Pattern Recognition 41 (2008) 3822 – 3833

Contents lists available at ScienceDirect

Pattern Recognition

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

2D face recognition based on supervised subspace learning from 3D models

Annalisa Franco [a,b*], Dario Maio [a,b], Davide Maltoni [a,b]

[a] C.d.L. Scienze dell'Informazione, Università di Bologna, Via Sacchi 3, 47023 Cesena, Italy
[b] DEIS, Università di Bologna, Viale Risorgimento, 2, 40136 Bologna, Italy

ARTICLE INFO

Article history:
Received 17 September 2007
Received in revised form 6 May 2008
Accepted 28 May 2008

Keywords:
Face recognition
3D face models
Subspace learning

ABSTRACT

One of the main challenges in face recognition is represented by pose and illumination variations that drastically affect the recognition performance, as confirmed by the results of recent face recognition large-scale evaluations. This paper presents a new technique for face recognition, based on the joint use of 3D models and 2D images, specifically conceived to be robust with respect to pose and illumination changes. A 3D model of each user is exploited in the training stage (i.e. enrollment) to generate a large number of 2D images representing virtual views of the face with varying pose and illumination. Such images are then used to learn in a supervised manner a set of subspaces constituting the user's template. Recognition occurs by matching 2D images with the templates and no 3D information (neither images nor face models) is required. The experiments carried out confirm the efficacy of the proposed technique.

1. Introduction

Face recognition is one of the most widely studied biometric problems and a huge literature exists on this issue [1]. For a long time only 2D images have been available for recognition and most of the existing techniques have been proposed in that context. The results of some recent face recognition competitions, i.e. the face recognition vendor test (FVRT) 2000 [2], FVRT 2002 [3] and FRVT 2006 [4], show that satisfactory results can be achieved when images are acquired under controlled pose and illumination conditions, but unfortunately the performance noticeably decrease in uncontrolled environments, thus limiting the diffusion of face recognition in some applications. The 3D-based face recognition has been introduced in recent years with the aim of overcoming this limitation. The use of 3D models (often in conjunction with 2D texture wrapped around), gives a richer representation of the face, and can lead to an improvement in the recognition performance, as confirmed by the results of the recent FRVT 2006 [4]. On the other hand, at today, the application of full 3D-based systems does not seem a viable solution for all the applications because of the cost and the constraints that many 3D acquisition devices impose (e.g. the 3D scanner used in FVRT 2006 required the user to stand perfectly steady for about 2 s and this can cause several acquisition errors).

In this work the 3D information is exploited only for enrolling the users and the recognition is performed on the basis of simple 2D images. The enrollment stage is here more complex than in 2D systems, because the creation of a 3D model and the generation of several virtual images are required, but usually the efficiency requirement for enrollment is not so strict. During the recognition stage, when the efficiency is a critical factor, this approach is very fast being based exclusively on 2D images.

The original contribution of our approach is neither the use of a combined 3D–2D technique nor the use of a 3D model for generating training samples; in fact, as discussed in Section 2, these ideas have been already introduced by other researchers. What makes our study novel is the use of a supervised subspace learning technique where samples of the same subject having similar (pose and light) characteristics are grouped and encoded by a specific subspace: more subspaces are created for each individual by partitioning in a supervised manner his entire 4D pose-and-light space. In fact, it is well know that the variability and complexity of the face pattern under pose and light changes cannot be represented by using a linear subspace, and we believe that training a non-linear model to learn in an unsupervised manner a face manifold is very complex. As demonstrated in Section 4 our piece-wise linear model, where "pieces" are automatically determined by the sample generation procedure, allows to control the system complexity and provides very interesting results.

* Corresponding author. Tel.: +39 547 338847; fax: +39 547 338890.
E-mail addresses: franco@csr.unibo.it, annalisa.franco@unibo.it (A. Franco), dario.maio@unibo.it (D. Maio), davide.maltoni@unibo.it (D. Maltoni).

0031-3203/$ - see front matter
doi:10.1016/j.patcog.2008.05.029

The paper is organized as follows: in Section 2 some related works are described, in Section 3 the new recognition approach is presented, the experiments carried out are reported in Section 4 and finally some concluding remarks are given in Section 5.

2. Related works

The idea of generating virtual views of an individual with different poses and illumination conditions has been initially proposed in the context of 2D face recognition. In Ref. [5] the authors propose to apply to a single training image a set of transformations on the basis of some prototypes describing the typical aspect of a face under a variety of conditions. Recognition is based on the use of template matching techniques to find the best mapping between the image to be recognized and the generated virtual views. The basic idea of using training prototypes for recognition has been successively extended in Ref. [6] where the concept of "linear object classes" is proposed. Linear objects are 3D objects whose shape can be represented as a linear combination of some prototypical objects, and these classes have the property that new views of any object of the class under uniform affine 3D transformations can be generated if the corresponding transformed views are known for the set of prototypes. The approach in Ref. [7] is aimed at transforming each image to be recognized into a prototype which is frontal-view and under frontal lighting. View synthesis is performed during training by shape from shading (SFS). In the training stage an LDA subspace is computed for each individual, starting from a set of frontal images. In the recognition stage each image is first transformed into a frontal-view and frontal-lighting image and then recognition is performed on the basis of its distance from the LDA subspaces representing the enrolled users. The technique proposed in Ref. [8] requires a set of training images with varying poses and illuminations from which new unseen poses can be generated by the generalized bas-relief (GBR) transformation. For each virtual view obtained, the illumination variations are modeled by the illumination cone [9] and recognition is performed by calculating the distance of the image from the illumination cones representing the different poses of each individual.

The main limitation of the 2D-based approaches is that, typically, a large number of training images with varying conditions are needed and such images are not always available in real applications. The 3D models, instead, offer a comprehensive representation of a face and, though the construction of the model is not a trivial task, make easier the enrollment process from the user's point of view; in fact, "few" images/data are needed for the model construction.

Recently several techniques for face recognition based either on 3D models or on the combination of 2D and 3D information have been proposed; interested readers can refer to Ref. [10] for a literature survey. Here we will concentrate only on those methods that exploit the 3D model only in the training stage and that perform recognition on the basis of 2D images. The authors of Ref. [11] propose a hybrid system based on an extension of the classical PCA (principal component analysis) called partial PCA (P^2CA) where each image is represented as a 2D matrix rather than a 1D vector. The 3D information used for training the P^2CA-based classifier consists of cylindrical 180° 3D texture images. In Ref. [12] a generic 3D face model is used, and each individual is represented by a set of shape and texture parameters describing the transformations needed to adapt the model to the specific training images. In the recognition stage the model is used to generate, by modifying the above parameters, a 2D image as similar as possible to the image to be recognized. The resulting transformation parameters are finally compared to those related to the enrolled users for recognition. Finally, in Ref. [13] a user-specific 3D model is obtained by adapting a generic 3D model to a set of training images of each individual. The resulting model is exploited to generate a number of training images with different poses and illumination conditions which are then used to train a component-based recognition system, i.e. a set of SVM classifiers trained with vectors obtained by concatenating features extracted from the principal face components (eyes, mouth, etc.). The component-based system is compared to a global approach where the whole face image is used for training the classifier and the experimental results clearly show the superiority of the component-based technique. This technique is probably the most related to our proposal and we will consider it in our experiments as a term of comparison to show that even a global approach, such as the one we are proposing, can achieve good performance provided that an appropriate user's representation is derived from the training images.

3. Supervised subspace learning from 3D models

Figs. 1 and 2 show a block scheme of the enrollment and recognition[1] stages of the approach proposed in this paper.

In the rest of this section the different steps involved in the enrollment and recognition procedures are described in detail.

3.1. Enrollment

In the enrollment stage a user's specific 3D model is constructed and subsequently used to generate a large number of training images with varying poses and illuminations. The images obtained are then organized into several subsets and finally a subspace is computed from each subset. A detailed description of the different steps is provided in the following subsections.

3.1.1. Acquisition of the data for creating the user 3D model

Different techniques can be used for data acquisition and 3D face model reconstruction [14,15]: (i) acquisition of two or more 2D images from standard cameras and subsequent registration for 3D reconstruction (i.e. shape from stereo [16] or SFS [17]) and (ii) acquisition of 3D data via either structured-light scanners where a specific light pattern is projected onto the face [18] or by means of laser scanners which project a laser beam [19]. Once the 3D structure of the face has been computed, typically a texture mapping step is required to obtain a complete and more realistic model. Each technique presents pros and cons and peculiar characteristics in terms of cost, acquisition time and accuracy. The focus of this paper is not on pure 3D face recognition, but on the use of the 3D model as a generative model for robust 2D face recognition; for this reason the technique here described does not rely on a particular data acquisition and reconstruction method: any approach able to create a reasonable 3D face model might be adopted. However, for some of our experiments we used a four-camera system [20] with the cameras grouped in two pairs, each producing a depth map of the related face region. The obtained depth maps are then merged into a unique map. The use of a four-camera system allows to obtain a quite accurate reconstruction, especially in presence of occlusions, with respect to the two- or three-camera systems. Fig. 3 shows an example of a reconstructed 3D model.

3.1.2. Generation of 2D virtual images

A number of 2D virtual images are generated through a ray-tracing approach where the face pose and the direction of the light are systematically varied. Face pose is controlled by the Yaw and

[1] The term recognition is here used in a generic way to denote both a face verification (1 : 1) and a face identification (1 : N). In the experiment section distinct results will be provided for the different cases.