A Study on Real-time Face Verification and Tracking with Segmented Common Vector

Dong-kyu Ryu, Min-ho Park and Ju-Jang Lee

Department of Robotics Program
Korea Advanced Institute of Science and Technology (KAIST)
Daejeon, Republic of Korea
(Tel : 82-42-350-8032; Fax : 82-42-350-5432)
(E-mail: ghostdz@naver.com)

Abstract: Recently, There has been much interest in automatically face recognition and tracking in many areas such as Intelligent Robotics, Military, Intelligent Transport System (ITS) and Smart device applications. However so far there has not been simultaneously face tracking and verification algorithms, so we propose the method in simultaneously verifying and tracking face which is segmented common vector method and It is theoretically based on discriminative common vector method and Fisher's LDA. The algorithm trains segmented and shift face image to obtain new face vector. The Gram-Schmidt orthogonalization is employed to calculate the orthogonal projection matrix. Our goal can be defined as the identification of individuals and real-time face tracking from video images simultaneously, so we use segmented common vector for two kind of different task at the same time.

Keywords: Discriminative Common Vector, Fisher's LDA, Gram-Schmidt orthogonalization**.**

I. INTRODUCTION

In computer-vision research area, there has been much demand in verification of human face in video and photo images due to intelligent robotics, military, smart phone and surveillance system. The problem of face verification can be defined as finding the identification of test faces from sample face images which was previously learned, however there are many complex problem in this task. For example, face data has numerous changeable factors such as 3-D pose, hair style, make up, and facial expression. In addition to these can be varied due to lighting, background, and scale changes. Thus there are many researches solving face verification problem as follow.

For face recognition, Harr-like feature and AdaBoost algorithm of Viola and Jones is very famous in face recognition. It is very fast and has high-rate accuracy [1]. In face identification tasks, the verifying face has many hard problems, so it is very important to extract common factors from various data sets for face identification.

To do this, eigen face of Turk and pentland is introduced, and this method uses Principal component analysis (PCA) which is used to find the best set of projection directions in the sample space. From here, the projection directions are called the eigen faces which is vector set having common factors. [2]

After proposed eigen face, several research have been more performed. The Linear Discriminant Analysis (LDA) method is proposed to overcome the limitations of the eigen face method. This method can find projection directions that on one hand maximize the distance between samples of different face and on the other hand minimize the distance between the sample of the alike face [3]. But, in face recognition task, we can't directly calculate projection directions, because dimension of the image samples is very larger than the number of samples which is known as the "small sample size problem"

To solve this problem, Chen proposed the Null space method based on the modified Fisher's LDA [4]. The Null space method is based on the modified Fisher's Linear Discriminant criterion,

$$J_{MFLD}(W_{opt}) = \arg\max_{W} \frac{|W^T S_B W|}{|W^T S_T W|} \quad (1)$$

For this method, all image samples are first projected onto the null space of S_W and then result in a new within –class scatter that is a zero matrix.

In the special case where $W^T S_W W = 0$ and $W^T S_W W \neq 0$, for all $w \in R^d \setminus \{0\}$, and then a better criterion is given as

$$J(W_{opt}) = \arg\max_{|W^T S_W W|=0} |W^T S_B W| = \arg\max_{|W^T S_W W|=0} |W^T S_T W| \quad (2)$$

The Discriminative Common Vector Method was proposed to extract the common properties of classes in the training set by eliminating the differences of the samples in each class, it gives more efficient algorithm to find Null space of within-class scatter matrix [5]. In the special case where $W^T S_W W = 0$ and, $W^T S_W W \neq 0$ for all $w \in R^d \setminus \{0\}$, and then a better criterion is given as

$$J(W_{opt}) = \underset{|W^T S_W W|=0}{\arg\max} |W^T S_B W| = \underset{|W^T S_W W|=0}{\arg\max} |W^T S_T W| \quad (2)$$

To find the optimal projection vectors w in the null space of S_W, we have to project the samples onto the null space of S_W and then obtain the projection vectors by performing PCA.

II. Segmented Common Vector

1. Definition of Segmented Common Vector

We propose the simultaneously face verifying and tracking method which is segmented common vector method. It is theoretically based on discriminative common vector method. However this method presents that several weak classifiers can more easily solve hard problem than one strong classifier such as face tracking problem. So that we have to make several feature vector from each one sample to create various sub classifiers, before the one sample was transferred to only one face feature vector.

2. Algorithm to calculate Segmented Common Vector

A. Span new image sample vector

In this section, we account for process in making segmented common vector in detail.

To make segmented common vectors, in advance we have to make segmented image to span new face vector. The size of segmented image is usually 60~90% than original face size. The size of segmented image is saved as cell size value and it used for face tracking. We should take apart scrap images to be overlapped, because the center of face has many information such as eyes, eyebrows, nose and mouse. The segmented feature extract manner is presented as below.

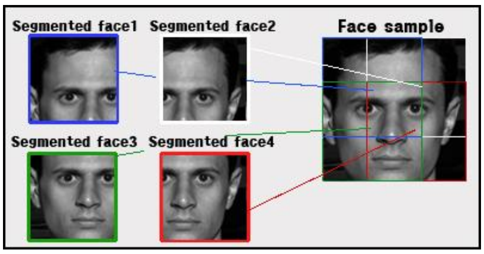

Fig.1 The Manner of Segmenting face image

In this paper, the number of segmented face feature has range from 4 to 9 features

Suppose that we have several people images, each people becomes each classes, thus the training set is composed of C classes.

To solve real-time tracking problem, new face sample set be generated by using shift-image method.

If we have P samples in each all classes, firstly we should shift face image to 1-3 pixels in order of left, right, top and bottom. This is called shifted sample set, so we can get $N = 5 \times P$ samples. And then we take segmented samples apart this N samples.

- *Definition of sample vectors* $x^C_{N,S}$

Let $x^i_{m,k}$ *be a* d *-dimensional column vector and* C *be denoted by i-th number of classes,* S *be denoted by k-th segment set in i-th classes and* N *is index of m-th sample from k-th segment set in i-th class.*

The segmented face vector sets is proposed as sub-class sets which made by segmenting face images from C classes as below.

$$sub\,class \; : \; x^i_k = span\{x^i_{1,k}, x^i_{2,k}, \ldots, x^i_{N,k}\} \quad (3)$$

$$i-th\ Class: \; x^i \supset x^i_1, \; x^i_2, \ldots, \; x^i_S$$

So we can have N x S samples from existing one class and C x S sub-classes which has N samples, then the number of total samples become $M = N \times S \times C$ in training set. Suppose that $d > M - C$. In this case, S_W, S_B, and S_T are newly defined as

$$S_W = \sum_{i=1}^{C}\sum_{k=1}^{S}\sum_{m=1}^{N}(x^i_{m,k} - \mu^i_k)(x^i_{m,k} - \mu^i_k)^T, \quad (4)$$

$$S_B = \sum_{i=1}^{C}\sum_{k=1}^{S}(\mu^i_k - \mu)(\mu^i_k - \mu)^T, \quad (5)$$

and

$$S_T = \sum_{i=1}^{C}\sum_{k=1}^{S}\sum_{m=1}^{N}(x^i_{m,k} - \mu)(x^i_{m,k} - \mu)^T = S_W + S_B \quad (6)$$

B. Span difference space

To span difference spaces B, we choose any one of the face vectors from the k-th segment sets in i-th class. Difference space is made by subtraction of vector.

$$b^i_{m,k} = x^i_{m+1,k} - x^i_{1,k}, \quad (7)$$

$$(i = 1, \ldots, C, \; k = 1, .., S, \; m = 1, \ldots, N-1)$$

The $b^i_{m,k}$ is difference vector from k-th segment in i-th class, so difference space is defined as B^i_k.

The complete difference subspace is represented as below

$$B = B^1_1 + B^1_2 + \ldots + B^1_S + B^2_1 + \ldots + B^C_S$$
$$= span\{b^1_{1,1}, \ldots, b^1_{N-1,1}, b^1_{1,2}, b^1_{2,2}, \ldots, b^C_{N-1,S}\} \quad (8)$$

To obtain orthonormal basis vectors $\beta_1, \ldots, \beta_{M-C}$, the difference vectors is orthonormalized by using the Gram-Schmidt-orthogonalization procedure.

Its Algorithm is explained as below.

C. Algorithm of learning Segmented Common vector

Step 1: Find the linearly independent vectors $b^{i}_{m,k}$ that span the difference subspace

$$B = span\{b^{1}_{1,1}, ..., b^{1}_{N-1,1}, b^{1}_{1,2}, b^{1}_{2,2}, ..., b^{C}_{N-1,S}\}$$

Step2: Apply the Gram-Schmidt orthogonalization procedure to obtain an orthonomal basis $\beta_1, ..., \beta_{M-(C\times S)}$ for B

Step 3: Choose any sample from each class and project it onto B to obtain common vectors by using

$$U = [\beta_1 \quad .... \quad \beta_{M-(C\times S)}], \qquad P = UU^T$$

$$x^{i}_{k,com} = \overline{P}x^{i}_{m,k} = x^{i}_{m,k} - Px^{i}_{m,k} = x^{i}_{m,k} - UU^T x^{i}_{m,k},$$

Step 4: Find the difference vectors that span B_{com}

$$b^{i}_{k,com} = x^{i}_{k+1,com} - x^{1}_{1,com}, \qquad i = 1, ..., C, \quad k = 1, ..., S$$

Step 5: Apply the Gram-Schmidt orthogonalization to obtain an orthonormal basis $\tilde{\omega}_1, ..., \tilde{\omega}_{(C\times S)-1}$ for B_{com}

Step 6: The projection matrix $W = [\tilde{\omega}_1, ..., \tilde{\omega}_{(C\times S)-1}]$ be used to obtain feature vectors in

$$\psi^{i}_{m,k} = W^T x^{i}_{m,k}, \quad m = 1, ..., N, \; i = 1, ..., C$$

We call the feature vectors $\psi^{i}_{m,k}$ segmented common vectors from k-th segmented in i-th class and they are used for classification of segmented face images.

4. Classifier of Segmented Common Vector

For classification, various classification methods can be used. However, in this special case which is real-time tracking, more fast method is needed to reduce calculation time. In this paper, Euclidean distance method is employed for fast classification.

After calculating projection matrix W, the segmented common vectors $\psi^{i}_{m,k}$ is spanned from samples $x^{i}_{m,k}$. Also mean of segmented common vector is calculated as below.

$$_{mean}\psi^{i}_{k} = \frac{1}{N}\sum_{m=1}^{N}\psi^{i}_{m,k} = \frac{1}{N}\sum_{m=1}^{N}W^T x^{i}_{m,k}, \qquad (9)$$

$$(i = 1, ..., C, \; k = 1, ..., S, \; m = 1, ..., N)$$

Each sub classes of segmented face sets have mean of segmented common vector which is a classifier of this method.

III. EXPERIMENTAL RESULTS

To classify the identification of face, each 10 samples of 7 people is learned by using OPENCV library. The samples is fragmented by 4 pieces and changed as segmented common vectors. The mean shift tracking algorithm [6] is used to track position of face, however segmented common vector is employed to calculate target candidates instead of color histograms which is usually adopted in face tracking. The experiments have been processed with a 3GHz Pentium PC and 2GB of memory under windows. Our implementation of tracker is made from visual studio 2008 environment.

The goal of experiments is to show how well our method tracks the optimal position of face. The tracked face has been changed by position and 3-D pose of face. The face sequence of 1200 frames is shown as below.

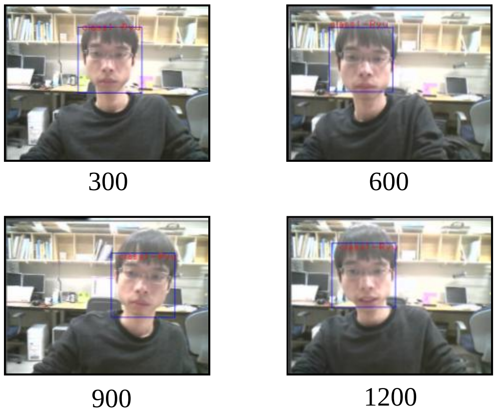

Fig.3 The experimental result

The experiment result shows that this method well follow the face of human, and face classification is done as class 1 from 7 classes. The tracker is very tough at various pose and has high precision of position face. The frame rate generally shows speed of algorithm, and, in this experimental, the frame rate of video was 14.5 per seconds. The tracker is very robust at illumination, since proposed method is not based on color feature, so it is also strong point of proposed method. The tracker can efficiently and successfully handle non-rigid and fast moving objects under different background.

IV. CONCLUSION

From experiments, the segmented common vector tracker successfully estimated position of face in complex background and can verify the identification face from several faces which was learned before.

However, the boundary of estimation should be needed for avoiding measurement error about position and size. Since proposed method is not based on color feature, the tracker is very robust at illumination and the tracker can efficiently and successfully handle non-rigid and fast moving objects under similar background. Our implementation of segmented common vector algorithm has proven to be robust to changes in shape, occlusion and color in experiment. However there is limitation of learning peoples, because $d > M-$ should be guaranteed in Null-space method.

REFERENCES

[1] PAUL VIOLA and MICHAEL J. JONES (2004) "Robust Real-Time Face Detection", International Journal of Computer Vision 57(2), 137-154

[2] M. Turk and A.P. Pentland, (1991) "Eigenfaces for Recognition", J. Cognitive Neuroscience, vol. 3, no. 1, pp. 71-86

[3] P.N Belhumeur, J.P. Hespanha, and D.J. Kriegman, (1997), "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection",IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 711-720

[4] L.-F. Chen, H.-Y.M. Liao, M.-T. Ko, J.-C. Lin, and G.-J. Yu,(2000) "A New LDA-Based Face Recognition System Which Can Solve the SmallSample Size Problem," Pattern Recognition, vol. 33, pp. 1713-1726

[5] Hakan Cevikalp, Marian Neamtu, Mitch Wilkes and Atalay Barkana (2005) " Discriminative Common Vectors for Face Recognition". IEEE TPAMI vol.27

[6] Dorin Comaniciu Visvanathan Ramesh and peter Meer (2000)"Real-Time Tracking of Non-Rigid objects using Mean shift", IEEE CVPR