Digital Signal Processing 31 (2014) 13–27

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Near infrared face recognition by combining Zernike moments and undecimated discrete wavelet transform

Sajad Farokhi [a,*], Siti Mariyam Shamsuddin [a], U.U. Sheikh [b], Jan Flusser [c], Mohammad Khansari [d], Kourosh Jafari-Khouzani [e]

[a] UTM Big Data Centre, Faculty of Computing, Universiti Teknologi Malaysia, 81310, Johor, Malaysia
[b] Digital Signal and Image Processing Research Group, Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310, Johor, Malaysia
[c] Institute of Information Theory and Automation of the Academy of Sciences of the Czech Republic, 182 08, Prague, Czech Republic
[d] Faculty of New Sciences and Technologies, University of Tehran, 14399-55941, Tehran, Iran
[e] Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Harvard Medical School, 02114, Boston, MA, USA

ARTICLE INFO

Article history:
Available online 29 April 2014

Keywords:
Zernike moments
Undecimated discrete wavelet transform
Decision fusion
Near infrared
Face recognition

ABSTRACT

This study proposes a novel near infrared face recognition algorithm based on a combination of both local and global features. In this method local features are extracted from partitioned images by means of undecimated discrete wavelet transform (UDWT) and global features are extracted from the whole face image by means of Zernike moments (ZMs). Spectral regression discriminant analysis (SRDA) is then used to reduce the dimension of features. In order to make full use of global and local features and further improve the performance, a decision fusion technique is employed by using weighted sum rule. Experiments conducted on CASIA NIR database and PolyU-NIRFD database indicate that the proposed method has superior overall performance compared to some other methods in the presence of facial expressions, eyeglasses, head rotation, image noise and misalignments. Moreover its computational time is acceptable for on-line face recognition systems.

1. Introduction

Face recognition (FR) has received much attention over the past decades due to its potential value for many applications as well as wide challenges such as illumination variations, facial expression, head rotation, eyeglasses and misalignments which result in a significant decrease of the accuracy of the best known techniques [1,2]. On the other hand, collecting sufficient prototype images which can cover all challenges is practically impossible. Hence proposing an accurate face recognition system which is robust to most of the variations is still a challenging problem in the field of face recognition. Many face recognition techniques have been developed over the past few decades, some of which can be found in [3–5]. Most of them have focused on visible face recognition due to the fact that face recognition is one of the primary activities of the human visual system. The main limitation of visible face recognition, however, is the high dependency of system performance on external light, angle of light and even skin color [6,7]. Various methods have been developed and introduced to solve the illumination problem by proposing illumination invariant face recognition [8–10].

Recently, researchers have investigated the use of near infrared (NIR) imagery for face recognition with satisfactory results [11–13]. Three advantages of near infrared images in comparison with visible imagery can be expressed as follows. First, near infrared images are scarcely influenced by natural light. Hence it is possible to take images under very dark illumination whereas visible cameras have deficiency in this case [12,14]. Second, the eye is not sensitive to near infrared illumination and thus can be used in a more flexible and possibly covert manner [9]. Third, face detection, based on the position of the eyes, can be made more accurate than visible images due to "bright pupil" effect which simplifies eye localization and face detection consequently [9,15]. As a result, an automatic and accurate face recognition system based on near infrared spectrum can be implemented more easily than visible imagery.

Different face representations have been proposed in near infrared domain which can be roughly classified into two main categories: global-based and local-based. In the global-based face representation, each dimension of the feature vector corresponds

* Corresponding author.
E-mail addresses: fsajad2@live.utm.my (S. Farokhi), mariyam@live.utm.my (S.M. Shamsuddin), usman@fke.utm.my (U.U. Sheikh), flusser@utia.cas.cz (J. Flusser), m.khansari@ut.ac.ir (M. Khansari), kjafari@nmr.mgh.harvard.edu (K. Jafari-Khouzani).

http://dx.doi.org/10.1016/j.dsp.2014.04.008
1051-2004/

to some holistic characteristics in the face and hence encodes the global information embodied in every part of facial images. In contrast, in the local-based face representations each dimension of feature vector contains the detailed information corresponding to a certain local region of facial images [16].

An advanced design of the active NIR camera and progressive local-based method to propose illumination invariant face recognition was introduced by Li et al. [17]. In this method, first, face images are produced by active near infrared (NIR) imaging system regardless of visible light in the environment. Second, local binary pattern (LBP) features are used to compensate monotonic transform and propose illumination invariant face recognition. Finally, statistical learning algorithms are used to extract most discriminative features. The main weakness of this study, however, is the high sensitivity of LBP features to noise, head rotation and misalignments [18,19].

He et al. offered a global feature-based method based on discrete wavelet transform (DWT) and two dimensional principal component analysis (2DPCA) [20]. Coefficients of low frequency component which contribute to global information of images are used as features for the proposed method. Despite its advantages, there are several weaknesses in this approach as follows. First, the number of images in training is high and the accuracy of a system with a small number of images in the training set is not examined. Second, discrete wavelet transform is used in this paper which is not translation invariant. In other words, shift of the image by an odd number of pixels may change the whole coefficients of wavelet transform [21,22]. Hence, its accuracy will be highly affected when misalignments occur [23].

Zhang et al. introduced a novel local-based face recognition method, namely directional binary code (DBC) to capture the directional edge information of NIR facial images [24]. They showed that DBC achieves better performance than LBP. But they did not consider the effect of noise and misalignments which are the grand challenges in face recognition systems. Some other prior works can be found in [25–28].

The researches to date have tended to develop a face recognition system based on local features which are believed very robust to eyeglasses, pose variations and facial expressions. However, many studies have shown that both local and global features are crucial for highly accurate face recognition systems [16,23,29,30]. The underlying reason is that local and global features play different roles in face recognition scenarios. While local features contain more detailed local information of images and correspond to finer representation, global features are proper for coarse representation because they reflect the information of the whole face.

Inspired by the works presented in [23] which is based on combination of local and global features, and trying to avoid the translation sensitivity of DWT, in this paper, a novel face recognition method based on the integration of undecimated discrete wavelet transform (UDWT), Zernike moments (ZMs) and spectral regression discriminant analysis (SRDA) is proposed. We expect better performance by combining local and global features.

Though the basic idea of the proposed method is somewhat similar to other methods which are based on a combination of local and global features, this study has made the following transcendent contributions:

- Unlike previous studies which are based on a combination of Fourier, wavelet or Gabor wavelet transform, this paper exploits undecimated discrete wavelet transform (UDWT) as local features and Zernike moments as global features. Our experimental results confirm the effectiveness of the proposed method in the near infrared domain compared to similar transformations.
- Unlike classic methods which use discrete wavelet transform (DWT), we use UDWT to enhance the performance of system and to highlight the effectiveness of UDWT. We show experimentally that the generated features by UDWT are more robust to facial expressions, noise and misalignments in comparison with generated features by DWT.
- Unlike the method in [23] which employs fixed weights for feature vectors, in this paper a new method for weighting process is proposed.
- Unlike the method in [23] which uses Principal Component Analysis (PCA) coupled with Linear Discriminant Analysis (LDA) as dimension reduction, in this paper a Spectral regression discriminant analysis (SRDA) is used for this purpose.
- Unlike the previous methods in NIR domain which examined proposed method's performance in the context of some of challenges [31], in this study comprehensive experiments and analysis encompassing all of challenges are done.
- The proposed method has high accuracy in the presence of eyeglasses whereas our previous work [28] has deficiency in this case.
- The proposed face recognition system is extensively evaluated on CASIA NIR database and PolyU-NIRFD database and the results show that our proposed method shows good performance for most common challenges such as facial expression, eyeglasses, head rotation, noise and misalignments. Therefore, it is an accurate method which can be the core for real-world face recognition systems.

The remainder of the paper is organized as follows. In Sections 2 and 3 a brief review of undecimated discrete wavelet transform and Zernike moments are provided. Spectral regression discriminate analysis is discussed in Section 4. The proposed algorithm is provided in Section 5. Experimental results are presented in Section 6. Finally, Section 7 concludes this paper.

2. Undecimated discrete wavelet transform

In the last two decades, wavelets have become very popular due to their flexibility, locality and their high ability to analyze image at different resolutions or scales [22,32]. They have been successfully used in many fields, such as, image processing, signal analysis and pattern recognition. The two dimensional wavelet transforms can be carried out as a set of filter banks including a low-pass and high-pass filter, each followed by downsampling by a factor of two ($\downarrow 2$). One of the major problems with discrete wavelet transform (DWT) is that it suffers from translation-sensitivity. In other words, a simple shift of the input signal may change all coefficients of wavelet transform [22,32]. This deficiency is visualized in Fig. 1.

UDWT arises as a good solution to deal with the DWT's translation sensitivity. It can be considered to be an approximation to the continuous wavelet transform which removes a downsampling process from the DWT to produce over complete representation. From the implementation point of view in the context of filter banks, the filtering process is done without any downsampling (decimation), so all bands keep the same size as the original image. The implementation process of UDWT is shown in Fig. 2. The shift-invariance is achieved by two main parts. In the first part, filter coefficients of selected wavelet are upsampled by a factor of $2^{(j-1)}$ in the jth level. In the second part, the coefficients of the approximation are convolved with an upsampled version of the original filters to generate coefficients of approximation and details in the next levels. The result is shift-invariant wavelet transform. It has been already used for face virtual pose generation, translation-invariant feature extraction and object tracking [33–35].