Expert Systems With Applications 47 (2016) 23–34

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Fully automatic face normalization and single sample face recognition in unconstrained environments

Mohammad Haghighat[a,*], Mohamed Abdel-Mottaleb[a,b], Wadee Alhalabi[b,c]

[a] *Department of Electrical and Computer Engineering, University of Miami, Coral Gables, FL 33146 USA*
[b] *Department of Computer Science, Effat University, Jeddah, Saudi Arabia*
[c] *Department of Computer Science, King Abdulaziz University, Jeddah, Saudi Arabia*

ARTICLE INFO

Keywords:
Face recognition in-the-wild
Pose-invariance
Frontal face synthesizing
Feature-level fusion
Canonical correlation analysis
Active appearance models

ABSTRACT

Single sample face recognition have become an important problem because of the limitations on the availability of gallery images. In many real-world applications such as passport or driver license identification, there is only a single facial image per subject available. The variations between the single gallery face image and the probe face images, captured in unconstrained environments, make the single sample face recognition even more difficult. In this paper, we present a fully automatic face recognition system robust to most common face variations in unconstrained environments. Our proposed system is capable of recognizing faces from non-frontal views and under different illumination conditions using only a single gallery sample for each subject. It normalizes the face images for both in-plane and out-of-plane pose variations using an enhanced technique based on active appearance models (AAMs). We improve the performance of AAM fitting, not only by training it with in-the-wild images and using a powerful optimization technique, but also by initializing the AAM with estimates of the locations of the facial landmarks obtained by a method based on flexible mixture of parts. The proposed initialization technique results in significant improvement of AAM fitting to non-frontal poses and makes the normalization process robust, fast and reliable. Owing to the proper alignment of the face images, made possible by this approach, we can use local feature descriptors, such as Histograms of Oriented Gradients (HOG), for matching. The use of HOG features makes the system robust against illumination variations. In order to improve the discriminating information content of the feature vectors, we also extract Gabor features from the normalized face images and fuse them with HOG features using Canonical Correlation Analysis (CCA). Experimental results performed on various databases outperform the state-of-the-art methods and show the effectiveness of our proposed method in normalization and recognition of face images obtained in unconstrained environments.

1. Introduction

Although face recognition has been a challenging topic in computer vision for the past few decades, most of the attention was focused on recognition based on face images captured in controlled environments. Capturing a face image naturally without controlling the environment, so-called in the wild (Huang, Ramesh, Berg, & Learned-Miller, 2007; Le, 2013), may result in images with different illumination, head pose, facial expressions, and occlusions. The accuracy of most of the current face recognition systems drops significantly in the presence of these variations, specially in the case of pose and illumination variations (Moses, Adini, & Ullman, 1994; Zhao, Chellappa, Phillips, & Rosenfeld, 2003).

Regardless of the face variations in pose, illumination and facial expressions, we humans have an ability to recognize faces and identify persons at a glance. This natural ability does not exist in machines; therefore, we design intelligent and expert systems that can simulate the recognition artificially (Haghighat, Zonouz, & Abdel-Mottaleb, 2015). Building deterministic or stochastic face models is a challenging task due to the face variations. However, normalization can be used in a preprocessing step to reduce the effect of these variations and pave the way for building face models. Pose variations are considered to be one of the most challenging issues in face recognition. Due to the complex non-planar geometry of the face, the 2D visual appearance significantly changes with variations in the viewing angle. These changes are often more significant than the variations of innate characteristics, which distinguish individuals (Zhang & Gao, 2009). In this paper, we propose a fully automatic single sample face

* Corresponding author. Tel.: +1 305 284 3291; fax: +1 305 284 4044.
E-mail addresses: haghighat@umiami.edu, haghighat@ieee.org (M. Haghighat), mottaleb@miami.edu (M. Abdel-Mottaleb), walhalabi@effatuniversity.edu.sa (W. Alhalabi).

http://dx.doi.org/10.1016/j.eswa.2015.10.047
0957-4174/

recognition method that is capable of handling pose variations in unconstrained environments. In the following two sections, we present a literature review of related methods and our contributions in this paper.

1.1. Related work

The Active Appearance Models (AAMs) proposed by (Cootes, Edwards, & Taylor, 1998; 2001) have been used in face modeling for recognition. After fitting the model to a face image, either the model parameters, the location of the landmarks, or the local features extracted at the landmarks are used for face recognition (Edwards, Cootes, & Taylor, 1998; Ghiass, Arandjelovic, Bendada, & Maldague, 2013; Hasan, Abdullaha, & Othman, 2013; Lanitis, Taylor, & Cootes, 1995) or facial expression analysis (Lucey et al., 2010; Martin, Werner, & Gross, 2008; Tang & Deng, 2007; Trutoiu, Hodgins, & Cohn, 2013; Van Kuilenburg, Wiering, & Den Uyl, 2005). For face recognition, (Guillemaut, Kittler, Sadeghi, & Christmas, 2006) and (Heo & Savvides, 2008) proposed using the normalized face images created by warping the face images into the frontal pose.(Gao, Ekenel, & Stiefelhagen, 2009) improved the performance of this technique using a modified piecewise affine warping. None of these methods, however, is fully automatic and they require a manual labeling or manual initialization.

(Chai, Shan, Chen, & Gao, 2007) assumed that there is a linear mapping between a non-frontal face image and the corresponding frontal face image of the same subject under the same illumination. They create a virtual frontal view by first partitioning the face image into many overlapped local patches. Then, a local linear regression (LLR) technique is applied to each patch to predict its corresponding virtual frontal view patch. Finally, the virtual frontal view is generated by integrating the virtual frontal patches. (Li, Shan, Chen, & Gao, 2009) proposed a similar patch-based algorithm; however, they measured the similarities of the local patches by correlations in a subspace constructed by Canonical Correlation Analysis. (Du & Ward, 2009) proposed a similar method based on the facial components. Unlike (Chai et al., 2007) and (Li et al., 2009), where the face image is partitioned into uniform blocks, the method in (Du & Ward, 2009) divides it into the facial components, *i.e.*, two eyes, mouth and nose. The virtual frontal view of each component is estimated separately, and finally the virtual frontal image is generated by integrating the virtual frontal components. The common drawback of these three patch-based approaches, (Chai et al., 2007; Du & Ward, 2009; Li et al., 2009), is that the head pose of the input face image needs to be known. Moreover, these methods require a set of prototype non-frontal face patches, which are in the same pose as the input non-frontal faces; hence, they cannot handle a continuous range of poses and are restricted to a discrete set of predetermined pose angles.

(Blanz & Vetter, 2003) proposed a face recognition technique that can handle variations in pose and illumination. In their method, they derive a morphable face model by transforming the shape and texture of example prototypes into a vector space representation. New faces at any pose and illumination are modeled by forming linear combinations of the prototypes. The morphable model represents shapes and textures of faces as vectors in a high-dimensional space. The knowledge of face shapes and textures is learned from a set of textured 3D head scans. This method requires a set of manually annotated landmarks for initialization and the optimization process often converges to local minima due to a large number of parameters, which need to be tuned. (Breuer, Kim, Kienzle, Scholkopf, & Blanz, 2008) presented an automatic method for fitting the 3D morphable model; however, their method seems to have a high failure rate (Asthana, Marks, Jones, Tieu, & Rohith, 2011).

(Castillo & Jacobs, 2009) used the cost of stereo matching as a measure of similarity between two face images in different poses. This method does not construct a 3D face or a virtual frontal view; however, using stereo matching, it finds the correspondences between pixels in the probe and gallery images. This method requires manual specification of feature points and in case of automatic feature matching, it is fallible in scenarios where an in-plane rotation is present between the image pair.

The method proposed by (Sarfraz & Hellwich, 2010) handles the pose variations for face recognition by learning a linear mapping from the feature vector of a non-frontal face to the feature vector of the corresponding frontal face. However, their assumption of the mapping being linear seems to be overly restrictive (Asthana et al., 2011).

(Asthana et al., 2011) used several AAMs each of which covering a small range of pose variations. All these AAMs are fitted on the query face image and the best fit is selected. The frontal view is then synthesized using the pose-dependent correspondences between 2D landmark points and 3D model vertices. (Mostafa, Ali, Alajlan, & Farag, 2012; Mostafa & Farag, 2012) constructed 3D face shapes from stereo pair images. These 3D shapes are used to synthesize virtual 2D views in different poses, *e.g.*, frontal view. A 2D probe image is matched with the closest synthesized images using the local binary pattern (LBP) features (Ahonen, Hadid, & Pietikäinen, 2006). The drawback of this method is the need for stereo images. In order to solve this problem, the authors developed another method where the 3D shapes are constructed using only a frontal view and a generic 3D shape created by averaging several 3D face shapes.

(Sharma, Al Haj, Choi, Davis, & Jacobs, 2012) proposed the Discriminant Multiple Coupled Latent Subspace method for pose-invariant face recognition. They propose to obtain pose-specific representation schemes so that the projection of face vectors onto the appropriate representation scheme will lead to correspondence in the common projected space, which facilitates direct comparison. They find the sets of projection directions for different poses such that the projected images of the same subject in different poses are maximally correlated in the latent space. They claim that the discriminant analysis with artificially simulated pose errors in the latent space makes it robust to small pose errors due to subjectś incorrect pose estimation.

(De Marsico, Nappi, Riccio, & Wechsler, 2013) proposed a face recognition approach, called “FACE”, in which an unknown face is identified based on the correlation of local regions from the query face and multiple gallery instances, that are normalized with respect to pose and illumination, for each subject. For pose normalization, the facial landmarks are first located by an extension of the active shape model (Milborrow & Nicolls, 2008) and then the in-plane face rotation is normalized using the locations of the eye centers. The rows in the best exposed half of the face are then stretched to a constant length. Then, the other side of the face image is reconstructed by mirroring the first half. The illumination normalization is performed using the Self-Quotient Image (SQI) algorithm (Wang, Li, Wang, & Zhang, 2004), in which the intensity of each pixel is divided by the average intensity of its $k \times k$ square neighborhood.

(Ho & Chellappa, 2013) proposed a patch-based method for synthesizing the frontal view from a given nonfrontal face image. In this method, the face image is divided into several overlapping patches, and a set of possible warps for each patch is obtained by aligning it with frontal faces in the training set. The alignments are performed using an extension of the Lucas–Kanade image registration algorithm (Ashraf, Lucey, & Chen, 2010; Lucas & Kanade, 1981) in the Fourier domain. The best warp is chosen by formulating the optimization problem as a discrete labeling algorithm using a discrete Markov random field and a variant of the belief propagation algorithm (Komodakis & Tziritas, 2007). Each patch is then transformed to the frontal view using its best warp. Finally, all the transformed patches are combined together to create a frontal face image. A shortcoming of this method is that they divide both frontal and non-frontal images into the same regular set of local patches. This division strategy results in the loss of semantic correspondence for some patches when the pose