Pattern Recognition Letters 30 (2009) 1096–1109

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

A bi-modal face recognition framework integrating facial expression with facial appearance

Pohsiang Tsai *, Longbing Cao, Tom Hintz, Tony Jan

Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia

ARTICLE INFO

Article history:
Available online 19 May 2009

Keywords:
Biometrics
Multimodal biometrics
Fusion techniques
Facial appearance
Facial expression
Human Computer Interface (HCI)

ABSTRACT

Among many biometric characteristics, the facial biometric is considered to be the least intrusive technology that can be deployed in the real-world visual surveillance environment. However, in facial biometric, little research attention has been paid to facial expression changes. In fact, facial expression changes have often been treated as noise that would degrade the recognition performance. This paper studies an innovative viewpoint: (1) whether facial expression changes, namely facial behavior, can be positively used for face recognition or not? (2) furthermore, can facial behavior be integrated with facial appearance for assisting the extra-personal separation to enhance face recognition performance? We propose a bi-modal face recognition framework which integrates facial expression with facial appearance. Substantial experiments on multiple facial appearance and facial expression data have been conducted. Our experimental results have validated that facial behavior can play a positive role in face recognition and can assist facial appearance in extra-personal separation in multiple modalities for personal identification improvement.

1. Introduction

Facial recognition although has gained much attention and has been reported successfully under certain conditions, it still suffers from many common problems of intra-personal variations under illumination, poses, and/or facial expressions which degrade the recognition performances.

Most algorithms (Belhumeur et al., 1997; Chen et al., 2000; Lu et al., 2003, 2005) in facial recognition are developed to cope with the singularity problem in the presence of these variations. Some papers (Martinez, 2002; Liu et al., 2003; Tsai and Jan, 2005; Bronstein et al., 2007) even consider facial expressions as noise that will degrade the system performance, and they attempt to build robust systems that are invariant to them. Combining multiple modalities[1] is an alternative way to cope with the limitations of such uni-modal biometric systems (Jain and Ross, 2004). A single modality can integrate with another or (many) biometrics to provide a more robust identification system. A special issue in 2007 presenting recent advances in biometric systems (Boyer et al., 2007) also addresses the advances in multimodal biometrics. Example literature that uses other biometrics in combination with facial biometric at a distance includes face + ear biometrics (Chang et al., 2003), 2D + 3D facial images (Chang et al., 2005), face + gait biometrics from multiple views (Shakhnarovich et al., 2001; Zhou and Bhanu, 2007), and visual + infra-red fusion (Toh et al., 2008).

Recognition of facial expression is an emerging trend in HCI affective computing for emotional or motivational analysis (Pantic and Rothkrantz, 2000a,b; Pantic et al., 2007; Pantic and Bartlett, 2007). Tsai et al. (2007) explore the empirical study for understanding facial expression, and Tsai et al. (2008) further analyze its potentiality for using it as another behavioral biometric. Their main contribution is to analyze human visual factors of facial behavior as these factors have potential applicability for biometric verification and identification in HCI.

Can intra-personal facial expression variations assist extra-personal separation? We assume that “intra-personal facial expression variations could assist extra-personal separation.” In this paper, we aim to explore the possibility of integrating facial appearance and facial expression characteristics based on the knowledge from psychophysiology (Knappmeyer et al., 2003) and computer science (Jain and Ross, 2004; Jain et al., 2006) for multimodal integration for HCI. Although the integration of different human biometric characteristics is not a new idea in facial recognition, combining human facial expression features is a new idea that is not yet exploited. The aim of the present paper is to propose a bi-modal fusion framework that (1) integrates the facial appearance and facial expression features, (2) validates our assumption that the facial behavior can not only be used for personal identification, but also can assist the extra-personal

* Corresponding author. Tel.: +61 2 9514 2353; fax: +61 2 9514 1807.
E-mail addresses: pohsiang@it.uts.edu.au (P. Tsai), lbcao@it.uts.edu.au (L. Cao), hintz@it.uts.edu.au (T. Hintz), jant@it.uts.edu.au (T. Jan).

[1] Biometric system in which two or more of the modality components occur in multiple (NIST, 2006).

0167-8655/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2009.05.008

separation, and (3) explores whether a combination of facial appearance and facial expression features may provide better performance than either one individually. *Note we do not attempt to propose a new algorithm to integrate multiple features.*

This new innovative idea of multimodal integrations can provide natural visual communication channel for humans and machines, and also has its potentiality of combining human visual cue from the low-cost camera technology in vision-based HCI at a distance.

This paper is organized as follows: Section 2 describes our proposed framework. Section 3 gives detailed analyses of different subspace methods and shows the experimental results. Section 4 gives some discussion. Finally, Section 5 presents our conclusions.

2. Face recognition using facial appearance and facial expression

We propose a bi-modal PCA fusion-level framework that integrates the facial appearance[2] and facial expression[3] features. PCA, which is first used to reduce the dimensionality of the original facial appearance and facial expression spaces, derives low dimensional features, which are then combined using a normalization procedure in order to form integrated features accounting for both facial appearance and facial expression information. Finally, the integrated features are processed by the FLD so as to enhance discriminant capability of the integrated feature space for face recognition (Duda et al., 2000).

2.1. Principal component analysis

PCA is a standard decorrelation technique that derives an orthogonal projection basis for dimensionality reduction and feature selection. Let a $X \in \mathbb{R}^{d \times n}$ be a random vector representing a facial appearance image or a facial expression dataset, where d and n are the dimensionality and numbers of either facial appearance or facial expression. The covariance matrix is then defined as follows (Bishop, 1995):

$$\sum_X = E\left\{[X - E(X)][X - E(X)]^T\right\}, \tag{1}$$

where $E(\cdot)$ is the expectation operator. This covariance matrix can be factorized in the form of $\sum_X = \Phi \Lambda \Phi^T$, where $\Phi = (\phi_1, \phi_2, \ldots, \phi_r) \in \mathbb{R}^{d \times d} (r \leqslant d)$ being the r orthonormal eigenvectors corresponding to the r largest diagonal eigen-values $\Lambda = diag\{\lambda_1, \lambda_2, \ldots, \lambda_r\} \in \mathbb{R}^{d \times d} (r \leqslant d)$. Thus, for any set of an original random vector $X \in \mathbb{R}^{d \times n}$, the reduced space of dimensionality $Y \in \mathbb{R}^{r \times n}$ can be obtained by projecting X onto the reduced eigen-feature space Φ as follows:

$$Y = \Phi^T X. \tag{2}$$

Hence, this reduced lower dimensional vector Y captures the most expressive features of the original data X.

2.2. Fisher linear discriminant analysis

While PCA is an unsupervised method that constructs the feature space without using categorical information, FLDA aims to find an "optimal" way to represent the feature vector space to maximize the discrimination between different class categories. Exploiting the class information can be helpful to the identification tasks. Let a $X \in \mathbb{R}^{d \times n}$ be a random vector representing a facial appearance image or a facial expression dataset, where d and n are the dimensionality and numbers of either facial appearance or facial expression. Let $\{\omega_i\}_{i=1}^c, \{n^i\}_{i=1}^c, \{\mu_i\}_{i=1}^c$, and $\boldsymbol{M}$ denote the classes, the number of images within each class, the means of the classes, and the grand mean. The within-class scatter and the between-class scatter covariance matrices are defined as follows (Bishop, 1995):

$$\sum_W = \sum_{i=1}^{c} \sum_{y \in \omega_i} E\left\{[y - \mu_i][y - \mu_i]^T\right\}, \tag{3}$$

$$\sum_B = \sum_{i=1}^{c} n^i E\left\{[\mu_i - \boldsymbol{M}][\mu_i - \boldsymbol{M}]^T\right\}. \tag{4}$$

Therefore, if $\sum_W$ is non-singular, the optimal projection Ψ_{opt} is chosen as the matrix with orthonormal columns which maximizes the ratio of the determinant of the between-class scatter matrix of the projected input samples to the determinant of the within-class scatter matrix of the projected input samples. The optimal projection Ψ_{opt} is defined as follows (Kirby and Sirovich, 1990; Duda et al., 2000):

$$\Psi_{opt} = \arg\max_{\Psi} \frac{|\Psi^T \sum_B \Psi|}{|\Psi^T \sum_W \Psi|} = \{\psi_i\}_{i=1}^{c-1}, \tag{5}$$

where ψ_i is the set of generalized eigen-vectors of $\sum_W^{-1} \sum_B$ (6) corresponding to the $(c-1)$ largest generalized eigen-values $\{\lambda_i\}_{i=1}^{c-1}$.

$$\sum_W^{-1} \sum_B \Psi = \Psi \Lambda. \tag{6}$$

Thus, for any set of an original random vector $X \in \mathbb{R}^{d \times n}$, the reduced space of dimensionality $Q \in \mathbb{R}^{(c-1) \times n}$ can be obtained by projecting X onto the reduced eigen-feature space Ψ_{opt} as follows:

$$Q = \Psi_{opt}^T X. \tag{7}$$

Hence, this reduced lower dimensional vector Q captures the most discriminant features of the original data X.

2.3. Fusion of facial appearance and facial expression features

The facial appearance features provide the textural information of a face, while the facial expression changes are geometrically encoded by the distance-based facial fiducial points that capture the changes of a face when it displays facial expressions or emotions. Let $\{X_i\}_{i=1}^m \in \mathbb{R}^{d_i \times n_i}$ represent the m modalities[4] of biometric characteristics, where d_i is the number of each modality's dimensionality, and n_i is the number of each modality, respectively. Using (1) one can derive the covariance matrices, $\left\{\sum_{X_i}\right\}_{i=1}^m \in \mathbb{R}^{d_i \times d_i}$, and the eigenvector matrices, $\{\Phi_i\}_{i=1}^m \in \mathbb{R}^{d_i \times d_i}$, of the m modalities, respectively. Finally, one should choose only a subset of principal components using (2) to derive the reduced lower dimensional subspaces $\{Y_i\}_{i=1}^m \in \mathbb{R}^{r_i}$, of the m modalities, respectively, to improve the generalization performance of the classifier.[5]

The reduced lower dimensional subspaces of $\{Y_i\}_{i=1}^m$ are then integrated using the normalization procedure to form the encoded information of the m modalities (Wechsler, 2007).[6] Hence, the new fused feature matrix is defined as follows:

$$U = cat\left(\left\{\frac{Y_i^T}{\|Y_i\|}\right\}_{i=1}^m\right)^T \in \mathbb{R}^{\sum_{i=1}^m r_i}. \tag{8}$$

[2] The facial appearance provides textural information of a face.

[3] Facial expression changes are geometrically encoded by distance-based facial fiducial points that capture the changes of a face when it displays facial expressions or emotions.

[4] $m = 2$ for bi-modalities.

[5] Note that r_i is chosen arbitrarily.

[6] Note that each reduced subspace is normalized to have unit norms before they are concatenated to form an augmented combined feature vectors.