Digital Signal Processing 22 (2012) 987–993

Contents lists available at SciVerse ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

A cascade face recognition system using hybrid feature extraction

Ping Zhang *, Xi Guo

Department of Mathematics and Computer Science, Alcorn State University, MS 39096-7500, USA

ARTICLE INFO

Article history:
Available online 11 July 2012

Keywords:
2D Complex Wavelet feature extraction
Ensemble classifier system
Face detection and recognition
Video image processing
Pattern recognition

ABSTRACT

A novel cascade face recognition system using hybrid feature extraction is proposed. Three sets of face features are extracted. The merits of Two-Dimensional Complex Wavelet Transform (2D-CWT) are analyzed. For face recognition feature extraction, it has proved that 2D-CWT compares favorably with the traditionally used 2D Gabor transform in terms of the computational complexity and features' stability. The proposed recognition system congregates three Artificial Neural Network classifiers (ANNs) and a gating network trained by the three feature sets. A computationally efficient fitness function of the genetic algorithms is proposed to evolve the best weights of the ensemble classifier. Experiments demonstrated that the overall recognition rate and reliability have been significantly improved in both still face recognition and video-based face recognition.

1. Introduction

Reliable and fast face recognition in the both still image-based and video-based applications has been extensively researched for years due to its enormously commercial and law enforcement applications. Some well-known methods have been reported in the literature. They include eigenfaces and Fisherfaces methods [1,2], Elastic Graph Matching (EGM) [3], robust Handsdorff distance measure for face localization and recognition [4,22], a combination strategy of neural networks for face recognition [23].

Many video-based face detection and recognition systems are conducted on the still-to-video scenario, which means that a subject's face images can be taken from still images in order to train classifiers beforehand (off-line). During the face recognition stage, the subject's face is detected and recognized online. How to quickly and accurately detect a non-titled frontal face and then recognize it from a video clip is a challenging research topic.

An important prerequisite for video-based face recognition is face detection. In order to capture frontal facial images timely and accurately, many face detection methods have been proposed. For example, Shih and Liu [5] proposed a novel face detection method by applying discriminating feature analysis (DFA) combining with support vector machine (SVM). Rowley, Baluja and Kanade [6] presented a neural network-based upright frontal face detection system, in which a retinally connected neural network examined small windows of an image and decided whether each window contained a face. A fast face detection method was proposed by Viola and Jones [7]. Face color information is an important feature in the face detection. Wu, Chen and Yachida [8] described a new method to detect faces in color images based on the fuzzy theory. In Ref. [9], authors used the quantized skin color regions for face detection. A survey of skin-color modeling and detection methods can be found in Ref. [10]. In addition, eye is another important feature for face detection and recognition. For instance, a robust method for eye features extraction on color image was reported in Ref. [11]. Using optimal Wavelet packets and radial basis functions for eye detection was introduced in Ref. [12]. Some face detection applications in the video clips can be found in Refs. [13,14]. A new and comprehensive survey on eye detection in the facial images was address in Ref. [24].

In the recent years, research on video-based face recognition remains a great interest for scientists and researchers worldwide due to the production of cheap and high resolution cameras and the simplification of the interface between computer and cameras. For instance, the face detection and tracking in a video using propagating detection probability method was proposed in Ref. [15]. Lee and Tsao [16] proposed the face recognizability measurement for visual surveillance. The video-based face recognition using adaptive hidden Markov modules was presented in Ref. [17], and an adaptive fusion of multiple matchers for face recognition was introduced in Ref. [18]. A framework for evaluating object detection and tracking in video: specifically for face, text, and vehicle objects was elaborated in Ref. [25]. Face recognition using dual-tree complex wavelet features was introduced in [26].

Ensemble classifiers have been used in many pattern recognition systems in order to increase recognition accuracy and reliability and at the same time to suppress error rate and rejection rate in the recognition system. In Ref. [27], authors proposed a cascade of boosted ensembles for face detection. An ensemble-based discriminant learning scheme with boosting for face recognition was presented in [28].

* Corresponding author.
E-mail addresses: pzhang@alcorn.edu (P. Zhang), xguo@mail.alcorn.edu (X. Guo).

1051-2004/$ – see front matter
http://dx.doi.org/10.1016/j.dsp.2012.07.003

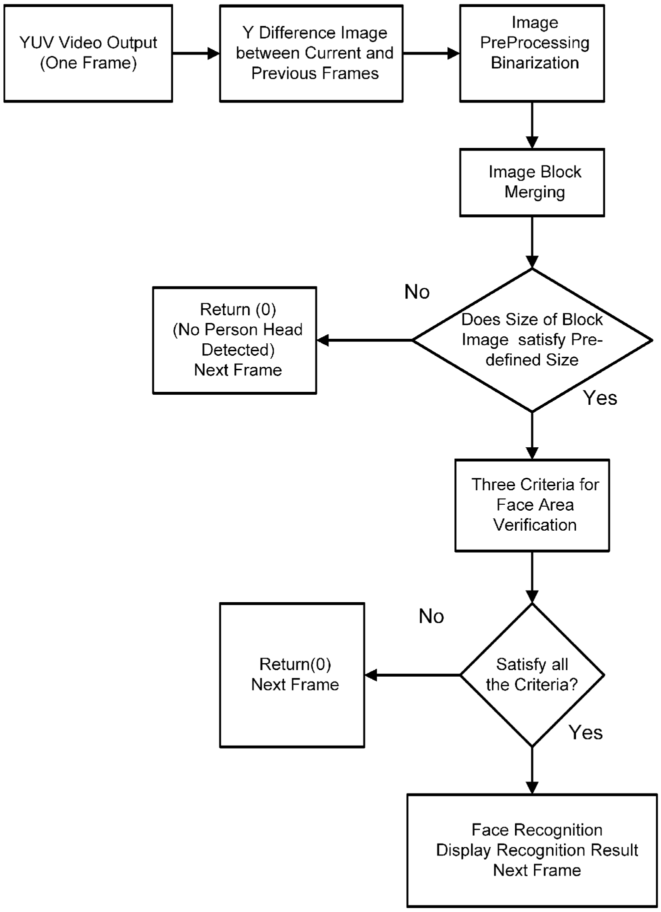

Fig. 1. Face detection and recognition schematic flowchart.

In this paper, a new cascade face recognition scheme is proposed. The paper is divided into four parts. In the first part, the flowchart of video-based face detection and recognition is drawn. In the second part, three sets of facial features are extracted. The merits of Two-Dimensional Complex Wavelet Transform (2D-CWT) is analyzed and the detail scheme of extracting facial feature using 2D-CWT is proposed. In the third part of this paper, an ensemble classifier scheme is used to congregate three individual Artificial Neural Network (ANN) classifiers trained by the three feature sets. A computationally efficient fitness function of genetic algorithms is presented and successfully used to evolve the best weights for the proposed ensemble classifier. In the last part, experiments will demonstrate that the proposed face recognition system has a significant improvement in terms of recognition rate and reliability.

2. Face detection and verification modules

In order to extract non-titled frontal facial images from a video clip, three face verification modules are serially processed. The flowchart of the proposed video-based cascade face detection and ensemble classifier system for face recognition is shown in Fig. 1.

Firstly, the spectra of face and non-face areas are analyzed and used for face skin verification. Secondly, a fast face symmetry verification algorithm is developed. Finally, three eye templates are chosen to further verify the non-titled frontal face. By doing so, there are two advantages: (1) computer recognizes only non-titled facial images in the video; (2) face recognition performance is increased accordingly. The detailed algorithms can be found in Ref. [35].

3. Feature extraction for face recognition

Feature extraction is one of the most important steps in designing a pattern recognition system. It requires that the extracted features have a small variation within a class and a strong discriminating ability among classes. In this section, three sets of features for face recognition are extracted as follows:

3.1. Complex Wavelet Transform for feature extraction

Complex Wavelet Transform (CWT) has been developed in order to keep Discrete Wavelet Transform (DWT)'s attractive attributions, such as approximate half-sample delay property, PR (orthogonal or biorthogonal), finite support (FIR filters), vanishing moments/good stopband, linear-phase filters, etc. [30,34]. Furthermore, CWT adds some new merits [31,33]: approximate shift invariance, good directional selectivity for 2D image, efficient order-N computation and limited redundancy. The computational complexity of CWT requires only twice that of DWT for 1D ($2m$ times for mD signal). These good properties have made CWT successfully applicable to image processing recently. The dual-tree CWT has shown a suitable solution to numerous applications, including pattern feature extraction and recognition [32].

2D Complex Wavelet Transform (2D-CWT) provides true directional selectivity and pixel-shifting insensitivity.

The six subband images of 2D-CWT can be represented by the following wavelet core functions:

$$\begin{aligned}
\psi_1(x,y) &= \psi(x)\psi(y)\\
\psi_2(x,y) &= \psi(x)\overline{\psi(y)}\\
\psi_3(x,y) &= \phi(x)\psi(y)\\
\psi_4(x,y) &= \psi(x)\phi(y)\\
\psi_5(x,y) &= \phi(x)\overline{\psi(y)}\\
\psi_6(x,y) &= \psi(x)\overline{\phi(y)}
\end{aligned} \tag{1}$$

where $\phi(x) = \phi_h(x) + j\phi_g(x)$ and $\psi(x) = \psi_h + j\psi_g(x)$. Both are complex functions. $\psi_i(x, y)$ $(i = 1, \ldots, 6)$ are six subbands of complex coefficients at ith level, which are oriented at angles of $\pm 75°$, $\pm 45°$, and $\pm 15°$.

2D-CWT can be implemented using a dual-tree structure. For each tree, its structure is similar to 2D-DWT, which has two decomposition operations on each level, namely row decomposition and column decomposition, except that the different filters are applied for perfect reconstruction and the output of subbands images is congregated into complex wavelet coefficients. Fig. 2 shows a 2D-CWT feature extraction scheme.

The dual-tree complex wavelet decomposition consists of two trees: Tree A and Tree B. The two trees have the same structure. In order to realize perfect reconstruction from decomposed subimages, a lowpass filter and a highpass filter at the first level need to be specially designed and denoted as h00, g00 for Tree A; h10, g10 for Tree B, which are called pre-filters. The other complex filters in the higher levels are set to h01 and g01 for Tree A, h11 and g11 for Tree B. For example, a facial image with size $N \times N$ is decomposed into four subband images: *LL*, *LH*, *HL*, *HH* at the first level of each tree and each of the subband images has a size of $N/2 \times N/2$. At the higher levels, the decompositions are based on *LL* subband image at the previous level. The complex wavelet coefficients will be used for feature extraction.