# Face Recognition Aiding Historical Photographs Indexing Using a Two-stage Training Scheme and an Enhanced Distance Measure

Ana Paula Brandão Lopes[1,2], Camillo Jorge Santos Oliveira[1,3], Arnaldo de Albuquerque Araújo[1]

[1]Computer Science Department – Federal University of Minas Gerais
Av. Antônio Carlos, 6627, Pampulha, CEP 31270–901, Belo Horizonte, MG, Brazil

[2]Exact and Technological Sciences Department – State University of Santa Cruz
Rodovia Ilhéus-Itabuna, km 16 – Pavilhão Jorge Amado, CEP 45600-000, Ilhéus, BA, Brazil

[3]Informatics Department – Pontifical Catholic University of Minas Gerais
Rua Rio Comprido, 4.580 - CEP 32.010-025, Contagem, MG, Brazil,

{paula, camillo, arnaldo}@dcc.ufmg.br

## Abstract

*In this work, we propose a 2D-PCA based face recognizer as a semi-automatic tool for helping indexing people in historical photographs. In the proposed recognizer we cope with the scarcity of training samples and the lack of precision of the detector using a training scheme in two stages. The first stage uses an external face database to compute an average face that is used as a reference either at the second training stage and at the recognition step. We also added an auxiliary distance measure we call* relative distance *to reorder the results generated by the original Euclidean-based distance measure for 2D-PCA. Experimental results with the ORL database as the external face database and a real collection of historical photographs have shown the viability of the proposed tool. These experiments also indicated that both improvements proposed were indeed able to increase recognition rates.*

## 1. Introduction

The digitization of historical and cultural materials has been widely used with preservation purposes. Also, the availability of digital copies has the potential to increase the access to a wider audience. But in order to turn this increased access into reality, all the digitized material need to be properly indexed, what normally results in a huge amount of manual work for historians and other related professionals.

Among the variety of these materials that can be digitized are historical photographs, and a particular item which is obviously worth indexing in that case is the people who appear in these photographs. In a typical setting, this is done manually by a historian with the aid of the associated textual annotations, when they exist, and visual inspection. In this paper, it is argued that face recognition (FR) techniques can be used to alleviate the workload on these professionals, by helping them in finding additional occurrences of selected people inside their photographs databases. In order to validate this idea, a face recognition approach was developed and applied to a collection of 1003 faces extracted from a real historical database under custody of the Minas Gerais State Public Archive[1]

In the case of historical photographs all the well-known issues that normally cause trouble to face recognition algorithms are present: varying poses, facial expressions and lighting conditions, presence of artifacts, partial occlusion and aging. Added to this is the poor quality of most original photographs. Also, some of them can be degraded by the effects of time or bad storage conditions. Figure 1 shows examples of some of these problems found on real faces extracted from the APM photographs collection.

An additional and important limitation of this kind of image set comes from the fact that typically there are few occurrences of each subject in the entire base. Actually, in most cases there is only one occurrence of each individual. It is well-known that this fact can seriously degrade the

1 Arquivo Público Mineiro (APM), http://www.cultura.mg.gov.br/?task=home&sec=5

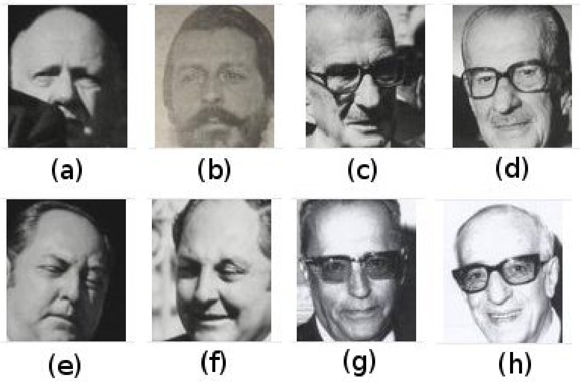


**Figure 1. Faces extracted from the APM photograph database: (a) shows the presence of occlusion, (b) shows a very bad quality image and an artifact (the beard), (c) and (d) are from the same person under different lighting conditions, (e) and (f) show the same subject under varied pose and expression, (g) and (h) show the aging effect.**

recognition rates of most face recognition algorithms, and in some cases entirely prevent their application.

In order to detect people's faces in the APM's photographs, the detector of [6] was applied and its outcomes resulted in another issue to be dealt with: the positioning of the detected faces inside the cropping area returned by the detector were not very precise, making the recognition task even more challenging. Figure 2 illustrates this problem.

**Figure 2. The issue of detection positioning: the first face is placed at the left side of the cropping area, the second one is centered and the last one is at the right side.**

In this work we propose a 2D-PCA [16] based face recognizer in which the training step is comprised of two different stages, in order to cope mainly with these two issues previously described: the lack of enough training samples, and the lack of precision of the detector in finding face positions.

The remaining text is organized as follows: Section 2 summarizes the background for this work; Section 3 details the proposed face recognizer; the experimental results are shown and discussed in Section 4; finally, some conclusions are drawn in Section 5.

## 2. Background

In [4], it is suggested that appearance based solutions are the most suitable in the setting described in the previous section. According to [1], the first appearance based FR method proposed was the Principal Component Analysis (PCA) method or Eigenfaces [13]. The PCA algorithm turns the image matrixes into linear vectors and then computes the covariance matrix for the vectors taken from the images in the training sample. This way, each individual face is represented exactly by a linear combination of the eigenfaces of this covariance matrix.

The features effectively used to describe the faces are selected by choosing the eigenfaces related to the greatest eigenvalues of the covariance matrix, which account for the most variance within the set of images. The underlying assumption is that the greatest eigenvalues indicate de most discriminative features in the modified space. Once the images are described this way, they are then compared with each other by computing the Euclidean distance between the transformed image vectors.

A major drawback of PCA based approaches is that they do not use previous class information (where each subject belongs to a class). To overcome this, a Linear Discriminant Analysis (LDA) [5] based method was proposed, but it is reported that it only achieves better performance than PCA when a great number of training samples is available. Both PCA and LDA gave rise to several variations applied to particular situations and have become common baselines for evaluating subsequent methods.

Some authors have suggested the usage of 3D models to deal with particular issues, as pose and illumination variations [8] or to enlarge the training base by the synthesis of new views from the combination of a generic 3D model with the real samples[2] [10]. The idea of synthesizing new views from a single example so to be able to use a classical FR technique is also explored in [14], whose synthesis method avoids the need for a 3D model. Such techniques, though, normally rely upon the precise positioning between the original face and the model, limiting their applications only to relatively small data sets.

A remarkable variation of PCA is called the two-dimensional PCA (2D-PCA) [16], which avoids the transformation of the image matrix into a linear vector by computing the covariance matrix directly from the matrix representing the image (gray level). Experiments show that 2D-PCA achieves better recognition rates, and has a lower time complexity in feature extraction. Yet, 2D-PCA is shown to be better at dealing with the few examples per subject issue.

A general framework to create a FR system when there is only one sample training per subject was proposed by [15]. In their work it is argued that any appearance based algorithm can be trained using a generic face database. The premise behind this proposal is that human faces, even when taken from different contexts, share many similarities. So, in their proposal the reduced feature space is learned from the external data set instead of from a subset of the target face data set. Provided the generic data base has enough training samples per subject, this overcomes the original problem regarding the number of training samples. After that, the target image base (the gallery) and the probe image are projected onto this space, and then a nearest neighbor classifier determines the identity of the probe.

The authors performed extensive experimental evaluations of this framework with a number of public face databases and FR methods, including the original PCA and some of its variations. Their results have shown that when there is only one training sample, their framework is indeed able to produce better results.

Finally, a relatively simple idea to mitigate the problem of the scarcity of training samples is to add the mirror images to the training set, as suggested in [17] and [7] *apud* [12].

In the next section, it is described how some of these techniques are mixed together and enhanced in the building of our face recognizer.

## 3. A face recognition approach for aiding people indexing

As it was said in the previous section, the 2D-PCA algorithm was chosen as the basis for the face recognizer proposed here. Its implementation was based on that of [3], modified to set the training and the projections steps apart from each other.

### 3.1. A two-stage training approach

In order to use this recognizer at the APM photographs database, a needed pre-processing step is to extract the faces from the photographs. This was made by applying the detector described in [6]. A first pass of these extracted faces through the detector has shown that head position and artifacts like hats and moustaches seemed to dominate the recognition over facial characteristics. In order to mitigate the effect of artifacts, a smaller cropping area was used, as illustrated in Figure 3.

**Figure 3. The first face is the originally detected, and the second one is cropped for recognition.**

To deal with the head positioning problem, the proposed algorithm computes a number of slightly different 'views' from every face, beginning with the position returned by the detector. Figure 4 shows the 'views' returned for one image as an example. One can observe that the same variations are computed for the mirror image, as suggested in [17] and [7]. This leads to 12 versions of each face. The two-stage training scheme which it is being proposed in this paper deals with the issue of choosing the best of these versions.

**Figure 4. All versions created from varying the position for extracting the face.**

The overall training scheme is summarized in Figure 5. In the first training stage, the entire publicly available Olivetti Research Laboratory (ORL) database [11] is used as the external training base. In the proposal described in [15], the transformation matrix returned by this training stage was directly used to project the gallery and the probe onto the modified space. In this work, however, the result of this first training stage is used to compute an "average face" in the 2D-PCA modified space. This average face is then used to choose the best versions among those previously computed from the detector output. These selected best versions are the closest ones to the average face.

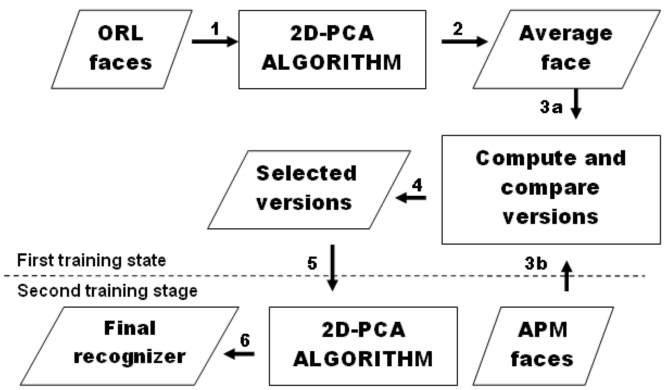


**Figure 5. The two stages of the proposed recognition scheme.**

The best versions are used in two different ways. Firstly, during the second stage of training, when the best version of each face of the database is used as the "real" training sample from which an entirely new recognizer was trained.

The best versions are also used in the recognizing step, as follows: for a given probe image, its $v$ best versions are compared to the entire search face base. The smallest of the $v$ distances computed for each comparing face is selected as the real distance between the probe and that comparing face. After that, the whole search face base is ordered by these selected distances.

### 3.2. Reordering the results by relative distances

The distance measure used so far was the same proposed by [16], which is the Euclidean distance adapted to a 2D matrix instead of a one-dimensional vector. However, it was observed that, in a number of times, given a pair of images of the same subject, $i_1$ and $i_2$, although $i_2$ appeared in the n-closest list of $i_1$ and $i_1$ appeared in the n-closest list of $i_2$, they were not *the* closest ones to each other. In other words, the recognition failed, even though it was very close to succeed. Also, most times, the other $n-1$ faces returned as the closest to $i_1$ and $i_2$ were not necessarily common between both lists. In other words, there were some images around $i_1$ at nearly the same distance as $i_2$, but in other directions of the 2D-PCA defined space, the same happening to $i_2$. This 'noise' eventually prevented the algorithm to return the correct answer. This situation is illustrated by Figure 6.

In order to use this information, the first $n$ images returned by the classical 2D-PCA distance measure are reordered using what we call the *relative distance*, as follows:

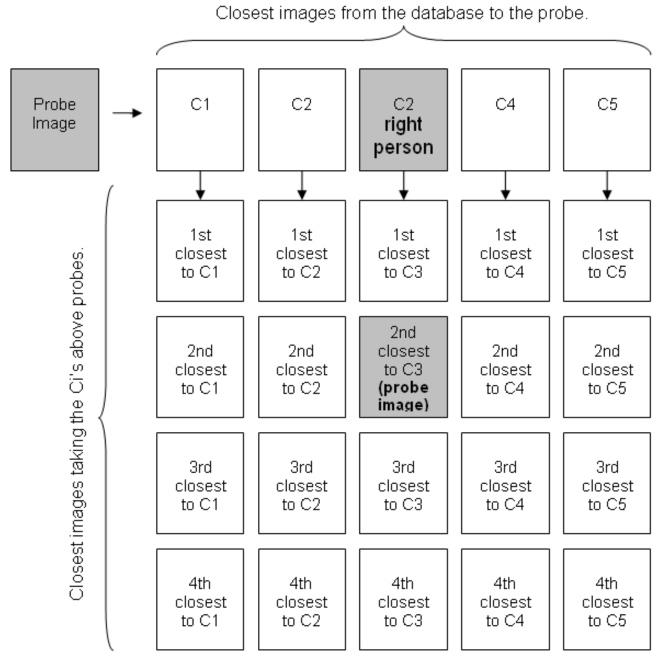


**Figure 6. Illustrates the situation tackled by the distance algorithm.**

- The initial relative distance for the first n-closest images of a particular probe image is given by its relative position in the list. In Figure 6, $C1$ would have initial relative distance equals to 1, $C2$ would have initial relative distance equals to 2, and so on.
- For each of the n-closest images in the list, their own n-closest are computed, using the $i_{th}$ returned image as a new probe.
- For each list of closest images, the probe image is looked for. If the probe is found at list $i$, the relative distance of the $i_{th}$ image to the probe is added by the position where it was found. Again, considering Figure 6, the relative distance of $C3$ would be added by 2, since the probe image was found in the second position of the closest images taking $C3$ as the probe.
- Otherwise, if the initial probe is not found, the relative distance of the $i_{th}$ image to the probe is added by a large value. This way, the images which appear in the closest list of each other have smaller relative distances, while the others have greater relative distances.
- After doing this to all $n$ images of the probe closest list, the images returned as the closest to the original probe are reordered by their relative distances.

This procedure can be viewed as adding some direction information, which is lost with the Euclidean distance. Algorithm 1 summarizes the relative distance computation.

**Algorithm 1** Computing relative distances and reordering the closest faces.

$closestFaces \leftarrow getClosest(probeImage)$
**for** $i = 1$ to $size(closestFaces)$ **do**
  $item \leftarrow closestFaces(i)$
  $relatDist(i) \leftarrow i$
  $itemClosests \leftarrow getClosest(item)$
  $foundPos \leftarrow find(probeImage, itemClosests)$
  **if** $foundPos\ is\ valid$ **then**
    $relatDist(item) \leftarrow relatDist(item) + foundPos$
  **else**
    $relatDist(item) \leftarrow relatDist(item) + greatN$
  **end if**
**end for**
$reorderByRelatDist(closestFaces)$

In the next section some experiments performed with this face recognition scheme are described and discussed, showing that either the proposed training scheme and the relative distances reordering can lead to better recognition results in the APM photographs.

## 4. Experimental Results and Discussion

In order to assure that the 2D-PCA implementation was in accordance with that of the original 2D-PCA article, it was applied to the ORL face database [11]. This database is comprised of 400 images, with 10 images of 40 different subjects. The experiment described in [16] was replicated: the first 5 images of each subject were used as the training set, the other 5 then made up the testing set. A recognition rate of 93.4% was achieved, matching the published result of 96% within a confidence interval of 95%.

The APM photographs collection under study is made up of over 6300 images already digitized. From these, roughly 50% contain people. The detector described in [6] was applied to these images. The false-positives and those images which were smaller than the ORL dimensions were manually removed. The remaining 1003 images were then scaled down to the ORL size ($112 \times 92$) pixels. Finally, visual inspection was used to identify those images having at least one counterpart in the faces collection. This inspection resulted in a manual classification of 106 images from 32 individuals. The remaining 887 images were considered as unique occurrences of different subjects. All images were converted to gray scale, and then enhanced by contrast expansion and Gaussian smoothing.

In order to evaluate the effectiveness of the proposed method, two measures were computed: the recognition rate when taking into account only the first nearest neighbor in the 2D-PCA space (top-1 recognition) and the recognition rate among the five nearest neighbors (top-5 recognition). These last ones could be considered as the suggestions that the system would offer to the user in the real scenario.

The number of versions used for the recognizing step is 4, the number of closest images used in the relative distance computation is 10 and the number of 2D-PCA components used is 10.

### 4.1. The role of different factors in recognition rate

In this paper, two main proposals are made: a) a two-stage training approach, and b) the reordering of the closest images by the relative distance. In order to evaluate if these two propositions indeed increased the recognition rate, a full $2^k$ factorial experimental project was performed [9].

In this experimental setting, the factors being studied are varied between two levels each, resulting in $2^k$ runnings, where $k$ is the number of factors. The goal here is to evaluate the role of each factor on the recognition rate variations. Four our purposes, the factors studied were the training approach (one or two stages), the relative distance (using it or not) and the size of the search base (530 and 636 images). This last factor was introduced in order to contrast the variations due to the algorithmic choices against the variation due to the search base size (which is further analyzed in next subsection).

The results of these experiments are shown in Table 1. For the training approach factor, the one-stage training using ORL base was considered the lowest level ($-1$) and the proposed two-stage approach was considered the highest level ($+1$). For the relative distance factor, not using it was the lowest level, and using it was the highest level. Finally, base size of 530 images was the lowest level and a base size of 536 images was the highest level for the base size factor. Within this convention, since the coefficients of the computed model were positive both for the training approach factor and for the relative distance factor, it can be concluded that going to the highest levels increased the recognition rate for these factors. In other words, the experimental setting indicates that using the two-stage training approach and the relative distance reordering indeed enhanced the recognition rates.

### 4.2. Recognition rates using search bases of different sizes

From the previous experimental setting it was observed that the total amount of images in the search base influenced

| $2^k$ Factorial Experimental Results | | | | |
|---|---|---|---|---|
| **Exp.** | **D** | **S** | **T** | **Rec.** |
| **1** | 1 | 1 | 1 | 39.6% |
| **2** | 1 | 1 | -1 | 35.5% |
| **3** | 1 | -1 | 1 | 39.3% |
| **4** | 1 | -1 | -1 | 38.3% |
| **5** | -1 | 1 | 1 | 33.6% |
| **6** | -1 | 1 | -1 | 34.6% |
| **7** | -1 | -1 | 1 | 36.5% |
| **8** | -1 | -1 | -1 | 36.5% |
| **Var.** | 53.1% | 21.4% | 6.7% | - |
| **Coeff.** | 1.44 | -0.91 | 0.51 | - |

**Table 1. The results for the $2^k$ factorial experimental setting. 'D' stands for the distance measure ($-1$ is the pure Euclidean, $1$ is using the relative distance reordering), 'S' stands for the search base size ($-1$ is a base with $530$ images and $1$ is with $636$) and 'T' stands for the training scheme ($-1$ means using only one stage with ORL base and $1$ means using our training scheme). 'Var' is the amount of variation in the recognition rates that can be attributed to each factor. The interaction between D and T is responsible for $14.9\%$ of the remaining variation. 'Coeff' are the coefficients of the computed model. Positive coefficients indicate that the recognition increases when the factor goes from the lowest to the highest level. Negative ones indicate decreasing rates.**

the recognition rates. In order to better estimate this effect, the top-1 and top-5 rates were computed with nine different search base sizes. The search bases were created as follows: the first one was made up only of the $106$ manually classified images; the second search base was built up from adding another group of $106$ images to the previous one. These last images were randomly selected from the remaining $887$ images from the original base. The third search base was composed by the original $106$ classified images added by another group of $(2 \times 106 = 212)$ images, and so on. Figure 7, shows the top-1 and top-5 recognition rates, where the number of unclassified images added to the search base is indicated in multiples of $106$.

It can be seen from this graph that the proposed algorithm is able to give ***at least*** one good suggestion to the user in about $50\%$ of times, if $5$ suggestions are considered or $40\%$ if only the first suggestion is taken into account.

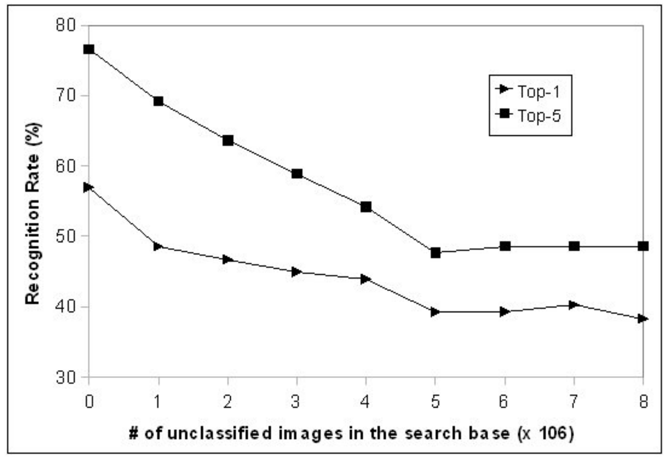


**Figure 7. Recognition rates for varied search base sizes. The horizontal scale ($x$) shows the number of images added to the original pre-classified search base, in multiples of $106$. As an example, for $x = 2$, the search base has $106 + (2 \times 106) = 318$ images.**

These are the values where the recognition rates seem to stabilize. Also, in a small search base, it can achieve a top-5 recognition rate of $76.6\%$.

Figure 8 shows some of the recognized images using a search base made up of $530$ images (point $4$ of the Figure 7). It can be observed that the proposed recognizer was able to cope with some challenging situations.

Finally, Figure 9 shows some top-5 results in this same search base, showing which suggestions would be given to the historian in each case.

These results indicate the viability of the initial proposal of this work, which was to use a FR system in aiding the annotation of historical photographs. They also suggest that the final system should use natural constraints for better searches. Examples of such constraints should be: searching inside a particular collection, searching only among the already-annotated images, or among the not-yet-annotated images, searching by date of annotation, and so on. Then, by using constraints like these, the system is much more likely to produce correct suggestions for the professional in charge of annotating the photographs.

## 5. Conclusion

This work was motivated by the idea of using face recognition techniques for aiding the indexing of historical photographs, by offering suggestions of possible other occurrences of the same person in the database.

**Figure 8. Some examples of correctly detected faces when each face is compared to a search base of $530$ images. In each pair of faces, the first one is the probe image, the second one is the closest image found to that probe.**

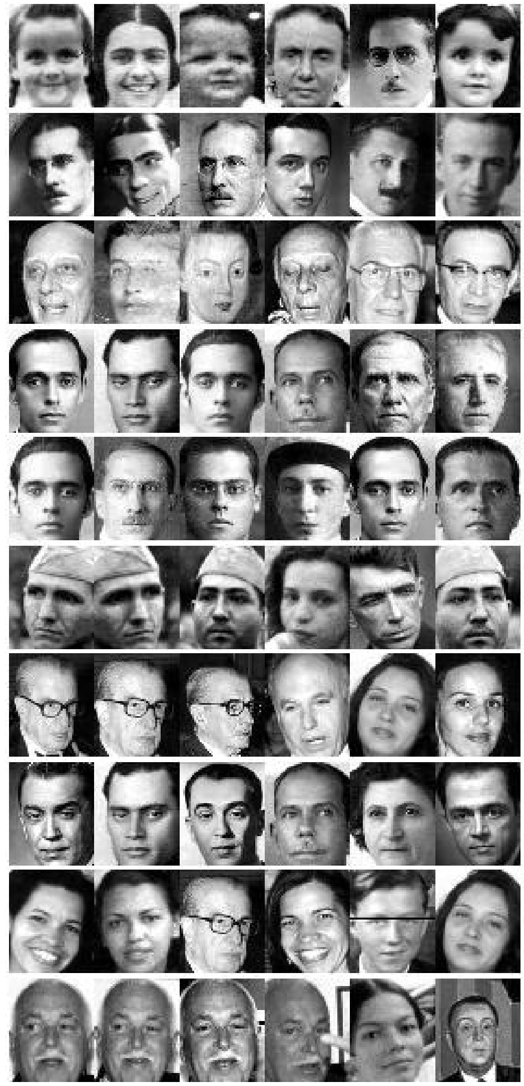

**Figure 9. Some top-5 results. In each line, the first face is the probe image and the next ones are the five closest to that probe.**

Face recognition is an open problem in the general case, but a number of applications benefit from working on images taken under relatively controlled conditions, carefully adjusted for the application purposes. From this point of view, the APM collection is a very challenging one, since it exhibits all characteristics that traditionally cause existing face recognition methods to fail, added to poor original image quality and an imprecise face detection.

Regardless of the great number of obstacles, the results have shown that it is possible to produce a reasonable number of correct suggestions to the annotators in the environment of a semi-automated tool, provided that the professionals are able to direct their searches toward some wisely constrained subset of the photographs.

Additionally, two propositions were made to improve the recognition rates: the first one was a training scheme in two stages, where an external database was used as a reference to adjust the real face database. The second proposition was the reordering of the closest images by the relative distance, computed by the inspection of the closest lists of the first returned images. The experimental results have shown that both propositions were able to increase the recognition rates relatively to the original schemes found in previous works.

An intended improvement of the final system is the establishment of a distance threshold to avoid too distant false positives. This way, we hope that the candidates of new occurrences of a person presented to the user will seem more reasonable. Another direction intended for further investigation is the validation of our algorithm against other databases, either standardized or historical ones.

## 6. Acknowledgments

The authors are grateful to CNPq and CAPES, Brazilian research funding agencies, for the financial support to this work.

## References

[1] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino. 2d and 3d face recognition: A survey. *Pattern Recogn. Lett.*, 28(14):1885–1906, 2007.

[2] V. Blanz and T. Vetter. Face recognition based on fitting a 3d morphable model. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(9):1063–1074, 2003.

[3] M. J. Brown. Singular value decomposition and 2d Principal Component Analysis of iris-biometrics for automatic human identification. Master's thesis, Russ College of Engineering and Technology at Ohio University, June 2006.

[4] R. Brunelli and T. Poggio. Face recognition: Features versus templates. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15(10):1042–1052, 1993.

[5] K. Etemad and R. Chellappa. Discriminant analysis for recognition of human face images (invited paper). In *AVBPA '97: Proceedings of the First International Conference on Audio- and Video-Based Biometric Person Authentication*, pages 127–142, London, UK, 1997. Springer-Verlag.

[6] D. L. Flam, C. J. dos Santos Oliveira, and A. de Albuquerque Araújo. A computer-aided system for indexing people in historical images. In *Proceedings of SIBGRAPI 2007 Posters*, 2007.

[7] S. Gutta and H. Wechsler. Face recognition using asymmetric faces. *Lecture Notes in Computer Science*, 3072:162–168, 2004.

[8] Y. Hu, D. Jiang, S. Yan, L. Zhang, and H. Zhang. Automatic 3d reconstruction for face recognition. pages 843–848, 2004.

[9] R. Jain. *The Art of Computer Systems Performance Analysis*. John Wiley and Sons, Inc., 1992.

[10] X. Lu, D. Colbry, and A. K. Jain. Three-dimensional model based face recognition. In *ICPR '04: Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04) Volume 1*, pages 362–366, Washington, DC, USA, 2004. IEEE Computer Society.

[11] F. Samaria and A. Harter. Parameterisation of a Stochastic Model for human Face Identification. In *IEEE Workshop on Applications of Computer Vision*, Sarasota (Florida), December 1994.

[12] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang. Face recognition from a single image per person: A survey. *Pattern Recogn.*, 39(9):1725–1745, 2006.

[13] M. Turk and A. Pentland. Face recognition using eigenfaces. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Maui, Hawaii, USA, 1991.

[14] T. Vetter. Synthesis of novel views from a single face image. Technical Report 26, Max–Planck Institut, Tübingen, Germany, February 1996.

[15] J. Wang, K. N. Plataniotis, J. Lu, and A. N. Venetsanopoulos. On solving the face recognition problem with one training sample per subject. *Pattern Recogn.*, 39(9):1746–1762, 2006.

[16] J. Yang, D. Zhang, A. F. Frangi, and J. yu Yang. Two-dimensional pca: A new approach to appearance-based face representation and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(1):131–137, 2004.

[17] Q. Yang and X. Ding. Symmetrical PCA in face recognition. In *International Conference on Image Processing (2)*, pages 97–100, 2002.