Metrological characterization of 3D biometric face recognition systems in actual operating conditions

Giovanni Betta[1], Domenico Capriglione[3], Mariella Corvino[1], Alberto Lavatelli[2], Consolatina Liguori[3], Paolo Sommella[3], Emanuele Zappa[2]

[1]*DIEI, University of Cassino and of Southern Lazio Cassino (FR), Italy*
[2]*Department of Mechanical Engineering, Politecnico di Milano, Via La Masa, 1, Milano, Italy*
[3]*DIIn, University of Salerno, via Giovanni Paolo II, 132, Fisciano (SA), Italy*

ABSTRACT
Nowadays, face recognition systems are going to widespread in many fields of application, from automatic user login for financial activities and access to restricted areas, to surveillance for improving security in airports and railway stations, to cite a few.
In such scenarios, the architectures based on stereo vision and 3D reconstruction of the face are going to assume a predominant role because they can generally assure a better reliability than solutions based on a single camera (which make use of a single image instead of a couple of images). To realize such systems, different architectures can be considered by varying the positioning of the pair of cameras with respect to the face of the subject to be identified, as well as both kind and resolution of camera considered. These parameters can affect the correct decision rate of the system in classifying the input face, especially in presence of image uncertainty.
In this paper, several 3D architectures differing in camera specifications and geometrical positioning of the camera pair (with respect to the input face) are realized and compared. The detection of facial features in the images is made by adopting a popular method based on the Active Appearance Model (AAM) algorithm. 3D position of facial features is then obtained by means of stereo triangulation. The performance of the realized systems has been compared in terms of sensitivity to the quantities of influence and related uncertainty, and of typical indexes for the analysis of classification systems. Main results of such comparison show that the best performance can be reached by reducing the distance between cameras and subject to be identified and by minimizing the horizontal angle between the plane containing the camera pair axis and the face to be identified.

1. INTRODUCTION

Facial Recognition (FR) systems are well-known computer vision applications whose aim is to identify (or verify) a person upon the measurement of some selected facial features [1]. Moreover, FR systems exploit digital images and photogrammetry to provide contact-less identification [2], [3]. Those systems spread widely due to an increasing demand in the fields of security assurance and automatic identity

verification [4] - [6]. In fact, it is possible to find applications showing good performances and good maturity [7] when operating under constrained conditions, so that the person to be identified lies in a controlled environment and faces the camera(s) with a specified orientation. However, the extension of FR systems to the generic unconstrained environment (variable lighting, variable orientation of the person toward the camera, moving person…) is still problematic [8], [9], since

person identification occurs with high uncertainty and low repeatability, which is the main problem in the forensic context. In order to tackle this critical issue, the research activity focused on the analysis of human perception with the aim of describing mathematically the processes that lead to the identification of a person [10] and then to implement in FR algorithms able to operate in uncontrolled environments [11]. The main principle of these algorithms is to focus on the analysis of a set of biometric features rather than on the analysis of the full 3D shape of a face. This is done by processing the image(s) of a person with suitable feature matching and extraction algorithms prior to stereo triangulation.

The most of the biometric FR systems proposed in literature are based on the analysis of the appearance of frontal and non-frontal faces acquired in different instants, without metrological issues [2], [7], [12]. Therefore, it is not possible for those systems to deal with pose variation or the natural variation of face expression. Conversely, the authors proposed a system that provides a fusion between the point-cloud 3D reconstruction and the pure image feature analysis by using a stereoscopic vision rig [13], [14]. The stereoscopic setup acquires synchronously from several different calibrated cameras the images of a person's face, so that it is possible to triangulate the feature points and express the biometric characteristics in the 3D space using couples of cameras grabbing the face from different orientation. Using different couples of synchronized stereo cameras, it is possible to evaluate the effect of the pose variability without the interaction of face expression variation, since images are acquired in synchrony from stereo pairs having different orientations towards the face.

In any case, the process of image formation is yet complex and stochastic, so that the generic image of the face is affected by an uncertainty contribution that propagates throughout all the processing stages. In addition, the final classification result is uncertain, up to a point that it may be impossible to find a perfect matching between images of the same person acquired in different moments [11]. Hence, there is a clear risk in accepting the generic classification results without a proper analysis of the uncertainty [15], [16].

In this context, the scientific literature proposes a good variety of approaches in order to limit or reduce the risk of an incorrect classification by the formulation of a priori uncertainty models [2], [8], [17] - [20]. In this way, given a set of uncertainty sources, it is possible to solve the problem in a Bayesan framework so that the most probable matching can be computed. However, these approaches have a limit since they can deal only with the supposed uncertainty sources, whereas the actual image can be characterized by other sources of uncertainty, which could be not included in the analysis model. This situation brings to a general overestimation or underestimation of uncertainty. Nonetheless, uncertainty propagation is not performed in the way advised by the ISO-GUM standard.

In addition, the evaluation of the performance of FR systems is still an open problem. Typically, this process relies on the estimation of the recognition reliability from a given database of biometric images. Furthermore, recognition reliability is judged with the help of synthetic indexes that express the probability of having a false positive or a false negative [16], [21], [22]. If perfectly run, this procedure is able to estimate the performance only for a given set of faces, thus it is hard to estimate the inference of such a result in the generic case. Oppositely, the authors proposed in previous papers a

metrological approach to the problem of characterizing FR systems [14], [23]. This has been done by analyzing the physical relations between the uncertainty in the final classification result and the uncertainty of input influence quantities [24]. The experimental activity demonstrated that the performance of these kinds of systems depends on several aspects from image acquisition to the classification procedure through the biometrical algorithm. The quantification of these uncertainty contributions lead the authors to formulate an original classification method which was able to improve the classification performance with respect to traditional score-based approaches [25], [26].

Even if the authors were able to propose an accurate metrology based method to evaluate the performances of FR systems, there are still open question in the design phase: what is the optimal layout of new FR systems? Does the face recognition algorithm interfere with the selection of the optimal vision rig? In literature, different approaches have been presented [27] - [29] to optimize the generic vision system, but only little is known on how to choose an architecture for a specific problem.

In this scenario, the aim of this work is to compare, from a metrological point of view, different architectures of a 3D system for face recognition. In particular, starting from the approaches proposed in [23]- [26] for evaluating the confidence level of the classifier output, a detailed study of the influence on the metrological performance of different camera configurations (in terms of different position with respect to the subject), different kind of cameras (in terms of sensor and resolution), and different poses (in terms of subject face rotation with respect to the couple of cameras) is reported in this paper. The comparison is based upon the analysis of uncertainty in both feature extraction and the classification process. This study is claimed to provide useful indications for system designers, whose task in practical applications is to select the best compromise between the system performances (in terms of correct classification rate), geometrical constraints (i.e. vision rig arrangement) and hardware constraints (i.e. cost of optics or acquisition hardware).

In the following, after a brief recall of the developed biometric system, the description of the processing for face recognition and the comparison of the metrological performance of different measurement configurations are shown.

2. THE BIOMETRIC SYSTEM

2.1. System for the images acquisition

As mentioned before, the image database was created with a peculiar multiple stereovision rig. The setup is described schematically in Figure 1 and consists of an aluminum frame carrying a stereovision matrix composed by 3 stereo pairs having 3 different nominal orientations toward the face: 0°, 5° and 10°. The cameras of the first row of the matrix have an attitude of −22.5° towards the observer, while the second row cameras have an attitude of +22.5° towards the observer. This choice leads to a relative angle (incidence) of 45° between each camera of the pair, which is a good compromise between accuracy in depth measurement (which increases with increasing incidence [30]) and avoiding the risk of view occlusion (which, conversely, is more probable with higher incidence of the stereo pair [31]). In addition, a seventh camera has been positioned at mid-height (cam 7 in Figure 1b),
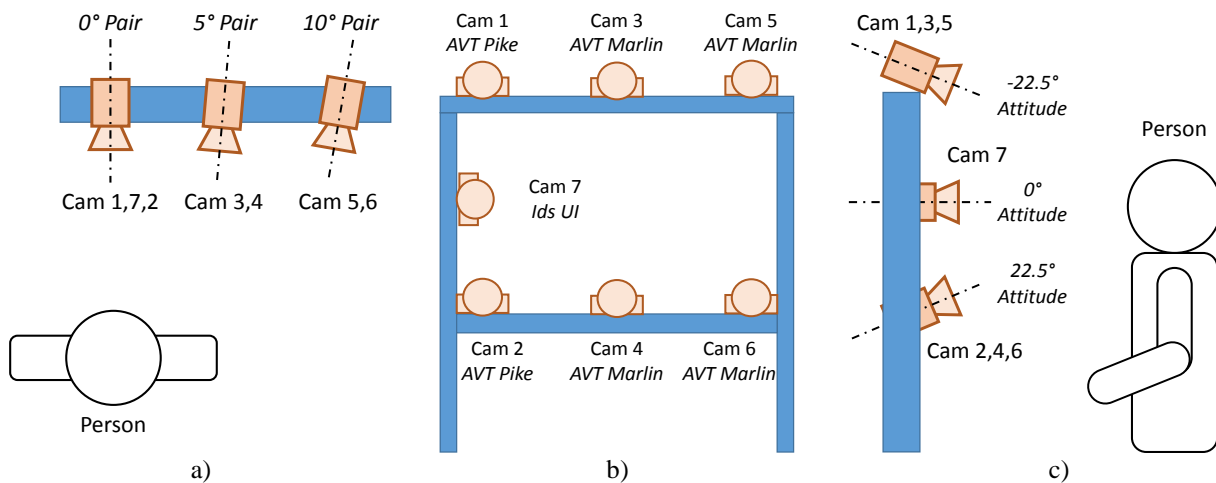
Figure 1. Design scheme of the vision rig used for FR purposes: a) upper view b) front view c) lateral view.

between the two rows with zero orientation and zero attitude, to collect passport-like photos that can be used for the validation of 2D recognition techniques. The configuration of the vision rig allows the field of view size to be approximately 300~400 mm, that means face registration can be performed with the subject standing about 1000 mm away from the cameras.

Stereoscopic 3D point reconstruction is obtained by coupling the view of each top camera with the corresponding bottom camera. The cameras have the characteristics listed in Table 1 and are all equipped with 25 mm focal length fixed optics. Camera calibration is performed pair-by-pair with the well-known Zhang method [32] by means of the Camera Calibration Toolbox from J.Y. Bouget. Image acquisition is performed with 3 IEEE1394 PCI adapters (for cameras 1 to 6) and a Gigabit Ethernet adapter (for camera 7). The synchronization of acquisition between the 7 cameras is provided by a common rising edge trigger source. Exposure time and aperture are regulated in order to have similar luminance with similar depth of focus. With the help of a LabVIEW Virtual Instrument, the images of all cameras are simultaneously recorded in bitmap format. The control system acquires sequences of images with a user-defined interval (default set 5 s), in order to obtain multiple images of the same person in about the same position but with small variation of face expression that naturally happens between each capture.

2.2. Biometric algorithm

The biometric algorithm considered is composed by two main steps:

a) Segmentation of the biometric features in the images acquired by the upper and lower camera with the help of an Active Appearance Model (AAM) algorithm [33].

b) Triangulation of the 2D points of the biometric features to retrieve a 3D mask representation of the biometric features.

Table 1. Characteristics of the vision equipment.

Camera ID	Model	Sensor
1, 2	AVT – Pike F-145B	1388X1038 px color CCD progressive
3, 4, 5, 6	AVT – Marlin F-131B	1280X1024 px CMOS sensor
7	IDS – GigE UI-5490SE-M	3840X2748 px CMOS sensor

The main phase, a), is the segmentation of 2D features out of a generic image of a human face. The scientific literature proposes several ways to perform this task and a good review can be found in [1] and [2]. The authors chose the AAM algorithm due to its effectiveness in matching both the shape and appearance of a given pattern [34]. The shape itself is defined as a 2D points set describing the geometry of a target body [35]. The AAM algorithm is able to deal with shape modification since it uses a dynamic model of shapes in the form of Principal Component Analysis (PCA), so that any given shape x_S can be expressed in the form:

$$x_S = \bar{x} + [\phi] \cdot b , \qquad (1)$$

where $\bar{x}$ represents the average shape, $[\phi]$ represents the matrix of principal components $[\phi_1|\phi_2|...|\phi_n]$ that express deformation modes and b a vector of real numbers that set the model deformable shape parameters.

The appearance is defined as the texture (a map of level of gray or color) of a portion of the target. The AAM algorithm deals with appearance variation with appearance PCA models that arrange all pixel intensity variations of the images around the mean shape. The PCA formulation gives for the general appearance vector g_a:

$$g_a = \bar{g} + [\phi_g] \cdot b_g , \qquad (2)$$

where $\bar{g}$ represents the average appearance, $[\phi_g]$ represents the matrix of principal components $[\phi_{g,1}|\phi_{g,2}|...|\phi_{g,n}]$ which express luminance modes and b_g a vector of real numbers that set the models variable luminance parameters.

Shape and Appearance models are based on the analysis of training images. So, once the model formulation is set, it is necessary to submit a sample of images where facial features have been manually annotated in order to define the Principal Components $[\phi]$ and $[\phi_g]$. The annotation process consists in tracing univocal landmarks that describe the most important facial traits on each image. In this work, the biometric description involves 58 landmarks that define the shape of jaw, mouth, nose, eyes, and eyebrows. The choice of the 58 points to be matched agrees with methodologies documented in other studies [36], [37], in order to be able to compare data.

Once the AAM is trained, it is able to find biometric landmarks in the generic image of one's face. Given the high difference in attitude, it has been necessary to train a model for the upper camera and a model for the lower camera. The

software implementation is provided by an Application Programming Interface dedicated to AAM provided by [38], [39].

Subsequently, for the generic face it is possible to match a set of points in the stereo pair and then to triangulate them with the epipolar constraint [28]. Eventually a 3D mask of a face is recorded.

Figure 2 schematizes the main phases for the realization of the 3D mask.

2.3. The identity database

The image acquisition system described before is then ready to be tested. The first step consists in the acquisition of a suitable database of facial images. The database was built by acquiring images of 117 volunteers, using the vision rig described in Section 2.1. For each of the 117 individual three series of images have been recorded:

- One set with subject's face oriented towards the point in the middle between cam1 and cam2 (0° neck rotation);
- One set with 10° neck rotation on the right hand side;
- One set with 20° neck rotation on the right hand side.

Neck rotation is controlled by asking the person to look toward a point that has been fixed at the wanted orientation with respect to the 0° neck rotation. In this sense, the values of neck rotation used in this work are to be considered as approximate values, with the only aim to get indications on the variation of the recognition accuracy with the neck rotation.

Each set contains 9 pictures of the same subject in the same position, acquired with a rate of 1 images every 5 s. In this way it is possible to record natural minor changes in head position and in facial expression and let the AAM model cover them. On the whole, 162 images for every person have been acquired. The position of the person with respect to the vision rig, as well as the relative distance between cameras and face, has not been controlled. This was chosen on purpose, since the aim of this work is to evaluate the performance of the system in operating conditions: modern identity verification systems, in fact, do not require constraining the person to a rigid frame and most of them work in completely unconstrained environments [3].

As previously underlined, AAM models should be trained properly upon a set of images, which is supposed to sample correctly the actual variability of face appearance. Therefore, as the number of facial images used for the training increases, the AAM model is able to cover a broader range of shape and texture variations. However, when the number of images used for training grows, the handling of such a great size of information becomes hard, the weight of random effects over deterministic phenomena becomes higher and the final reliability of face recognition becomes lower [13]. A good compromise has been found building a model using 200 images from 50 people belonging to the database of 117 individuals. Note that images from all the couples of cameras and with the 3 different head rotation (0°, 10° and 20°) were used to create the model.

It is not convenient to train the AAM model with the same images that will be used in personal identification (i.e. to test the accuracy of the results). Because of this reason, images used for AAM model training were not included in the database used for the evaluation of the performances of the FR.

3. PROCESSING FOR FACE RECOGNITION

3.1. Evaluation of scores

In order to perform identity verification, the 3D mask of a person is compared with each of the 3D masks included in an identity database. The comparison outputs a value named *score*, which is computed by means of the weighted sum of squared differences between a collection of masks stored in an identity database. Therefore, for the *i-th* mask stored in the database the *i-th* score, S_i is given by:

$$S_i = \frac{\sum_{k=1}^{n}(W_k*(V_{k,i}-V_{k,ref})^2)}{n} , \qquad (3)$$

where, $V_{k,i}$ are the coordinates of the *k-th* point for the *i-th* individual, W_k is the weight of the *k-th* point of the mask. Then, the score S_i represents the sum of squared discrepancy between the 3D coordinates of the mask to be recognized and the corresponding coordinates of each mask in the database. Prior to the evaluation of point-to-point distances ($V_{k,i} - V_{k,ref}$), a roto-translation is computed in order to move the coordinate frame of one mask to be onto the coordinate frame of the other mask with a rigid motion. The roto-translation allows the compensation of differences in position and orientation of the subject with respect to the stereoscopic system in acquisition. The analysis of the variance of the identification parameters calculated on identifying the same person at different neck rotation conditions generates the weights W_k. The procedure for weight estimation is: images of the same individual are submitted to FR, for the *k-th* landmark the variance Var_k of
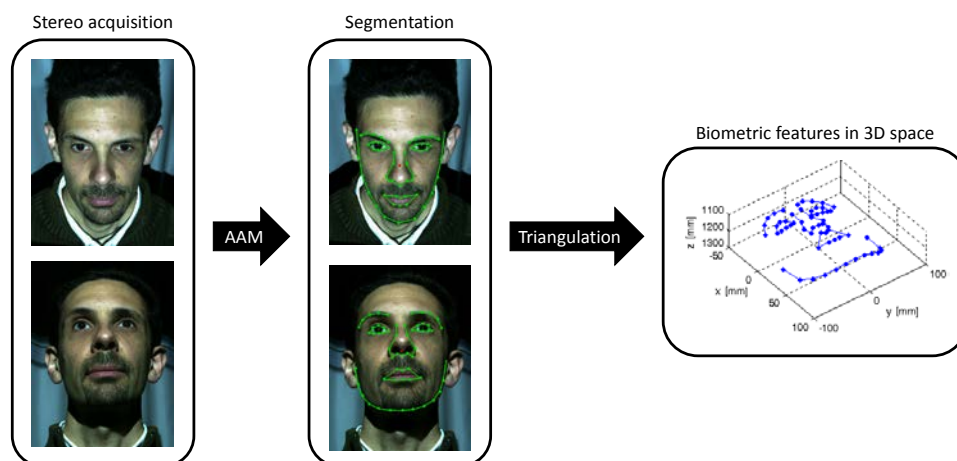

Figure 2. The schematic of biometric algorithm with featured detected on stereoscopic images and the obtained 3D mask.

the 3D coordinates is estimated (Var_k is an estimate of identification uncertainty), the weight for the k-*th* landmark is defined as the inverse of Var_k [13].

3.2. Sources of score variability and bias

As previously described, the evaluation of S_i (by means of (3)), results from a suitable sequence of operations performed starting on the input images. In particular, they are:

i. 2D image acquisition;
ii. 2D images processing by means of AAM algorithm;
iii. triangulation;
iv. roto-translation.

Each one of these operations could introduce uncertainty and/or bias on the S_i estimate [24], [40].

As for step *i*, the acquisition conditions, in terms of luminance, lens defocus, presence of motion on the image itself (due for example to vibration of the system during the image acquisition or to the subject movement), and of face pose angle can modify the images with respect to ones contained in the identity database, thus changing the positions of the landmarks on 2D pair of images, and consequently on the resulting 3D mask. As an example, Figure 3 shows such effects for the case of luminance, motion blur and lens defocus: the interpolation of the landmarks significantly changes as the motion blur or lens defocus is present, whereas it is weakly influenced by luminance. As a consequence, it is expected that S_i will be affected by such quantities of influence, in particular by motion blur and lens defocus. As an example, Figure 4 reports the scores evaluated at different levels of lens defocus and when the subject #1 of the identity database and the pair cam1-2 are considered. For a sake of clarity, the scores evaluated considering only the first seven subjects (of the identity database) have been reported. As you can see, if the level of
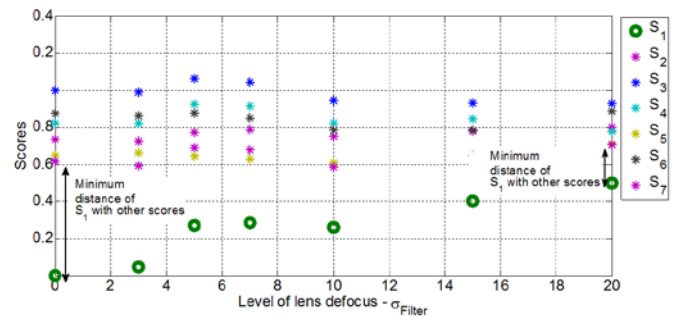
Figure 4. Effects of the lens defocus on the scores.

lens defocus increases, the value of the score, S_1, increases, whereas the distances between the classes (subjects of the identity database) decrease, thus making worse the system performance in terms of correct decision rate. Similar trends have been observed when other input subjects or other pairs of cameras (3-4 and 5-6) are considered.

Then, the image uncertainty propagates in the next step (step *ii*). With reference to the AAM algorithm, the uncertainty in the building of the Shape Model and the Appearance Model [20] determines the accuracy of feature localization. Since the weights, W_k, of (3) are defined in the training phase on the basis of a number of a pair of images for each of all subjects, an intrinsic score variability should be considered to take into account such aspect. As an example, Figure 5 reports, for a subject of the database, the nine images used for the training of the AAM, together with the scores. The first image on the left in Figure 5 is used as a reference.

As for step *iii*, the triangulation is realized by means of the method proposed by Zhang [27] and a residual error is mainly due to the non-ideality of the calibration phase.

As for step *iv*, the roto-translation is made by means of an iterative procedure and the residual error is mainly linked to the noise on the 3D features of the masks.

3.3. Uncertainty modelling for Score

The effects of triangulation (step *iii*) and roto-translation (step *iv*) are here modelled by means of residual systematic *Score* greater than zero also for the correct class, whilst the measurement uncertainty introduces variability on the *Score* of the correct class that strictly depends on the image acquisition condition (step *i*) and on the accuracy of AAM (step *ii*). Therefore, in the following, all the systematic effects are directly included in the measured *Score*, whilst only the measurement uncertainty (due to the acquisition process and to the accuracy of AAM) is considered.

As for the uncertainty component related to step *i*, a statistical approach [26] is considered to estimate the uncertainty model corresponding to each quantity of influence. In particular, as for the quantities of influence luminosity, lens defocus and motion blur, new couples of artificial images are used to achieve further images characterized by the desired values of the quantities of influence. The new couples of images were generated by applying suitable digital filtering on the reference images (i.e. the ones contained in the identity database, hereinafter denoted as *database A*). For each pair of cameras, several values of variation (with respect to the reference image contained in the identity database) have been considered in order to generate the new images (see Table 2). The variations were applied to luminosity (11 different values), lens defocus (7 different values) and motion blur (11 different
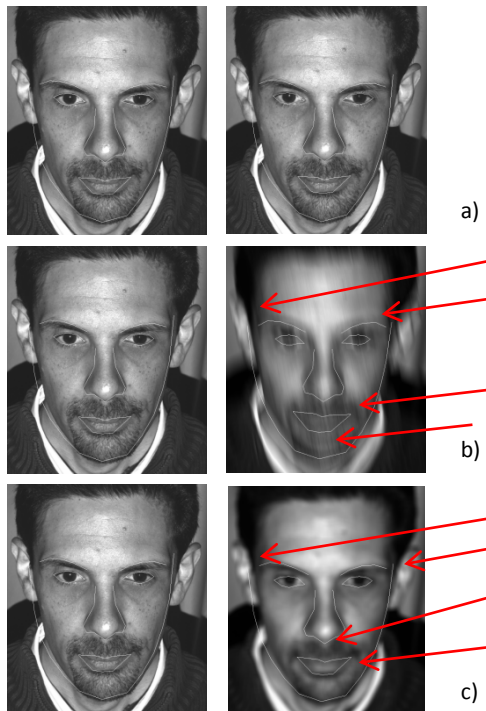

Figure 3. Effects of the acquisition conditions (the red arrows point the areas mainly affected by the quantity of influence):
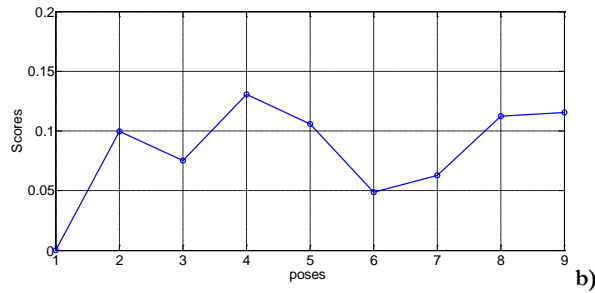a) luminance, b) motion blur, c) lens defocus.

Figure 5. Definition of W_k: a) Images acquired by Camera 1 for subject #1, b) Scores evolution (the first image of the subject is used as reference).

values). As for the image luminosity, the values considered are referred to a 8-bit grey scale, whereas, as for lens defocus and motion blur the values considered are referred to the standard deviation (in pixels) of the Gaussian filter employed.

On the contrary, as for the pose face angle, for each subject, the images acquired at different neck rotation (see Section 2.3) were considered, obtaining the image types summarized in Table 2.

In this way new databases (one for each pair of cameras, 1-2, 3-4, and 5-6, hereinafter denoted as *database B*) of images designed for the uncertainty estimation are achieved.

Then, for each considered condition, the procedure adopted for uncertainty estimation for each pair of cameras is the following:

- for each subject of these new databases, the 3D features are evaluated applying the AAM algorithm and the triangulation;
- the *Scores* with respect to the recorded data in the identity database of the subject itself is estimated;
- on the *Scores* obtained for all the subjects, a statistical analysis is made, and the uncertainty is estimated according to the ISO-GUM, by using the simple model described in the following equation.

$$(u_S)_l = \sqrt{\frac{(\mu_S^2)_l}{3} + (\sigma_S^2)_l} \ , \tag{4}$$

where, $(u_S)_l$ is the contribution due to the l-th quantity of influence on the *Score* uncertainty, $(\mu_S)_l$ and $(\sigma_S)_l$ are respectively the mean and the sample standard deviation of the measured *Scores* related to the *database B*. Then, for each quantity of influence, the identification of the relationship between $(u_S)_l$ and the value of l can be based on a large experimental analysis and on the fitting of simple models onto

Table 2. Values of variations of the quantities of influence with respect to the reference images of the identity database.

Quantity of Influence	Values	Units
Luminosity	[-30, -10, -3, 0, +3, +10, +30]	Grey levels
Lens defocus (σ_{Filter})	[0, 3, 5, 7, 10, 15, 20]	Pixels
Motion blur (σ_{Filter})	[0, 3, 9, 15, 17, 19, 23, 27, 30, 50, 70]	Pixels
Pose angle	[0, 10, 20]	Degree

experimental data [26].

As for the uncertainty related to step *ii*, $(u_S)_{AAM}$, we pose:

$$(u_S)_{AAM} = \sqrt{\frac{(\mu_S^2)_{AAM}}{3} + (\sigma_S^2)_{AAM}} \ , \tag{5}$$

where $(\mu_S)_{AAM}$ and $(\sigma_S)_{AAM}$ are respectively the mean and the sample standard deviation of all of the measured *Scores* achieved for each subject when the nine images (related to the *database A*) are considered.

As for the overall uncertainty on the scores, u_S, all of the quantities of influence are considered uncorrelated with the other ones, then the combined uncertainty on the score is evaluated as:

$$u_S = \sqrt{\sum_{l=1}^{M}(u_S)_l^2 + (u_S)_{AAM}^2} \ , \tag{6}$$

where M is the number of the considered quantities of influence. Applying these models, the uncertainty of the *Scores* is evaluated. The hypothesis made of uncorrelated quantities has been also verified through experiments.

3.4. Classification procedure

Starting from the *Scores* and the related uncertainty, u_S, estimated as described in the previous sections, the classification procedure provides the subject identification and the corresponding related confidence level (CL). The output of such a procedure is a classification list in which all the possible identified subjects are included with their CL.

To these aims, at first, the probability that the input subject belongs to each *j-th* class, P_j, is evaluated; then, on the basis of the obtained probability, the classification list is created with a selection of the probable classes; and finally, the confidence level of each class in the list is evaluated. In more detail, the decision algorithm (see Figure 6) is composed by three steps [23], [25]:

- STEP A: for each subject, j, present in the database, the measured Score, S_j, is used together with the corresponding uncertainty, u_S, in order to evaluate the probability, P_j, that the input subject is the subject j. In particular, P_j represents the probability that the Score of the *j-th* subject is equal to zero given a measured value (S_j), considering the score as a random variable (see Figure 6), whose standard deviation is equal to u_S [25];

STEP A

STEP B: The classification list is composed only by the subjects with a probability, P_j, greater than a threshold, **TH**.

STEP C: $K = \sum P_j;$ $\quad CL_j = \dfrac{P_j}{K}$

ith subject with CL_i

jth subject with CL_j
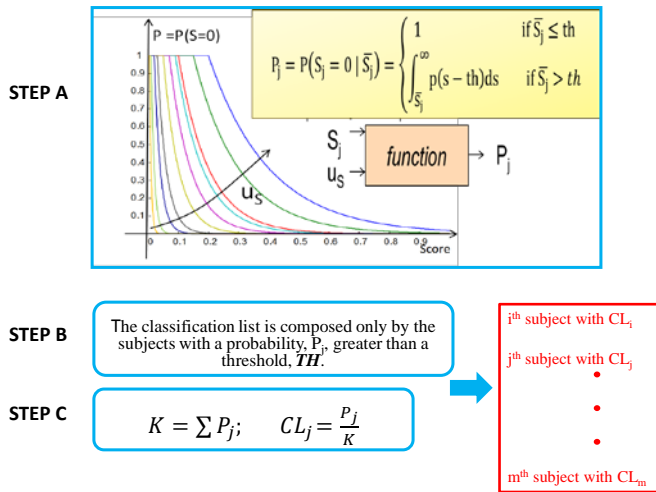
•
•
•

mth subject with CL_m

Figure 6. Main steps of the classification procedure.

- STEP B: using all the estimated probabilities the classification list is created including only the subjects whose probability is greater than a threshold, TH;
- STEP C: the confidence level, CL_j, of the subjects in the classification list is evaluated as $CL_j = P_j/K$ where K is equal to the sum of all the P_j greater than TH.

As for the selection of TH, it should be chosen as best trade-off between sensitivity and selectivity required by the specific application. Indeed, it is expected that the higher is TH, the higher is the selectivity and the lower is the sensitivity of the system. Therefore, as an example, on the basis of the requirements of the application in terms of True Acceptance Rate (TAR evaluated as True Positive/all positive) and of the false acceptance rate (FAR evaluated as False Positive/all negative) the value of TH could be selected [26].

4. PERFORMANCE COMPARISON

In this section, the comparison of the three considered system architectures (i.e. Cam1-2, Cam3-4, Cam5-6) is reported. In particular, at first, the comparison of all uncertainty components involved in (4) and (5) is made. Then, the recognition performance of the systems is presented by comparing suitable figures of merit, typically adopted for the analysis of classification systems [40].

4.1. Score uncertainty estimation

In this section the uncertainty components related to the acquired 2D images (luminosity, lens defocus, motion blur and pose angle) and to the accuracy of AAM are reported for each pair of cameras.

In particular, Figure 7 shows the value of $(u_S)_l$ calculated by means of (4) and Table 3 reports the value of $(u_S)_{AAM}$ estimated by means of (5).

Looking at Figure 7 and Table 3, some considerations can be drawn:
- For each pair of cameras, the image luminosity is the quantity of influence corresponding to the lowest uncertainty value ($(u_S)_{LUM}$), whereas the pose angle represents the quantity responsible of the highest uncertainty values ($(u_S)_{ANGLE}$).
- Except for the pose angle, for each quantity of influence and whatever be its deviation from the corresponding

value on the reference image, the lowest value of uncertainty is achieved with the pair Cam1_2, whereas the pair Cam5_6 shows always the highest value of uncertainty.

These results are mainly due to the smaller distance (from the subject to be identified) and orientation of the pair Cam1_2 with respect to the other pairs (see Figure 1).
- As for the pose angle, pair Cam1_2 shows the worst performance. These results are due to the different orientation of camera pairs from the subject to be identified. In particular, the pair Cam1_2 is the most well aligned with the person to be identified, thus it results as the more sensitive to the pose angle.
- The uncertainty due to the AAM algorithm, $(u_S)_{AAM}$, is comparable with the uncertainty due to the luminosity of the acquired images ($(u_S)_{LUM}$, while it is lower than the
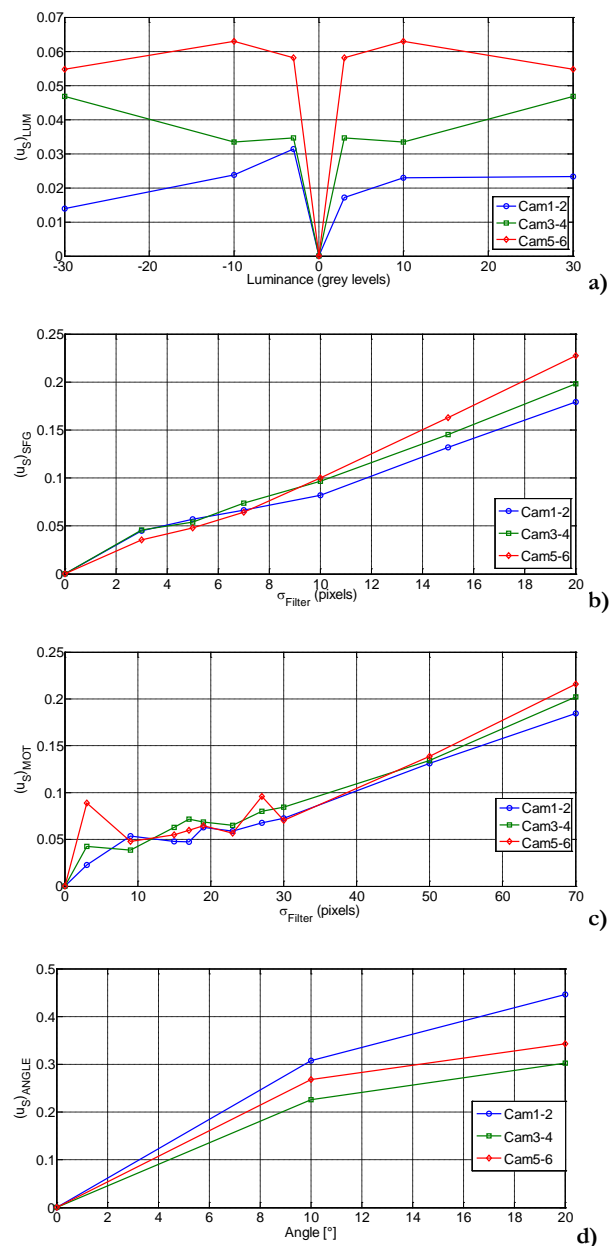

Figure 7. $(u_S)_l$ versus the quantities of influence:
a) luminosity, b) lens defocus, c) motion blur, d) pose angle.

Table 3. Values of $(u_S)_{AAM}$ for the pairs of cameras.

	Cam1_2	Cam3_4	Cam5_6
$(\mu_S)_{AAM}$	0.140	0.139	0.17
$(\sigma_S)_{AAM}$	0.006	0.005	0.02
$(u_S)_{AAM}$	0.080	0.080	0.10

uncertainty due to the other quantities of influence.

4.2. Classification performance

In order to evaluate and compare the overall classification performance of the three system architectures, the following figures of merit have been considered.

- *Correct classification (CC)*: the right class is identified with the highest CL value;
- *Abstention (AB):* the right class is in the classification list but there are some other classes with the same *CL*, thus the system does not provide any decision;
- *Missed classification (MC)*: the classification list is empty;
- *Wrong classification (WC)*: either the correct class is not in the classification list or it is present without the highest value of CL.

These performance indexes have been evaluated by considering the following values of TH: 0.3, 0.5, 0.7. In this way, it is possible to compare the three system architectures for different degrees of selectivity and sensitivity.

The recognition procedure has been run by considering the images of *database B*. Table 4 compares the considered figures of merit for different values of TH and camera pairs.

As expected, whatever the pair of cameras, the value of TH affects the values of the figures of merit. In particular:

- the lower is TH, the lower is the MC percentage and the higher are the AB and WC percentages;
- the best performance in terms of CC is observed for the considered intermediate value of TH.

In addition, for a given value of TH:

- Even if pairs Cam1-2 and Cam3-4 are based on different hardware (see Table 1), they show similar performance in terms of all the considered figures of merit. Therefore, it seems that the type of sensor weakly influence the performance of the system.
- Pair Cam5-6 shows the best performance in terms of CC and MC only when TH=0.7, whereas, it seems the

Table 4. Performance comparison of the pairs of cameras for different TH.

TH	Figure of merit	Cam1_2	Cam3_4	Cam5_6
0.3	CC [%]	84.7	82.3	70.9
	AB [%]	3.3	3.6	3.4
	MC [%]	1.0	2.3	2.3
	WC [%]	11.0	11.8	23.3
0.5	CC [%]	88.0	85.7	87.5
	AB [%]	1.1	1.1	1.6
	MC [%]	7.6	10.0	6.2
	WC [%]	3.3	3.1	4.6
0.7	CC [%]	77.7	76.8	84.2
	AB [%]	0.2	0.5	0.7
	MC [%]	21.8	22.3	14.4
	WC [%]	0.3	0.3	0.7

worst solution for TH=0.3 and TH=0.5. Therefore, since the pairs Cam3-4 and Cam5-6 share the same hardware (see Table 1), the different performance can be imputable to their distances from the input subject (in the case of Cam5-6 such a distance is the largest one).

Consequently, depending on the performance constraints required by the application or the trade-off to be satisfied, a designer/user can select the most suitable solution in terms of both arrangement of pair of cameras and value of TH.

5. CONCLUSIONS

The paper has compared the metrological performance of different architectures for face recognition based on 3D features. The study has been conducted by considering a popular algorithm, the AAM, for facial feature detection and then stereo triangulation for 3D position measurement and by considering the classification procedure proposed by authors in previous papers. The work took into account the main causes of uncertainty generally affecting the performance of face recognition systems with a direct impact on the reliability of the final decision in the classification stage in terms of Correct Classification, Abstention, Missed Classification and Wrong Classification percentages.

In particular, the following main results can be drawn:

- for the considered sensor resolutions the kind of hardware does not influence the metrological and the classification performance of the system;
- the distance between the subject (to be recognized) and the pair of cameras generally affects the metrological performance of the system. In particular, the pair Cam1_2 shows slightly better performance in terms of uncertainty on the score due to the variation of image luminosity, lens defocus and motion blur (see Figures 7a-7c);
- the horizontal angle between the plane containing the camera pair axis and the face to be identified, generally affects the metrological performance of the system in terms of uncertainty on the score (see Figure 7d). Such phenomenon is more evident for the pair Cam1_2 because it is the pair most well aligned with the person to be identified, thus it is more sensitive to the pose;
- as for the classification performance, the best performance is shown by pairs Cam1_2 and Cam3_4 if low values of TH are considered (i.e. highest selectivity and the lowest sensitivity), whereas the pair Cam5_6 shows best performance if high values of TH are considered. Such results confirm that the distance between cameras and the subject affects the performance, and that a designer could also prefer the pair Cam5_6 if the specific application constraints require the lowest selectivity and the highest sensitivity (i.e. a high value of TH).

The quantification of such kind of information could be very useful also for system designers and/or users, which, in practical applications, have often to select the best trade-off between camera arrangements and system performance.

Future developments will investigate the possibility of exploiting more pairs of camera at a time with the aim of fusing the information extracted, looking for improving the system classification performance.

REFERENCES

[1] S. Li e A. Jain, Handbook of Face Recognition, Springer London, 2011.

[2] D. Yadav e A. Bhattacharya, "Face identification methodologies in videos", Communication Technologies (GCCT), 2015 Global Conference on, 2015.

[3] A. Danelakis, T. Theoharis e I. Pratikakis, "A survey on facial expression recognition in 3D video sequences," Multimedia Tools and Applications, vol. 74, n. 15, pp. 5577-5615, 2015.

[4] D. White, P. Jonathon Phillips, C. Hahn, M. Hill e A. O'Toole, "Perceptual expertise in forensic facial image comparison," Proceedings of the Royal Society B: Biological Sciences, vol. 282, n. 1814, 2015.

[5] Z. Sufyanu, F. Mohamad e A. Ben-Musa, "A proposed integrated human recognition for security reassurance," American Journal of Applied Sciences, vol. 12, n. 2, pp. 155-165, 2015.

[6] Y. Park, G. Joung, Y. Song, N.-S. Yun e J. Kim, "A study on the safety management system of a passenger ship using biometrics," Journal of Nanoelectronics and Optoelectronics, vol. 11, n. 2, pp. 194-197, 2016.

[7] M. Hassaballah e S. Aly, "Face recognition: Challenges, achievements and future directions," IET Computer Vision, vol. 9, n. 4, pp. 614-626, 2015.

[8] M. Haghighat, M. Abdel-Mottaleb e W. Alhalabi, "Fully automatic face normalization and single sample face recognition in unconstrained environments," Expert Systems with Applications, vol. 47, pp. 23-34, 2016.

[9] H. Yang e I. Patras, "Mirror, mirror on the wall, tell me, is the error small?," 2015.

[10] R. Larsen, K. Hilger, K. Skoglund, S. Darkner, R. Paulsen, M. Stegmann, B. Lading, H. Thodberg e H. Eiriksson, "Some issues of biological shape modelling with applications," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 2749, pp. 509-519, 2003.

[11] M. Vidhyalakshmi e E. Poovammal, "A survey on face detection and person re-identification," Advances in Intelligent Systems and Computing, vol. 410, pp. 283-292, 2016.

[12] C. Wöhler, 3D Computer Vision: Efficient Methods and Applications, Springer London, 2012.

[13] E. Zappa, R. Testa, M. Barbesta e M. Gasparetto, "Uncertainty of 3D facial features measurements and its effects on personal identification," Measurement: Journal of the International Measurement Confederation, vol. 49, n. 1, pp. 296-307, 2014.

[14] G. Betta, D. Capriglione, M. Corvino, C. Liguori e A. Paolillo, "Face based recognition algorithms: A first step toward a metrological characterization," IEEE Transactions on Instrumentation and Measurement, vol. 62, n. 5, pp. 1008-1016, 2013.

[15] D. White, J. Dunn, A. Schmid e R. Kemp, "Error rates in users of automatic face recognition software," PLoS ONE, vol. 10, n. 10, 2015.

[16] Bundesamt für Sicherheit in der Informationstechnik, "Study: An investigation into the performance of facial recognition systems relative to their planned use in photo identification documents – BioP I" Public final report. Available on line: http://www.bsi.bund.de.

[17] M. De-la-Torre, E. Granger, R. Sabourin e D. Gorodnichy, "An adaptive ensemble-based system for face recognition in person re-identification," Machine Vision and Applications, vol. 26, n. 6, pp. 741-773, 2015.

[18] A. Punnappurath, A. Rajagopalan, S. Taheri, R. Chellappa e G. Seetharaman, "Face Recognition Across Non-Uniform Motion Blur, Illumination, and Pose," IEEE Transactions on Image Processing, vol. 24, n. 7, pp. 2067-2082, 2015.

[19] Y.-T. Chou, S.-M. Huang e J.-F. Yang, "Class-specific kernel linear regression classification for face recognition under low-resolution and illumination variation conditions," Eurasip Journal on Advances in Signal Processing, vol. 2016, n. 1, pp. 1-9, 2016.

[20] A. Fathi, P. Alirezazadeh e F. Abdali-Mohammadi, "A new Global-Gabor-Zernike feature descriptor and its application to face recognition," Journal of Visual Communication and Image Representation, vol. 38, pp. 65-72, 2016.

[21] S. A. Rizvi, P. Phillips e H. Moon, "Verification protocol and statistical performance analysis for face recognition algorithms," 1998.

[22] P. Wang, Q. Ji e J. Wayman Jr., "Modeling and predicting face recognition system performance based on analysis of similarity scores," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, n. 4, pp. 665-670, 2007.

[23] G. Betta, D. Capriglione, M. Gasparetto, E. Zappa, C. Liguori e A. Paolillo, "Managing the uncertainty for face classification with 3D features," In proceedings of 2014 IEEE International Instrumentation and Measurement Technology Conference (I2MTC 2014), p. 412-417, Montevideo, Uruguay, 12-15 May, 2014.

[24] G. Betta, D. Capriglione, M. Corvino, M. Gasparetto, E. Zappa, C. Liguori e A. Paolillo, "Metrological performance comparison of biometric system architectures for 3D face recognition," XXI IMEKO World Congress "Measurement in Research and Industry", September 4, 2015, Prague, Czech Republic.

[25] G. Betta, D. Capriglione, M. Corvino, C. Liguori e A. Paolillo, "A proposal for the management of the measurement uncertainty in classification and recognition problems," IEEE Transactions on Instrumentation and Measurement, vol. 64, n. 2, pp. 392-402, 2015.

[26] G. Betta, D. Capriglione, M. Gasparetto, E. Zappa, C. Liguori e A. Paolillo, "Face recognition based on 3D features: Management of the measurement uncertainty for improving the classification," Measurement: Journal of the International Measurement Confederation, vol. 70, pp. 169-178, 2015.

[27] S. Zhang, Handbook of 3D Machine Vision: Optical Metrology and Imaging, CRC Press, 2013.

[28] R. Hartley e A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, 2004.

[29] H. Wechsler, Reliable Face Recognition Methods: System Design, Implementation and Evaluation, Springer US, 2009.

[30] D. Szalóki, K. Csorba e G. Tevesz, "Optimizing camera placement in motion tracking systems," 11th International Conference on Informatics in Control, Automation and Robotics, ICINCO 2014, 2014.

[31] X. Chen e J. Davis, "An occlusion metric for selecting robust camera configurations," Machine Vision and Applications, vol. 19, n. 4, pp. 217-222, 2008.

[32] Z. Zhang, "A flexible new technique for camera calibration," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 22, n. 11, pp. 1330-1334, 2000.

[33] T. Cootes, G. Edwards e C. Taylor, "Active appearance models," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 1407, pp. 484-498, 1998.

[34] G. Edwards, T. Cootes e C. Taylor, "Advances in Active Appearance Models," Proceedings of the 1999 7th IEEE International Conference on Computer Vision (ICCV'99), 1999, Kerkyra, Greece.

[35] G. Edwards, C. Taylor e T. Cootes, "Interpreting face images using active appearance models," 3rd IEEE International Conference on Automatic Face and Gesture Recognition, FG 1998, Nara, Japan.

[36] E. Zappa, P. Mazzoleni e Y. Hai, "Stereoscopy based 3D face recognition system," 10th International Conference on Computational Science 2010, ICCS 2010; Amsterdam, Netherlands.

[37] E. Zappa e P. Mazzoleni, "Reliability of personal identification base on optical 3D measurement of a few facial landmarks," 10th International Conference on Computational Science 2010, ICCS 2010; Amsterdam, Netherlands.

[38] M. Stegmann, "The AAM-API: An open source Active Appearance Model implementation," Medical Image Computing and Computer-Assisted Intervention, MICCAI 2003 - 6th International Conference Proceedings; Montreal, Canada, 2003.

[39] M. Stegmann, B. Ersbøll e R. Larsen, "Fame - A Flexible Appearance Modeling Environment," IEEE Transactions on Medical Imaging, vol. 22, n. 10, pp. 1319-1331, 2003.

[40] F. Tortorella, An optimal reject rule for binary classifiers, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 1876 LNCS, 2014, pp. 611–620.