Pattern Recognition Letters 36 (2014) 235–242

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Face recognition on partially occluded images using compressed sensing

A. Morelli Andrés, S. Padovani, M. Tepper, J. Jacobo-Berlles *

Departamento de Computación, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina

ARTICLE INFO

Article history:
Available online 13 August 2013

Keywords:
Face recognition
Compressed sensing
Partial occlusion

ABSTRACT

In this work we have built a face recognition system using a new method based on recent advances in compressed sensing theory. The authors propose a method for recognizing faces that is robust to certain types and levels of occlusion. They also present tests that allow to assess the incidence of the proposed method.

1. Introduction

Face detection and recognition are issues that are being widely studied due to the large number of applications they have. At present, we can find face recognition systems in social networking sites, photo management software and access control systems, to name a few. Face recognition presents several difficulties. The image of the human face can have large intra-subject variations (changes in the same individual) that make it difficult to develop a recognition system. Variations may arise, for example, from head position variation when taking the picture, differences in lighting, facial expression (laughter, anger, etc.), occlusion of parts of the face due to the use of accessories such as lenses, sunglasses and scarves, facial hair (mustache, beard, long hair, etc.), and changes in facial features due to aging. On the other hand, inter-subject variations (differences between individuals) can be very small between two people with similar traits, making correct identification difficult. Presently there are various methods of face recognition. Among the most popular are Eigenfaces (Turk and Pentland, 1991) and Active Appearance Model (Cootes et al., 2001; Kahraman et al., 2007; Stegmann et al., 2003). However, when the image is occluded, those methods that extract global features (holistic features) (as Eigenfaces and Fisherfaces) cannot be applied. There are many approaches that deal with occlusion in face recognition. Among them we can mention the one proposed by Shermina and Vasudevan (2012) that propose block comparison and thresholding to detect occlusion in the query image and use Empirical Mode Decomposition (Huang et al., 1998) (EMD) to normalize facial expression. In Lang and Jing (2011) , Guillamet and Vitria (2002) and Lee and Seung (1999) the use of Non-negative Matrix Factorization is explored because the locality of this approach lends itself to deal with occlusions. In Chiang and Chen (2011) a occlusion resistant face recognition method is proposed in which the query image is first normalized to a common shape and then, its texture is reconstructed by using PCA for each specific person in the database, which in turn allows to identify the occluded pixels as the ones that are very different from the query image. While methods that use local features may not be affected by occlusion, (Wright et al., 2009) has shown that useful information is lost when only local features are extracted. There, the authors propose a method for recognizing faces that is robust to certain types and levels of occlusion. This method is based on recent advances in the study of statistical signal processing, more specifically in the area of compressed sensing (Candès et al., 2006; Candès and Tao, 2006; Candès et al., 2008).

As this method fails when it reaches a 33 % of occlusion in one connected region, the same work proposes to partition the problem into smaller problems. To this end, the image is partitioned into smaller blocks and each block is then processed separately. Doing this entails several disadvantages, the main ones are that holistic features are lost and that the optimal way of partitioning the image cannot be determined. Ideally, it would best to detect which areas are occluded in the image and then discard them at the time of recognition. There are other methods which seek to detect occlusion. For example, Lin and Tang (2007) presents a method that uses a Bayesian filter and Graph-Cut to do Face Inpainting to restore the occluded sections. Moreover, Zhou et al. (2009) presents another method that detects and recognizes occlusion using Markov Random Fields, but with a high computational cost. In this work the authors propose a method for recognizing faces that is robust to certain types and levels of occlusion. An assessment of this method shows that it obtains a better performance than the aforementioned methods.

* Corresponding author. Fax: +54 11 4576 3359.
E-mail address: jacobo@dc.uba.ar (J. Jacobo-Berlles).

0167-8655/$ - see front matter © 2013 Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.patrec.2013.08.001

2. Face recognition

First, for the sake of completeness, the model from Wright et al. (2009) is addressed in this section. In it, face recognition is modeled as an optimization problem in which the query face is represented as a sparse linear combination of the faces in a dictionary.

All the images to be used have the same size *width* × *height*. These images represent a point in $\mathbb{R}^m$, where $m = width \times height$, which is obtained by stacking their column vectors.

It has been shown that images of the same person under different lighting conditions and expressions fall (approximately) in a linear subspace of $\mathbb{R}^m$ of much lower dimension, called *faces subspace* (Basri and Jacobs, 2003; Belhumeur et al., 1997; Lee et al., 2005).

A **dictionary** of n atoms is a matrix $A \in \mathbb{R}^{m\times n}$ where each of the n columns is an image in $\mathbb{R}^m$ of a known face.

Let $A \in \mathbb{R}^{m\times n}$ be a dictionary with n atoms, and let $y \in \mathbb{R}^m$ be a query image. We represent image y by a linear combination of the atoms in A, so $y = Ax$, where $x \in \mathbb{R}^n$ is the vector of the coefficients used in the linear combination. Our goal is to find the most sparse solution, i.e. the one that has the largest number of null terms and uses the least number of atoms in the dictionary.

In practice, the number of atoms n is larger than the image size m, so the system is under-determined and the solution x may be not unique. Recent studies in sparse representations and in signal sampling (Donoho, 2004) show that, the most sparse solution to the problem can be found using the l^1 norm ($||\cdot||_1$). We then consider $\hat{x}_1 = \min ||x||_1$ subject to $y = Ax$ as the sought solution.

When the person to be evaluated is not represented in the dictionary, the solution x is usually dense and non-zero coefficients are distributed over the atoms of different people in the dictionary. However, if the individual to be evaluated is in the dictionary, most nonzero coefficients of x will correspond to that person's atoms.

In order to establish a rule to decide when the recognition is satisfactory, (Wright et al., 2009) defines a coefficient $SCI(x)$ (Sparsity Concentration Index) of a vector $x \in \mathbb{R}^n$ as $SCI(x) \triangleq \frac{k\cdot\max_j \|\delta_{W_j}(x)\|_1/\|x\|_1 - 1}{k-1} \forall j = 1,\ldots,k$, where $W_1, W_2, \ldots, W_k$ with $k > 1$, are the different classes or individuals belonging to the dictionary A and $\delta_{W_j}(x)$ is a function of $\mathbb{R}^n \to \mathbb{R}^n$ that makes zero those coefficients of x that do not correspond with atoms of the class W_j.

If all the nonzero coefficients of x are concentrated in a single class, then $SCI(x) = 1$. Conversely, if all the nonzero coefficients of x are uniformly distributed among all classes, then $SCI(x) = 0$. Thus, a threshold on the *SCI* can determine whether or not a person is in the dictionary of faces.

To determine the identity of a person matched in the dictionary, we define the residue

$$r_{W_j}(y) \triangleq \|y - A\delta_{W_j}(x)\|_2 \tag{1}$$

We consider the person's identity as defined by the class W_j with the lowest residue, i.e. $Identity(y) \triangleq \arg\min_{W_j} r_{W_j}(y)$.

When the face to be recognized is partially occluded, large errors appear that have to be modeled specifically. Consequently, one can think of occlusion as an error $e$that affects a portion of the image.

$$y = Ax + e = \begin{bmatrix} A, \\ I \end{bmatrix}\begin{bmatrix} x \\ e \end{bmatrix} = Bu \tag{2}$$

where $B = [A, I] \in \mathbb{R}^{m\times(n+m)}$ and $u = \begin{bmatrix} x \\ e \end{bmatrix}$ and e has nonzero components only in the occluded portion of the image. The location of these errors is not known and its magnitude is completely arbitrary. However, we assume that the portion affected by the occlusion is small – less than 33% (Wright et al., 2009) – relative to the whole image. The system (2) is indeterminate and, if there is any solution for u, that solution is not unique. The most sparse solution $u_0 = [x_0, e_0]$, obtained using the norm l^0 is sparse enough so as to allow us to use the norm l^1 and obtain the same solution. So we reformulate the system (2), extending it to:

$$\hat{u}_1 = \min ||u||_1 \quad \text{subject to } Bu = y \tag{3}$$

Then, in order to calculate the identity of a detected individual, we need to reformulate the residue Eq. (1). The new equation becomes:

$$r_{W_j}(y) \triangleq \|y - \hat{e}_1 - A\delta_{W_j}(\hat{x}_1)\|_2 \tag{4}$$

where

$$\hat{u}_1 = \begin{bmatrix} \hat{x}_1 \\ \hat{e}_1 \end{bmatrix}$$

This model works when the occlusion is scattered throughout the whole image, as it is the case of impulsive noise and random pixel corruption. But when the occlusion is concentrated in one place, this method begins to fail when the image is occluded in more than 33% (Wright et al., 2009). A way to improve recognition in this case is to partition the image into blocks to be processed separately.

Once the identities for each block are obtained, a voting system is used to determine the identity of the person. In this way, it is expected that, when certain zones are occluded, the remaining ones will allow us to obtain the right identity. Also, it is possible to detect the partitions that are most affected by occlusion by calculating the SCI coefficient of each of them.

More specifically, the image to be evaluated y and each image of the face dictionary are partitioned into L blocks of size $a \times b$. From the partitioned dictionary the matrices: $A^{(1)}, A^{(2)}, \ldots, A^{(L)} \in \mathbb{R}^{p\times n}$, with $p = ab$, are obtained, and for each block the system: $y^{(l)} = A^{(l)}x^{(l)} + e^{(l)}$, con $y^{(l)}, x^{(l)}, e^{(l)} \in \mathbb{R}^p$ is formulated. In the same manner, we calculate for each block the system of Eq. 3, obtaining:

$$\hat{u}_1^{(l)} = \min ||u||_1 \quad \text{subject to } B^{(l)}u = y^{(l)} \tag{5}$$

where

$$u \in \mathbb{R}^{n+p} \wedge u = \begin{bmatrix} x^{(l)} \\ e^{(l)} \end{bmatrix}$$

$$B^{(l)} = \left[A^{(l)} I\right]$$

The main problem with this approach is to decide the best partition in order that occlusions do not affect many blocks. There is also another problem: this method uses only local characteristics of each partition, not considering the global information present in them.

Due to these inconveniences, in the next section we present an alternative approach for heavily occluded images.

To overcome this limitation, one can detect the occluded zone and then exclude it at the time of the recognition phase. In the next section we propose a method based on the same basis to detect occluded areas of the image.

3. Occlusion detection

Let us consider an image, such as that shown in Fig. 1(a). Let us suppose also that somehow we get another picture of the same subject with the same pose, expression and lighting, but without occlusion (Fig. 1(b)). Then, taking the absolute value of the difference between them we obtain Fig. 1(c) in which the non-null pixels are the ones that were affected by the occlusion. By applying a threshold τ, we get the Fig. 1(d), in which a good approximation to the occluded area is obtained. Fig. 1(d) presents several interest-