XXI IMEKO World Congress "Measurement in Research and Industry"
August 30 – September 4, 2015, Prague, Czech Republic

METROLOGICAL PERFORMANCE COMPARISON OF BIOMETRIC SYSTEM ARCHITECTURES FOR 3D FACE RECOGNITION

G. Betta[1], *D. Capriglione*[1], *M. Corvino*[1], *M. Gasparetto*[2], *C. Liguori*[3], *A. Paolillo*[3], *E. Zappa*[2]

[1] DIEI, University of Cassino and of Southern Lazio Cassino (FR), Italy, e-mail: betta@unicas.it
[2] DME, Politecnico di Milano via La Masa, 1, Milano, Italy, e-mail: emanuele.zappa@polimi.it
[3] DIIn, University of Salerno, via Giovanni Paolo II, 132, Fisciano (SA), Italy, e-mail: tliguori@unisa.it

Abstract – Different biometric system architectures based on 3D images can be realized for face recognition. In such systems the triangulation of images provided by a couple of 2D cameras is employed to achieve the 3D features of a face. To setting up the system, different positioning of cameras as well as both kind and resolution of camera could be considered. These parameters can affect the correct decision rate of the system in classifying the input face, especially in presence of image uncertainty.
In previous papers, the authors proposed an original approach for the estimation of the confidence level of results provided by classification systems for face recognition. Such approach is here adopted in order to compare several 3D architectures differing in camera specifications and geometrical positioning.

Keywords**:** face recognition; measurement uncertainty; classifier; decision support systems, 3D image features.

1. INTRODUCTION

Face Recognition (FR) is an even more widespread technique for subject identification in many contexts such as access control and security research [1], [2].

The working of such systems is generally based on a pre-existing database (often identified as training set) containing the subjects of interest for the application. Then, an input image acquired by one or more cameras is processed to extract some features and compare them with the corresponding ones extracted by the images contained in the database. Afterwards, a maximum likelihood approach is generally adopted to classify and recognize the input subject.

However, the input images are generally affected by uncertainty, which propagates through all the processing stages up to the final decision (i.e. the classification result). By this way, even the output result is affected by uncertainty, thus generating a risk in accepting a classification result. In literature, several techniques have been proposed to reduce this risk [3]-[9] by exploiting the knowledge of the uncertainty with a priori approaches.

Nevertheless, these approaches are not always optimal since each new case can be characterized by a different uncertainty, consequently a fixed value of uncertainty can give rise to its overestimation or underestimation. In other words, these approaches do not consider the propagation of the measurement uncertainty according to ISO-GUM [10] (that can be derived starting from input measured data).

As for the characterization of recognition systems, it is typically based on the evaluation of recognition reliability indexes that express the probability of a false positive and of a false negative [11]-[13]. On the contrary, in previous papers the authors tackled the problem of the metrological characterization of face recognition classification systems by analysing the existing relationship between the overall uncertainty of the final results and the input influence quantities. It has been shown that the performance of these kinds of systems depends on several aspects from image acquisition to the classification procedure through the biometrical algorithm. In [14], [15] the authors have proposed an original method for the evaluation of the measurement uncertainty in face recognition systems and for exploiting such a quantity in the classification phase. The proposed method allows to improve also the classification performance with respect to a traditional approach [16], [17].

As for 3D systems, from a designer point of view, different approaches have been presented in the literature, for optimizing the performance of vision systems [18], [19], but the variety of the proposals reported in literature does not help a system designer to choose an architecture for a specific problem.

In this framework the authors propose to adopt the approach developed in previous papers [15]-[17] for the comparison, from a metrological point of view, of different architectures of a 3D system for face recognition. To this aim, a suitable experimental setup involving couples of cameras characterized by different features (in terms of resolution and colour) and geometrical positioning, has been developed. The results of such a study could be very useful for system designers which, in practical application, have often to select the best trade-off between geometrical constraints (i.e. arrangement of cameras) and system performance (in terms of correct decision rate in the classifying the input subject).

In the following, after a brief recall of the developed classification procedure, and of the experimental setup, the comparison of the metrological performance of different measurement configurations is shown.

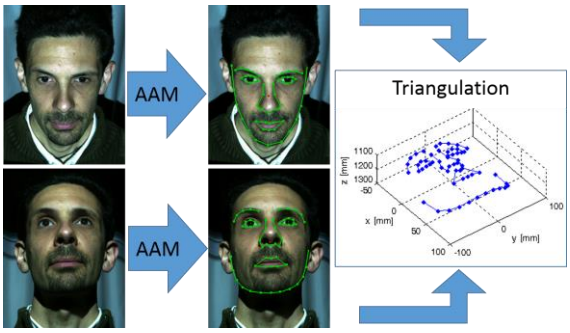

Fig. 1. The schematic of biometric algorithm with featured detected on stereoscopic images and the obtained 3D mask.

2. THE USED BIOMETRIC APPROACH

2.1. Biometric algorithm

The facial images obtained by means of a couple of stereoscopic cameras are analyzed with the AAM-API software to automatically detect the 2D coordinates of a set of landmarks [20]. A 3D mask is obtained triangulating the two 2D masks (see Figure 1). The 3D mask of a person to be identified is finally compared with each of the 3D masks included in a database and a score is computed for each comparison.

According to studies carried out in a previous paper [21], 58 landmarks demarcating seven areas of the face have been used in this work: jaw, mouth, nose, eyes, and eyebrows, as shown in Figure 2b, where an example of 3D mask is also shown. The output of the biometric algorithm is a *Score* for each subject in the database. Given a 3D mask to be recognized and the set of weights W_k (k=1,..,n, where n is the total number of points in the mask), the score, S_i, for each *i-th* mask of the database is computed as:

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

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* basically 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 that evaluation of point to point distances, a roto-translation is computed in order for the coordinate frame of one mask to be moved onto the coordinate frame of the other mask with a rigid motion. The roto-translation allows to compensate for differences in position and orientation of the subject with respect to the stereoscopic system in acquisitions.

2.2. Bias and uncertainty sources

The biometric algorithm can be schematized in five serial main steps, each one introducing uncertainty and/or bias, able to affect the whole system performance:

(i) 2D image acquisition;
(ii) 2D image processing by means of AAM algorithm;
(iii) triangulation;
(iv) roto-translation;
(v) Score evaluation.

As for the 2D image acquisition, item (i), the luminance, the lens defocus, or the presence of motion on the image itself (due for example to vibration of the system during the image acquisition or to the subject movement) are significant uncertainty causes.

Then, the image uncertainty propagates in the next step (item (ii)). With reference to the considered biometric algorithm, an AAM approach was used, then the uncertainty in the building of the Shape Model and the Appearance Model [20] determine the accuracy of feature localization.

As for the (iii), the triangulation is realized by means of the method proposed by Zhang [22] and a residual error is mainly due to the non-ideality of the calibration phase.

As for (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.

Finally, (item (v)), the *Score* is evaluated by means of equation (1).

The effects of the accuracy of AAM, triangulation and roto-translation is 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 acquisition condition. In the following, all the systematic effects present are directly included in the measured *Score*, whilst only the measurement uncertainty (due to the acquisition process) is considered.

As for the uncertainty on the *Score*, u_S, it depends mainly on the uncertainty of the 2D coordinates and, as highlighted in previous papers [14]-[17], the main quantities of influence can be related to luminance, defocus and motion blur. In order to quantify this uncertainty according to the ISO-GUM, a simple model was used in associating each quantity of influence to u_S. Denoting with $(u_S)_i$ the contribute due to the *i-th* quantity of influence on the *Score* uncertainty, we have posed:

$$(u_S)_i = \sqrt{\frac{(\mu_S^2)_i}{3} + (\sigma_S^2)_i} \qquad (2)$$

where $(\mu_S)_i$ and $(\sigma_S)_i$ are respectively the mean and the sample standard deviation of the measured *Scores* due to the *i-th* influence quantity.

All of the quantities of influence are considered uncorrelated with the other ones, then the combined uncertainty on the score is evaluated as:

$$u_S = \sqrt{\sum_{i=1}^{M} u_i^2} \qquad (3)$$

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 quantities uncorrelated has been also verified through experiments.

3. THE CLASSIFICATION SCHEME

The output of classification scheme is a classification list containing all the subjects, in which the unknown subject can be recognized, each one characterized by a confidence level (CL). The procedure implementing the classification scheme reported in Fig.2 [16], [17], [23] is composed by three main steps.

- The *Biometric Algorithm* processes two 2D images in order to evaluate the Score, Si (i=1,..,N), of each of the N subjects of the database.

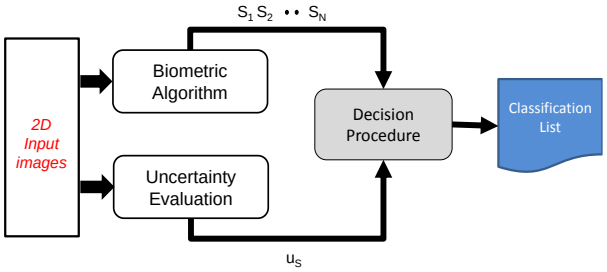

Fig. 2. The proposed classification scheme

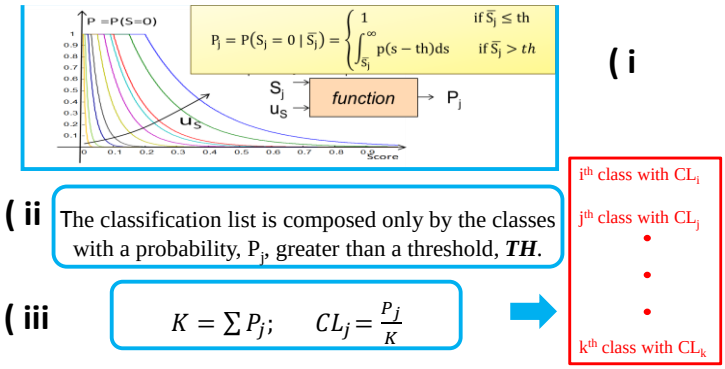

Fig.3 The Decision Procedure

- The Uncertainty Evaluation of the measured Score, u_S is made analyzing the two images and applying the relationships experimentally evaluated (see Section 5).
- The *Decision Procedure* uses the measured Scores and the associate uncertainty in order to create the classification list. In Figure 3, the decision procedure is sketched: (i) for each class, *j*, the probability, *Pj*, that the unknown subject belongs to it is evaluated, as the probability that the Score of the *j-th* class is equal to zero given a measured value *Sj** using probability density function of the Score, *p(s)*, [17]; (ii) the classification list is created considering only the probable classes, namely those classes which show a probability greater than a suitable **TH**; (iii) for each class of the classification list the confidence level evaluated [16], [17].

4. EXPERIMENTAL SET-UP

The biometric approach and the classification procedure described in the previous sections were applied to a database of images obtained with a multi-camera vision system that allows to simultaneously acquire images of the face from different points of view. From each image of the database the set of 2D facial features position described in section 2.1 is extracted by means of the AAM technique and then the 3D point set is estimated by means of stereo triangulation. The image acquisition system is provided of 6 cameras, as shown in Figure 3; all the cameras are synchronized by a common trigger to allow the acquisition of pictures of the same individual from different points of view at the same time. In this way, data referred to images of the face acquired in any position are perfectly comparable because they are acquired in the same instant of time and therefore the differences among them are due only to the camera position, while no change in facial expression occurs. The six cameras are arranged in three pairs: each pair of cameras is vertically aligned and placed at 0°, 5° and 10° referring to the axis of the face. Since all the cameras are acquired synchronously and were calibrated with a unique calibration

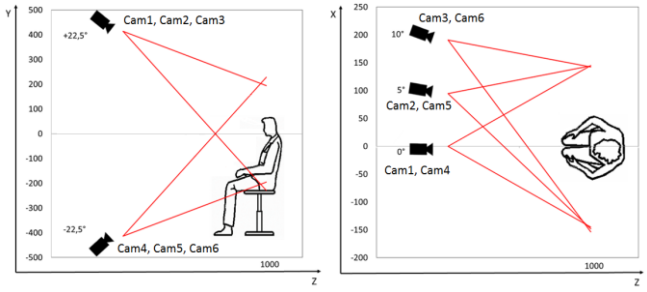

Fig. 4. The proposed multi-camera vision system

Fig. 5. Example of images collected in the database.

procedure, they all share a unique reference system and therefore it is possible to obtain 3D measurements by means of processing images acquired by any set of two or more cameras arbitrarily selected from the six composing the acquisition device. In previous works, the authors analysed the performances of the recognition system in the hypothesis of 3D measurements obtained with couples of vertically aligned cameras (Cam1 and 4 or Cam2 and 5 or Cam3 and 6) [23]. Since a perfect vertical alignment of cameras is not always possible in practice, in this work the case of freely coupled cameras is also considered, in order to qualify the corresponding recognition capabilities.

The database includes facial images of 117 volunteers. In the image acquisition, the environmental conditions (in particular the lighting) are controlled in order to ensure that the images are acquired without appreciable shadows on the face, light reflections or motion blur. During the image acquisition the volunteers were requested to look at a reference point marked in the middle between camera 1 and camera 4 in order to obtain frontal images in camera 1 and 4, images horizontally angled of 5° for cameras 2 and 5 and finally images angled 10° for cameras 3 and 6.

5. EXPERIMENTAL RESULTS

In this section, the results related to nine different configurations of the set-up architecture are reported. For each couple of cameras, at first, the uncertainty on the score, u_S, is evaluated, then the performance of the proposed classification scheme for all the configurations are evaluated.

5.1 Score Uncertainty

In order to define the relationships between the influence quantities and the Score experimental tests were carried out. Starting from the original database new images are generated simulating on turn the luminosity variation, defocus and motion blur (10 different intensities of motion simulated blur, 4 levels of intentional defocus, 4 luminosity values) [16], [17]. Then the contribute due to the *i-th* quantity of influence on the *Score* uncertainty, $(u_S)_i$ is evaluated considering eq.2.

In Fig. 6 the obtained trends are summarized.

As for the influence of the *Luminosity*, For each couple, the uncertainty is almost constant for each kind of luminosity variation, but, since the values are all very low, a constant overestimated value of 0.05 can be considered for all the configurations, including camera 1 and 0.07 for the other ones. Even though the uncertainty is overestimated, the overall performance does not get worse.

The contribution of the Defocus can be modelled with a second order polynomial model, which fits the observed data well for all the configurations. The measured values are very similar to one another, when camera 1 is used (namely couple 1-2, 1-4 and 1-6 show quite the same results), analogously for camera 3 and 5. A little worsening is observed for camera 3 with respect to camera 1 and for camera 5 with respect to camera 3. The influence of the lowest positioned camera is less evident.

In order to describe the effect of the motion blur (generally due to the movement between the subject and the camera), a linear model can be used for all the

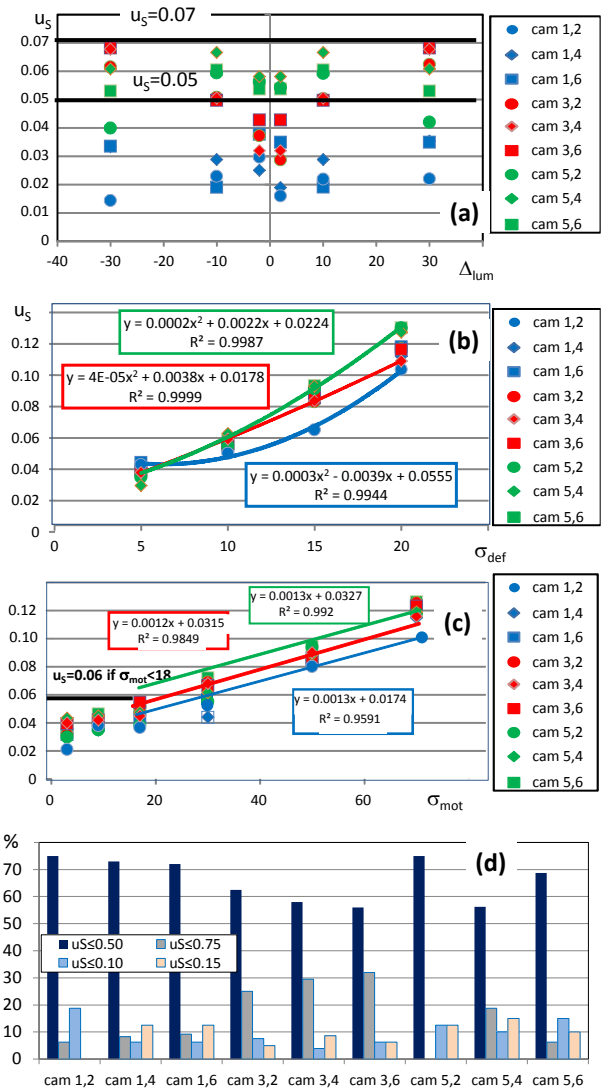

Fig. 6. a) u_S versus luminosity, b) u_S versus defocus, c) u_S versus motion blur, d) histograms of the uncertainties obtained in the different configurations

configurations. The uncertainty values increase when the high camera moves away from the central axis.

The distribution of the obtained uncertainty is reported in Figure 6d where the relative frequency histogram of the estimated u_S on the whole dataset is shown; this distribution behaves well over all the expected working conditions (both for range and frequencies). The previous results are confirmed.

5.2 Classification performance

In order to evaluate the overall classification performance of the proposed classification scheme, the so created datasets (see Fig. 6d for the distributions of the datasets uncertainty) are fed to the classification scheme. The classification results are grouped in different categories (see Fig.7):

- *Correct classification (CC)*: the right class is identified with the highest CL value.
- *Long classification list (LCL):* the right class in the classification list but there are some other classes with the same *CL*.
- *Missed classification (MC)*: the classification list is empty;

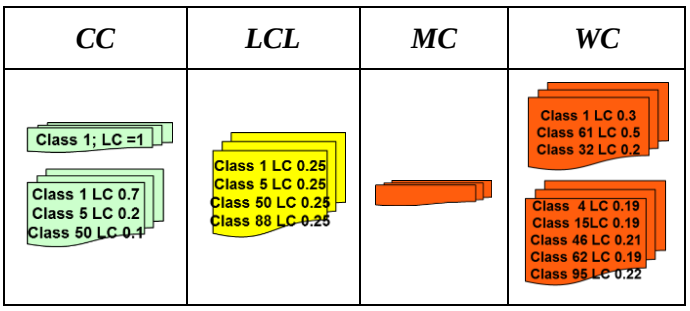

Fig. 7. Example of Classification Lists for a subject of the class 1 for the different Category.

- *Wrong classification (WC)*: either the correct class is not in the classification list or it is present without the highest value of CL.

Fig. 8 reports the percentage of each category in the experimental tests for all the considered architectures. The reported results demonstrate the utility of the proposed classification strategy, since for all the solutions the correct decision is reached in more than the 85 % of cases, including cases with high uncertainty (greater than 0.75, see Fig. 6d).

Moreover, figure 8 shows that the couple of cameras 1-2 shows the best overall classification performance while the couple of cameras 5-6 is characterized by the worst one. Considering the other solutions, intermediate results are obtained, in particular the percentage of Long Classification List increases when farther cameras are used, and using the camera 5 the number of wrong classification increases a little.

6. CONCLUSIONS

The paper has compared the metrological performance of different architectures for face detection systems based on 3D features. The study has been conducted by considering a popular algorithm based on 3D features (the AAM) and main causes of uncertainty generally affecting the performance of face recognition algorithms.

The method proposed by the authors in previous papers has been adopted here to evaluate the uncertainty on the score and suitable figures of merit associated to the classification performance. The achieved results quantify the performance of the considered system configurations highlighting that the distance between the subject (to be recognized) and the pair of cameras, as well as the horizontal angle with respect to the vertical axis of the face, generally affect the metrological performance of the system with a direct impact on the reliability of the final decision in the classification stage. The quantification of such kind information could be very useful also for system designers, which, in practical applications, have often to select the best trade-off between cameras arrangement and system performance. The followed approach can be used also for exploiting other configurations including more than two cameras.

REFERENCES

[1] W. Zhao, R. Chellapa, P.J. Phillips, A. Rosefeld, "Face recognition: A literature survey", Journal of ACM Computer Survey, 2003, pp.399–458.

[2] Y. Hea, L. Zhao, C. Zou, "Face recognition using common faces method", Pattern Recognition, 2006, pp. 2218 2222.

[3] G. Betta, D. Capriglione, F. Crenna, M. Gasparetto, C. Liguori, A. Paolillo, G.B. Rossi, E. Zappa, "Face-based recognition techniques: proposals for the metrological characterization of global and feature-based approaches", Measurement Science and Technology, 22 (2011) 124005.

[4] E. Zappa, R. Testa, M. Barbesta, M. Gasparetto, "Uncertainty of 3D facial features measurements and its effects on personal identification", Measurement.

[5] F. Crenna, E. Zappa, L. Bovio, R. Testa, M. Gasparetto, G.B. Rossi, "Implementation of perceptual aspects in a face recognition algorithm," Journal of Physics: Conference Series 459 (2013) 012031 doi:10.1088/1742-6596/459/1/012031.

[6] R.-C. David, C.-A. Dragos, R.-G. Bulzan, R.-E. Precup, E. M. Petriu, amd M.-B. Radac, "An approach to fuzzy modeling of magnetic levitation systems", International Journal of Artificial Intelligence, vol. 9, no. A12, pp. 1-18, Oct. 2012.

[7] A. S. Aisjah and S. Arifin, "Maritime weather prediction using fuzzy logic in Java sea for shipping feasibility", Intern. Journal of Artificial Intelligence, vol. 10, n. S13, pp. 112-122, Mar. 2013.

[8] D. Zhang, Q.-G. Wang, L. Yu, and H. Song, "Fuzzy-model based fault detection for a class of nonlinear systems with networked measurements", IEEE Trans. on Instrument. and Measurement, vol. 62, no. 12, pp. 3148 3159, Dec. 2013.

[9] R. R. Yager, "On the fusion of imprecise uncertainty measures using belief structures," Information Sciences, Vol. 181, n. 15, 2011, pp. 3199–3209.

[10] JCGM 100, "Evaluation of measurement data — Guide to the expression of uncertainty in measurement", ISO-GUM, 2008.

[11] S.A. Rizvi, P.J, Phillips, H. Moon, "A verification Protocol and Statistical Performance Analysis for Face Recognition Algorithms", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 1998, pp. 833-838.

[12] Peng Wang, Qiang Ji, Wayman J.L., "Modeling and Predicting Face Recognition System Performance Based on Analysis of Similarity Scores", IEEE Trans. on Pattern Analysis and

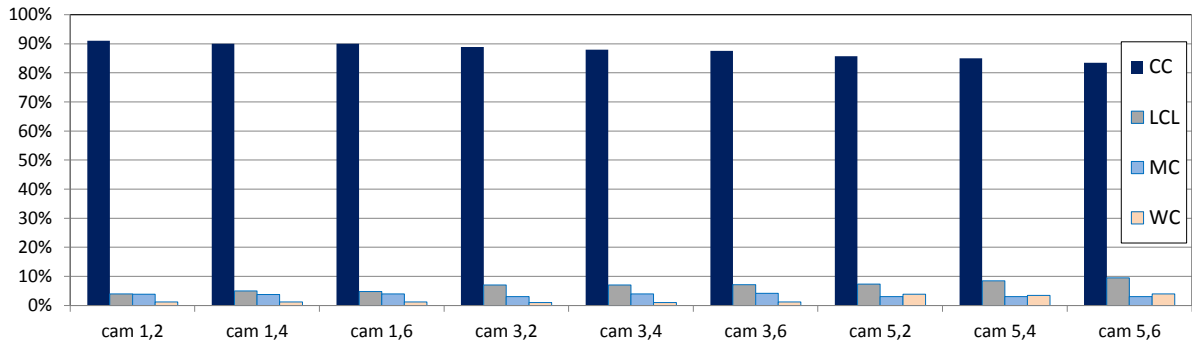

Fig. 8. Performance indexes of the nine system architectures.

Machine Intelligence, Vol. 9, Issue 4, April 2007, pp. 665-670.
[13] 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.
[14] G. Betta, D. Capriglione, M. Corvino, C. Liguori, A. Paolillo, "Face based recognition algorithms: A first step toward a metrological characterization", IEEE Trans. on Instrument. and Measurement, 2013, 62 (5), art. no. 6493520, pp. 1008-1016.
[15] G. Betta, D. Capriglione, M. Gasparetto, C. Liguori, A. Paolillo, E. Zappa, "Managing the uncertainty for face classification with 3D features", IEEE Proceedings of I2MTC14, pp. 412-417, 2014.
[16] G. Betta, D. Capriglione, M. Corvino, C. Liguori, A. Paolillo, "A Proposal for the management of the measurement uncertainty in Classification and Recognition problems", IEEE Trans. on Instrument. and Measurement, Vol. 64, Issue 2, 2015, pp. 392-402.
[17] G. Betta, D. Capriglione, M. Gasparetto, C. Liguori, A. Paolillo, E. Zappa, "Face recognition based on 3D features: Management of the measurement uncertainty for improving the classification", Measurement, 70 (2015), pp. 169-178.
[18] R. Hartley, P. Sturm, "Triangulation", Computer Vision and Image Understanding, 68(2):146–157, 1997.
[19] S. Y. Chen, Y. F. Li, "Vision sensor planning for 3-D model acquisition," IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics, Vol. 35, No. 5, Oct. 2005, pp. 894 – 904.
[20] T.F. Cootes, J.E. Gareth, J.T. Christopher, "Active appearance models," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, n.6, 2001, pp. 681-685.
[21] E. Zappa, P. Mazzoleni, Y. Hai, Stereoscopy based 3D face recognition system, ICCS 2010, Computational Science University of Amsterdam, May 31 - June 2, 2010.
[22] Z. Zhang, "A flexible new technique for camera calibration", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 22, n.11, 2000, pp. 1330-1334.
[23] G. Betta, M. Corvino, D. Capriglione, M. Gasparetto, C. Liguori, A. Paolillo, E. Zappa, "Performance Evaluation of Face Classification Systems", Proceedings of 2nd ICOCOE, 2015.