DOI: 10.1478/AAPP.932A4

AAPP | Atti della Accademia Peloritana dei Pericolanti
Classe di Scienze Fisiche, Matematiche e Naturali
ISSN 1825-1242

Vol. 93, No. 2, A4 (2015)

BIOMETRICS EMPOWERED AMBIENT INTELLIGENCE ENVIRONMENT

ANDREA F. ABATE [a*], FABIO NARDUCCI [a], AND STEFANO RICCIARDI [a]

(communicated by Liliana Restuccia)

ABSTRACT. The paper presents a face recognition system based on 3D features with the aim of verifying the identity of subjects accessing a controlled Ambient Intelligence Environment and customizing all the services accordingly. The proposed approach relies on stereoscopic face acquisition and 3D mesh reconstruction avoiding non-automated and expensive 3D scanners, unsuitable for real time applications in general. A bidimensional feature descriptor is extracted from each 3D mesh. It consists in a color image transferring face's 3D features in a 2D space. An automatic weighting mask of each authorized person improves the robustness of recognition in presence of diverse facial expressions and beard. The experiments conducted show high average recognition rate and a measurable effectiveness of both flesh mask and expression weighting mask.

1. Introduction

Information and Communication Technologies are increasingly entering in all aspects and sectors of our life. This contributes to a wide variety of innovative scenarios where people interact with electronic devices, which are embedded in the environment and are sensitive and responsive to the presence of users. Indeed, the request for more sophisticated services, tailored to user's specific needs, has characterized the domotic research since its first examples of "intelligent buildings" (e.g., computer aided security and fire safety systems). The result of the evolution of the original concept of home automation is known as Ambient Intelligence (Aarts and Marzano 2003). It refers to an environment viewed as a "community" of smart objects powered by computational capability and high user-friendliness. Each object is capable of recognizing and responding to the presence of different individuals in a seamless, non-intrusive and often invisible way. In such context, person sensing and recognition become two crucial tasks to provide customized services. This scenario offers the opportunity to exploit the potential of the user's face as a non-intrusive biometric identifier to regulate access to the controlled environment as well as to adapt the provided services to the user's preferences.

Biometric recognition (Maltoni *et al.* 2009) refers to the use of distinctive physiological (e.g., fingerprints, face, retina, iris) and behavioural (e.g., gait, signature) characteristics,

called biometric identifiers, for automatically recognizing individuals. Compared to traditional token or knowledge-based methods, biometric identifiers are generally harder to misplace, forge, or share, thus resulting more reliable for person recognition. Moreover, the person's recognition by biometric traits is more convenient (e.g., access to services without providing a Personal Identification Number), ensures better security (e.g., difficult to forge access). All these reasons make biometrics very suited for Ambient Intelligence applications, and this is especially true for the user's face that is one of the most common methods of recognition that humans use in their visual interactions. Moreover, face features allow to recognize the user in a non-intrusive way without any physical contact with the sensor. A generic biometric system could operate either in verification or identification modality, better known as one-to-one and one-to-many recognition (Perronnin and Dugelay 2003). Since we want to recognize authorized users accessing the controlled environment or requesting a specific service, the Ambient Intelligence application proposed in this paper is focused on a one-to-one recognition. We present a face recognition method based on 3D features to verify the identity of subjects accessing the controlled Ambient Intelligence Environment and to customize all the services accordingly. In other words, the purpose is to add a social dimension to man-machine communication thus contributing to make such environments more attractive to the human user. The proposed approach relies on stereoscopic face acquisition and 3D mesh reconstruction that does not depend on highly expensive and non-automated 3D scanning (typically time-demanding and non-suitable for real time applications). Each subject is enrolled, by matching a bidimensional feature descriptor, extracted from the 3D mesh, with the previously stored correspondent template. This descriptor is a normal map, namely a color image in which RGB components represent the normals to the face geometry. A weighting mask, automatically generated for each authorized person, improves recognition robustness to a wide range of facial expression.

2. Ambient Intelligence and biometric identifiers

Ambient Intelligence (AmI) worlds offer exciting potential for richer interactive experiences. The Ambient Intelligence Environment can be defined as the set of actuators and sensors composing the system together with the domotic interconnection protocol. People interact with the environment by the embedded electronic devices that are sensitive and responsive to the users' preferences. This is made possible by creating and manipulating user's profiles obtained by learning physical and emotional status that are inferred from user's biometric features. Privacy issues represent a key aspect of AmI environments. Bohn et al. observe that "by virtue of its very definitions, the vision of ambient intelligence has the potential to create an invisible and comprehensive surveillance network, covering an unprecedented share of our public and private life" (Bohn *et al.* 2005). The privacy risks of AmI are great because of the highly sensitive recording and encoding of personal information as well as the scale on which such information is recorded (Brey 2005). Introducing a robust and reliable biometric recognition ensures a higher level of privacy preservation at the cost of constraining the user to claim for his/her identity achieved by a proper pre-registration. However, being rather unobtrusive and working from reasonable distance, face represents one of the most adequate biometry that can be efficiently exploited in the context of AmI environments. To design Ambient Intelligent Environments, many

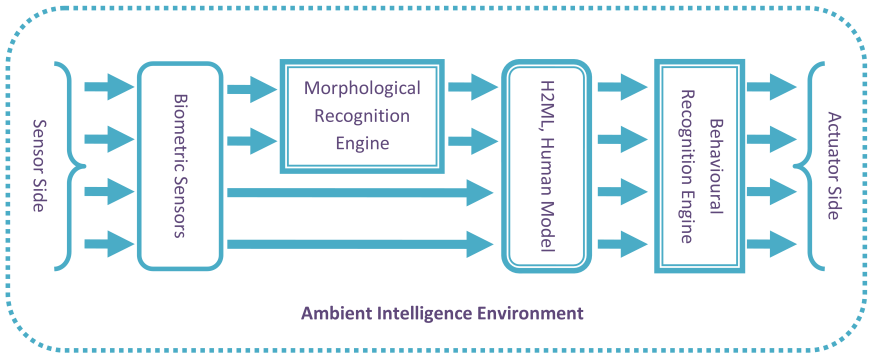

FIGURE 1. Ambient Intelligence Architecture

methodologies and techniques have to be merged together originating many approaches reported in recent literature (D. J. Cook, Augusto, and Jakkula 2009; Sadri 2011). We exploit a framework aimed to gather biometrical and environmental data, described in (Abate, Ricciardi, and Sabatino 2007; Acampora *et al.* 2005) to test the effectiveness of face recognition systems to aid security and to recognize the emotional user status. This AmI system's architecture is organized in several sub-systems, as depicted in Figure 1, and it is based on the following sensors and actuators: temperature sensors (internal and external) and temperature actuator, luminosity sensor (internal and external) and luminosity actuator, proximity sensors for person's presence, an infra-red camera to capture user's thermal images and a set of RGB cameras to capture information about gait and facial features. *Biometric Sensors* are mainly used to gather user's biometrics (e.g., temperature, gait, position, facial expression). Parts of this information is handled and semantically organised by the *Morphological Recognition Subsystem* (*MRS*). The resulting description, together with the remaining biometrics previously captured, are arranged in a hierarchical structure based on XML technology thus leading to a new mark-up language, called *H2ML* (*Human to Mark-up Language*) representing user status at a given time. The *Behavioural Recognition Engine* (*BRE*), starting from sequences of H2ML descriptions tries to find a match to a particular user behaviour for which the system is able to provide suitable services. The *Service Regulation System* (*SRS*) is in charge of regulating the available services. It consists in an array of fuzzy controllers exploited to achieve hardware transparency and to minimize the fuzzy inference time. This approach is particularly suited to exploit biometric technologies to capture user's physical info gathered in a semantic representation describing a human in terms of morphological features.

3. Main approaches to 3D face recognition

As highlighted by various surveys (Bowyer, K. Chang, and P. Flynn 2004; Jafri and Arabnia 2009), face recognition represents a research topic for which "the variety and sophistication of algorithmic approaches explored is expanding". Especially for 3D face recognition, the main challenges are to improve the recognition accuracy, enhance the

robustness to facial expressions and, more recently, design higher efficient algorithms. The various methods proposed so far can be categorized in four classes: holistic, region-based, hybrid and multimodal. The first includes the algorithms that perform face comparison at a global level while the second the algorithms that compare homologous regions between faces. The third class contains the algorithms that exploit both the previous approaches. The last one consists of all algorithms that use both 2D and 3D features for the comparison, fusing together the results of both modalities of face matching. Many holistic methods are based on Principal Component Analysis (PCA) applied either to depth images (Hesher, Srivastava, and Erlebacher 2002; Pan *et al.* 2005) or to both color and depth channels (Tsalakanidou, Tzovaras, and Strintzis 2003). Other authors combine 3D and 2D similarity scores obtained comparing 3D and 2D profiles (Beumier and Acheroy 2000), or extract a feature vector combining Gabor filter responses in 2D and point signatures in 3D (Wang and Chua 2005). Canonical surfaces have been exploited to mitigate the effects of facial expressions on recognition accuracy (A. M. Bronstein, M. M. Bronstein, and Kimmel 2005, 2006). Morphable models, and elastic registration have also been used (Amberg, Knothe, and Vetter 2008), though the computational cost involved is relevant. Among region-based approaches, C. Xu *et al.* (2004) aim to divide face in sub-regions using nose as the anchor, PCA to reduce feature space dimensionality and minimum distance for matching. They also proposed a method to face partitioning based on the intersection between spheres of increasing radius and the face scans (D. Xu *et al.* 2008). Another major research trend is based on Iterative Closest Point (ICP) algorithm, which has been exploited in many variations for 3D shape aligning, matching or both. The first example of this kind of approach to face recognition has been presented by Medioni and Waupotitsch (2003), while other authors (K. I. Chang, Bowyer, and P. J. Flynn 2005) proposed to apply ICP to a set of selected subregions instead (Faltemier, Bowyer, and P. J. Flynn 2007, 2008). Iso-geodesic stripes and 3D Weighted Walkthroughs (3DWWs) have been proposed by Berretti, Del Bimbo, and Pala (2010) proving to be accurate in terms of recognition and robust to intra-class variations. The methods belonging to the hybrid and multimodal categories aim at improving the precision of recognition by combining well established techniques like PCA, LDA and ICP and/or operating at a 2D and 3D level, to overcome the limits of the individual approaches. The work by Mian, Bennamoun, and Owens (2007) represents a good example of this approach, producing the best score on the FRGC v2.0 contest.

The basic idea behind the system proposed is to represent user's facial surface by a digital signature called normal map. A normal map is an RGB color image providing a 2D representation of the 3D facial surface. Given a mesh, the normals to each polygon is represented by a RGB color pixel. To this aim, we project the 3D geometry onto 2D space through spherical mapping. The result is a bidimensional representation of original face geometry which retains spatial relationships between facial features. Colour information coming from face texture are used to mask regions possibly covered by beard according to their relevance, resulting in a 8 bit grayscale filter mask (Flesh Mask). Then, a variety of facial expressions are generated from the neutral pose through a rig-based animation technique in order to cope with expression variations. The corresponding normal maps are used to compute a further 8 bit grayscale mask (Expression Weighting Mask). At this time, the two grayscale masks are multiplied and the resulting map is used to augment with extra

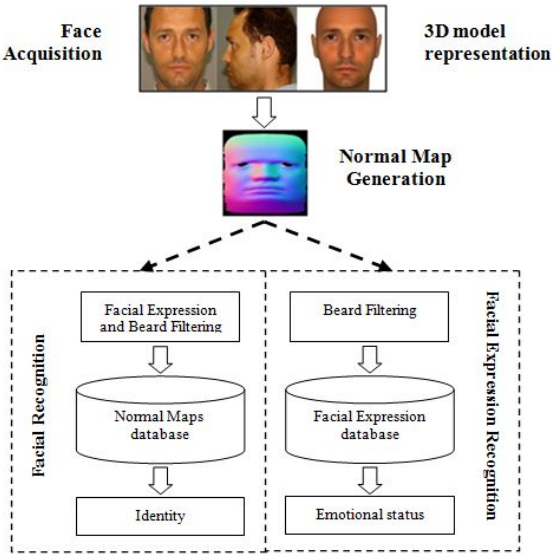

FIGURE 2. Facial and Facial Expression Recognition workflow

8 bit per pixel the normal map, resulting in a 32 bit RGBA bitmap (Augmented Normal Map). The whole process (see Figure 2) is discussed in depth in the following section.

4. Detailed description of proposed approach to face recognition

Since the proposed method works on 3D polygonal meshes, we firstly need to acquire actual faces and to represent them as polygonal surfaces. The Ambient Intelligence context, in which we are implementing face recognition, requires fast user enrolment to avoid annoying waiting time. Many 3D face recognition methods use laser or structured light scanners to produce a range image of the face. This kind of devices offer a high resolution in the captured data. The main drawbacks are the long acquisition time and their low tolerance to motion during capturing, two aspects that make these systems not suitable for real-time applications. Moreover, laser scanning could not be harmless to the eyes. For all these reasons we opted a stereoscopic approach for 3D mesh reconstruction based on the work by Enciso *et al.* (1999) as it requires simpler equipment: a couple of digital cameras shooting at high shutter speed from two slightly different angles with strobe lighting. This simple setup offers higher chances to be adopted in a real-time context despite a lower accuracy of the shape obtained compared to that achieved by using 3D scanners. Experimental results proved it to be sufficient for a reliable recognition with a mean recognition rate above 98% (see section 5). Moreover, it offers additional advantages such as a precise mesh alignment in 3D space (achieved by the warping approach), facial texture generation from the two captured orthogonal views and its automatic mapping onto the reconstructed face geometry.

Since the 3D polygonal mesh is an approximation of the actual face shape, polygon normals describing local curvature of the captured face could be considered as a signature.

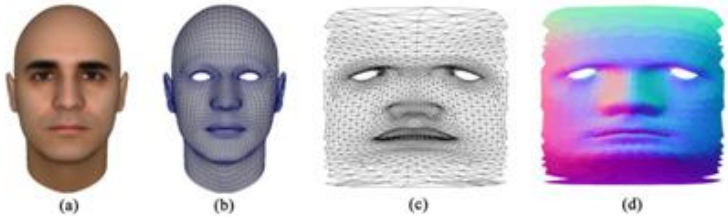

FIGURE 3. (a) 3D mesh model, (b) wireframe model, (c) projection in 2D spatial coordinates, (d) normal map

As shown in Figure 3, we intend to represent these normals by a color image transferring 3D features of the face in a 2D space. In order to preserve the spatial relationships between facial features, we project 3D coordinates of vertices onto a 2D space using a spherical projection, which is geometrically centred in the nose tip. After that, we can store normals of mesh M in a bidimensional array N using mapping coordinates. In other words, each pixel represents a normal coded in RGB values (Abate *et al.* 2006). We refer to the resulting array as the *Normal Map N* of the mesh M that is the signature we use for the identity verification.

Given N_A and N_B the normal map from input subject and another normal map previously stored in the reference database respectively, we compute the angle between each pairs of normals (represented by colors of pixels) with corresponding mapping coordinates:

$$\theta = \cos^{-1}(r_{N_A} \cdot r_{N_B} + g_{N_A} \cdot g_{N_B} + b_{N_A} \cdot b_{N_B}) \tag{1}$$

where components r, g and b are previously normalized from spatial domain to color domain $0 \leq r_{N_A}, g_{N_A}, b_{N_A} \leq 1$ and $0 \leq r_{N_B}, g_{N_B}, b_{N_B} \leq 1$. Therefore, the value θ, with $0 \leq \theta < \pi$, is the angular difference between the pixels of coordinates (x_{N_A}, y_{N_A}) in N_A and (x_{N_B}, y_{N_B}) in N_B. The difference normal map D stores the information from equation (1) resulting in a gray-scale image. At this point, the histogram H of the difference map D is analysed to estimate the similarity score between N_A and N_B. On the X axis we represent the resulting angles between each pair of comparisons (sorted from 0 degree to 180 degree), while on the Y axis we represent the total number of occurrences found. The curvature of H represents the angular distance distribution between mesh N_A and N_B, thus implying that two faces are similar if the distribution shows very high values on small angles (see Figure 4). We define a similarity score through a weighted sum between H and a Gaussian function G:

$$similarityscore = \Sigma_{x=0}^{k} \left(H(x) \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} \right) \tag{2}$$

where θ and k can be used to change the sensibility of the recognition. To reduce the effects of residual face misalignment during acquisition and sampling phases, we calculate the angle θ using a k×k (usually 3×3 or 5×5) matrix of neighbour pixels.

The presence of beard with variable length can cover a significant portion of the face surface in a subject previously enrolled without it (or vice-versa). This could lead to a measurable difference in the overall or local 3D shape of the face mesh (see Figure 5). In

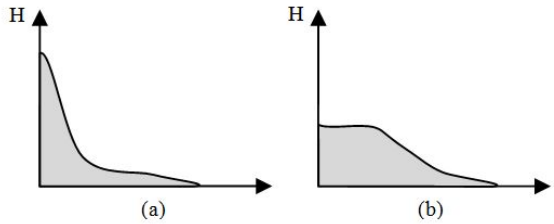

FIGURE 4. Example of histogram H to represent the angular distances. (a) Shows a typical histogram between two similar Normal Maps, while (b) between two different Normal Maps

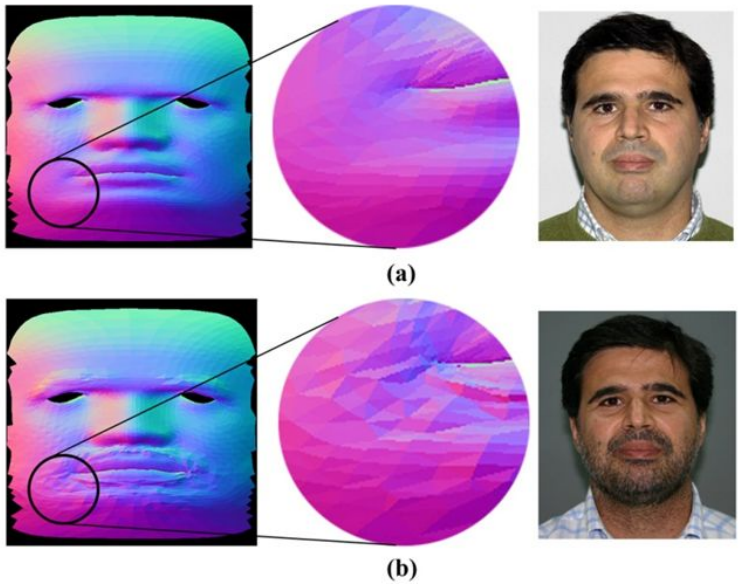

FIGURE 5. Normal maps of the same subject enrolled in two different sessions with and without beard.

such cases, the recognition accuracy could be significantly affected implying, for instance, a misleading False Rejection Rate (FRR). To address the variability of this biometric feature, we detect the non-skin regions by generating a mask obtained from color data of the captured face texture, eventually disregarding them during the comparison. Discriminating skin from beard/moustaches/eyebrows is rather simple to achieve by simple flesh hue thresholding in the HSB color space (Abate *et al.* 2005). Compared to the RGB color model, the hue component of each pixel is much more invariant to lighting conditions during capturing. On the other side, factors like facial morphology, skin conditions or pathologies, race and others can determine a significant difference in hue ranges of skin region. This asks for an

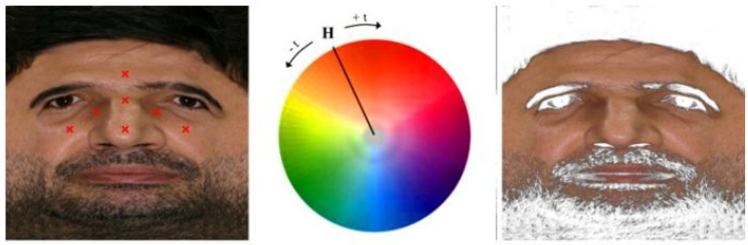

FIGURE 6. Flesh Hue sampling points (a), Flesh Hue Range (b) non-skin regions in white (c).

case-by-case definition of range to obtain a valid mask for each possible user. We adopt a solution based on the use of a set of specific hue sampling spots. The absolute coordinates of these regions are empirically calculated to be representative of flesh's full tonal range and possibly distant from critical regions like eyes, lips and typical beard and hair that, covering the face, does not contribute to skin color detection.

This is possible because each face mesh and its texture are centred and normalized during the reconstruction process (i.e. the median axis of a face is always centred on the origin of 3D space with horizontal mapping coordinates equal to 0.5), otherwise no reliable comparison on normal maps would be possible. To infer the flesh hue we either 2D or 3D technique can be used to locate the main facial features (eye, nose and lips) thus positioning the sampling spots according to them. However, these approaches do not work reasonably well in all possible conditions. Therefore, for each sampling spot we consider the neighbourhood of the texel by using a 5×5 matrix centred in the sampling spot texel and averaging them to minimize the effect of local image noise. As local anomalies of skin color (e.g., moles, scars) or even improper positioning could lead to wrong color picking for some sampling spots, we calculate the Flesh Hue Value (FHV) as the median of all resulting hue values of all sampling spots. It represents the center of the range of the valid flesh hue values. We therefore consider like skin region all texels whose hue value is within the range: $-t \leq \mathrm{FHV} \leq t$, where t is a hue tolerance (experiments found that it could be set below 10 degree as shown in Figure 6b). Once the skin region has been selected, it is filled with pure white while the remaining pixels are converted to a grayscale value depending on their distance from the selected flesh hue range (the more the distance the darker the value).

To improve the facial recognition method and to address facial expressions, we opt to use an expression weighting mask, which is a subject specific pre-calculated mask aimed to assign different relevance to different face regions. This mask, which size is equal the size of normal and difference maps, contains 8 bit weight information for each pixel encoding the local rigidity of the face surface (obtained by an analysis of a pre-built set of facial expressions of the same subject). Indeed, for each enrolled subject, expression variations (see Figure 7) are compared to the neutral face thus obtaining a difference map.

The average map, obtained by such set of difference maps of a specific user, represents its expression weighting mask. More precisely, given N_0 the neutral face of a

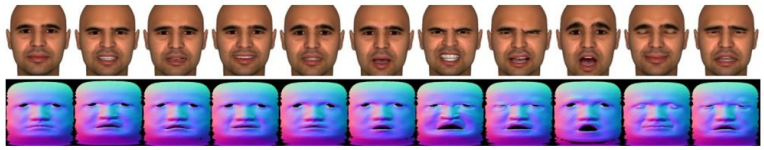

FIGURE 7. An example of normal maps of the same subject featuring a neutral pose (leftmost face) and different facial expressions.

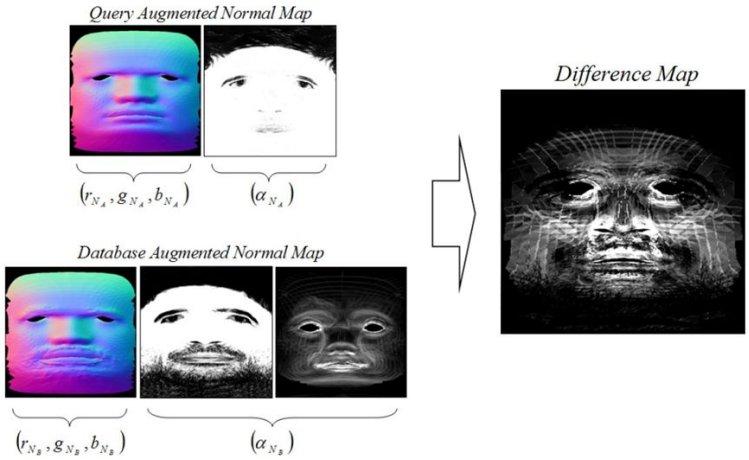

FIGURE 8. Comparison of two Normal Maps using Flesh Mask and the resulting Difference Map (c).

generic user with its normal map and the set $\{N_1, N_2, \ldots, N_n\}$ of normal maps (the expression variations), we compute the difference maps $\{D_1, D_2, \ldots, D_n\}$ resulting from $\{N_0 - N_1, N_0 - N_2, ..., N_0 - N_n\}$. The average of the set $\{D_1, D_2, \ldots, D_n\}$ is the expression weighting mask which is multiplied by the difference map in each comparison between two faces. We generate the expression variations through a parametric rig-based deformation system previously applied to a prototype face mesh (morphed to fit the reconstructed face mesh (Enciso *et al.* 1999)). This fitting is achieved via a landmark-based volume morphing where the transformation and deformation of the prototype mesh is guided by the interpolation of a set of landmark points with a radial basis function. The accuracy of the rough mesh fitting is then improved by an optimization of the surface by minimizing a cost function based on the Euclidean distance between vertices. Once complete, we can augment each 24 bit normal map by multiplying the Flesh Mask and Expression Weighting Mask normalized to 8 bit (see Figure 8). The resulting 32 bit per pixel RGBA bitmap can be conveniently managed via various image formats like the Portable Network Graphics format (PNG) which is typically used to store for each pixel 24 bit of colour and 8 bit of alpha channel (transparency). In any comparison between two faces, the difference map is computed on the first 24 bit of color info (normals) and multiplied to the alpha channel (filtering mask).

5. Experimental Results

One of the main aims in experiments was to test the performance of the proposed method in a realistic operative environment. Therefore, we decided to build a 3D face database from the face capture station used in the domotic system described above. The capture station featured two digital cameras with external electronic strobes shooting simultaneously with a shutter speed of 1/250 sec. while the subject was looking at a blinking led to reduce posing issues. More precisely, every 3D model of a face in the gallery was created deforming a pre-aligned prototype polygonal face mesh to closely fit a set of facial features extracted from front and side images of each individual enrolled in the system. Indeed, for each enrolled subject a set of corresponding facial features were extracted by a structured snake method from the two orthogonal views. After correlating them, we used facial features to guide the prototype mesh warping, performed through a Dirichlet Free Form Deformation. The two captured face images are aligned, combined and blended resulting in a color texture precisely fitting the reconstructed face mesh. The prototype face mesh used in the dataset has about 7K triangular facets, a level of detail we observed to be adequate for face recognition. This is mainly due to the optimized tessellation which privileges key area such as eyes, nose and lips whereas a typical mesh provided by 3D scanner features almost evenly spaced vertices. Another remarkable advantage involved in the warp-based mesh generation is the ability to reproduce a broad range of face variations through a rig-based deformation system. This technique is commonly used in computer graphics for facial animation (Blanz and Vetter 1999) and is easily applied to the prototype mesh linking the rig system to specific subsets of vertices on the face surface. Any facial expression could be mimicked opportunely combining the effect of the rig controlling lips, mouth shape, eye closing or opening, nose tip or bridge, cheek shape, eyebrows shape, etc. The facial deformation model we used is based on the work by Lee, Terzopoulos, and Waters (1995) and the resulting expressions are anatomically correct.

We augmented the 3D dataset of each enrolled subject through the synthesis of fifteen additional expressions selected to represent typical face shape deformation due to facial expressive muscles, each one included in the weighting mask. The fifteen variations to the neutral face are grouped in three different classes: “good-mood”, “normal-mood” and “bad-mood” emotional status (see Figure 9).

We acquired three set front-side pair of face images from 235 different persons in three subjective facial expression to represent “normal-mood”, “good-mood” and “bad-mood” emotional status respectively (137 males and 98 females, age ranging from 19 to 65). For the first group of experiments, we obtained a database of 235 3D face models in neutral pose (represented by “normal-mood” status) each one augmented with fifteen expressive variations. Experimental results are generally good in terms of accuracy, showing a Recognition Rate of 100% using the expression weighting mask and flesh mask, the Gaussian function with $\theta = 4.5$ and $k = 50$ and normal map sized 128×128 pixels. These results are generally better than those obtained by many 2D algorithms but a more meaningful comparison would require a face dataset featuring both 2D and 3D data. To this aim we experimented a PCA-based 2D face recognition algorithm (Martínez and Kak 2001; Moon and Phillips 2001) on the same subjects. We have trained the PCA-based recognition system with frontal face images acquired during several enrolment sessions (from 11 to 13 images for each

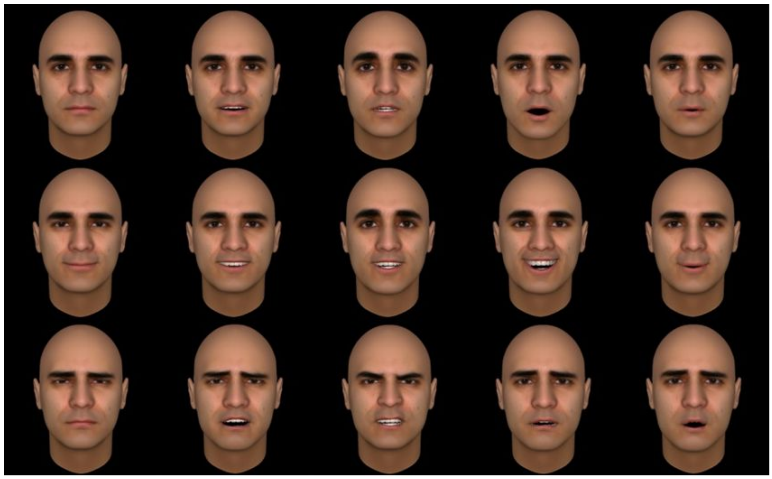

FIGURE 9. Facial expressions grouped in normal-mood (first row), good-mood (second row), bad-mood (third row).

subject), while the probe set is obtained from the same frontal images used to generate the 3D face mesh for the proposed method. This experiment has shown that our method produce better results than a typical PCA-based recognition algorithm on the same subjects. More precisely, PCA-based method reached a recognition rate of 88.39% on gray-scaled images sized to 200×256 pixels, proving that face dataset was really challenging.

Figure 10 shows the precision/recall improvement provided by the expression weighting mask and flesh mask. The results shown in Figure 10a were achieved comparing in one-to-many modality a query set with one expressive variation to an answer set composed by one neutral face, ten expression variations and one face with beard. In Figure 10b are shown the results of one-to-many comparison between subject with beard and an answer set composed of one neutral face and ten expressive variations. Finally for the test reported in Figure 10c the query was an expression variation or a face with beard, while the answer set could contain a neutral face plus ten associated expressive variations or a face with beard. The three charts clearly show the benefits involved with the usage of both expressive and flesh mask, especially when combined together.

The second group of experiments has been conducted on FRGCv2 (only shape considered) to test the method's performance with respect to Receiver Operating Characteristic (ROC) curve which plots the False Acceptance Rate (FAR) against Verification Rate (1 - False Rejection Rate or FRR) for various decision thresholds. The 4007 faces provided in the dataset have undergone a pre-processing stage to allow our method to work effectively. The typical workflow included: mesh alignment using the embedded info provided by FRGC dataset such as outer eye corners, nose tip, chin prominence; mesh subsampling to one fourth of the original resolution; mesh cropping to eliminate unwanted detail (hair, neck, ears, etc.); normal map filtering by a 5×5 median filter to reduce capture noise and artifacts. Figure 11 shows the resulting ROC curves with typical ROC values at FAR =

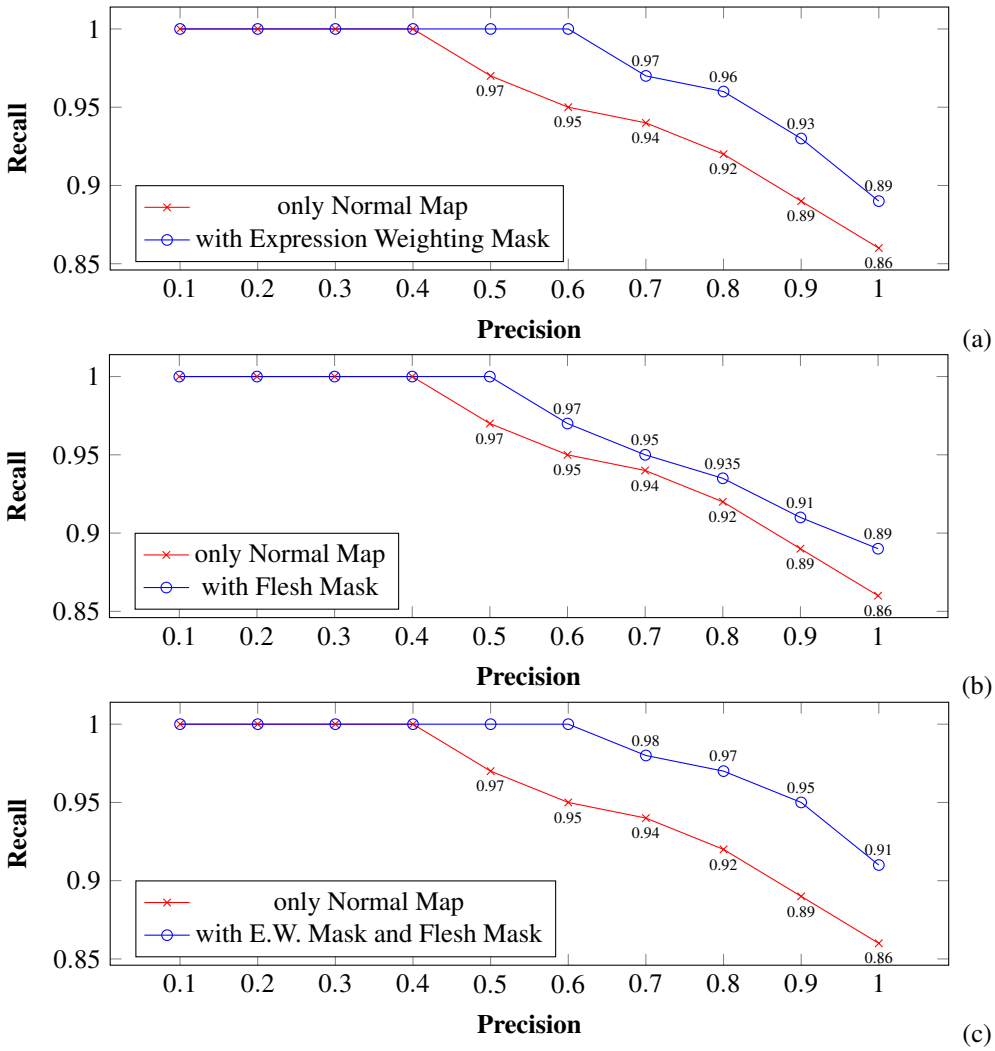

FIGURE 10. Precision/Recall Testing with and without Expression Weighting Mask and Flesh Mask to show efficacy respectively to (a) expression variations, (b) beard presence and (c) both.

0.001. The Equal Error Rate (EER) measured on all two galleries reaches 5.45% on our gallery and 6.55% on FRGC dataset.

A comparison with some among the best performing face recognition methods available in literature is shown in Table 1. In terms of recognition performance, the proposed method achieves a relevant rate, outperforming some recent proposals in literature. Indeed, a comparison simply based on recognition accuracy has limited value in the application considered. To this regard, relevant aspects for AmI applications can be: the time to

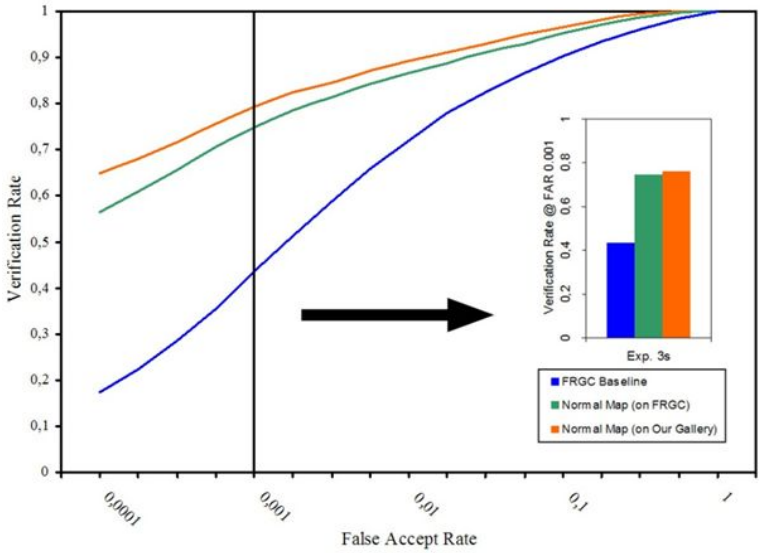

FIGURE 11. Comparison of ROC curves and Verification Rate at FAR=0.001.

TABLE 1. Comparison with state-of-the-art methods on FRGC dataset

METHOD	RECOGNITION RATE	ROBUST TO FACIAL HAIR	EXPRESSION RECOGNITION
Our proposal	98%	Yes	Yes (97.5%)
MS-eLBP 2012 (Huang *et al.* 2012)	97.6%	No	No
Wang, Liu, and Tang (2010)	98.4%	No	No
Mpiperis, Malassiotis, and Strintzis (2008)	86%	No	Yes (90.5%)
Mian, Bennamoun, and Owens (2007)	96.2%	No	No
J. A. Cook, Chandran, and Fookes (2006)	94.6%	No	No
Li and Barreto (2006)	80%	No	Yes (92.5%)
Passalis *et al.* (2005)	89.5%	No	No

reconstruct the 3D representation of the face, the robustness to everyday face changes (e.g., facial hair), recognition of expressions and similar. All cited methods work on a 3D representation of the face. However, the added value of our proposal is the design of a fast and reliable 3D reconstruction pipeline. The average time for mesh reconstruction is of about 0.5 seconds on a P4/3.4 Ghz based PC. Moreover, compared to others, the proposed recognition method results robust to facial hair, which represents a particular problem in unconstrained environments like AmI applications. Concerning facial expressions, our proposal shows a significant high average recognition rate of 97.5%.

Finally, we have tested the method in order to statistically evaluate the reliability of the method of recognizing the "emotional" status of the user. To this aim, we have performed a

TABLE 2. Recognition rate of the proposed method according to the "emotional" status of the user.

RECOGNITION RATE		
normal-mood	good-mood	bad-mood
98.3%	100%	97.%

one-to-one comparison of a probe set of 3D face models representing real subjective mood status captured by camera (three facial expressions per person) with three gallery set of artificial mood status generated automatically by controlling rig-based deformation system (fifteen facial expression per person grouped as shown in Figure 9). As shown in Table 2, the results are very interesting, because the mean recognition rate on "good-mood" status gallery is 100% while on "normal-mood" and "bad-mood" status galleries is 98.3% and 97.8% respectively (according to our observation, this is mainly due to the propensity of people to make similar facial expressions for "normal-mood" and "bad-mood" status).

6. Conclusions

We presented a 3D face recognition method applied to an Ambient Intelligence Environment. The proposed 3D face recognition method based on 3D geometry and color texture showed high average recognition rate and a measurable effectiveness of both flesh mask and expression weighting mask. It also proved to be capable of improving robustness to presence/absence of facial hair. As the acquisition system requires the user to look at a specific target to allow a valid face capture, we are working on a multi-angle camera arrangement, to make this critical task less annoying and more robust to a wide posing range. Ongoing research will also implement a true multimodal version of the basic algorithm with a second recognition engine dedicated to the color info (texture) which could further enhance the discriminating power.

References

Aarts, E. H. L. and Marzano, S., eds. (2003). *The new everyday: Views on ambient intelligence*. Rotterdam: 010 Publishers.

Abate, A. F., Nappi, M., Ricciardi, S., and Sabatino, G. (2005). "Fast 3D face recognition based on normal map". In: *Image Processing, 2005. ICIP 2005. IEEE International Conference on*. Vol. 2, pp. II–946-9. DOI: 10.1109/ICIP.2005.1530213.

Abate, A. F., Nappi, M., Ricciardi, S., and Sabatino, G. (2006). "Multi-modal face recognition by means of augmented normal map and PCA". In: *Image Processing, 2006 IEEE International Conference on*, pp. 649–652. DOI: 10.1109/ICIP.2006.312414.

Abate, A. F., Ricciardi, S., and Sabatino, G. (2007). "3D Face Recognition in a Ambient Intelligence Environment Scenario". In: *Face Recognition*. Ed. by K. Delac and M. Grgic. InTech. Chap. 1. DOI: 10.5772/4828.

Acampora, G., Loia, V., Nappi, M., and Ricciardi, S. (2005). "Human-based models for smart devices in ambient intelligence". In: *Industrial Electronics, 2005. ISIE 2005. Proceedings of the IEEE International Symposium on*. Vol. 1, pp. 107–112. DOI: 10.1109/ISIE.2005.1528896.

Amberg, B., Knothe, R., and Vetter, T. (2008). "SHREC'08 entry: Shape based face recognition with a Morphable Model". In: *Shape Modeling and Applications, International Conference on*, pp. 253–254. DOI: 10.1109/SMI.2008.4547993.

Berretti, S., Del Bimbo, A., and Pala, P. (2010). "3D face recognition using isogeodesic stripes". *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **32**(12), 2162–2177. DOI: 10.1109/TPAMI.2010.43.

Beumier, C. and Acheroy, M. (2000). "Automatic face verification from 3D and grey level clues". In: 11th Portuguese Conference on Pattern Recognition, pp. 95–101. URL: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.32.855.

Blanz, V. and Vetter, T. (1999). "A Morphable Model for the Synthesis of 3D Faces". In: *Proceedings of the 26th annual conference on computer graphics and interactive techniques*. SIGGRAPH '99. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., pp. 187–194. DOI: 10.1145/311535.311556.

Bohn, J., Coroamă, V., Langheinrich, M., Mattern, F., and Rohs, M. (2005). "Social, economic, and ethical implications of ambient intelligence and ubiquitous computing". In: *Ambient Intelligence*. Ed. by W. Weber, J. Rabaey, and E. Aarts. Berlin: Springer, pp. 5–29. DOI: 10.1007/3-540-27139-2_2.

Bowyer, K. W., Chang, K., and Flynn, P. (2004). "A survey of approaches to three-dimensional face recognition". In: *Proceedings of the 17th International Conference on Pattern Recognition*. (Cambridge). Vol. 1. IEEE, pp. 358–361. DOI: 10.1109/ICPR.2004.1334126.

Brey, P. (2005). "Freedom and privacy in ambient intelligence". *Ethics and Information Technology* **7**(3), 157–166. DOI: 10.1007/s10676-006-0005-3.

Bronstein, A. M., Bronstein, M. M., and Kimmel, R. (2005). "Three-dimensional face recognition". *International Journal of Computer Vision* **64**(1), 5–30. DOI: 10.1007/s11263-005-1085-y.

Bronstein, A. M., Bronstein, M. M., and Kimmel, R. (2006). "Robust expression-invariant face recognition from partially missing data". In: *Computer Vision – ECCV 2006*. Ed. by A. Leonardis, H. Bischof, and A. Pinz. Vol. 3953. Lecture Notes in Computer Science. Berlin: Springer, pp. 396–408. DOI: 10.1007/11744078_31.

Chang, K. I., Bowyer, K. W., and Flynn, P. J. (2005). "Adaptive rigid multi-region selection for handling expression variation in 3D face recognition". In: *Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*, pp. 157–157. DOI: 10.1109/CVPR.2005.567.

Cook, D. J., Augusto, J. C., and Jakkula, V. R. (2009). "Ambient intelligence: Technologies, applications, and opportunities". *Pervasive and Mobile Computing* **5**(4), 277–298. DOI: 10.1016/j.pmcj.2009.04.001.

Cook, J. A., Chandran, V., and Fookes, C. B. (2006). "3d face recognition using Log-Gabor templates". In: *BMVC06 Proceedings*. Vol. 2, pp. 769–778. URL: http://www.macs.hw.ac.uk/bmvc2006/papers/286.pdf.

Enciso, R., Li, J., Fidaleo, D. A., Kim, T.-Y., Noh, J.-Y., and Neumann, U. (1999). "Synthesis of 3D faces". In: *Proceedings of the 1st USF International Workshop on Digital and Computational Video (DCV '99)*. (Tampa, Florida, USA, 1999), pp. 8–15.

Faltemier, T. C., Bowyer, K. W., and Flynn, P. J. (2007). "Using a Multi-Instance Enrollment Representation to Improve 3D Face Recognition". In: *Biometrics: Theory, Applications, and Systems, 2007. BTAS 2007. First IEEE International Conference on*, pp. 1–6. DOI: 10.1109/BTAS.2007.4401928.

Faltemier, T. C., Bowyer, K. W., and Flynn, P. J. (2008). "A region ensemble for 3-D face recognition". *Information Forensics and Security, IEEE Transactions on* **3**(1), 62–73. DOI: 10.1109/TIFS.2007.916287.

Hesher, C., Srivastava, A., and Erlebacher, G. (2002). "Principal component analysis of range images for facial recognition". In: *Proceedings of the International Conference on Imaging Science, Systems, and Technology (CISST)*, pp. 62–68.

Huang, D., Ardabilian, M., Wang, Y., and Chen, L. (2012). "3-D Face Recognition Using eLBP-Based Facial Description and Local Feature Hybrid Matching". *Information Forensics and Security, IEEE Transactions on* **7**(5), 1551–1565. DOI: 10.1109/TIFS.2012.2206807.

Jafri, R. and Arabnia, H. R. (2009). "A Survey of Face Recognition Techniques". *Journal of Information Processing Systems* **5**(2), 41–68. DOI: 10.3745/JIPS.2009.5.2.041.

Lee, Y., Terzopoulos, D., and Waters, K. (1995). "Realistic Modeling for Facial Animation". In: *Proceedings of the 22Nd Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '95. New York, NY, USA: ACM, pp. 55–62. DOI: 10.1145/218380.218407.

Li, C. and Barreto, A. (2006). "An Integrated 3D Face-Expression Recognition Approach". In: *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*. Vol. 3, pp. III–III. DOI: 10.1109/ICASSP.2006.1660858.

Maltoni, D., Maio, D., Jain, A. K., and Prabhakar, S. (2009). *Handbook of fingerprint recognition*. London: Springer. DOI: 10.1007/978-1-84882-254-2.

Martínez, A. M. and Kak, A. C. (2001). "PCA versus LDA". *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **23**(2), 228–233. DOI: 10.1109/34.908974.

Medioni, G. and Waupotitsch, R. (2003). "Face modeling and recognition in 3-D". In: *Analysis and Modeling of Faces and Gestures, 2003. AMFG 2003. IEEE International Workshop on*, pp. 232–233. DOI: 10.1109/AMFG.2003.1240848.

Mian, A. S., Bennamoun, M., and Owens, R. (2007). "An efficient multimodal 2D-3D hybrid approach to automatic face recognition". *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **29**(11), 1927–1943. DOI: 10.1109/TPAMI.2007.1105.

Moon, H. and Phillips, P. J. (2001). "Computational and performance aspects of PCA-based face-recognition algorithms". *Perception* **30**(3), 303–321. DOI: 10.1068/p2896.

Mpiperis, I., Malassiotis, S., and Strintzis, M. G. (2008). "Bilinear models for 3-D face and facial expression recognition". *Information Forensics and Security, IEEE Transactions on* **3**(3), 498–511. DOI: 10.1109/TIFS.2008.924598.

Pan, G., Han, S., Wu, Z., and Wang, Y. (2005). "3D face recognition using mapped depth images". In: *Computer Vision and Pattern Recognition – Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*, pp. 175–175. DOI: 10.1109/CVPR.2005.560.

Passalis, G., Kakadiaris, I., Theoharis, T., Toderici, G., and Murtuza, N. (2005). "Evaluation of 3D Face Recognition in the presence of facial expressions: an Annotated Deformable Model approach". In: *Computer Vision and Pattern Recognition - Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*, pp. 171–171. DOI: 10.1109/CVPR.2005.573.

Perronnin, F. and Dugelay, J.-L. (2003). "An Introduction to biometrics and face recognition". In: IMAGE 2003, Learning, Understanding, Information Retrieval, Medical. (Cagliari, Italy, June 9–10, 2003). URL: http://www.eurecom.fr/publication/1550.

Sadri, F. (2011). "Ambient intelligence: A survey". *ACM Computing Surveys (CSUR)* **43**(4), 36. DOI: 10.1145/1978802.1978815.

Tsalakanidou, F., Tzovaras, D., and Strintzis, M. G. (2003). "Use of depth and colour eigenfaces for face recognition". *Pattern Recognition Letters* **24**(9-10), 1427–1435. DOI: 10.1016/S0167-8655(02)00383-5.

Wang, Y. and Chua, C.-S. (2005). "Face recognition from 2D and 3D images using 3D Gabor filters". *Image and Vision Computing* **23**(11), 1018–1028. DOI: 10.1016/j.imavis.2005.07.005.

Wang, Y., Liu, J., and Tang, X. (2010). "Robust 3D Face Recognition by Local Shape Difference Boosting". *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **32**(10), 1858–1870. DOI: 10.1109/TPAMI.2009.200.

Xu, C., Wang, Y., Tan, T., and Quan, L. (2004). "Automatic 3D face recognition combining global geometric features with local shape variation information". In: *Automatic face and gesture recognition, 2004. Proceedings. Sixth IEEE international conference on*, pp. 308–313. DOI: 10.1109/AFGR.2004.1301549.

Xu, D., Hu, P., Cao, W., and Li, H. (2008). "SHREC'08 entry: 3D face recognition using moment invariants". In: *Shape Modeling and Applications, 2008. SMI 2008. IEEE International Conference on*, pp. 261–262. DOI: 10.1109/SMI.2008.4547997.

[a] Università degli Studi di Salerno
Dipartimento DISTRA
Via Giovanni Paolo II, 84084 Fisciano (SA), Italy

* To whom correspondence should be addressed | Email: abate@unisa.it

Communicated 30 May 2013; manuscript received 10 May 2014; published online 19 November 2015.