



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.

© 2013 Published by Elsevier B.V.

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), Guillaumet 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).

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 $width \times height$. 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 $SCI(x) \triangleq \frac{k \cdot \max_j \|\delta_{W_j}(x)\|_1 / \|x\|_1 - 1}{k-1} \forall j = 1, \dots, k$, where W_1, W_2, \dots, 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 $\mathbb{R}^n \rightarrow \mathbb{R}^n$ 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

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

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.

$$y = Ax + e = \begin{bmatrix} A \\ I \end{bmatrix} \begin{bmatrix} x \\ e \end{bmatrix} = Bu \quad (2)$$

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:

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

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

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

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: $A^{(1)}, A^{(2)}, \dots, A^{(L)} \in \mathbb{R}^{p \times n}$, 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:

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

where

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

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

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-

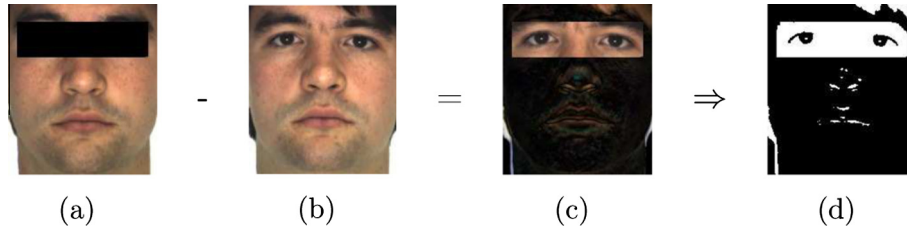


Fig. 1. (a) Synthetically occluded image. (b) Another image of the same subject, the pose, expression and lighting similar. (c) Result of the difference between (a) and (b). (d) Threshold results with $\tau = 50$ image (c).

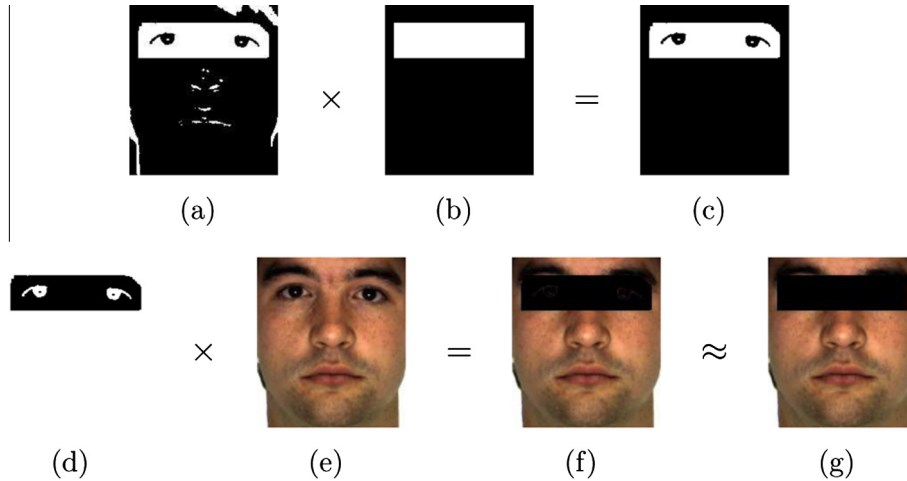


Fig. 2. (a) Mask obtained in 1(d). (b) Synthetic occlusion mask added to Figure (g). (c) Result of multiplying (b) and (a), in order to get the occlusion detected on the eyes zone only. (d) Occlusion detected in the eye area. (e) Original image without occluding. (f) Result occlude the image (e) with (d). (g) Original image synthetically occluded. It can be seen that it is not very different from the image (f).

ing features to note. The first is that there are small sections erroneously identified as occlusion in the area of the mouth and nose. This can be reduced if instead of comparing pixel by pixel, pixel neighborhoods had been compared. Second, we can see that in the upper right corner of the image some hair is marked as occlusion. The hair covering the front portion of the image does not appear in Fig. 1(a) but in Fig. 1(b). That is, the detected occlusion included covered parts in both images. Third and finally, we can see that the area of the iris in the eyes is not marked as occlusion. This is because the black area of the synthetic occlusion Fig. 1(a) is not an atypical value for this person's iris. Fig. 2 shows an example of this.

The synthetically occluded image in Fig. 2(g) shows the ideally occluded area as a black rectangle on the eyes. However, the algorithm marks as not occluded some areas of the eyes of the individual. To answer the question whether this result makes sense, we do the following test. First we get the section corresponding to the eyes zone detected as occlusion. This is done by multiplying (a) and (b) obtaining (c). Using an inverted version of his mask (Fig. 2(e)) considering its foreground as transparent, we multiply it with the non-occluded picture (e) obtaining Fig. 2(f). Comparing it with the original image (g) we can see there are virtually no observable differences. We therefore conclude that the detected mask for the eye area is correct, because by occluding the detected area we obtain a result that is very similar to the original. The problem of detecting occlusion is to determine which pixels in an image are showing a portion of a face and which pixels are not. We can think of occluded pixels as

outliers in a face image. To detect the occlusion is then to detect those pixels with outlier values for the face image. The proposed method is to use local features of images to generate a new image that is similar to the one we want to recognize, but without occlusion. After obtaining this image, we find the difference between this image and the query image and then apply a threshold τ . If the generated image is sufficiently similar to the query image, this procedure will yield the outlier pixels present in the occluded image.

The proposed method uses only a non-occluded fragment F of the database images to generate an image similar to the one we want to recognize. We consider fragment F as a possibly non-contiguous subset of image pixels. The image generated by using fragment F has no occlusions because database images have no occlusions. After obtaining this image, we find the difference with the query image and then apply a threshold τ . If the generated image is sufficiently similar to the query image, this procedure will yield the occluded regions, i.e. the outlier pixels present in the query image.

Let us consider that there is a non-occluded fragment F in the query image. Using only this fragment of the database images, Algorithm 1 reconstructs the fragment of the query image, and later in the final step reconstructs the complete query image using the same coefficients x used to reconstruct the fragment F . By using only a fragment of the images, we obtain a less accurate result because not all of the non-occluded pixels are used, but this is addressed later by the iteration in Algorithm 2 using different sets F_i and choosing the best result.

Let us see in detail in Algorithm 1.

Algorithm 1.

Algorithm to generate an image similar to y using pixels from fragment F only

Require: F , a set of pixels

- 1: Let y_F be the fragment F of the query image y .
- 2: Let A_F be the dictionary that results from selecting fragment F in each image of dictionary A .
- 3: Using Eq. (2), image y_F and face dictionary A_F , we calculate coefficients x and error e so $y_F = A_F x + e$.
- 4: **return** $y' = Ax$

We seek then a set of pixels F containing a large portion of non-occluded pixels to use Algorithm 1 in order to obtain an image y' similar to the query image y to finally obtain the occlusion present in the latter. To obtain the sets of pixels F we consider the following four options: to use a sample of uniformly distributed pixels, to take a sample of random pixels, to take random blocks of contiguous pixels, and to use predefined regions of the image. For the fourth option, we take into account the most common occlusions that may arise.

Occlusion usually happens in blocks, i.e. it is usually concentrated in one or more portions of the image. To explain this, we will use as an example Fig. 3(a) which presents a block occlusion of 50%. In this image occlusion is concentrated in a single block, but then conclusions will get will be valid even if occlusion is fragmented into several blocks. For readability, from now on we will mark in red and blue the pixels that are occluded and non-occluded respectively, as shown in Fig. 3(b).

Let us consider then the first method to obtain a set of pixels F , which involves taking an uniformly distributed sample of image pixels. For example, in Fig. 3(c) we can see that pixels were chosen scattered throughout the image. This method has the advantage that it captures well the global features of the image, however it has two major disadvantages. The first is that it is not good at capturing local image characteristics. The second and main disadvantage is that the proportion of occluded pixels in the set F is the same (or very similar) as in the original image. So, if the query image has a large number of pixels occluded (as in the example shown in the figure, where 50% of the pixels are occluded), the set F also will, resulting in Algorithm 1 yielding a large error and in a resulting image that differs substantially from the query image in the non-occluded areas.

Our second method is a somewhat naive attempt to solve the problems of the first method and uses random selection of pixels. As shown in Fig. 3(d), if the number of pixels selected is large enough, it is very likely that the selected pixels will be scattered

throughout the image. This makes this method to have the same advantages and disadvantages as the first one. The third method is to select adjacent pixel blocks at random. The larger these pixel blocks are, fewer blocks we need for a given size of F . Then, the more blocks we have and the smaller they are, then set F will have pixels scattered throughout the image. If instead there are few large blocks, the pixels of set F will be more concentrated in some regions of the image. In Figs. 3(e)–(g) are examples of blocks of pixels selected at random. Considering that the occlusion is usually concentrated in some regions of the image, by choosing different positions of a few blocks of pixels, we obtain sets of pixels with different ratios of occluded pixels. For example, Fig. 3(e) has 70% of occluded pixels, Fig. 3(f) has the 30% and Fig. 3(g) has none. However, the fact that the set of pixels F is concentrated in one region can have its negative effect on Algorithm 1 making it a little less stable. Unfortunately there is no way to determine the optimal number of blocks, this is why it is advisable to test with multiple block sizes and quantities. The fourth and last method is to use a predefined sets of pixels, taking into account the most common occlusions that can occur. That is, if many of the test images have dark glasses, it is desirable to try a set of pixels F that does not include pixels near the eye position. This is not difficult, since we assume that all images are aligned. As expected, this method is only to be used combined with others, because it is not possible to know all types of occlusion that may arise. Since we have multiple sets of pixels, we apply the Algorithm 1 several times, one for each set of pixels. We said earlier that in algorithm 1 we use a set of pixels with a high proportion of occluded pixels, the resulting image is most likely to be very different from the query image. Given this, among all images obtained y'_1, \dots, y'_n we will look for one that meets $y'_{max} = \arg \max_{y'_i} S_\tau(y, y'_i)$, where $S_\tau(y, y'_i)$ is the number of pixels p such that $|y(p) - y'_i(p)| \leq \tau$ and τ is a given threshold.

Algorithm 2.

Occlusion detection algorithm.

Require occluded query image y , faces A , different sets of pixels F_1, F_2, \dots, F_n and threshold τ

- 1: **for** $i = 1 \dots n$ **do**
- 2: Obtain y'_i using Algorithm 1 and set F_i
- 3: Let $d_i = |y - y'_i|$
- 4: Let $u_i = \text{threshold}(d_i, \tau)$
- 5: **end for**
- 6: Let $max = \arg \max_i S_\tau(y, y'_i)$
- 7: Let F_{n+1} be the set of pixels u_{max} .
- 8: Obtain y'_{n+1} using Algorithm 1 and set F_{n+1}
- 9: Sea $d_{n+1} = |y - y'_{n+1}|$
- 10: **return** $u_{n+1} = \text{threshold}(d_{n+1}, \tau)$

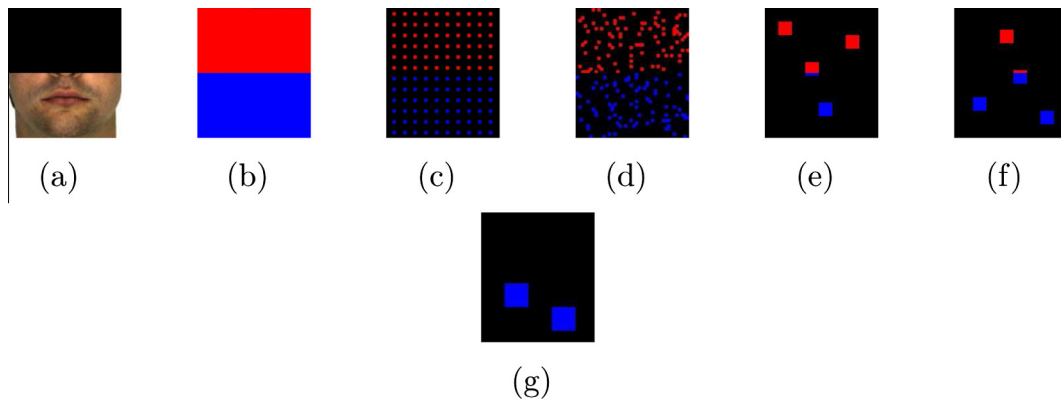


Fig. 3. Different forms of obtaining a fragment F (see text).

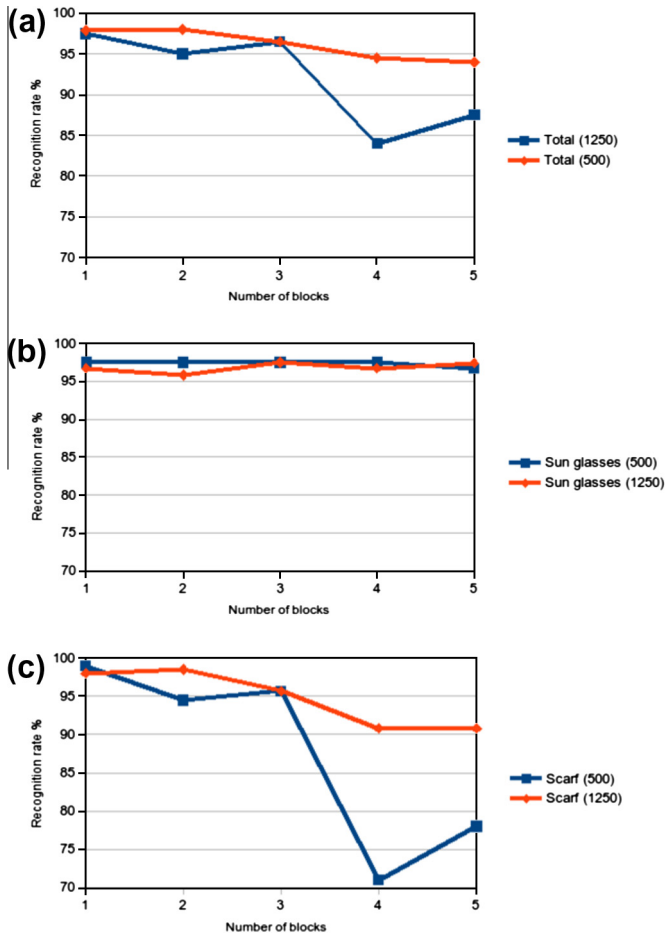


Fig. 4. Recognition rates obtained by detecting and then excluding occluded parts. Here, the threshold was set to $\tau = 20$ and the number of blocks of contiguous pixels varied from 1 to 5 for the sets F_i in Algorithm 2. For the blue lines, the sets F_i have 500 pixels approximately, while for the green lines those sets have 1250 pixels. (a) recognition rate obtained with several numbers of blocks of contiguous pixels. (b) the same as in (a), but only for images where the occlusion consists of sun glasses. (c) the same as in (a), but only for images where the occlusion consists of scarves.

The Algorithm 2 takes n fragments F_i randomly generated using a distribution function. Then, for each of these inputs makes a deterministic calculation, using the result of each individual calculation (for each input) to form a final result. Specifically, this Algorithm 2 requires a dictionary of faces A and multiple sets of pixels

F_1, F_2, \dots, F_n generated by taking pixel blocks in a random manner and/or using some predefined sets that are convenient for the most common types of occlusion. Using these sets and a threshold τ (whose value can be obtained experimentally), it calculates the occluded area of an image y .

The idea of the Algorithm 2 is to generate multiple images y'_i (step 2) and then select the one with the highest similarity S_τ (step 6). Once this is done we have a first detection of the occlusion u_{max} . Even though this result is useful, we have found that if we use the non-occluded pixels as the set of input pixels (steps 7 and 8), then the final result is much more accurate and robust. This may be due to the fact that the intermediate result u_{max} has a low percentage of occluded pixels and also that these pixels are not located in a few regions of the image, but instead are scattered throughout the regions where there is no occlusion.

Once obtained the occluded sections of the image, we can use the rest for recognition, using the aforementioned methods.

4. Results

In this section we present the tests performed using the recognition algorithm described in Wright et al. (2009), together with the occlusion detection Algorithm 2.

Our method is made up of two stages. The first one deals with occlusion detection and the second one deals with recognition of the individual using non-occluded zones only. Even though we are interested in the result of the second stage, its success depends on the success at the first stage.

In Algorithm 2, the resolution of the system in step 2 is done in polynomial time by standard linear programming methods (Wright et al., 2009). Other factors that affect complexity are block size, number of blocks and number of fragments F_i . Clearly, computational complexity is linear in the product of these magnitudes. Some representative CPU times are: 335.3 s for eight blocks of 12×12 pixels and 50 fragments F_i and 404.6 s for six blocks of 15×15 pixels and the same number of fragments. The hardware used was an Intel Cores i5 860 processor with an 2.85 MHz clock and a cache of 8 M.

This section does not assess the performance of each individual step, but that of the whole of both stages. In other words, we will consider a successful case when a partially occluded image is correctly recognized by the algorithm.

The faces database used for testing is the AR Database (Martinez and Benavente, 1998). From this base we selected 60 people, and from each one 8 non-occluded images were used for the faces dic-

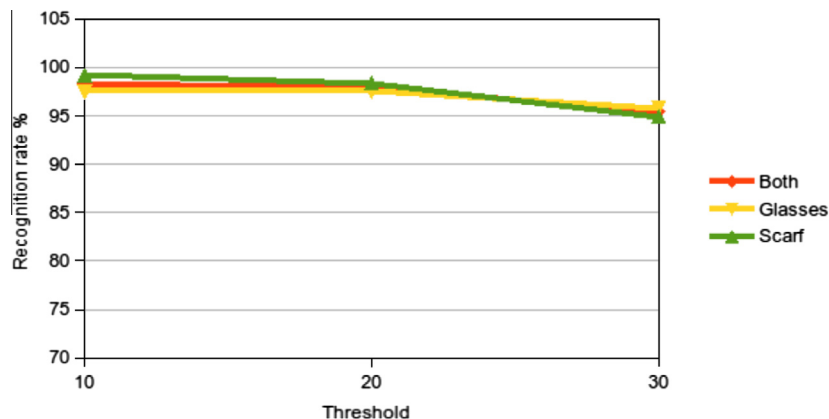


Fig. 5. Recognition rates obtained by detecting and then excluding occluded parts. In these experiences two blocks of contiguous pixels were used, with the 15, 5% of the pixels in the sets F_i and varying the threshold τ .

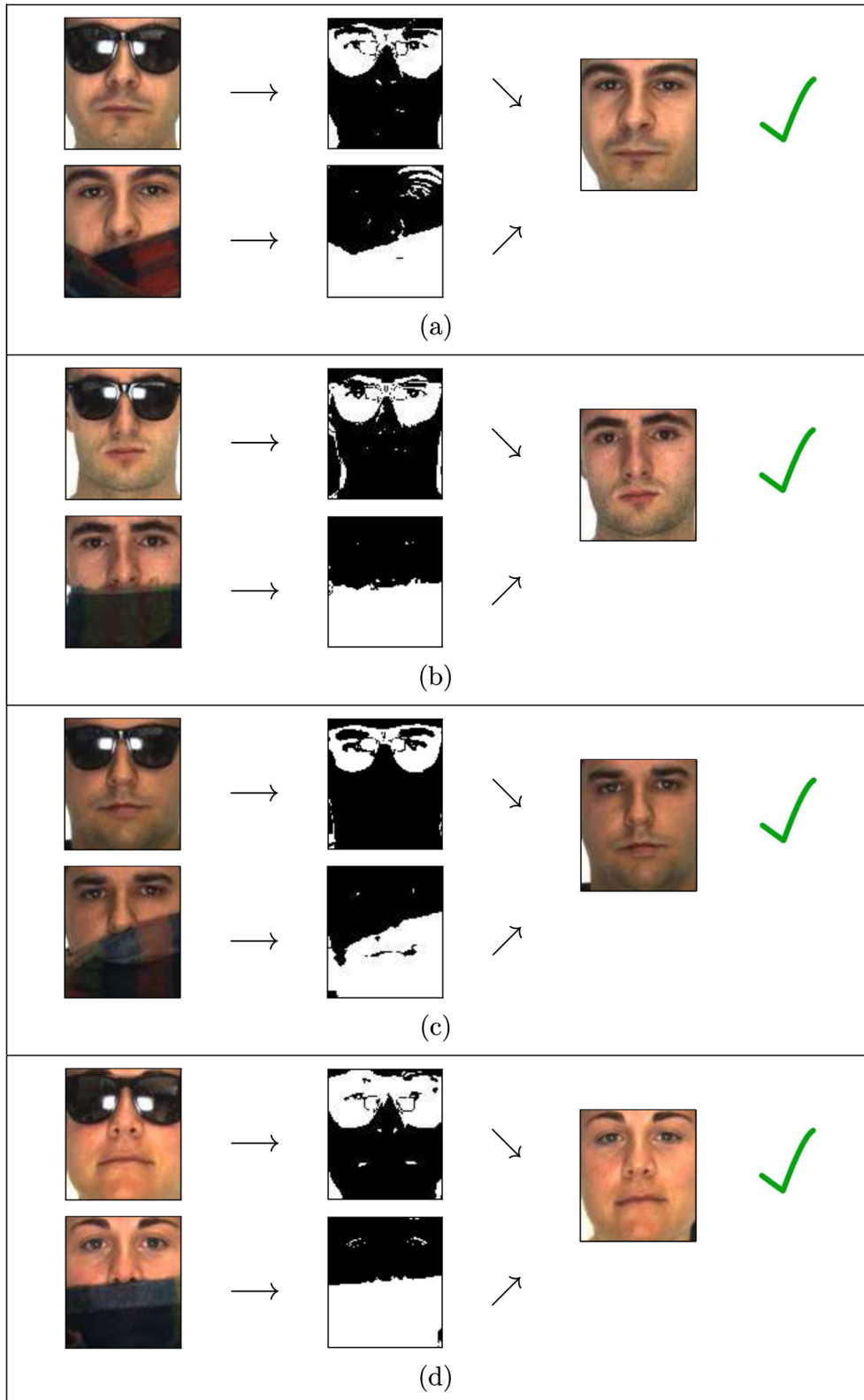


Fig. 6. Reconnection with occluded images of four individuals, (a)–(d). Here are shown two occluded images for each individual. Then, the occluded zone detected is shown. Finally the corresponding non-occluded images are shown. It can be seen that all individuals were recognized correctly by the algorithm.

tionary and four occluded images were used for test. From the test images, two are occluded by sunglasses, and two by a scarf.

The proposed occlusion detection approach has several parameters. Among them is the threshold τ , the number of pixels of the sets F_i in Algorithm 2 in Section 3 and how these sets are obtained.

In Section 3 we presented four methods to obtain those sets, from which only two were useful for us. These methods are: “Use predefined sets of pixels, considering the most common occlusions that can occur” and “Take blocks of contiguous pixels randomly.” The first one was not used in our measurements in or-

der not to give the detector any previous knowledge about the existing occlusion in the faces database. The second method requires another parameter: the number of blocks of pixels.

Our first experiment was then to set threshold $\tau = 20$ and vary the number of pixel blocks. This was done using sets of 500 pixels (6.25%) and of 1250 pixels (15.5%) from a total of 8064 pixels.

Fig. 4 shows the recognition rate obtained using different quantities of blocks of contiguous pixels. The blue lines represent the tests made using 500 approximately; while the green lines represent tests in which the block size is approximately 1250 pixels.

It can be seen that, as the number of blocks grows, the recognition rate for the occlusion by scarves begins to fall (Fig. 4(c)). This happens because, as there are more blocks, the probability that all of them are in a non-occluded zone is much smaller. It can also be observed that the increase in the number of blocks does not affect so much the images occluded by sun glasses. This happens because these images have a smaller proportion of occluded pixels, and thus the probability that the blocks of pixels do not have occluded pixels is smaller.

With respect to the size of the sets F_i , it is worthy to mention that the experiences that had approximately 15% of the total number of pixels happened to be more stable than those that had only 6, 25%.

The next experiment was to set the number of blocks of pixels to 2 and to 1250 pixels in the sets F_i (15, 5% of the total image) and then vary the value of threshold τ .

Fig. 5 shows the results of varying the threshold τ . Empirically, we got the best results using $\tau = 10$. Using a lower threshold introduces too much false negatives in the detection of occluded pixels, i.e. non-occluded pixels considered as occluded.

In Fig. 6 we included some occluded faces. The algorithm detected the occluded zone and recognized the person using the rest of the image.

Observing carefully the images in Fig. 6 that show the occluded zones detected, we can see that some of them have small non-occluded regions which were detected as occluded and vice versa. As the recognition algorithm is robust with respect to small occluded zones, these small errors do not affect the recognition rate. This is confirmed by the high recognition rate obtained.

Finally, in Table 1 we see comparatively higher recognition rates obtained for the different types of occlusion tested. Also included are the recognition rates obtained by implementing the two solutions proposed in Wright et al. (2009) (Simple Model and Partitioning Model). By Simple Model, we refer to the model proposed by Wright et al. (2009) in Eq. 3 and by Partitioning Model we refer to the one presented by the same authors in Eq. 5. To test the three models we used the same database described previously in this section and we implemented the three of them using the same math library to resolve the systems involved. We tested all the models using gray-scale and color images. In the particular case of the Simple Model, we tested with different image scales. In the Partitioned and Occlusion Detection models we used the original resolution of the images.

Even though our paper was primarily focused on occlusion robustness in face detection, other tests were also carried out but no thorough comparisons were made. Illumination variations were assessed using the extended Yale B database with good results,

similar to the ones obtained by Wright. This is not surprising because, as there are no occlusions in this database, the performance is basically the same. The same happened with noise corruption. On the other hand, occlusion was tested using the AR database, which includes gesture variation. Non-occluded images were used as database and occluded images were used as queries. The results can be seen in Table 1.

The advantages of the proposed method over Wright's method can be summarized as follows: our method is not dependent of the position of the occlusion, because it detects the occluded pixels and then makes the recognition without using them. In this way our method tries to find and also use all the non-occluded pixels. On the other hand, Wright's method selects 8 big fixed blocks. Some of them are partially occluded, and these partially occluded blocks do not contribute to decide the person's identity, even when they possess a considerable percentage of non-occluded pixels.

Our method also does not make assumptions on the position of the occlusion. So, our method is adaptive in this sense.

In conclusion, the method proposed in this paper presents an improvement in face recognition in the presence of occlusion compared to the performance obtained by the method proposed in Wright et al. (2009). The performed experiments have shown that successful recognition is strongly linked to the success of the occlusion detection. Likewise, one of the factors determining the occlusion detection is the choice of the set of pixels F , which can be a subject for further research.

5. Conclusions

In this paper the authors propose a method for recognizing faces that is robust to certain types and levels of occlusion. We approach the problem of face recognition proposing a novel method which is based on recent advances in the study of signal processing, more specifically in the area of *compressed sensing*. The recognition can be performed even when the image of the subject's face is partially occluded. Even though the method proposed in Wright et al. (2009), which subdivides the recognition problem into smaller problems, improves the recognition rate, it is not flexible enough to adapt to the different distributions of occlusion that may arise. For this reason, we propose a new method to detect the occluded areas and exclude them from the recognition process using only areas which provide information on the identity of the person. Experiments have shown that successful recognition is strongly linked to the successful detection of the occlusion. Likewise, one determining factors in occlusion detection is the choice of the set of pixels F , which can be a subject of further investigation.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.patrec.2013.08.001>.

References

- Basri, R., Jacobs, D.W. 2003. Lambertian reflectance and linear subspaces. IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (2), 218–233. URL <http://dx.doi.org/10.1109/TPAMI.2003.1177153>
- Belhumeur, P.N., Hespanha, J., Kriegman, D.J. 1997. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7), 711–720. URL <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.10.3247>
- Candès, E.J., Romberg, J.K., Tao, T. 2006. Stable signal recovery from incomplete and inaccurate measurements. Communications on Pure and Applied Mathematics 59 (8), 1207–1223. URL <http://dx.doi.org/10.1002/cpa.20124>
- Candès, E.J., Tao, T., 2006. Near-optimal signal recovery from random projections: Universal encoding strategies? IEEE Transactions on Information Theory 52 (12), 5406–5425.

Table 1

Recognition rates obtained with different face recognition methods in presence of occlusion.

Algorithm	Sunglasses	Scarves	Total
Simple model	69, 49%	12, 71%	41, 10%
Partitioned model	86, 44%	95, 76%	91, 10%
Occlusion detection	97, 46%	99, 15%	98, 31%

- Candès, E.J., Wakin, M.B. 2008. An introduction to compressive sampling. *Signal Processing Magazine, IEEE* 25 (2), 21–30. URL <http://dx.doi.org/10.1109/MSP.2007.914731>
- Chiang, C.-C., Chen, Z.-W., 2011. Recognizing partially-occluded faces by recovering normalized facial appearance. *International Journal of Innovative Computing* 7 (11), 6219–6234.
- Cootes, T.F., Edwards, G.J., Taylor, C.J., 2001. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23, 681–685.
- Donoho, D.L., 2004. For most large underdetermined systems of linear equations the minimal ℓ_1 -norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics* 59, 797–829.
- Guillamet, D., Vitria, J. 2002. Classifying faces with non-negative matrix factorization. *Proceedings of the 5th Catalan Conference for Artificial Intelligence (CCIA 2002)*, Castello de la Plana (2002) 24–31.
- Huang, N.E., Long, S.R., Shen, Z., 1998. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 454 (1971), 903–995, e. a..
- Kahraman, F., Kurt, B., Gokmen, M. 2007. Robust face alignment for illumination and pose invariant face recognition. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1–7.
- Lang, L., Jing, X., 2011. Application of non-negative sparse matrix factorization in occluded face recognition. *Journal of Computers* 6 (12), 2675–2679.
- Lee, D.D., Seung, H.S. 1999. Learning the parts of objects by non-negative matrix factorization. *Nature* 401, 788. URL <http://hebb.mit.edu/people/seung/papers/ls-lpnm-99.pdf>
- Lee, K.-C., Ho, J., Kriegman, D.J. 2005. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27 (5), 684–698.
- Lin, D., Tang, X. 2007. Quality-driven face occlusion detection and recovery. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1–7.
- Martinez, A.M., Benavente, R. 1998. The AR face database. Technical Report 24, CVC.
- Shermina, J., Vasudevan, V. 2012. Recognition of the face images with occlusion and expression. *International Journal of Pattern Recognition and Artificial Intelligence* 26 (03), 1256006. URL <http://www.worldscientific.com/doi/abs/10.1142/S021800141256006X>
- Stegmann, M.B., Ersbøll, B.K., Larsen, R. 2003. FAME – a flexible appearance modelling environment. *IEEE Transactions on Medical Imaging* 22 (10), 1319–1331. URL <http://www2.imm.dtu.dk/pubdb/p.php?1918>
- Turk, M., Pentland, A. 1991. Eigenfaces for recognition. *Journal of Cognitive Neuroscience* 3 (1), 71–86. URL <http://citeseer.ist.psu.edu/contextsummary/61360/0>
- Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y. 2009. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31 (2), 210–227.
- Zhou, Z., Wagner, A., Mobahi, H., Wright, J., Ma, Y. 2009. Face recognition with contiguous occlusion using markov random fields. In: *International Conference on Computer Vision (ICCV)*.