Reliability of personal identification base on optical 3D measurement of a few facial landmarks

Emanuele Zappa, Paolo Mazzoleni*

* Politecnico di Politecnico di Milano, Dipartimento di Meccanica, via La Masa 34, 20156 Milan, Italy

Abstract

Technologies related to identity recognition have found widespread application in the last years; among these, face recognition is one of the most promising and probably the most studied. The main goal of this work is to verify if it is possible to obtain a reliable identification of people, starting from a few 3D measurements of the face shape. To this purpose, an algorithm of face recognition based only on the 3D localization of some significant points of the face, starting from a pair of stereoscopic photographs, was developed and applied on a 51 people database. The localization of these landmarks is carried out manually: the goal of the work is the evaluation of the reliability of a few-point based algorithm, not its automation. The obtained results show a recognition percentage grater than 90%, an encouraging outcome for a preliminary work.

Keywords: forensic science; face; biometric; identification; recognition; 3D;

1. Introduction

Nowadays capacity of recognizing the identity of a person is becoming increasingly important in more and more areas and the different technologies related to this issue are finding widespread applications. The most common uses are in the forensic, legal and security fields, but also medicine, biometric sciences, computer graphic and many more trivial applications are approaching this aspect.

Many researches have been carried on in the identity recognition field and they have brought to the development of different technologies. A review of all the available biometric techniques is present in [1] and [2], while detailed analysis of the most popular ones can be found for example in [3] for fingerprints and [4] for face-based recognition. The latter is currently one of the most studied and the main topic of this work.
Two different approaches of face recognition are present in literature [5]: the so called “global” codes (top down approach), based on the investigation of the whole face at the same time, as a single unit, and “feature base” algorithms (bottom up approach), where the independent localization and analysis of the different features of the face is the first step of the recognition. This last approach is adopted in this work.

The main goal of this research is to verify if it is possible to obtain a reliable identification of people, based only on the 3D localization of some significant points of the face, manually located in a pair of stereoscopic photographs. Preliminary works in this field [6] showed encouraging results.

As mentioned, the recognition is done considering the position of some specific points of the face, called “repere points” or “landmarks”, easy to be found and characterized by the fact that their relative position is almost fix during the life of a person [7]. That is due to the fact that these points are supported by bones or parts of cartilage particularly stable: it means that the positions of these zones in a person are almost independent from the age, the haircut, the presence of beard and moustache and so on.

Biometric techniques are generally based on the extraction of biometric measurements from the face of a subject to be recognized and the successive comparison of these measures with corresponding ones contained in a reference database. These measures can be 2D (drawn from photographs), or 3D (three-dimensional reliefs of the face, made for example with rangefinder scanners or stereoscopic systems), as described in [8]. It’s important to notice that sensors that requires the person to remain motionless for several seconds are in general unusable in a real environment [9], despite the high accuracy achievable.

There is also mixed 2D/3D approaches, where information obtained from 3D scans of faces are projected onto 2D images (see for example [9] and [10]). In this paper the applied technique is 3D/3D type: measured, using stereoscopy, the 3D positions of a set of landmarks of the face, the recognition is done comparing this set of 3D points against a collection of 3D masks belonging to a database.

The choice of adopting a 3D acquisition is linked to the more potentiality of such an approach: using a stereoscopic camera does not only mean to double the information grabbing two images at the same time, but gives the possibility to triangulate the positions of the features and reconstruct their 3D coordinates in the space, analogously to what happens with the binocular human system [11]. The 3D data measured with the stereoscopic system are independent from the original position of the individual and this allows to work correctly also when the subject’s head is not exactly frontal: this is very important thinking to the impossibility to grab all the images with exactly the same angle and at the same distance from the cameras [12]. Furthermore, the use of a stereoscopic approach allows overcoming an intrinsic limit of 2D-based recognition algorithms: the incapability to distinguish between a real person and a picture of him placed in front of the camera.

2. Acquisition and Location of Repere Points

2.1. Acquisition System Design

The acquisition system is basically composed by a frame that supports a pair of digital cameras; the person to be recognized sits in front of the cameras, as shown in Fig. 1.

![Stereoscopic acquisition system](image_url)
The characteristics of the sensors (dimensions and resolution) and of the optics (focal length) of the cameras and their mutual position (distance and angles) are parameters that have to be tuned in order to optimize the system mainly in terms of measurement volume and corresponding uncertainty.

A software able to simulate the stereoscopic system and to numerically reproduce the image formation process has been developed. This algorithm has been used in order to simulate all the possible combinations of the geometric parameters in order to determine the configuration able guarantee the best resolution of the measurements ensuring a sufficient vision volume.

The result of the optimization process is the choice of two identical Allied Marlin F-131B cameras, equipped with a Sony Type 2/3 inches progressive scan CCD sensors (1024 x 1280 px). The selected cameras are grayscale ones, in order to avoid the interpolation procedure linked to the use of the Bayer filter, which is usually adopted in color sensors [13]. With reference to Fig. 1, the following values for the geometrical parameters have been found: \( \alpha = 70^\circ, f = 18 \text{ mm}, L = 1000 \text{ mm} \).

Note that, in order to optimize the working volume of the stereoscopic system for the specific application of face measurements, the cameras were angled of 90\(^\circ\) around the optical axes (i.e. with the longest side of the sensor in the vertical direction).

Another key parameter in the image acquisition process is the illumination: the whole face has to be correctly enlightened, so that neither the presence of shadows could make difficult to recognize some repere points nor peaks of brightness would saturate the sensors. A lamp of 500 W, screened with a diffusive surface, is positioned between the two cameras, exactly in front of the person and provides the main lighting, while other two dimmer lights prevent the formation of shades (Fig. 1). Just behind the person a black curtain has been put with the goal of making uniform the background and reducing possible reflections (Fig. 1).

2.2. Calibration of the Stereoscopic System

Using a stereoscopic vision system, the measurement of the 3D position of points is basically done triangulating every acquired point so that it is possible, known the geometry of the system, to trace his position in the space.

The calibration of the system consist in the determination of the parameters necessary to pass from the 2D coordinates of a point in both the sensors to a 3D coordinates in a real world reference system. The calibration procedure adopted in this work is the so called Direct Linear Transform (DLT) [11,14,15].

This procedure is based on the knowledge of the 3D coordinates of a set of calibration points and the images of these points on the two cameras. Providing that the points are distributed in the whole working volume, it is possible to estimate the full set of calibration parameters, including the ones needed to compensate the optical aberration.

3. Acquisition and Location of Repere Points

3.1. Acquisition

The acquisition through the stereoscopic system is based on a software developed in Microsoft Visual Basic®. The code is able to visualize the image framed by each camera in real time and to acquire in series the desired number of image pairs. The two cameras are synchronized using a trigger signal that allows to get the two samples at the same time.

For every volunteer the acquisition has been done in two different poses, grabbing 10 pairs of images in each posture, in order to create a database. A total of 51 testers has been scanned. For every tester an additional pair of images has been sampled in a third pose: these data are the ones that will be compared with the ones in the database during the recognition.

The volunteers were asked to remain in a natural pose, without wincing or smiling.
3.2. Location of the Landmarks

Biometric literature [7,16] reports a huge number of repere points spread in the whole head of an individual (Fig. 2).

Fig. 2 Repere points in a head

The problem is to select the most significant ones for the purpose of face-based identification: for example all the points not belonging to the face (1, 2-3, 4-5) can be easily hide by the hair and their localization could be impossible. The front view, or one slightly angled, as in our system, is the one in which a higher number of landmarks can be found. Some marker points (e.g. 35 and 6-7) are subjected to a wide movement as the expression of the face changes and so should be discarded; others (e.g. 21, 22-23, 9-10, 24-25 and 26) could be difficult to be located without touching the face. Furthermore the vertical limits of the eyes (15-16 and 17-18) are strictly linked to how much the eyes are opened and for this reason should not be included in the analysis. It is however important that the chosen repere points would be scattered in the whole face and not concentrated in a single zone, in order to make the recognition more reliable. Finally, to obtain the 3D position of the points it is fundamental that they are contemporary visible from the two cameras (for this reason we discard for example points 13-14).

After all these considerations, the points shown in Fig. 3 have been chosen.

| n° | name             |
|----|------------------|
| 1  | Exocathion (right) |
| 2  | Endocathion (right) |
| 3  | Endocathion (left) |
| 4  | Exocathion (left) |
| 5  | Alare (right)    |
| 6  | Subalare         |
| 7  | Alare (left)     |
| 8  | Chelion (right)  |
| 9  | Chelion (left)   |

Fig. 3 Example of manual selection of the nine chosen repere points in a stereoscopic image
To analyze the images, a software was developed in Matlab® to show one by one on the screen the pairs of acquired pictures; for each image the user is asked to zoom in the different zones and to manually select the position of the landmarks (Fig. 3).

The 3D position of every point is reconstructed through triangulation, using the DLT algorithm. The 3D points obtained in this way allow to generate a 3D “mask”.

At the end of this process we obtain 20 3D “masks” for each of the 51 volunteers, for a total of 1020 3D masks.

At this stage of the research it is not important to develop a code able to automatically locate the repere points, but instead to try to understand if it make sense to do it; in other words it will be showed if the 3D localization of some marker points is a firm basis to develop a face recognition algorithm.

For the same reason, at this time is not so important the exact number of the selected repere points; if necessary some others can be add in future. The goal of this research is the evaluation of the reliability of a few-point based algorithm in comparison, for example, with a 3D scan, where thousands of points have to be analyzed for a single face.

3.3. Differences Among Users

The main aspect to be consider in tests that request the contribute of the user is how much the results are related to the experimenter. Some investigations have been done in order to verify if different people can obtain sensible dissimilar results. All the tests will not be presented in this article, but important remarks have to be underlined. From the experiments, the system appears to be sufficiently user-independent, different skilled users obtained satisfactory similar results. The main problem arises with non-expert users: a training period of a few hours, in which non-expert users were asked to annotate images of 10 people (i.e. 400 picture because of repeated image acquisition for each person) in which the 9 points have to be selected, appears to be appropriated in order to reach a sufficient repeatability in the selection of the landmarks.

4. Face Recognition Algorithm

Providing that the repere points have been previously selected, the personal recognition process requires three main tasks:

- to create a database of 3D biometric information starting from the coordinates of repere points in the stereoscopic images (in our tests the database contains 51 European people, 20 to 30 years old);
- to find a method to compare a new sample with all the ones preset in a database;
- to define a threshold parameter which defines if the comparison gives positive or negative response. This parameter is derived only from the database and it is independent from the new sample tested.

For every scanned person we generated 20 3D masks, one from every pair of images. Using 20 masks and averaging allows reducing the random component of the measuring uncertainty, mainly linked to the manual selection of the features.

Let us consider now a single volunteer. Every pair of images allows generating a 3D mask, so we have 20 3D masks. For every mask we create a matrix containing the distances between all the points (Table 1).

### Table 1 3D Distances matrix in millimeters

| Person n, mask m | P1 | P2 | … | P9 |
|------------------|----|----|---|----|
| P1               | 0  | \(d_{12}\) | \(d_{13}\) | \(d_{19}\) |
| P2               | \(d_{21}\) | 0  | \(d_{23}\) | \(d_{29}\) |
| …                | \(d_{\ldots1}\) | \(d_{\ldots2}\) | 0  | \(d_{\ldots9}\) |
| P9               | \(d_{91}\) | \(d_{92}\) | \(d_{93}\) | 0  |
So we obtain 20 “distances matrixes” for each person in the database. Obviously all the matrices are symmetric (\(d_{ij} = \text{distance between point } i \text{ and point } j \text{ is the same as } d_{ji}\)) and all the terms in the main diagonal are equal to zero (distance between a point and itself).

Averaging term by term these 20 matrices we obtain an “average distances matrix”. If we now subtract to every distances matrix the average distances matrix we obtain 20 “delta matrices”, containing the differences between every distance and its mean value. Averaging all the terms of every delta matrix (excluding the diagonal values, which are zero by definition) we obtain 20 values, called “average delta”; the mean value of these 20 terms is defined “personal average delta”, \(\overline{\delta}\), while their maximum “personal max delta”, \(\delta_M\) (Fig. 4).

![Fig. 4 Personal average delta and personal max delta](image)

Considering now all the volunteers, we get 51 personal average deltas \(\overline{\delta}_i\) and 51 personal max deltas \(\delta_{M,i}\) where \(i=1\ldots51\) represents the index of the person; both the maximum of the \(\overline{\delta}_i\) and \(\delta_{M,i}\) values in the database can be chosen as the threshold (Fig. 5): we suppose that no one could have the comparison parameter against himself higher than the threshold value and, at the opposite, every control versus a different person will never give a lower response respect to the same threshold. At this step we do not define a unique threshold value, but we keep both of them. In the next chapter the choice of the threshold will be formalized.

![Fig. 5 Personal average delta \(\overline{\delta}_i\) and personal max delta \(\delta_{M,i}\): thresholds](image)

Note that in an ideal case, in which the landmarks are perfectly fix and there are no errors in their manual selection, all the distances matrices coincide with the mean one, so the 51 personal average and max deltas are identically equal to zero.

After the definition of these parameters we are no more interested in all the collected data: only the average distances matrix is stored for every single volunteer. So our database is composed by 51 matrices and two different threshold values.

We have now to define the comparison parameter (Fig. 6).
If we want to compare a new person (X) with one in the database (Y), we calculate his distances matrix, as explained before. Subtracting to this matrix the mean one of a person present in the database, we obtain a delta matrix: the value obtained averaging all the terms of this matrix has been chosen as the parameter used in the comparison. In case we have more acquisition of the new person, we can use, instead of a single distances matrix, the average one.

So the recognition operates as follow: we have a database, a threshold derived from it and a new sample. Comparing one by one the new distances matrix with all the ones in the database, we obtain 51 comparison parameters. If the comparison parameter is higher than the threshold the two faces are expected to belong to two different people while if it is lower they are likely to be the same person.

5. Results

The previously described algorithm has been exploited in order to actuate the real recognition.

Let us consider once again the single volunteer. A new pair of images (not used for the database creation) is loaded and the user is asked to select the landmarks. At this point the software computes the 3D position of the points, using the mentioned DLT method. After the reconstruction of the 3D mask, the algorithm actuate the recognition with the method explained before, comparing one by one the biometric data in the database and the new set. First of all the code scan all the database to find the most resembling mask, that is the one which minimize the comparison average delta.

Then the algorithm states that the most resembling mask in the database and the new mask belong to the same person if the comparison parameter is lower than the threshold.

In the “face recognition algorithm” paragraph, two different values has been indicated as possible threshold; the “max” method obviously gives a higher value as threshold with respect to the “average” one. If we select a too low value, we increases have a larger probability to run into what is known, in statistic, as “type 1 error”, i.e. the risk to refuse the result of the recognition even if it is true. At the opposite a too high threshold imply that we will always identify the tested person as the most similar in the database, even when it is not true (type “2 error”). In this work the optimum value has been found, with a trial and error iterative procedure, between the two boundaries (max and average), trying to reach a sufficient score in the recognition without increasing too much the limit.

The recognition has been repeated for all the 51 volunteers. The results are show in Table 2, in which:

- **YES** – correct: the algorithm recognizes the right person;
- **YES** – wrong: the algorithm recognizes a wrong person;
- **NO** – correct: the algorithm states that the person to be recognized is not in the database, but indicates the right person as the most similar one;
- **NO** – wrong: the algorithm states that the test is not in the database and indicates the wrong person as the most similar one;

|                      | Percentage |
|----------------------|------------|
| YES – correct        | 91%        |
| YES – wrong          | 7%         |
| NO – correct         | 2%         |
| NO – wrong           | 0%         |

Table 2 Results of the recognition
The results show a percentage of correct recognition higher than 90%.

It is important to stress that all the volunteer belong to the same age bracket (20 to 30 years) and to the same ethnic group (European): the percentage of recognition will probably increases with a less homogeneous sample.

6. Conclusions

The design and calibration of a stereoscopic acquisition system has been presented; subsequently that system has been adopted to collect a sufficiently wide database of faces. A face recognition algorithm, based on the manual selection of 9 repere points, has been then developed and tested: it reports encouraging results, with a fraction of recognition of 9 out of 10.

This algorithm has to be consider a preliminary work, since its application requests the interaction of the user, but it gives important information for future studies. The approach of extracting from a very complex stereoscopic image the 3D coordinates of only 9 landmarks and use them in order to actuate the recognition appear a simple but firm base over which the research can continue.

Acknowledgements

This work has been partially supported by the Italian Ministry of University and Research (MIUR) by a PRIN grant.

References

1. Ratha, N.K., Senior, A., Bolle, R.M.: Automated Biometrics. In: Proc. of ICAPR, Rio, Brazil, pp.445-454 (1997)
2. Wang, P.S.P, Yanushkevich, S. N. Y.: Biometric Technologies and Applications. In: Proc. of IASTED, Austria, pp. 226-231 (2007)
3. Neumann, C., Champod, C., Puch-Solis, R., Egli, N., Anthonioz, A., Meuwly, D., Bromage-Griffith, A.: Computation of Likelihood Ratios in Fingerprint Identification for Configurations of Three Minutiae, Journal of Forensic Science, 51(6):1255-66 (2006)
4. Zhao, W., Chellapa, R., Phillips, P.J., Rosefeld, A.: Face Recognition: A Literature Survey, ACM Computing Surveys (CSUR), v.35 n.4, p.399-458, (2003)
5. Heisele, B., Ho, P., Wu, J., Poggio, T.: Face recognition: component-based versus global approach, Computer vision and image understanding, vol. 91, no. 1–2, pp. 6–21, (2003)
6. Sala, R., Zappa, E., Cigada, A.: Personal Identification Through 3D Biometric Measurements Based on Stereoscopic Image Pairs. In: IMS 2006 - IEEE International Workshop on Measurement Systems for Homeland Security, Contraband Detection and Personal Safety Alexandria, VA, USA, 18-19 (2006)
7. Farkas, L.: Anthropometry of the Head and Face in Medicine, Elsevier North Holland, Inc., (1981)
8. Abate, A., Nappi, M., Riccio, D., Sabatino, G.: 2D and 3D face recognition: A survey, Pattern Recognition Letters 28, pp. 1885–1906 (2007)
9. Bowyer, K., Chang, K., Flynn, P.: A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition, Computer Vision and Image Understanding 101 1–15 (2006)
10. Berretti, S., Del Bimbo, A., Pala, P.: Using Geodesic Distances for 2D-3D and 3D-3D Face Recognition. In: Image Analysis and Processing Workshops, pp. 95 – 100 (2007)
11. Hartley R, Zisserman A, Multiple view geometry in computer vision, II ed., Cambridge, Cambridge University Press, (2003)
12. Goos, M., Alberink, I., Ruijrok, A.: 2D/3D image (facial) comparison using camera matching, Forensic Science International, 163, pp. 10–17 (2006)
13. Sharma, G., Digital colour image handbook, CRC Press (2003)
14. Abdel-Aziz, Y., Karara, H.: Direct linear transformation into object space coordinates in close range photogrammetry, Symposium on Close-Range Photogrammetry, Urbana, IL, pp. 1–18 (1971)
15. Chen, C., Armstrong, C., Raftopoulos, D.: An investigation on the accuracy of three-dimensional space reconstruction using the direct linear transformation technique, Journal of Biomechanics 27, 493-500 (1994)
16. Bailenson, J.: Examining Virtual Busts: Are Photogrammetrically-Generated Head Models Effective for Person Identification?, Presence: Teleoperators and Virtual Environments, Volume 13, Issue 4 (2004)