

# Uncertainty of 3D facial features measurements and its effects on personal identification

E. Zappa\*, R. Testa, M. Barbesta, M. Gasparetto

*Department of Mechanical Engineering, Politecnico di Milano, Via G. la Masa 1, 20156 Milano, Italy*

## *Article history:*

Received 5 June 2013

Received in revised form 19 September 2013

Accepted 28 November 2013

Available online 6 December 2013

## 1. Introduction

Facial recognition techniques are being more and more important because of the multitude of their possible applications [1]. In [2–5] the most relevant techniques developed during last years have been reported, demonstrating good results and a certain degree of maturity when operating under constrained conditions. However, they still suffer high uncertainty if used in everyday situations of the practical life; due to this reason some authors analyzed the human perception [6] with the purpose of identifying human cognitive processes that operate during personal recognition, with the idea of implementing them in novel algorithms [7]. Most of the state-of-the-art recognition techniques are based on the analysis of some biometric measurements, provided by processing facial images with features extraction algorithms [8,9].

Several studies have been conducted on face identification with the purpose of evaluating the main issues that can affect the reliability of these systems (i.e. illumination, pose, expression and occlusion). However, most of the databases proposed in literature [10–12] do not contain stereoscopic acquisition of the faces but only set of frontal and non-frontal faces acquired in different instant of time. Relying on these databases it is therefore not possible neither obtain 3D information nor compare the impact of pose variation on recognition rate, as facial images acquired in different instants of time are affected by the natural variation of facial expression. In this work images of the individuals faces, acquired synchronously from different calibrated cameras, have been analyzed in order to estimate the pose effect on recognition based on 3D features. The purpose of this work is to evaluate the 3D facial feature extraction performed with AAM models and to study their measurement uncertainty when applied on facial images. Furthermore, facial features processing algorithms have been implemented to maximize the system reliability

---

\* Corresponding author. Tel.: +39 0223998445.

E-mail address: [emanuele.zappa@polimi.it](mailto:emanuele.zappa@polimi.it) (E. Zappa).

(i.e. weighting different facial features according to the corresponding estimation accuracy and use multiple images acquisition to improve the recognition reliability). Similar studies [13] calculating Euclidean distance of 3D facial features achieved 96.2% of correct recognition, based on a 26 subjects database. In [14] a feature based approach is presented, demonstrating 96.3% and 91.7% recognition reliability rates, using images databases developed in [10,11] with 40 and 15 individuals respectively. In the works [15,16] the reliability of 97.3% and 96.6% has been reached respectively, analyzing images of the 130 people database developed in [17]. In [18] a stereo system and a 3D model have been used for the face recognition of 100 individuals in different positions, with a correct recognition rate of 98%. Although these studies include non-frontal images in their database, they do not investigate the exact error rate in different pose conditions. In the presented paper, impact of individuals' position on the identification rate has been calculated considering three different stereoscopic systems that synchronously acquire the volunteer's face from three different orientations. In this work, according to [19,20] and in order to compare data with results obtained in previous work [21,22], 58 points have been used to describe the facial geometry. One of the most popular algorithms of feature extraction available in literature is the Active Appearance Model (AAM), developed by Cootes et al. [23], which is based on the combined use of two different statistical models: the Shape model and the Appearance model. The AAM has been used in literature to extract 2D biometric features and several studies have implemented this algorithm in 2D face recognition [3,24]. As known in literature, 3D recognition techniques show higher reliability than 2D techniques [25,26]. For this reason in previous works [21,22] it was decided to use a stereoscopic vision system to obtain images of individuals. In order to create a system able to analyze a high number of individuals and to evaluate its performance on a wide statistical sample, a database of facial images has been constituted.

In particular, in order to analyze recognition reliability on images acquired from different points of view (not frontal images), individuals were simultaneously framed from different directions using different synchronized stereoscopic visions.

This system has been used to acquire image sequences of 117 individuals in order to evaluate measurement repeatability and uncertainty, also with a minimal expression variability (the volunteers were requested to maintain a neutral expression, however the face geometry undergoes to minor expression changes due to the natural motion of the face). In particular, for every 117 people, 5 repetitions of the registered acquisitions have been analyzed.

The proposed instrument is also tested by reproducing a recognition system oriented to control the access of authorized and unauthorized people. In particular, among images of 117 individuals, 90 faces have been used to create a database of persons and to verify the identification both with the same 90 individuals and with remaining 27 persons. When analyzing individuals out of the database, they could be associated to the similar person in

the database, relying on minimization of geometry discrepancy. To avoid this type of error, called False Acceptance case [27], a minimum similarity threshold has been proposed.

## 2. Facial features extraction technique

The subject recognition is performed through a prior dimensionality reduction of facial images. Using image processing algorithms it is possible to select more relevant biometrics to distinguish individuals [28]. In digital image processing, the feature extraction methods allow to select most relevant information for the case and to represent them in a lower dimensionality space.

### 2.1. 2D features detection

In these years, several algorithms for facial features extraction have been developed, of which most important ones are presented in [29]. One of the most interesting model-based methods [30–33] is the AAM [23]. An open source code of this algorithm has been implemented in the software AAM-API developed by Stegmann [34]. The AAM algorithm is based on Principal Component Analysis [35] and, after it has been trained with images of the same type of the ones to be analyzed, permits to create a model that allows to detect features from any image of the same type. In this section, AAM algorithm will be presented. It is based on two principal functions: 2D Shape model and Appearance model. Although the description of AAM algorithm is out of the scope of this work, a brief summary of the algorithm is given here; more detailed information can be found in [23].

#### 2.1.1. The 2D Shape Model

The shape is defined as a 2D points set describing the shape of a target body. During the shape models creation, shapes traced on images of bodies belonging to the same family are submitted to the Procrustes Analysis in order to align these shapes to a common reference system and to make the application of the Principal Component Analysis (PCA) possible.

The PCA generates the shape variation basis  $\phi = (\phi_1 | \phi_2 | \dots | \phi_t)$  that can represent any  $x$  shape of analyzed bodies starting from a mean shape  $\bar{x}$ ; vector  $b$  is a real number set that models deformable shape parameters.

$$x = \bar{x} + \phi b \quad (1)$$

#### 2.1.2. The Appearance model

The Appearance is defined as the texture of a portion of the target. The Appearance model arranges all pixel intensity variations of the images on the mean shape. During the creation of this model, all the training images are transformed into images of the same shape and dimensions and the normalization of the texture is performed to avoid lighting or luminosity changes in pictures. Thereafter, appearance model is elaborated by performing PCA on the training images. Similarly as the shape model, the

appearance model consists of the grey level vector of the mean appearance  $\bar{g}$ , the variation basis  $\phi_g$  and a group of grey level parameters  $b_g$ .

$$g = \bar{g} + \phi_g b_g \quad (2)$$

## 2.2. AAM model annotation

Shape and Appearance models building is based on the analysis of training images, i.e. a sample of images where facial features have been manually annotated. The annotation process of an image consists in tracing different landmarks that outline the most important facial traits on various images. In particular, 58 landmarks demarcating seven areas of the face has been used in this work (three closed path and four open): jaw, mouth, nose, eyes, and eyebrows, as shown in Fig. 1. The choice of the 58 recorded points is made according to the method of record of previous works carried out in agreement with other studies [21,22], in order to make the database mergeable and to be able to compare data. The annotations made on images obtained with the top camera and the bottom camera are shown in Fig. 1. At a later stage, is possible to fit this model on new images and to automatically obtain biometric features.

## 2.3. AAM model creation

The created AAM model has to include shape and texture information of the face which may vary in repeated acquisitions depending on subject expression and position. In order to have an accurate detection of the features, the parameters of the created models may not cover a too large variation of the shape. In particular, using the system presented in Section 4, when retracting the individuals with stereoscopic cameras, in other words from two different points of view, a big camera angle in the vertical plane induces big shape modifications. Therefore, the experiences gained in literature [36] and in this work have shown the need of having two models: one for the top cameras and one for bottom cameras. In Fig. 2, three principal variation modes are shown: the first mode represents the rotation of the head in the horizontal plane, the second mode represents the rotation in the vertical plane and third mode represents skin variation.



Fig. 2. Bottom cameras model visualization.

## 2.4. Feature extraction

All presented operations are preliminary developed to create the two described models, which will be fitted on facial images in order to automatically detect the landmarks. The AAM fitting algorithm iteratively varies both model parameters until the difference between real face image and simulated face image is minimized. The results of this operation is a 58 2D points mask for top images and another 58 2D points mask for bottom images that will be triangulated [37] in order to build a 3D mask. For each individual analyzed, the 3D points mask has been built by triangulating two 2D masks, each one extracted with the AAM algorithm. The computational cost of this procedure varies proportionally to the complexity of the created model. Processing time optimization has not been implemented as it was not the purpose of this work; with the currently implemented software solution the average time required for the complete recognition of one individual is

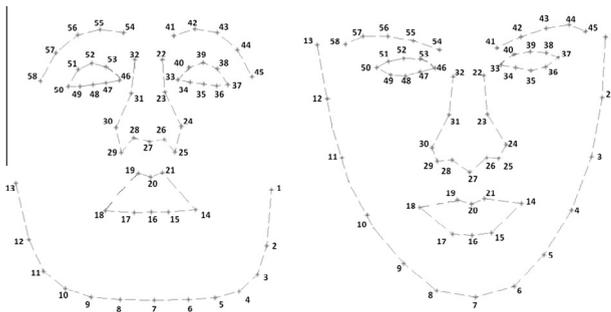


Fig. 1. Representation of 58 annotated points on bottom image (left) and top image (right).

10.3 s, where 5.1 s are required to extract the 2D features from each image (resolution  $1280 \times 1024$  pixel), and 0.1 s for the comparison with the 117 individuals database. Analysis has been performed with a PC provided of a Intel i7-3.40 GHz processor and 8 GB RAM.

### 3. Design and development of the acquisition system

The most important objective of this work is to collect a large sample of facial images and to use it as a database to assess and refine different recognition techniques.

In the first stage of this project a database of images has been created in order to constitute a robust statistical basis for the verification of obtained results. From the experience gained during the test of the AAM-API code [38], it has been experimented the need of instructing the shape and texture models, which constitute the heart of the algorithm, with the use of adequately large statistical sample. Furthermore, as shown in previously conducted studies [39], the presence of small rotations of the face introduces an important error in recognition rate. For this reason, recognition uncertainty in function of different angles of the face has been evaluated. To do that, it was important to have facial images acquired from different viewpoints in the same instant of time. Indeed, since the face is a deformable body, the geometry of the face may slightly vary even in consecutive instants, according to different expressions.

To acquire necessary information for the creation of the database described in the following chapter 4, a multi-camera vision system has been developed, allowing to simultaneously retract faces from different points of view. The system is provided of three pairs of cameras, which perform synchronous acquisitions of the same individual from different points of view, as the subject was turning the head. In this way, data referred to images of the face in frontal and rotated positions are perfectly comparable. The system developed is schematized in the diagram of Fig. 3a and shown in Fig. 3b. Each pair of cameras is vertically aligned and placed at  $0^\circ$ ,  $5^\circ$  and  $10^\circ$  referring to the axis of the face (see Fig. 4), as indicated in Fig. 3a with the tags cam1-2, cam3-4 and cam5-6. A seventh camera is placed at mid-height (cam 7 in Fig. 3b), between two central cameras, to collect images that will form the database and that can be used for the validation of 2D recognition techniques.

Stereoscopic vision is obtained by coupling each top camera with the corresponding bottom camera, with the intent of minimizing the nose undercut issue. Since the measurement error decreases in the third dimension when the angle between cameras increases [40], an angle of  $45^\circ$  has been used, which results to be the best compromise [21] because it maximizes the angle and avoids occlusion phenomenon in the two views. The configured system allows having a field of view size of approximately  $300 \times 400$  mm, large enough to contain the subject's face placed 1000 mm away from the system.

As described above, the system consists of a pair of AVT – Marlin F-131B (equipped with  $2/3''$   $1280 \times 1024$  pixel CMOS sensor) cameras, and two pairs of AVT – Pike F-145B (Sony  $2/3''$   $1388 \times 1038$  pixel CCD progressive scan

sensor) cameras. The seventh camera is the IDS – GigE UI-5490SE-M (Aptina  $1/2''$   $3840 \times 2748$  pixel CMOS sensor). All cameras are equipped with 25 mm focal length lenses and connected to a computer via three firewire IEEE1394 acquisition cards (for cameras 1 to 6) and a Gigabit Ethernet card (for camera 7). A common trigger generates the signal to synchronize all cameras. A LabView software allows simultaneous management of seven cameras and images recording in bitmap format. The control system allows to acquire sequences of images with a user-defined interval (default set 5 s), in order to obtain multiple images of the same person in the same nominal position but allowing small changes of expression that naturally happens.

According to [41], each pair of cameras has been calibrated using the Zhang's method [37,42] and Camera Calibration Toolbox [43], which provide to determine internal and external calibration parameters.

### 4. Database building

The acquisition system described above has been used to collect several facial images, in order to test the reliability of the recognition judgment as a function of the acquisition angle and considering natural variability of the expression and head position.

As mentioned in Section 2.1, in the proposed system the 2D features detection is based on the AAM algorithm, which describes biometric properties of the face using two statistical models built on facial images. According to [36] it is necessary to adequately train these models, in order to have good flexibility in both of them. The AAM model has to be trained considering also images with angled face and a wide range of facial characteristics in order to automatically recognize them. Increasing the number of pictures used to train the AAM, increases the number of recognizable subjects and the model adaptability to different positions. A good model flexibility permits to cover a wide range in terms of shape and texture variation and to represent a large genetic diversity, pose variations and expression changes. On the other hand if a too broad range of condition is used, the reliability of features recognition is reduced. For this reason, one model for the top camera images and one model for the bottom camera images have been realized.

The database includes facial images of 117 volunteers; for each one of them, 3 series of acquisitions are carried out:

- 1 Set with the subject in frontal position ( $0^\circ$  rotation) as shown in Fig. 5a.
- 1 Set with  $10^\circ$  rotation of the head, as shown in Fig. 5b.
- 1 Set with  $20^\circ$  rotation of the head, as shown in Fig. 5c.

Each set contains 9 pictures of the same subject in the same position, acquired with a rate of 1 images every 5 s. to allow natural minor changes in head position and in facial expression.

Therefore, 6 cameras frame 9 times the individual placed in 3 positions that is  $6 \times 9 \times 3 = 162$  images

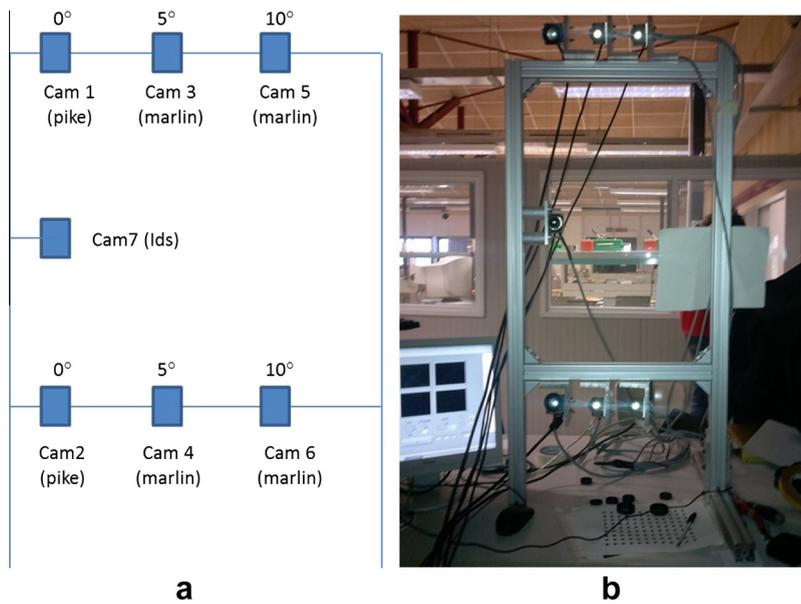


Fig. 3. The proposed multi-vision system: (a) development scheme and (b) picture.

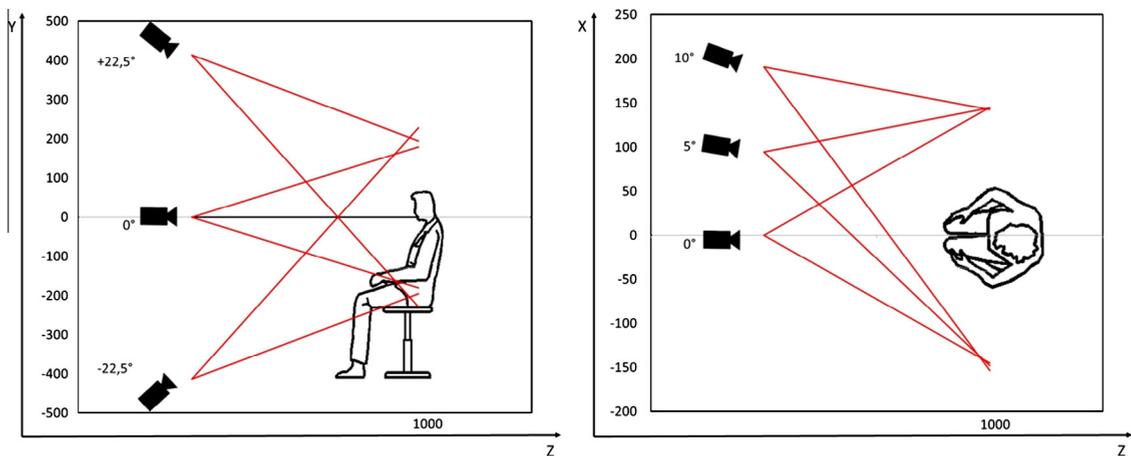


Fig. 4. System design scheme.

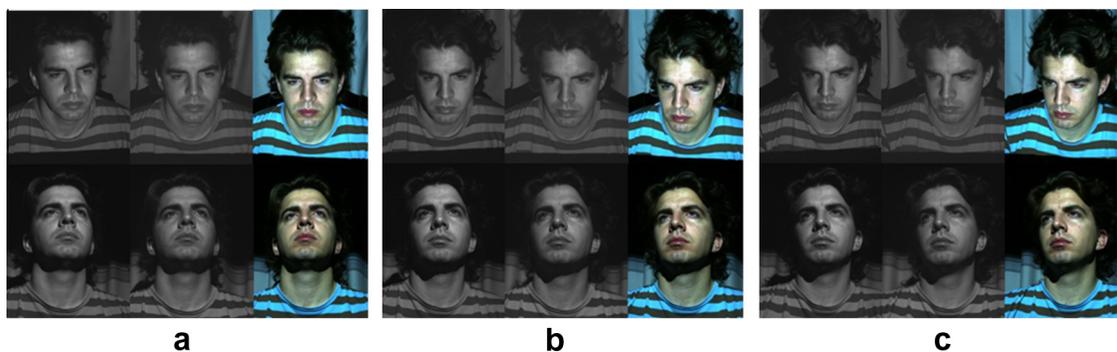


Fig. 5. Images collected in the database: the subject is framed in three different positions: (a) 0° rotated, (b) 10° rotated, (c) 20° rotated.

recorded for each person. Since some acquisitions were compromised due to quick movements of volunteers that induces the blurred effect of the image due to not negligible motion during the exposure time, for all 117 individuals it was possible to analyze only five repetitions of the nine. The first repetition was used to train the AAM model, and other four to verify the recognition.

## 5. Analysis and results

In this part, the process of personal identification will be presented, starting from two stereoscopic images of the face, passing through landmarks extraction with AAM and concluding with aspects of 3D shape construction and mask comparison with those in the database.

### 5.1. Model development

The first phase of the process is the annotation of training images to build the variability model of the AAM software. As previously explained, the annotated images provide the statistical base for the training of shape and texture models built by the Active Appearance Model (AAM). In this stage, the following annotations have been carried out:

- For 50 volunteers, the two images obtained with cameras 1 and 2 (Fig. 3) with the subject in frontal position has been annotated.
- For 19 of them, 18 pair of images has been annotated, using images of cameras 1 to 6 (Fig. 3) in three different position of the subject (0°, 10°, 20° head rotation).

With a heuristic approach to the problem, first group of images is used to introduce model adaptability to human biological diversity, second group provides the shape modification caused by head rotation.

In conclusion, the AAM models have been built using about 200 images. These images are used only to train the AAM and the other four images of the same acquisition set are used to test the recognition.

At a later stage, the facial images were analyzed with AAM in order to automatically extract the 2D facial masks. As the feature detection software works in two dimensions (i.e. images), it is necessary to elaborate separately two stereoscopic images to detect the landmarks prior to triangulate. In [22], according to [36], it has been experienced the need of two different AAM models, the first for the images of the top cameras and another for the images of the bottom cameras.

After obtaining the 2D masks from each pair of images, it is possible to triangulate homologous landmarks and to calculate a 3D masks. These operations have been undertaken for all 117 volunteer and 117 masks are collected in a database containing the 3D landmarks coordinates. Similarly, in the recognition process, from the pair of stereoscopic pictures of the person to be recognized, two 2D masks are extracted with AAM and then triangulated to obtain the 3D mask. The recognizing 3D mask will be

compared with the masks belonging to the database as detailed below.

### 5.2. Landmarks weighting according to measurement uncertainty

Since the recognition process is based on geometric comparison of 3D masks, one of the main issues in the recognition process is the variation of landmarks coordinates due to facial expression changes. There are some parts of the face that are more stable and less affected by facial expressions, such as eyes and nose, while other points which position is strongly correlated with the expression assumed by the individual [44].

Moreover the face AAM model does not detect all landmarks with the same accuracy; in particular, their identification is affected by high uncertainty in those areas of the face where it is not evident the variation of texture, such as the profile of the jaw. The worst results are found for those parts of the jaw located close to the ear, as the derivative of the facial surface respect to the image plane is high and a small shift in plane, caused by an inexact detection of landmarks, produces a wide variation in depth. These points are also affected by occlusion issues in rotated face images.

The variability of these landmarks could deceive the subject recognition and, for this reason, it seems reasonable to weight differently each of the 58 points, according to their uncertainty (and therefore their reliability in personal identification).

For this reason, 58 different weights have been assigned to each of 58 landmarks, using high weights for features extracted with good accuracy, and low weights for points detected with low accuracy.

In order to estimate weights for each one of the 58 landmarks, the variance between repeated measurements of the face has been calculated analyzing 9 masks obtained from 9 repeated pictures of the same subject, for all the 117 subjects in the database. Since small rotations of the face may occur between consecutive acquisitions, it is necessary to align one mask to the other one before to calculate average position and relative variance for each landmark. An arbitrary 3D mask is chosen as reference and all the other ones are roto-translated in order to minimize the distance of corresponding points.

The first mask of the subject is taken as reference.

- (1) Each of the other 8 masks of the  $i$ -th subject is iteratively aligned to the reference mask applying a roto-translation, based on minimization of the summed squared distance between corresponding points.
- (2) After roto-translation the variance of the 58 points coordinates in the nine masks of the  $i$ -th subjects is computed as:

$$\text{Var}(U)_{i,k} = \frac{\sum_{j=1}^f (U_{i,j,k} - \overline{U}_{i,k})^2}{f} \quad (3)$$

where:

- $1 < j < f = 9$  image repetitions for the same subject;
- $U_j$  mask coordinates;

- $\bar{U}$  coordinates of the reference mask;  
 $1 < k < p = 58$  landmarks;  
 $1 < i < n = 117$  people.

- (3) This operation is repeated for all the subjects in the database ( $i = 1-117$ ) and the result is the matrix  $Var(U)_{i,k}$  of  $58 \times 117$  values of the variance corresponding to each landmark for each individual.
- (4) The variances  $Var(U)_{i,k}$  are averaged between the 117 individuals and the vector  $\overline{Var}_k$  of 58 elements is obtained.

$$\overline{Var}_k = \frac{\sum_{i=1}^n Var_{i,k}}{n} \quad (4)$$

- (5) Finally, the weights of each  $k$ -th point are calculated as the inverse of the average variance of point  $U_j$  and then normalized to make their sum equal to 58.

$$Weight_k = \frac{1}{\overline{Var}_k} \quad (5)$$

$$Weight_{norm,k} = \frac{Weight_k}{\sum_{k=1}^{58} Weight_k} * 58 \quad (6)$$

In this way, the variance of each point associated with each individual was obtained. Fig. 6 shows an example of landmarks variability in one person.

In Table 1 the average weights are shown, gathered in face areas. The results confirm what was intuitively suggested: there is more stability on the eyes; the mouth is subject to large modifications due to facial expression changes, while the outline of the jaw presents the highest variability.

The individual recognition procedure can be summarized in the followings process:

- (1) A pair of top and bottom images are respectively analyzed through the top and bottom AAM models and 2D landmarks are extracted.

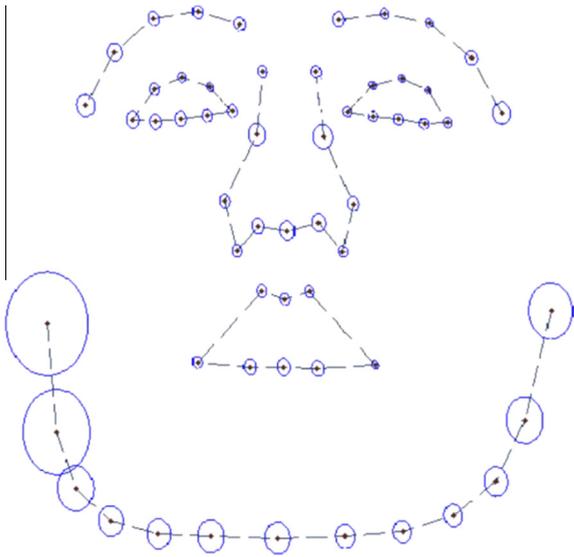


Fig. 6. Landmarks position and their variance representation proportionally to the circles radius.

Table 1  
Averaged weight by areas of the face.

Area	Weight
Jaw	0.23
Mouth	0.51
Nose	1.10
Eyes	1.87
Eyebrow	0.96

- (2) Each pair of 58 homologous points are triangulated and the 3D mask is obtained.
- (3) The mask is compared with  $i$ -th one in the database by calculating the Summed Weighted Error (SWE) as described in Eq. (7):

$$SWE_i = \frac{\sum_{k=1}^{58} (Weight_k * (U_{k,i} - U_{k,ref})^2)}{58} \quad (7)$$

where  $U_{k,i}$  are the coordinates of the  $k$ -th point for the  $i$ -th individual.

- (4) The mask is iteratively rotated and translated to the  $i$ -th mask, until it matches as well as possible (by minimizing the SWE).
- (5) The comparison is repeated for all  $i = 117$  masks in the database and the 117 SWE final values are calculated.
- (6) Correspondence of the face is verified by searching the mask of the database with the minimum associated SWE.

### 5.3. Reliability of the recognition system

The evaluation of the recognition reliability has been conducted by comparing 117 new masks, obtained from a new pair of pictures, with the 117 masks of the database. In order to prove the goodness of these results, this analysis was repeated four times, using four different pair of images. In other words, this test has been carried out on  $(117 * 117) * 4 = 54,756$  comparisons.

Error rates in the recognition test on four images of each subject (acquired at interval of 5 s), are shown in Table 2. The reported results show high variability of the recognition in various images. In fact, the recognition process was highly influenced by high uncertainty of the worst landmarks that can deviate calculated discrepancy between masks. In Table 3 results of analogous analysis performed using weights are reported in order to prove their effectiveness.

From the obtained results, it is clear that with the use of rotated cameras, (maintaining the face oriented to camera 1 and 2, or rather a  $0^\circ$  head rotation) the reliability of recognition decreases, due to uncertainty in the detection of some landmark. Furthermore, with the use of weights, the error rate significantly decreases, both for frontal and angled images.

Similar tests are performed on images of the face  $10^\circ$  rotated and shown in Table 4. These results can be compared with those taken in frontal position by the pair of cameras 5 and 6 ( $10^\circ$  cameras in Table 3) showing an average error slightly higher than the latter. Their comparison is motivated because these two cases reproduce symmetric

**Table 2**

Error rate in recognition test, in four trials for 117 \* 117 comparisons, without using weights.

Trial	Without weights, 0° face rotation		
	0° Cameras	5° Cameras	10° Cameras
1	15.4	19.7	23.1
2	23.1	34.2	28.2
3	34.2	28.2	29.9
4	22.2	33.3	35.0
Average error (%)	23.7	28.8	29.1

**Table 3**

Error rate (%) in recognition test using weights: comparing results with those in Table 2 is possible to observe considerable failure reduction.

Trial	Use of weights, 0° face rotation		
	0° Cameras	5° Cameras	10° Cameras
1	2.6	5.1	4.3
2	3.4	6.8	6.8
3	4.3	5.1	10.3
4	5.1	9.4	8.5
Average error (%)	3.8	6.6	7.5

images. Analyzing the standard deviation of recognition error in both cases, it is possible to prove the compatibility of data. In fact, in the first case, average is 7.5% and standard deviation equal to 2.6%, whereas in the second one, average is 9.0% and standard deviation is 1.5%, which means that their confidence intervals overlap.

#### 5.4. Reliability improvement

Indicating the system recognition reliability as the ability to perform a correct recognition, it is possible to define the reliability rate as  $R = 100 - \text{error rate}$ .

Information obtained from each one of three pairs of cameras can be combined in order to improve the recognition reliability, as if three systems simultaneously worked. In particular, identifications performed with 5° and 10° cameras (cameras 3 and 4 and cameras 5 and 6 in Fig. 3) and frontal cameras (1 and 2) were compared. It is possible to interpret it as a new system composed of three instruments with reliability values  $R_{0^\circ}$ ,  $R_{5^\circ}$ ,  $R_{10^\circ}$  and with overall reliability rate  $R_{tot}$  (Eq. (8)), as schematized in Fig. 7. Each of these recognition instruments has reliability estimated by means of the results shown in Table 3:

- Reliability 1-2 cameras =  $100 - 3.8 = R_{0^\circ} = 96.2\%$ .
- Reliability 3-4 cameras =  $100 - 6.6 = R_{5^\circ} = 93.4\%$ .
- Reliability 5-6 cameras =  $100 - 7.5 = R_{10^\circ} = 92.5\%$ .

**Table 4**

Error rate (%) in recognition test using weights with 10° rotated subjects.

Use of weights, 10° face rotation	
Trial	0° cameras
1	10.3
2	7.7
3	10.3
4	7.7
Average error (%)	9.0

The developed algorithm, permits to compare the recognition reliability achieved from the pairs of cameras using the following logic: only if both pairs of 5° and 10° cameras recognize the same person and this is different from the one recognized by the central cameras, the result from the angled cameras are considered more reliable; otherwise recognition obtained from central pair of cameras is considered correct. We can represent the overall system with the diagram in Fig. 7, where the total reliability [45] is expressed by the formula (8).

$$R_{tot} = 1 - (1 - R_{0^\circ}) * (1 - (R_{5^\circ}) * (R_{10^\circ})) = 99.48\% \quad (8)$$

It is possible to appreciate a theoretical reliability improvement of the system, by decreasing the error rate from 3.8% to 0.52%.

The error rates of the overall system in four experimental tests have been reported in Table 5, with an actual reliability increase from 96.2% to 98.9%.

However, the recognition solution with three stereoscopic systems requires the use of a complex system composed of six calibrated cameras and configured to perform the synchronous acquisition. In order to obtain a reliability improvement, reducing the complexity of the structure, multiple tests using only one pair of frontal cameras (cam1 and cam2) was performed. Hence, the results obtained from the four trials previously described (Table 3) were considered as repeated biometric measurements and the information were combined in a unique recognition judgement as described below.

Referring to Table 6, it is possible to have 5 cases: in the case #1 the identification was correct and coherent in all four tests, whereas in case #2 recognition was exact and coherent in three trials out of four. Similarly, in the case #3 the correspondence was found to be exact and coherent in two tests out of four. It is clear that negative result on all four tests (incorrect and incoherent shown in case #5) have never occurred, but in the worst case (#4) one individual was recognized in just one trial out of four, that is without any coherence with other trials.

In the case of repeated tests with cameras 1 and 2 only (Fig. 3), it is possible to have a structurally less complex system, which allows obtaining reliability improvement by simply matching multiple acquisitions of the face. For example, if recognition is considered correct when occurring cases #1 and #2 (i.e. the cases of coherence at least in three tests out of four), the error rate was 3% and the correct recognition rate is therefore 97%, which is a result closely aligned with the state of the art systems described in Section 1.

#### 5.5. Recognition threshold

The developed identification system recognizes an individual basing on the evaluation of the SWE, Summed Weighted Error, selecting the most resembling mask in the database (i.e. minimum SWE). If we suppose to use a recognition system in order to supervise admissions in a restricted area, during a real identification process it is possible to have five situations:

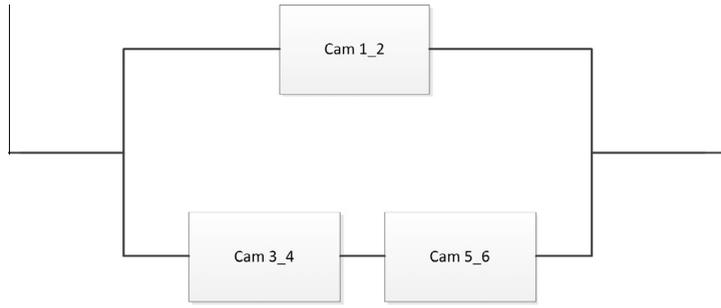


Fig. 7. Reliability scheme of the proposed system.

Table 5

Error rate (%) for the proposed system, in four trials on 117 \* 117 comparisons.

Trial	0° + 5° + 10° Cameras
1	0.0
2	0.9
3	0.9
4	2.6
Average error (%)	1.1

Table 6

Recognition results comparing 4 trials.

Case #	Identification in 4 trials	People correctly recognized	Percentage (%)
1	100% Correct	102	87
2	75% Correct	12	10
3	50% Correct	2	2
4	25% Correct	1	1
5	0% Correct	0	0
		117	100

- Access admitted (positive recognition):
  - (1) The person is correctly recognized and present in the database.
  - (2) Unauthorized person is accepted by mistake as authorized (False Acceptance) [27].
  - (3) An authorized subject is recognized as another member of the database (False Identification).
- Access denied (negative recognition):
  - (4) The person is correctly rejected because not present in the database.
  - (5) The system erroneously rejects an authorized person, by giving a false alarm (False Rejection) [27].

Consequently, three main types of percentage errors can occur in a recognition system [27] when analyzing authorized and unauthorized individuals:

Unauthorized person:

- FAR (False Acceptance Rate) is the percentage of false acceptances, this event occurs when the system erroneously accepts unauthorized users.

Authorized person:

- FRR (False Rejection Rate) is the percentage of false rejections, when authorized users are erroneously rejected as they were an external individual.
- FIR (False Identification Rate) is the percentage of falsely assigned identifications that occurs when an internal user is admitted as another one.

In order to evaluate these three types of error, among the 117 individuals in the database, two groups were created: one composed of 90 authorized people and the other one with 27 unauthorized persons.

Therefore, following analysis are carried out by evaluating SWE obtained in the access attempt of 90 authorized and 27 impostors as follows:

- False Acceptance Rate is calculated by collecting SWE values in the recognition of 27 individuals not belonging to the database.
- False Rejections are generated by discarding greater SWE of Correct Recognitions in the 90 \* 90 comparison (Fig. 8).
- False Identifications cases extracted from Incorrect Recognition cases, as recognizing people have been associated to other individuals of the database when comparing 90 \* 90 persons.

Plotting on a histogram of occurrences the SWE calculated for users recognitions (Figs. 8 and 9 respectively), it is possible to visualize its occurrences distribution [46]. It is intuitive to assume the SWE average value is lower in the case of authorized correct recognition cases than in impostor recognition cases.

In the histogram in Fig. 8 occurrences of individuals correctly recognized are highlighted; the 94.4% of the showed data lie below the SWE value of 4 mm<sup>2</sup>.

Similarly, data regarding the impostors recognition are highlighted in the histogram of Fig. 9. Below the SWE value of 4 mm<sup>2</sup> it can be found a percentage of False Acceptance Rate of 13.9%.

It is important to note that the probabilities of correct and incorrect identifications of internal members are complementary because they refer to the analysis of the same subjects sample.

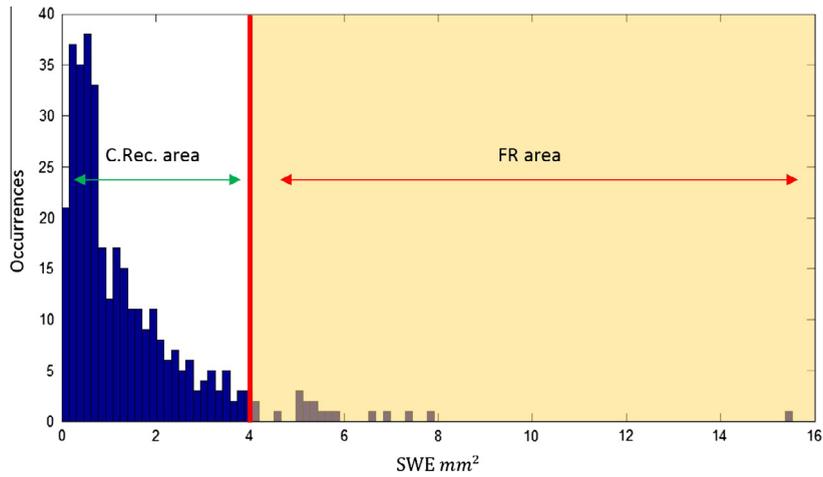


Fig. 8. SWE values distribution of correctly recognized authorized members and false rejection cases.

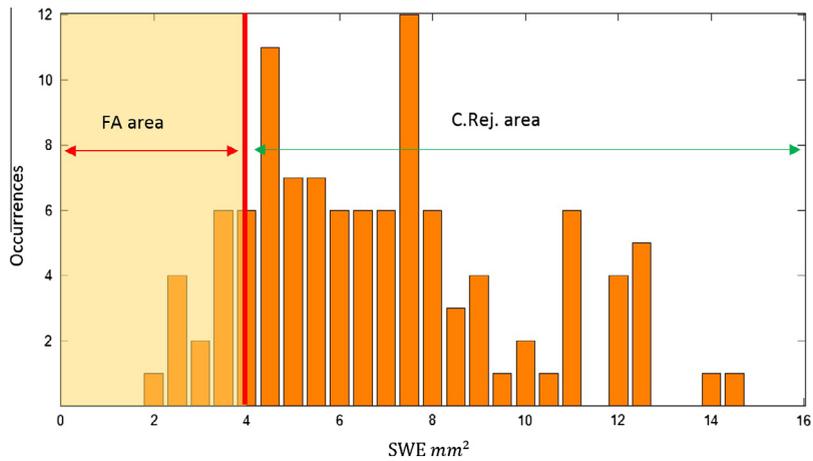


Fig. 9. SWE values distribution in impostors recognition cases.

Correct identifications represent the 96.2% of all  $90 * 4 = 360$  internal individual recognitions (Table 3), whereas False Identifications constitute only the 3.8% (average error Table 3) of the  $90 * 4 = 360$  internal recognition trials. As False Identifications constitute a very small data collection, they are not represented on a histogram but their occurrences are shown in Table 7.

Introducing a data acceptance threshold, it is possible to reduce the FAR and the FIR, at the cost of generating False Rejection cases (i.e. increasing FRR). By adjusting the threshold value, it is possible to modify the ratio between False Rejection and False Acceptance cases, (i.e. to vary the selectivity of the system). The threshold introduction can define a tolerance degree in the recognition process: if the threshold is high the system will have a large tolerance with a high number of False Acceptances, otherwise, imposing a low value, it will be very restrictive and it will have a high number of False Rejections. The choice of the threshold value depends by the required tolerance.

Different threshold levels have been applied on SWE values and FAR, FIR and FRR percentage as a function of

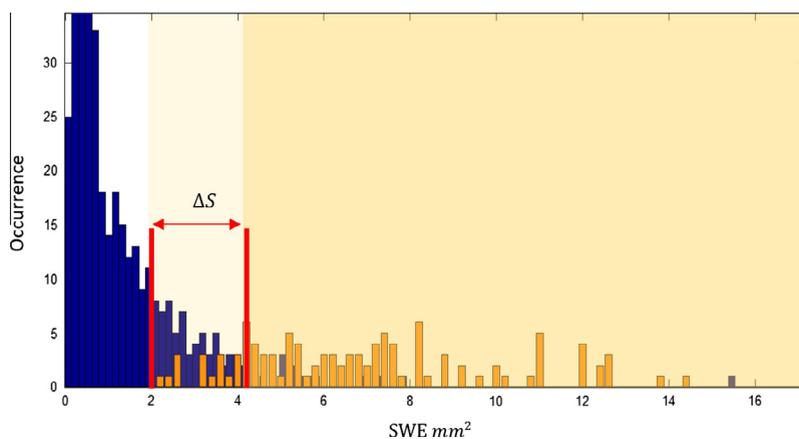
Table 7

False Recognition Rate (FRR), False Acceptance Rate (FAR) and False Identification Rate (FIR) as a function of SWE threshold.

SWE threshold $mm^2$	2	2.5	3	3.5	4
FAR (%)	0	1.8	4.6	8.3	13.9
FIR (%)	0	0	0.2	1.1	1.3
FRR (%)	22.7	15.8	12.0	7.7	5.6

the threshold are collected in Table 7. In Fig. 10 the threshold effect is shown: the left area of the first curve represents the Positive Recognitions (Correctly Recognized people + False Identifications + False Acceptances) while on the right there are Negative Recognitions cases (Correctly Rejected people + False Rejections).

The SWE threshold value may vary depending to the required tolerance degree for the recognition system: if the threshold is set to  $2 mm^2$  the percentage of False Rejections is 22.7%; however, the identification is almost totally free of False Acceptances probability. Contra, if a higher



**Fig. 10.** By superposing correct recognition and impostors recognition diagrams, it is possible to highlight the impact of the SWE threshold in a variability range  $\Delta S$ . At the left of this threshold, SWE recognition values are considered positive; SWE values at the right of the threshold are considered negative.

threshold is used, the possibility of False Acceptances increases. The minimum value of summed errors (FRR + FAR + FIR) is equal to 16.8%, obtained assuming a  $3 \text{ mm}^2$  threshold.

The numerical values of the thresholds reported above were computed for the specific case of 58-point masks used in this work. Of course, considering a different points set would lead to different numerical values for the threshold, nevertheless the concept of the threshold and the procedure to estimate its numerical value would be the same as the one proposed above. In particular the value of the threshold depends on the SWE values (see Fig. 10) and, as previously described in this section, it is possible to fix different values, according to the desired recognition policy (i.e. to minimize FAR or FRR ...).

## 6. Conclusions

Facial recognition techniques still suffer of lack of reliability when applied under real conditions. In this work the impact of face rotation on recognition performance is estimated using a stereoscopic multi-camera vision system, which allows to frame the subject from three different point of view:  $0^\circ$ ,  $5^\circ$ ,  $10^\circ$  rotation referring to the head axis.

With the 7-cameras stereoscopic system, facial images of 117 individuals have been collected in a database composed of 3 series of frames acquired in 3 different head positions, repeated 9 times. Hence, 22113 pictures are recorded.

The proposed recognition system has been tested on  $117 * 117 * 4 = 54,756$  comparisons, demonstrating an initial average error rate of 23.7% using frontal cameras, 28.8% using  $5^\circ$  rotated cameras and 29.1% using  $10^\circ$  rotated cameras. Weighting the facial feature with a weight inversely proportional to the measuring uncertainty, average errors decrease respectively to 3.8%, 6.6% and 7.5%. By matching recognition results provided by three pair of cameras it is possible to further reduce the error to an average value of 1.1%. However, this solution requires a complex six-cameras acquisition system, which is hardly applicable in

most of the practical conditions. Due to this reason, a similar analysis has been made by matching information obtained using one pair of cameras in four repeated trials. If we consider satisfying a correct results coherence in at least three recognition of the four, it is possible to appreciate a 3% error. Similar works, reported in literature, showed comparable error rates; some of which have been discussed in Section 1. Finally the use of measurement threshold on the SWE value to reduce False Acceptance Rate has been contemplated. The analysis in the effectiveness of the SWE threshold was done considering a subset of 90 individuals of the database as “authorized people” and the other 27 people as “non-authorized”. It has been observed that the SWE average in rejected individuals is larger than the SWE in correctly recognized individuals. Introducing a SWE threshold of  $4 \text{ mm}^2$  it is possible to reduce to 13.9% the False Acceptance Rate producing only the 5.78% of False Rejection Error. If a zero-FAR system [47] is required, the threshold has to be placed at  $2 \text{ mm}^2$ .

## Acknowledgment

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

## References

- [1] Thomas Huang, Ziyong Xiong, Zhenqiu Zhang, *Face recognition applications*, Handbook of Face Recognition, Springer, London, 2011. pp. 617–638.
- [2] Rabia Jafri, Hamid R. Arabnia, A survey of face recognition techniques, *J. Inform. Process. Syst.* 5 (2) (2009) 41–68.
- [3] Wenyi Zhao et al., Face recognition: a literature survey, *Acn Computing Surveys (CSUR)* 35 (4) (2003) 399–458.
- [4] A.S. Tolba, A.H. El-Baz, A.A. El-Harby, Face recognition: a literature review, *Int. J. Signal Process.* 2 (2) (2006) 88–103.
- [5] Xiaoyang Tan et al., Face recognition from a single image per person: a survey, *Pattern Recogn.* 39 (9) (2006) 1725–1745.
- [6] F. Crenna, G.B. Rossi, L. Bovio, Measurement of the perceived similarity in face recognition, in: Proc. of the XX IMEKO World Congress – Metrology for Green Growth, September 9–14, 2012, Busan, Republic of Korea, ISBN 978-89-950000-5-2-95400.

- [7] F. Crenna, E. Zappa, L. Bovio, R. Testa, M. Gasparetto, G.B. Rossi, Implementation of perceptual aspects in a face recognition algorithm, *J. Phys.: Conf. Ser.* 459 (1) (2013). IOP Publishing.
- [8] Yang Meng, Lei Zhang, Gabor feature based sparse representation for face recognition with gabor occlusion dictionary. *Computer Vision-ECCV 2010*, Springer, Berlin Heidelberg, 2010. pp. 448–461.
- [9] Stan Z. Li, *Handbook of Face Recognition*, Springer-Verlag London Limited, 2011.
- [10] The At & T Database of Faces, <<http://www.uk.research.att.com/facedatabase.html>>.
- [11] The Yale database, <<http://cvc.yale.edu/>>.
- [12] The FERET database, <<http://www.itl.nist.gov/iad/humanid/feret/>>.
- [13] A.N. Ansari, M. Abdel-Mottaleb, 3D face modeling using two views and a generic face model with application to 3D face recognition, in: *Proceedings IEEE Conference on Advanced Video and Signal Based Surveillance*, IEEE, 2003, pp. 37–44.
- [14] Mohamed Aly, Face recognition using SIFT features. *CNS/Bi/EE report*, 2006, 186.
- [15] J. Cook, V. Chandran, S. Sridharan, C. Fookes, Face recognition from 3D data using iterative closest point algorithm and Gaussian mixture models, in: *Proc. 2nd Internat. Symposium on 3D Data Processing, Visualization and Transmission (3DPVT 2004)*, Thessaloniki, Greece, 6–9 September, 2004.
- [16] M.O. Irfanoglu, B. Gokberk, L. Akarun, 3D shape-based face recognition using automatically registered facial surfaces, in: *Proc. 17th Internat. Conf. on Pattern Recognition (ICPR2004)*, Cambridge, 2004, pp. 183–186.
- [17] Charles Beumier, Marc Acheroy, Automatic 3D face authentication. In: *Proc. Image and Vision Computing*, vol. 18(4), 2000, pp. 315–321. Database, <[http://www.sic.rma.ac.be/~beumier/DB/3d\\_rma.html](http://www.sic.rma.ac.be/~beumier/DB/3d_rma.html)>.
- [18] G. Medioni, R. Waupotitsch, Face recognition and modeling in 3D, in: *Proc. IEEE Internat. Workshop on Analysis and Modeling of Faces and Gesture Recognition (AMFG'03)*, Nice, France, October, 2003, pp. 232–233.
- [19] Leslie G. Farkas, *Anthropometry of the Head and Face in Medicine*, Elsevier, New York, 1981.
- [20] Jeremy N. Bailenson et al., Examining virtual busts: Are photogrammetrically generated head models effective for person identification?, *PRESENCE: Teleoperators & Virtual Environ* 13.4 (2004) 416–427.
- [21] Emanuele Zappa, Paolo Mazzoleni, Reliability of personal identification base on optical 3D measurement of a few facial landmarks, *Proc. Comput. Sci.* 1 (1) (2010) 2769–2777.
- [22] Emanuele Zappa, Paolo Mazzoleni, Yumei Hai, Stereoscopic based 3D face recognition system, *Proc. Comput. Sci.* 1 (1) (2010) 2521–2528.
- [23] Timothy F. Cootes, Gareth J. Edwards, Christopher J. Taylor, Active appearance models, *Pattern Anal. Machine Intelligence*, IEEE Trans. 23 (6) (2001) 681–685.
- [24] Gareth J. Edwards, Timothy F. Cootes, Christopher J. Taylor, Face recognition using active appearance models, *Computer Vision—ECCV'98*, Springer, Berlin Heidelberg, 1998. pp. 581–595.
- [25] Mohd Farid Ariff Mohd et al., Geometric and radiometric characteristics of a prototype surveillance system, *Measurement* (2012).
- [26] Abate, F. Andrea, et al., 2D and 3D face recognition: a survey, *Pattern Recogn. Lett.* 28 (14) (2007) 1885–1906.
- [27] Ted Dunstone, *Biometric System and Data Analysis: Design, Evaluation, and Data Mining*, Springer, 2009.
- [28] Remo Sala, Emanuele Zappa, Alfredo Cigada, Personal identification through 3D biometric measurements based on stereoscopic image pairs, measurement systems for homeland security, contraband detection and personal safety, in: *Proceedings of the 2006 IEEE International Workshop on*. IEEE, 2006.
- [29] Elham Bagherian, Rahmita Wirza O.K. Rahmat, Facial feature extraction for face recognition: a review, in: *International Symposium on Information Technology*, 2008, ITSIm 2008, vol. 2, IEEE, 2008.
- [30] Xinbo Gao et al., A review of active appearance models, systems, man, and cybernetics, Part C: applications and reviews, *IEEE Trans.* 40.2 (2010) 145–158.
- [31] Tony Ezzat, Tomaso Poggio, Facial analysis and synthesis using image-based models, *Automatic Face and Gesture Recognition*, 1996, in: *Proceedings of the Second International Conference on*. IEEE, 1996.
- [32] Michael J. Jones, Tomaso Poggio, Multidimensional morphable models, in: *Sixth International Conference on Computer Vision*, 1998. IEEE, 1998.
- [33] Andreas Lanitis, Christopher J. Taylor, Timothy F. Cootes, Automatic interpretation and coding of face images using flexible models, *Pattern Anal. Machine Intell.*, IEEE Trans. 19 (7) (1997) 743–756.
- [34] Mikkel B. Stegmann, *The AAM-API: an open source active appearance model implementation*. *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2003*, Springer, Berlin Heidelberg, 2003. pp. 951–952.
- [35] Ian T. Jolliffe, *Principal Component Analysis*, vol. 487, Springer-Verlag, New York, 1986.
- [36] Timothy F. Cootes, K. Walker, Christopher J. Taylor, View-based active appearance models, *Automatic Face and Gesture Recognition*, 2000, in: *Proceedings Fourth IEEE International Conference on*. IEEE, 2000.
- [37] Zhengyou Zhang, Flexible camera calibration by viewing a plane from unknown orientations, *Computer Vision*, 1999. The Proceedings of the Seventh IEEE International Conference on, vol. 1, Ieee, 1999.
- [38] Betta G., Capriglione D., Crenna F., Rossi G. B., Gasparetto M., Zappa E., Liguori C., Paolillo A. (2011), Face-based recognition techniques: proposals for the metrological characterization of global and feature-based approaches, *Measurement Science and Technology*, 22 (2012), 124005.
- [39] Xiaozheng Zhang, Yongsheng Gao, Face recognition across pose: a review, *Pattern Recogn.* 42 (11) (2009) 2876–2896.
- [40] Xing Chen, James Davis, Camera placement considering occlusion for robust motion capture, *Computer Graphics Laboratory*, Stanford University, Tech. Rep 2.2.2 (2000): 2.
- [41] Songxiang Gu, J.E. McNamara, M.O. Ward, Karen Johnson, Michael A. Gennert, M.A. King, Calibration Accuracy Evaluation for Stereo Reconstruction, *Nuclear Science Symposium Conference Record*, 2006. IEEE, vol. 6, October 29 2006–November 1 2006, pp. 3242, 3246.
- [42] Zhengyou. Zhang, A flexible new technique for camera calibration, *Pattern Anal. Machine Intell.*, IEEE Trans. 22 (11) (2000) 1330–1334.
- [43] Jean-Yves Bouguet, Camera calibration toolbox for matlab, 2004.
- [44] Giovanni B. Rossi, Birgitta Berglund, Measurement involving human perception and interpretation, *Measurement* 44 (5) (2011) 815–822.
- [45] Robert A. Harrison, *Handbook of Reliability Engineering*, in: Igor Alekseevich Ushakov (Ed.), Wiley, New York, 1994.
- [46] F. Crenna, G.B. Rossi, L. Bovio, Probabilistic characterization of face measurement, in: F. Pavese, M. Bar et al. (Eds.), *Advanced mathematical and Computational Tools in Metrology and Testing IX-Series on Advances in Mathematics for Applied science*, vol. 84, 2012, pp. 90–101, ISBN 978-981-4397-94-0.
- [47] Ramji M. Makwana, Vishvjit K. Thakar, Narendra C. Chauhan, Fuzzy measure based adaptive methods for illumination invariant face recognition, *Int. J. Comput. Intell. Appl.* 10 (03) (2011) 313–333.