Abstract

Face verification in the wild, remains a challenging problem. This paper makes two contributions: first, for improving face recognition in the wild, at least in terms of pose variations, we propose a method for aligning faces by employing single-3D face model as reference produced by FaceGen Modeller. Second, we develop a novel face descriptor based on Gabor Filters. The proposed descriptor relies on combination of Gabor magnitude and Gabor phase informations into an unified framework, which is capable to overcome standard representations in the most popular benchmark “Labeled Faces in the Wild” (LFW). This compact descriptor has a better recognition performance, reaches an accuracy of 97.29% on the LFW dataset.

© 2017 The Authors. Published by Elsevier B.V.

Keywords: Face Recognition, Gabor Filter, Pose Correction, Support Vector Machine, RBF Kernel, LFW;

1. Introduction

In face verification, images are presented in pairs and the task is to verify if they belong to the same or different persons. Face verification has recently gained lot of popularity owing to few public benchmark datasets being available. Fig. 1 shows example of such a task. Applications of this task are in search and authentication domains.

Brahim AKSASSE*, Hamid OUANAN and Mohammed OUANAN

Department of Computer Sciences, ASIA Team
Moulay Ismail University,
Faculty of Science and Techniques
BP 509 Boutalamine 52000 Errachidia, Morocco

* Corresponding author. Tel.: +212672345351; fax: +212535574485.
E-mail address: baksasse@yahoo.com

10.1016/j.procs.2017.06.108
such as entertainment, human machine interaction, homeland security, and video surveillance, access control to user authentification schemes in e-commerce, e-health, and e-government services. There are many challenges in dealing with this applications listed such as variation in illumination, variability in scale, location, orientation and pose. Furthermore, facial expression, facial decorations, partial occlusion and lighting conditions change the overall appearance making it harder to recognize faces. Fig. 2 shows some examples of these types of challenges. Face recognition is really a series of several related problems: face detection, face normalization, feature extraction, and feature matching. As a human, your brain do all of this tasks automatically and instantly. Computers are at least not yet able of this sort of high-level generalization, so we have to teach them how to do each task in this process separately. To do so, we need to build a pipeline where we solve each step of face recognition separately and pass the result of the current step to the next step (see figure 3).

The main objective of this work is to propose a reliable framework insensitive to challenges listed above, in particular capable to “identify faces from a side view” as well as when the person is directly facing the camera in the
picture to approach human-level performance in this domain. In summary, the novelty of this paper comes from: (i) an effective 3D face alignment module; (ii) effective representation for describing faces using Gabor Filters. We also investigate various approaches to effectively reduce their dimension while improving their performance further; and (iii) extensive performance evaluation studies.

2. Related Works

The conventional pipeline of a typical face verification system requires these steps: face detection, facial landmark detection, alignment, representation and classification. However, several papers focus on a few of these aspects in order to improve the overall system performance. In this work, we have focused on both alignment and the representation steps. In this section, we briefly review some recent related works on face alignment and face representation in the context of face verification.

State-of-the-art face alignment: Aligning faces in under in-the-wild conditions is still a most difficult problem that has to account for many factors like non-rigid face expressions and pose. Recently, some techniques bearing capable to compensate for these difficulties, which can be roughly divided into two main categories: (i) part-based methods which represent the face by using a set of local image patches extracted around of the predefined landmark points and (ii) Holistic methods which use the whole texture of face as representation. The most-well known techniques and produced good results: In the first category methods like Active Shape Models (ASMs)\(^2\) and Constrained Local Models (CLMs)\(^2\). In the second category, methods like Active Appearance Models (AAMs)\(^3\) and 3D Deformable Models (3DMs)\(^4\). However, no complete solution is currently present in the context of face recognition in the wild because the accuracy of those detection and localization landmarks algorithms degrades as the yaw or pitch angle of the face increases.

Brief review of recent face verification approaches: Representing face images has been an important topic in computer vision and image processing. The diversity of feature extraction methods is surprising. In this section, we look at some methods which produced better performance over large scale database like LFW and FERET face databases. The authors of the paper\(^5\) proposed a facial image representation giving better results on FERET database\(^6\), this method rely on Gabor filters (GFs) and Zernike moments (ZMs), where GFs is used for texture feature extraction and ZMs extracts shape features, in other hand, a simple Genetic Algorithm (GA) is applied to select the moment features that better discriminate human faces under several pose and illumination conditions. Next, the augmented extracted feature vectors are projected onto a low-dimensional subspace using Random Projection\(^7\) (RP) method. The authors of the paper\(^8\) proposed a regularization framework to learn similarity metrics for face verification in the wild. This method achieves a good results on the (LFW) database\(^1\). In the paper\(^9\), the authors proposed a joint Bayesian approach based on the classical Bayesian face recognition approach proposed by Baback Moghaddam et al.\(^10\). This approach achieved 92.4% accuracy on the LFW dataset. Another interesting approach is Fisher vector encoding performs well on LFW. However, the accuracy of those algorithms degrades on extreme poses of face like profile. This show the need of techniques capable to compensate large pose variation.

3. The proposed framework

3.1. Pose Correction

In this section, we describe briefly our method used for 3D pose correction: In this method, we use the same textured 3D face model as reference to align all query images. This 3D face model is produced by FaceGen Modeller\(^11\). We begin by rendering this reference model in a fixed, frontal view. We refer to this as the reference frontal view \(I_R\) (see equation 1) which serves as our reference coordinate system, 68 facial landmarks \(p_i = (x_i, y_i)^T\) are detected in this image using the method\(^12\), selected for its accuracy in real world face photos. For each point detected we associate the 3D coordinates \((P_i = (X_i, Y_i, Z_i))^T\) . Given a query image \(I_Q\), it is processed by first running the Viola-Jones detector\(^13\). We again use\(^14\) to detect the same 68 landmarks in \(I_Q\), giving us points \((p_i = (x_i, y_i))^T\) . Using these, we form correspondences \((p_i, P_i)^T\) from 2D pixels in the query photo to 3D points on the model. We then compute specific \(3 \times 4\) camera matrix \(C_M\) by selecting suitable intrinsic and extrinsic camera parameters, using a standard calibration method.
\[ p^* = C_M P \]  
\[ C_M = A_M [R_M t_M] \]  
\[ C_Q = A_Q [R_Q t_Q] \]  

Where \( A_M \) : is the intrinsic matrix, \( R_M \) : is the rotation matrix, and \( t_M \) : is the translation vector.

**Input:** Query image \( I_Q \), textured 3D face model, rendered frontal view of this model \( (I_R) \).

**Output:** Frontalized Face

**Step 1:** Facial feature points \( (p_i = (x_i, y_i)^T) \) detected in the query image \( I_Q \).

**Step 2:** Same facial feature points \( (p_i' = (x_i', y_i')^T) \) will be detected in \( I_R \) and their correspondence points \( (P_i = (X_i, Y_i, Z_i)^T) \) on the surface of the model.

**Step 3:** Seek 2D-3D correspondences between points \( (p_i, P_i) \).

**Step 4:** Estimation the query 3x4 camera matrix \( C_Q \) used to capture the query image \( I_Q \).

**Step 5:** Back-projection query intensities to \( I_R \) (equation 3).

**Step 6:** Estimation of visibility due to non-frontal poses by symmetry.

**Step 7:** Face patches Extraction and classification.

**Step 8:** Final frontalized crop canonical view.

Fig. 4: An overview of the face alignment proposed method

3.2. Face Representation

The frequency and orientation representations of Gabor filters are similar to those of the human visual system and they have been found to be particularly appropriate for texture representation. Gabor filters have been widely used in pattern analysis applications. The most important advantage of Gabor filters is their invariance to illumination, rotation, scale, and translation. Furthermore, they can withstand photometric disturbances, such as illumination changes and image noise.

A 2D Gabor function \( g(x, y) \) and its Fourier transform \( G(u, v) \) are as follows:

\[ g(x, y) = \frac{1}{2\pi \sigma_x \sigma_y} \exp[-\frac{1}{2} \left( \frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) + 2\pi i \omega x] \]  
\[ G(u, v) = \exp[-\frac{1}{2} \left( \frac{(u-\omega)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2} \right)] \]  

where: \( \sigma_u = \frac{1}{2\pi \sigma_x} \) and \( \sigma_v = \frac{1}{2\pi \sigma_y} \)

\[ g_{mn}(x, y) = a^{-m}G(x', y') \]  

where \( a \geq 1; \ x' = a^{-m}(x \cos \theta + y \sin \theta) \) and \( y' = a^{-m}(y \cos \theta - x \sin \theta) \),

for \( m = 0,1,...M-1 \) and \( n = 0,1,...N-1 \), \( M \) is the number of resolutions and \( N \) is the number of orientations.
The feature extraction procedure can then be defined as a convolution operation of the face image \( I(x, y) \) with the Gabor filter \( G(u, v) \). The result of this operation is a complex image defined by the amplitude and the phase for each pixel of the image:

\[
\psi_{u,v}(x, y) = I(x, y) * G_{u,v}(x, y)
\]

Based on this equation the magnitude and the phase responses of the convolution operation can be computed as follows:

\[
A_{u,v}(x, y) = \sqrt{\text{Re}(\psi_{u,v}(x, y))^2 + \text{Im}(\psi_{u,v}(x, y))^2}
\]

\[
\varphi_{u,v}(x, y) = \arctan\left(\frac{\text{Im}(\psi_{u,v}(x, y))}{\text{Re}(\psi_{u,v}(x, y))}\right)
\]

The great number of Gabor-based face recognition approaches found in the literature rely solely on the magnitude information when constructing the Gabor face representation and discard the phase information of the convolution output image. The magnitude face representation vector is computed by taking the following steps:

**Input:** Face Image \((128 \times 128)\) pixels.

**Output:** The validity of the Identity claim

**Step 1:** Face Frontalization

**Step 2:** Gabor filter construction with bank of 40 filters.

**Step 3:** Gabor features derived from the Gabor filter magnitude response (similarly from the Gabor filter phase) are computed for all frequencies \((u = 4)\) and orientations \((v = 8)\) (GMFR).

**Step 4:** Downsampling by a factor \((\rho = 64)\) the computed GMFRs (similarly GPFRs).

**Step 5:** The downsampled GMFRs are normalized using an appropriate normalization procedure.

**Step 6:** The downsampled and normalized GMFRs (similarly GPFRs) in vector form are concatenated to form the augmented GMFR vector.

**Step 7:** Project the augmented feature vectors into a subspace by using Kernel Fisher Analysis (KFA).

Fig. 5: An overview of the Magnitude Gabor Descriptor

Despite the downsampling procedure, the size of the descriptors presented still reside in a very high-dimensional space. In this paper, the KFA technique\(^5\) is applied to the augmented Gabor magnitude face and augmented Gabor phase representation vectors to obtain a compact representation based Gabor-Magnitude and a compact representation based Gabor-Phase. Then, we use the SVM classifier\(^6\) based on RBF kernel to classify GM+KFA feature and GP+KFA feature extracted from face images. Finally, we combine both matching score at the matching score. The accuracy \(\delta\) of the method proposed is computed using the following expression:

\[
\delta = (1 - \gamma)\delta_{GM} + \gamma\delta_{GP}
\]

Where \(\delta_{GM}\) denotes the accuracy obtained from Gabor magnitude features, \(\delta_{GP}\) denotes the accuracy obtained from Gabor phase features and \(\gamma \in ]0, 1[\) denotes the fusion parameter.

### 4. Experiments and tests

The performance of the proposed approach is assessed by conducting experiments on the well-known LFW dataset which contains 13,233 images of 5,749 people downloaded from the Web. This database, cover large variations including: different subjects, poses, illumination, occlusion etc. For evaluation, we have used the standard protocol which defines 3,000 positive pairs and 3,000 negative pairs in total and further splits them into 10 disjoint subsets for cross validation. Each subset contains 300 positive and 300 negative pairs, portraying different people. We compare the mean accuracy of the proposed approach with some methods which achieve state of the art and other commercial systems. The results are summarized in Table 1.
The performance of the proposed approach is assessed by conducting experiments on the well-known LFW dataset. The results are summarized in Table 1. Comparing the mean accuracy of the proposed approach with some methods which achieve state-of-the-art and commercial systems, our approach shows that our proposed method achieves good results on the LFW dataset, which contains faces with full pose, illumination, and other difficult conditions. It is robust, especially in the presence of large head pose variations. It can be seen from the table that our approach performs well comparatively to other methods and commercial systems.

5. Conclusion

In this paper, we have presented a new face verification approach. Our new approach was evaluated on the LFW dataset. Experimental results demonstrate that the performance of the proposed approach is much better than the some methods which achieve state-of-the-art and commercial systems. The gap between our proposed approach and human performance on the LFW benchmark is less than 1%. In the future, we will apply our approach to video processing. We believe that our method will demonstrate competitive performance.

References