonstrates the gray image split into 30 blocks.

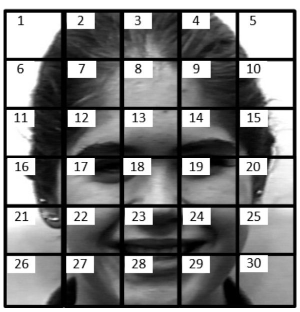

Figure 2. Sample image is split into 30 blocks

$$I = \begin{pmatrix} X_{11} & \cdots & X_{1p} \\ \vdots & \ddots & \vdots \\ X_{p1} & \cdots & X_{pp} \end{pmatrix} \quad (1)$$

The collected block set is denoted as:

$$X = \{X_1, X_2, ..., X_{K \times N}\}$$

We compute the average block of all blocks: $\bar{X} = \frac{1}{K \times N} \sum_{i}^{K \times N} X_i$

Next, the sample block is subtracted with average block to get mean centered image $X_i = X_i - \bar{X}$.

The final step in preprocessing stage is to normalize matrix X_i in order to obtain the final matrix by Frobenius norm $X_i = \frac{X_i}{\|X_i\|_F}$.

2.2. Bilateral two-dimensional fractional principal components analysis

2.2.1. Row-direction 2D-FPCA

The input matrix X_i is the element of set X, which is normalized in previous stage. The Fractional Covariance Matrix (FCM) dimensionality has $\left(\frac{n}{p} \times \frac{n}{p}\right)$ where Rd donates row-direction computation.

$$C_{r(Rd)} = \frac{1}{K \times N} \sum_{i=1}^{K \times N} (X_i^T)^{(r)} \times (\mathrm{X}_i)^{(r)} \quad (2)$$

Where let $X_i = (a_i^{kl})_{\frac{m}{p} \times \frac{n}{p}}$ and $X_i^{(r)} = ((a_i^{kl})^r)_{\frac{m}{p} \times \frac{n}{p}}$ ($k = 1 ... \frac{m}{p}$, $l = 1 ... \frac{n}{p}$) and $(0 < r < 1)$.

2.2.2. Column-direction 2D-FPCA

Similarity to Column-base 2D-FPCA matrix, the FCM dimensionality has $\left(\frac{m}{p} \times \frac{m}{p}\right)$ where Cd denotes Column-direction computation for the equation.

$$C_{r(Cd)} = \frac{1}{K \times N} \sum_{i=1}^{K \times N} (X_i)^{(r)} \times (X_i^T)^{(r)} \quad (3)$$

From Column-direction and Row-direction FCM, we can compute the optimal projection matrices. As we see in Eq. (2), (3), the row-direction and column-direction have $\left(\frac{n}{p} \times \frac{n}{p}\right)$ and $\left(\frac{m}{p} \times \frac{m}{p}\right)$ dimensions respectively, thus, they are smaller than $(n \times m)$ dimensions in original PCA.

2.2.3. Computing eigenvectors and constructing patterns

Let $R \in \mathbb{R}^{\frac{n}{p}} \times \mathbb{R}^d$ and $C \in \mathbb{R}^{\frac{m}{p}} \times \mathbb{R}^d$ be the left- and right-multiplying projection matrix.

$$\begin{aligned} R_{\frac{n}{p} \times d} &= [\mathrm{r}_1, \mathrm{r}_2, ..., \mathrm{r}_d] \\ \mathrm{C}_{\frac{m}{p} \times d} &= [\mathrm{c}_1, \mathrm{c}_2, ..., \mathrm{c}_d] \end{aligned} \quad (4)$$

For $(d \times d)$ projected block Y_i, the bilateral projection is formulated as follow:

$$Y_i = C^T X_i R \quad (5)$$

The optimal projection matrices, C and R in Eq. (5) can be computed by solving the minimization criterion:

$$J(C,R) = \min \sum_{i=1}^{K\times N} \left\| X_i - CY_iR^T \right\|_F^2 \quad (6)$$

When $\|\bullet\|_F$ is Frobenius norm of a matrix. To solve the optimization problem in Eq. (6), we use the iterative algorithm proposed by Kong *et la* [4].

The pattern is made from the product between the column base r_u and the row base c_v (in Eq. (4)) as follow Eq. (7). Therefore, we have $(d \times d)$ patterns whose size are $(\frac{m}{p} \times \frac{n}{p})$, equal to the size of block.

$$E_{uv} = r_u c_v^T \ (1 \le \mathrm{u}, \mathrm{v} \le \mathrm{d}) \quad (7)$$

For instance, the Figures 3 demonstrates the pattern of Cohn-Kanade dataset.

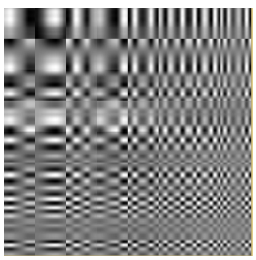

Figure 3. Pattern of Cohn-Kanade dataset

2.3. Filtering an image with patterns

There are two main stages in this process. Let be the gray-level image $(m \times n)$ I and a pattern $\{r,c\}$. Firstly, we filter the image I with r_c and obtain a response matrix T. We extract a $\frac{m}{p} \times 1$ neighbor column vector x centering at pixel (i,j) in I. Next, we calculate the response value in T by dot product x and r together.

$$x = \left[\mathrm{I(i, j}-\frac{1}{2}(\frac{m}{p}-1),...,\mathrm{I(i,j)},...,\mathrm{I(i, j}+\frac{1}{2}(\frac{m}{p}-1) \right]^T \quad (8)$$

$$T(i,j) = \langle r, x \rangle \quad (9)$$

The computation cost of this stage is $O(mnpd)$.

Secondly, we filter the response matrix T with c and obtain the final response matrix Z. We extract a $\left(\frac{n}{p} \times 1\right)$ neighbor row vector y centering at pixel (i,j) in T. Then, we calculate the response value in Z by dot product y and c together.

This step also have computation cost $O(mnd\frac{m}{p})$.

$$y = \left[\mathrm{T(i}-\frac{1}{2}\left(\frac{n}{p}-1\right),\mathrm{j}),...,\mathrm{T(i,j)},...,\mathrm{T(i}+\frac{1}{2}\left(\frac{n}{p}-1\right),\mathrm{j}) \right]^T \quad (10)$$

$$Z(i,j) = \langle c, y \rangle \quad (11)$$

Finally, we have a filter response Z with pattern $\{r,c\}$. For $(p \times p)$ patterns, the computational cost of the filtering stage is $O(mnd\frac{n}{p} + mnd\frac{m}{p})$.

2.4. Using 2-Dimensional Statistical Texture Features for extracting features

With 2D-STF, all filter responses are treated as weights contributing to pattern matching.

Firstly, $Z(\mathrm{i,j})$ is normalized by Frobenius norm.

$$Z(\mathrm{i,j}) = \frac{Z(\mathrm{i,j})}{\left\| rc^T \right\|_F} \quad (12)$$

Next, compute the element k of feature vector by:

$$f_k = \sum_R |Z_k|, k = 1,2,...,m \quad (13)$$

Where m is the number of patterns and $|Z_k|$ is the determinant of matrix Z_k.

In the final step, we have a feature vector normalized by dividing for sum of magnitude of all patterns.

$$F = \frac{[\mathrm{f}_1, \mathrm{f}_2, ..., \mathrm{f}_m]^T}{\sum_{k=1}^m f_k} \quad (14)$$

Therefore, we have patterns from statistical subspace and the data of SVM model.

2.5. Classification

Suppose that we a get face from a testing image A by using Haar-like algorithm. Similar to the preprocessing stage in training, the image is enhanced by histogram equalization.

In the next step, we have patterns already in training phase (Eq. (2), (3), (4)). Thus, we use the patterns to construct the filter response for testing image by using Eq. (8), (9), (10), (11).

The third step is to extract STF features (Eq. (12), (13), (14)) from testing image. Finally, we give these features into learning model to predict. For the classification problem, we adopt SVM as learning model because its efficiency.

3. ANALYSIS

We split image into blocks in order to analyze the distribution of each part to whole image (Eq. (1)). The patterns indicate the correlation of blocks (by computing the covariance matrix of block set in Eq. (2), (3)) and thus it enhances the corners and edges of image. Performing on data with parameter r $(0 < r < 1)$ (Eq. (2), (3)) is in order to produce dense values of each image play an essential role in enhancement data. The number of patterns makes the subspace more variability than the subspace constructed from a whole image. For above reasons, statistical subspace means that the space representing images is constructed by blocks and then the features are extracted from statistical texture.

Table 1. The accuracy of statistical subspace model with varying parameters on Cohn-Kanade dataset (%)

r parameter	The number of eigenvector									
	2	4	6	8	10	12	14	16	18	20
0.1	66.03	97.94	99.14	99.69	99.46	99.53	99.52	99.52	99.47	99.68
0.2	66.74	98.03	99.28	99.69	99.52	99.61	99.61	99.66	99.62	99.63
0.3	67.26	98.10	99.30	99.71	99.59	99.53	99.54	99.52	99.55	99.63
0.4	68.08	98.17	99.34	99.67	99.58	99.59	99.63	99.69	99.63	99.61
0.5	68.71	98.23	99.37	99.63	99.53	99.53	99.56	99.56	99.51	99.66
0.6	69.52	98.19	99.50	99.71	99.60	99.57	99.67	99.60	99.51	99.69
0.7	70.09	98.32	99.52	99.74	99.54	99.56	99.61	99.66	99.57	99.66
0.8	70.42	98.37	99.48	99.71	99.60	99.62	99.64	99.62	99.56	99.69
0.9	70.61	98.39	99.47	99.67	99.63	99.64	99.68	99.63	99.54	99.58
1.0	70.84	98.40	99.47	99.63	99.62	99.67	99.62	99.70	99.62	99.67

By the idea of histogram of oriented gradients (HOG) [8], 2D-STF calculates the histogram of edge orientations within an image (Eq. (14)). Thus, we emphasize the distributions of all block rather than the maximum filter response of previous methods such as BBPTF of Zeng *et al* [9], HOG [8] .

4. EXPERIMENTS

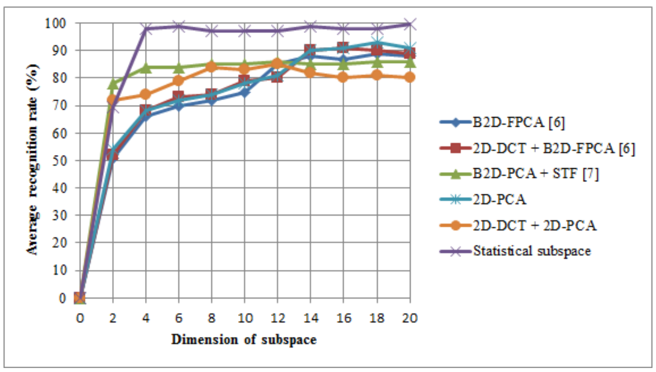

Figure 4. Classification performance of different methods with varying dimensions

In this section, we investigate the performance of B2D-FPCA and STF features on the problem of facial expression recognition. We use Cohn-Kanade dataset [10] to compare with other approaches and examine the efficiency of statistical subspace in recognition problem. The Cohn-Kanade dataset consists of 100 university students aged from 18 to 30 years, which contains 7 types of facial expressions: angry, disgust, fear, happy, sadness, surprise and neutral. For our experiments, we collect 30 faces in Cohn-Kanade that consist 6 emotions (each emotion has 10 images). We do not use the neutral emotion because we just interest in the variation of situation on facial images. All experiments are implemented on a PC machine with processor Core i5 2.4 GHz and 4GB RAM under Visual Studio 2012 and OpenCV version 2.4.6.

In the Table 1, we examine a degree of variability in whenever the parameter r and the number of dimensions change. We select 30 images for training phase and 30 images for testing phase. In the majority of instances, the optimal parameter r for recognition problem is about $(0.6 < r < 0.9)$. This parameter produces frequency values of each image and tries to clarify features in facial image texture. According to experimental results, it has a very effective outcome.

We compare the performance of our proposed model and other methods with varying dimension of subspace (Fig. 4). The dimension number d is in range of 0 to 20. It is apparent that the statistical subspace method has higher rate than others do. When $d = 4$, the performance rate of our proposed model is higher than B2D-PCA+STF model about 10% and higher than B2D-FPCA 20%. Between 4 and 20 the dimensions number changes, the rate of our model is higher than other methods a 8%.

Table 2. The accuracy of our proposed method with r = 0.6 on Cohn-Kanade dataset

Training/testing	2/8	4/6	5/5	6/4	8/2
$d = 2$	84.80	68.73	69.52	64.17	68.73
$d = 4$	99.79	97.72	98.19	99.60	97.80
$d = 6$	99.86	99.28	99.50	99.60	99.24
$d = 8$	99.86	99.45	99.71	99.45	99.50
$d = 10$	99.73	99.53	99.60	99.52	99.43

The Table 2 shows the accuracy of our proposed model in k-fold validation with $r = 0.6$ and d in range of 2 to 10 and data divided into 10 parts. The experiments results show that the accuracy is high with small dimension number (d = 4, 6, 8). If we use higher the number of dimensions, the phenomenon of overfitting will occur. This is the reason we select the small statistical subspace instead of higher subspace.

5. CONCLUSION

We proposed statistical subspace based on the fundamentals of statistical approach to construct the efficient feature descriptor. Moreover, we examine the density of value of input matrix, retain the 2-dimensional attitude of image by adopting bilateral PCA and discuss about the effectiveness of statistical subspace on facial recognition problem. We determined parameters of feature descriptor and analyze the importance of distribution of patterns in recognition problem. In future work, we will apply this framework to various databases in order to improve it more convincing.

6. ACKNOWLEDGEMENTS

The National Foundation funded this research for Science and Technology Development (NAFOSTED) of Vietnam.

7. REFERENCES

[1] V. Bettadapura, "Face Expression Recognition and Analysis: The State of the Art", *Tech Report, arXiv:1203.6722*, April 2012.

[2] Patil, A. M., S. R. Kolhe, and P. M. Patil, "2d face recognition techniques: A survey" *International Journal of Machine Intelligence* 2.1, 2010.

[3] Yang, J., Zhang, D., Frangi, A. F., and Yang, J. Y., "Two-dimensional PCA: a new approach to appearance-based face representation and recognition", *Pattern Analysis and Machine Intelligence*, IEEE Transactions on 26.1, 2004, 131-137.

[4] H. Kong, L. Wang, E.K. Teoh, X.C. Li, J.G. Wang, R. Venkateswarlu, "Generalized 2D principal component analysis for face image representation and recognition," *Neural Networks 5,* pp. 585–594, 2005.

[5] Chaobang Gao, Jiliu Zhou, Qiang Pu, "Theory of fractional covariance matrix and its applications in PCA and 2D-PCA", *Expert Systems with Applications*, vol. 40 (13), pp. 5395-5401, 2013.

[6] Thai Hoang Le, Hung Phuoc Truong, Ha Thi Thanh Do, and Duc Minh Vo, "On approaching 2D-FPCA technique to improve image representation in frequency domain," Proceedings of the Fourth Symposium on Information and Communication Technology, pp.172-180, 2013.

[7] Dong Wang, Huchuan Lu, and Xuelong Li., "Two dimensional principal components of natural images and its application" *Neurocomputing*, pp. 2745-2753, October 2011.

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

[9] Xiang-Yan Zeng, Yen-Wei Chen, Zensho Nakao, Hanqing Lu, "Texture representation based on pattern map", *Signal Processing*, Volume 84, Issue 3, pp. 589-599, March 2004.

[10] Kanade, Takeo, Jeffrey F. Cohn, and Yingli Tian, "Comprehensive database for facial expression analysis," *Automatic Face and Gesture Recognition*, Fourth IEEE International Conference, 2000.