Randomized Face Recognition on Partially Occluded Images

Ariel Morelli Andres, Sebastian Padovani, Mariano Tepper,
Marta Mejail, and Julio Jacobo

Departamento de Computación, Facultad de Ciencias Exactas y Naturales,
Universidad de Buenos Aires, Argentina
{amorelli,spadovan,mtepper,marta,jacobo}@dc.uba.ar

Abstract. In this work we propose a new method for face recognition
that successfully handles occluded faces. We propose an innovative im-
provement that allows to detect and discard occluded zones of the face,
thus making recognition more robust in the presence of occlusion. We
provide experimental results that show that the proposed method per-
forms well in practice.

1 Introduction

Face detection and recognition 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 come, 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 for face recognition. Among the most
popular are Eigenfaces [1] and Active Appearance Model [2,3,4]. However, when
the face is occluded, methods that extract global features (holistic features) (such
as Eigenfaces and Fisherfaces [5]) cannot be applied. While methods that use
local features are less affected by occlusion, useful information might be lost
when only local features are extracted [6].

In a seminal paper, Wright et al. [6] proposed 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 [7,8,9]. The method fails when 33% of the image is
occluded in a single connected region. A straightforward solution is to partition

the problem into smaller subproblems. 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 being that holistic features are lost and that the best way of partitioning the image can be hardly determined a priori.

Ideally, it would be best to detect which areas are occluded in the image and then discard them for recognition. There are other methods which seek to detect occlusion. For example, Lin et al. [10] presented a method that uses a Bayesian filter and Graph Cuts to do Face Inpainting to restore the occluded sections. Zhou et al. [11] proposed to detect and recognize occlusion using Markov Random Fields, but at a high computational cost. In this work we propose a method for detecting occlusion based on compressed sensing that obtains a better performance than the afore mentioned methods.

The rest of the paper is structured as follows. In Section 2 we present the face recognition method. Then we proceed to explain the proposed approach to handle occlusions in Section 3. We experimentally check the good performance of the proposed approach and provide concluding remarks in Section 4.

2   Face Recognition

In this section we model face recognition as an optimization problem in which we want to represent the query face as a sparse linear combination of the faces in a dictionary.

All the images to be used have the same size $width \times hight$. As usual, we assume that images are correctly cropped and aligned to the right size and position, respectively. These images represent a point in $\mathbb{R}^m$, where $m = width \times hight$, 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* [5,12,13].

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 person's face, whose identity is known. Let us assume that there are k ($1 < k \le n$) different classes or individuals in A. We denote by W_j the images (columns fo A) corresponding to the j-th class.

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 uses the least number of atoms in the dictionary. This can be achieved by adding a constraint on the l^0 norm of x, unfortunately the resulting problem is non-convex. However, by using the l^1 norm, a sparse solution is obtained, while maintaining convexity [14]. We then define the problem as finding

$$\hat{x}_1 = \min \|x\|_1 \quad \text{subject to} \quad y = Ax.$$

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, [6] defines a coefficient $SCI(x)$ (Sparsity Concentration Index) of a vector $x \in \mathbb{R}^n$ as

$$SCI\left(x\right) \triangleq \frac{k}{k-1} \cdot \left(\max_{1 \leq j \leq k} \left\|\delta_{W_j}\left(x\right)\right\|_1 / \left\|x\right\|_1 - 1\right),$$

where $\delta_{W_j} : \mathbb{R}^n \to \mathbb{R}^n$ sets to zero those coefficients of x that do not correspond to atoms of class W_j.

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

To determine the identity of a person already found in the dictionary, we define the residual $r_{W_j}(y) \triangleq \left\|y - A\,\delta_{W_j}(x)\right\|_2$. We consider the person's identity assigned to the class with the lowest residual, i.e. $\arg\min_j r_{W_j}(y)$.

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

$$y = Ax + e \tag{1}$$

where e has nonzero components only in the occluded portion of the image. The location of these errors is not known and their magnitudes are completely arbitrary. However, we assume that the portion affected by the occlusion is not overwhelmingly large with respect to the image size m.

To overcome this limitation, we develop a method that first detects the occluded zone and then performs recognition by using only the non-occluded parts.

3   Occlusion Detection

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 a face image. We propose to use local image features 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 original 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.

To generate this image, let us consider that there is a non-occluded and possibly non-connected fragment F in the image domain. The procedure for generating this reconstructed image is described in Algorithm 1. We start by building
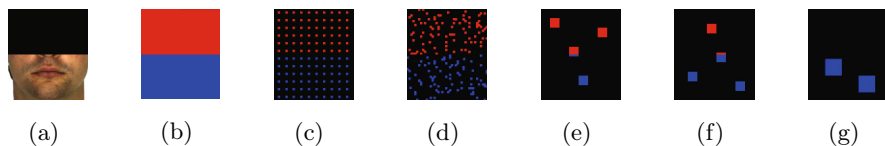
Fig. 1. An image and its occlusion map. Pixels in blue correspond to a non-occluded part, while pixels in red are occluded. We illustrate different ways of obtaining a fragment F.

the subimage y_F and the subdictionary A_F, where only the rows corresponding to pixels in F are kept. Then we try to recognize the identity of the person using the procedure described in the previous section. Of course, since we are only using partial information from the image, we obtain a less accurate result. We also obtain the vector of coefficients x that was used to represent y_F using A_F (step 3 in Algorithm 1). Using these coefficients, we obtain a reconstructed image $y' = Ax$, similar to the query image.

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 image y_F and face dictionary A_F, we compute the sparse code x and error e so that $y_F = A_F\, x + e$.
 4: **return** $y' = Ax$

The question now is: how to obtain an appropriate set of pixels F to be used algorithm 1? To this end, we consider the following four options: (1) to use predefined regions of the image, (2) to pick pixels from a fixed grid, (3) to sample pixels at random, and (4) to generate a sample of random blocks of contiguous pixels.

Let us first observe that occlusion usually occurs in blocks, i.e. it is usually concentrated in one or more connected portions of the image. Fig. 1(a) serves as a simple example, presenting a block occlusion of 50%. In this image, occlusion is concentrated in a single block, but in the following we make no assumptions in this respect.

The first method uses 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 determine all types of occlusion that may arise.

Let us consider then the first randomized method for obtaining a set of pixels F, which involves considering pixels in a fixed uniform grid, as depicted in Fig. 1(c). The advantage of this method is that it captures well the global features of the image, however it clearly has two major disadvantages. First, since local traits usually span several pixels, the method fails at capturing these 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. Then, Algorithm 1 will yield a large error and the resulting reconstructed image will differ substantially from the query image in the non-occluded areas.

Our third method is a somewhat naive attempt to solve the problems of the first method and uses random selection of pixels. As shown in Figure 1(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 fourth method consists in selecting blocks of adjacent pixels at random. The more blocks we have and the smaller they are, then set F will have pixels scattered throughout the image. Contrarily, the larger these pixel blocks are, the fewer blocks we need for a given size of F. Moreover, if we use just a few large blocks, the pixels of set F will be more concentrated in some regions of the image, thus increasing the chance of not hitting the occluded region. Figures 1(e), 1(f) and 1(g) show 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, F contains 70% of occluded pixels in Figure 1(e), only 30% in Figure 1(f) and none in Figure 1(g).

However, the fact that the set of pixels F is concentrated in one region can have a negative effect on Algorithm 1 making it a less stable. Unfortunately there is no efficient way to determine the optimal number of blocks, this is why it is advisable to test with multiple block sizes and quantities. We cope with this effect by selecting multiple such sets $F_1, F_2, \ldots, F_t$ in a Monte Carlo-like approach, to finally select the one that achieves the best results.

Since we have multiple sets of pixels, we apply the Algorithm 1 several times, one for each set of pixels. We already stated that if 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 obtained images $y'_1, \ldots, y'_t$ we select y'_{max} such that $max = \arg\max_i \#P_i$, where $P_i = \{p ~, ~ |y(p) - y'_i(p)| \leq \tau\}$, $\#S$ denotes the cardinal of set S, and τ is a given threshold. Algorithm 2 summarizes the proposed approach.

Algorithm 2. Occlusion detection algorithm

Require: occluded query image y, faces A, different sets of pixels $F_1, F_2, \ldots, F_t$ and threshold τ

1: **for** $i = 1 \ldots t$ **do**
2:     Obtain y_i' using algorithm 1 and set F_i
3:     Let $P_i = \{p , |y(p) - y_i'(p)| \leq \tau\}$
4: **end for**
5: Let $max = \arg\max_i \#P_i$
6: Obtain y_{max}' using Algorithm 1 and fragment $F = P_{max}$
7: **return** u $= |y - y_{max}'| \leq \tau$

---

4   Results and Conclusions

In this section we present tests performed using the recognition algorithm described in Section 2, together with the proposed occlusion detection.

For testing purposes, we use the publicly available AR Database [15]. From this base we selected 60 people. From each individual, 8 non-occluded images were used for the faces dictionary and 4 occluded images were used for test: 2 occluded by sunglasses and 2 by a scarf.

The proposed occlusion detection approach (Algorithm 2) has several parameters. The main ones being the threshold τ, the number of pixels in each set F_i and how these sets are obtained. The sets F_i are selected by taking contiguous pixel blocks at random. We do this in order not to give any previous knowledge about occlusion to the detector.

We start by fixing the threshold $\tau = 20$ and vary the number of blocks. This was done for sets of 500 pixels (6.25%) and 1250 pixels (15.5%) from a total of 8064 pixels in the image. Figure 2(a) shows the obtained recognition rates .

As the number of blocks increases, the recognition rate with scarves begins to decay. This is due to the fact that, having more blocks, the probability that all blocks are in a non-occluded area is much lower. It can also be seen that the increase in the number of blocks does not affect so much images occluded with sunglasses. This phenomenon occurs because images with sunglasses have a lower proportion of occluded pixels than those with scarves and thus the probability of a pixel blocks containing non-occluded pixels is also smaller.

With regard to the number of pixels in each set F_i, we obtain the best and more stable results by using around 1250, that is 15% of the total number of pixels.

The second experiment involves fixing the number and the size of pixel blocks to 2 and 1250 pixels, respectively and then change the threshold τ. We obtained the best results using $\tau = 10$ (see Figure 2(b)).

Figure 3 presents an example of the detected occlusions. The performance is globally satisfactory. However, when analyzing carefully the occlusions maps, we see that there are small non-occluded regions which were detected as occluded and vice versa. Since the recognition algorithm in Section 2 is robust to small occluded areas, these small errors do not affect the recognition rate.

(a)   (b)
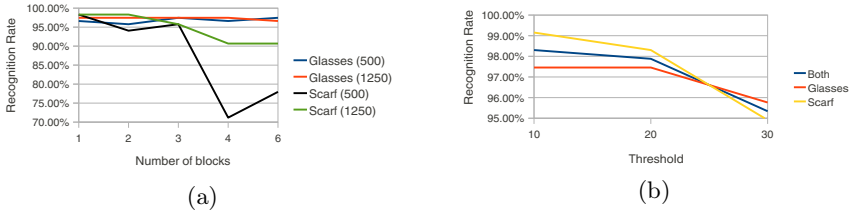
Fig. 2. Recognition rates obtained using occlusion detection. (a) We set $\tau = 20$, and for fragments F_i of 500 and 1250 pixels approximatively, we tested different number of blocks of contiguous pixels. We measured the recognition rate obtained in images with sunglasses and scarves. (b) We use fragments F_i with two blocks and 1250 pixels per block and test for different values of τ.
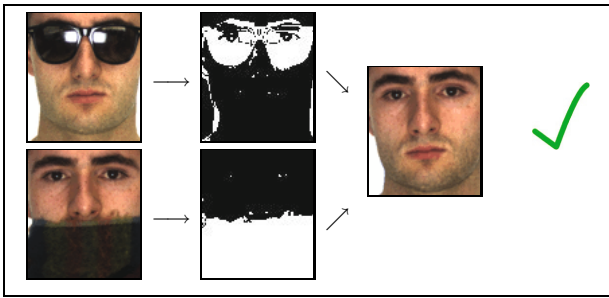
Fig. 3. Recognition in the presence of occlusion. On the left: two images from the same person, with different degrees of occlusion. On the center, the detected occlusions (in white). On the right, the non-occluded image of the recognized person.

Table 1. Recognition rates obtained with different face recognition methods in presence of occlusion

Algorithm	Sunglasses	Scarves	Total
Simple Model [6]	69.49%	12.71%	41.10%
Partitioned Model [6]	86.44%	95.76%	91.10%
Proposed method	**97.46%**	**99.15%**	**98.31%**

Finally, we compare our results with the state-of-the-art solutions proposed by Wright et al. [6]. They propose two different methods to perform face recognition in the presence of occlusion: the Simple and Partitioned Models. We obtain much higher recognition rates for the two different types of occlusion tested.

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 [6]. Experiments have shown that successful recognition is strongly linked to the success of the occlusion detection. Likewise,

one of the factors with a major impact in occlusion detection. Using blocks that are more adapted to the geometry of human faces can be a subject for further research.

References

1. Turk, M., Pentland, A.: Eigenfaces for recognition. Journal of Cognitive Neuroscience 3(1), 71–86 (1991)
2. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. IEEE Transactions on Pattern Analysis and Machine Intelligence 23, 681–685 (2001)
3. Stegmann, M.B., Ersbøll, B.K., Larsen, R.: FAME - A Flexible Appearance Modelling Environment. IEEE Transactions on Medical Imaging 22(10), 1319–1331 (2003)
4. Kahraman, F., Kurt, B., Gokmen, M.: Robust face alignment for illumination and pose invariant face recognition. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1–7 (2007)
5. Belhumeur, P.N., Hespanha, J., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(7), 711–720 (1997)
6. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 31(2), 210–227 (2009)
7. Candès, E.J., Romberg, J.K., Tao, T.: Stable signal recovery from incomplete and inaccurate measurements. Communications on Pure and Applied Mathematics 59(8), 1207–1223 (2006)
8. Candès, E.J., Tao, T.: Near-optimal signal recovery from random projections: Universal encoding strategies? IEEE Transactions on Information Theory 52(12), 5406–5425 (2006)
9. Candès, E.J., Wakin, M.B.: An introduction to compressive sampling. IEEE Signal Processing Magazine 25(2), 21–30 (2008)
10. Lin, D., Tang, X.: Quality-driven face occlusion detection and recovery. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1–7 (2007)
11. Zhou, Z., Wagner, A., Mobahi, H., Wright, J., Ma, Y.: Face recognition with contiguous occlusion using markov random fields. In: International Conference on Computer Vision, ICCV (2009)
12. Basri, R., Jacobs, D.W.: Lambertian reflectance and linear subspaces. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(2), 218–233 (2003)
13. Lee, K.C., Ho, J., Kriegman, D.J.: Acquiring linear subspaces for face recognition under variable lighting. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(5), 684–698 (2005)
14. Donoho, D.L.: For most large underdetermined systems of linear equations the minimal, L1-norm solution is also the sparsest solution. Communications on Pure and Applied Mathematics 59, 797–829 (2004)
15. Martinez, A.M., Benavente, R.: The AR face database. Technical Report 24, CVC (1998)