Interpreting 3D Faces for Augmented Human-Computer Interaction

Marinella Cadoni, Enrico Grosso, Andrea Lagorio, and Massimo Tistarelli

University of Sassari
Computer Vision Laboratory
Porto Conte Ricerche, Tramariglio, Alghero, Italy
{maricadoni,grosso,lagorio,tista}@uniss.it

Abstract. Human-machine interaction requires the ability to analyze and discern human faces. Due to the nature of the 3D to 2D projection, the recognition of human faces from 2D images, in presence of pose and illumination variations, is intrinsically an ill-posed problem. The direct measurement of the shape for the face surface is now a feasible solution to overcome this problem and make it well-posed. This paper proposes a completely automatic algorithm for 3D face registration and matching based on the extraction of stable 3D facial features characterizing the face and the subsequent construction of a signature manifold. The facial features are extracted by performing a continuous-to-discrete scale-space analysis. Registration is driven from the matching of triplets of feature points and the registration error is computed as shape matching score. A major advantage of the proposed method is that no data pre-processing is required. Despite of the high dimensionality of the data (sets of 3D points, possibly with the associate texture), the signature and hence the template generated is very small. Therefore, the management of the biometric data associated to the user data, not only is very robust to environmental changes, but it is also very compact. The method has been tested against the Bosphorus 3D face database and the performances compared to the ICP baseline algorithm. Even in presence of noise in the data, the algorithm proved to be very robust and reported identification performances in line with the current state of the art.

Keywords: Face recognition, 3D faces, Pattern recognition, Scale-space theory, Geometric invariants.

1 Introduction

The interaction among humans is always driven by direct visual perception which conveys several information including the identity, gender, age and emotional state. As such, the analysis and recognition of human faces is of paramount importance to devise a proper interaction of a computer system with humans. The management of identities implies the construction of compact biometric templates, requiring a limited storage and minimal computational resources. This may seem unfeasible when dealing with high dimensional data, such as dense

C. Stephanidis (Ed.): Universal Access in HCI, Part II, HCII 2011, LNCS 6766, pp. 535–544, 2011.

3D face shape representations. The geometric approach proposed in this paper is aimed at minimizing the required storage for the face template by extracting and processing a limited number of characteristic 3D points. The resulting template requires only a few KBytes of data. The information contained in the 3D face shape is exploited to devise a robust and accurate identification system through the alignment of the shapes and the computation of their similarity. The Iterative Closest Point (ICP) algorithm [1] has proven to be very effective to accurately register (or match) 3D face scans, but an approximate initial alignment of the two point sets is required to bootstrap the algorithm. For this reason, an accurate and efficient face registration is always mandatory to perform face recognition. Therefore, in this paper 3D face recognition is tackled as a by product of the registration of 3D point sets. The algorithm is based on the extraction of facial features characterizing the face and the subsequent construction of a signature manifold. Registration is driven from the matching of triplets of feature points. After registration two different processes are performed: the registration error is computed first as shape matching score, secondly the coarse registration is refined by using the Iterative Closest Point (ICP) technique [1]. The final match score is determined by the registration error computed after the last iteration.

The proposed algorithm was tested on the Bosphorus database [2], particularly with faces under different poses. Previous works on this database have concentrated on landmarks detection robust to occlusions and noise. The algorithm proposed in this paper significantly outperform the benchmarks algorithms based on automatic features extraction. To demonstrate the efficiency of the algorithm in real application scenarios, several experimental tests are performed and the results compared to those obtained with the ICP baseline algorithm. Even in presence of noise in the data, the algorithm proved to be very robust and reported identification performances in line with the current state of the art.

2 Extraction of 3D Facial Features

2.1 Scale-Space Theory and 3D Face Analysis

Human faces can be characterized from 3D information just by registering the data from two individuals and measuring the goodness of fit. This process requires to identify anchor points on the faces which are similar for all faces but also to locate 3D features which may be highly distinguishing. Considering a 3D face scan as a smooth surface, both kind of points are either local maxima or minima of the Gaussian curvature. Our aim is then to find an algorithm to extract local maxima and minima of curvature, with a given approximation.

The scale-space theory [6], originally proposed to describe the gray level variations in 2D intensity images, can be applied to 3D face scan to optimally select all "common" points, namely 3D features, to be extracted from a set of 3D faces. Given a scale-space representation of the face, we can characterize the face at each scale by means of the Gaussian curvature at each point. Due to computational time and memory limits, the scale can not be varied continuously, nor can the cloud of points be model with a parametrized surface [7] . This problem is

overcome by extracting, for each 3D scanned point p_i, an approximation of the Gaussian curvature computed on the set of spherical neighbors $N_{p_i}(r_j)$, centered at the point p_i and of increasing radius r_j. The scale step, i.e. the difference between the radii of two consecutive neighbors is chosen on the basis of the sampling density of the scan. In the performed experiments the scale step was determined by constraining, on average, the difference between two neighbors to be equal to 10 points. Given a 3D point p_i and the 3D neighbors $N_{p_i}(r_j)$, an approximation of the Gaussian curvature can be obtained by computing the Principal Components of $N_{p_i}(r_j)$. The eigenvalues $\lambda_0 \leq \lambda_1 \leq \lambda_2$ and the respective eigevectors v_0, v_1, v_2 corresponding to the principal directions, are computed. The absolute value of the curvature is then defined as $\mathcal{C}(p_i, r_j) = \frac{2|(p_i - p_g) \cdot v_0|}{d_m^2}$, where p_g is the center of gravity of the neighbor $N_{p_i}(r_j)$ and d_m^2 is the mean of distances $|p_i - p_j|,\ p_j \in N_{p_i}(r_j)$. The surface normal $\nu(p_i, r_j)$ at the point p_i at scale r_j is computed as the principal direction corresponding to the smallest eigenvalue λ_0.

2.2 Extraction of Features at Multiple Scales

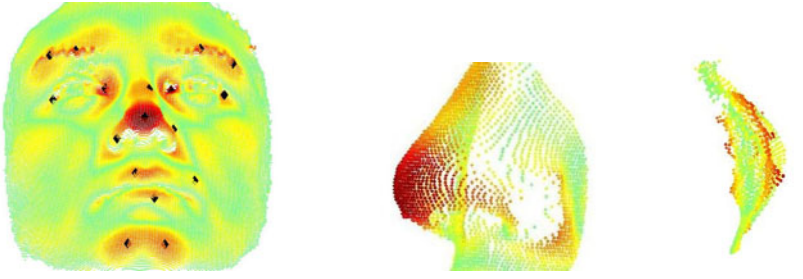
The scale-space 3D feature extraction is based on the following steps:

- Two extreme values for the search radius are set $(r_s\ ,\ r_e)$, determined on the basis of anthropometric facial measures. In all experiments the two radii were set empirically to 6mm and 22mm.
- The scale step σ_s is defined to partition the interval (r_s, r_e) into a set of $n_\sigma = \frac{r_s - r_e}{\sigma_s} + 1$ intervals of equal length.
- For each point p_i of a face scan, the curvature $\mathcal{C}(p_i, r_j)$ is computed for $j = s, s + \sigma_s, s + 2\sigma_s, \ldots, e$. The curvature values are then interpolated to produce a function $\mathcal{C}(p_i) : [s, e] \rightarrow \mathbb{R}$. A median filter is applied to smooth the curve, and the scale $\sigma_m(p_i)$ for which the curvature $\mathcal{C}(i) = \mathcal{C}(p_i, \sigma_m(p_i))$ reaches a maximum is computed. The normal ν_i at point p_i is determined as $\nu(p_i, \sigma_m(p_i))$.

For each point p_i of the face scan an optimal curvature value $\mathcal{C}_i$ and an optimal normal vector $\nu(i)$ are obtained. The face edges are first detected and marked to be excluded from the successive processing. Given $r = (r_e - r_s)/2$, and for each p_i in the face scan, p_i is defined to be a local maxima or minima of the curvature if $|\mathcal{C}_i|$ is the largest of all $|\mathcal{C}_k|$ for $p_k \neq p_i \in N_{p_i}(r)$.

While the number of features is naturally bounded by the radius r, up to 12 points of highest curvature are selected amongst them. In figure 1 (a), the projected surface of a sample 3D face scan is shown. The surface color encodes the curvature values computed at the fixed scale $(r_e - r_s)/2$. The marked points on the surface represent the extracted 3D features.

In most 3D acquisition devices the sampling density of the face scan is lower exactly in those areas where curvature variation occurs. This non-uniform sampling often leads to occlusions and may impair the extraction of feature points. Despite the occlusions and noise in the data, preprocessing of the data has been



(a) (b) (c)

Fig. 1. (a) Feature points extracted from a sample face scan from the Bosphorus database, (b) Subsampled nose area, (c) Noisy eye area

carefully avoided. It is worth stressing that all results presented in the experimental section were obtained without applying any kind of data preprocessing. This allows to better evaluate the performance on face registration and matching as related to the raw data only and not to the quality of any pre-processing step.

3 Registration of 3D Facial Scans

The registration algorithm is based on the Moving Frame Theory [9]. The procedure that leads to the generation of the invariants and the signature are discussed in full detail in [10]. Only the fundamental issues are discussed here.

Given a surface F, the Moving Frame Theory defines a framework (and an algorithm) to calculate a set of invariants, say $\{I_1, \dots I_n\}$, where each I_i is a real valued function that depends on one or more points of the surface. By construction, this set contains the minimum number of invariants that are necessary and sufficient to parametrize a "signature" $S(I_1, \dots, I_n)$ that characterizes the surface up to Euclidean motion. The framework offers the possibility of choosing the number of points the invariants depend on, and this determines both the number n of invariants we get and their differential order. The more the points the invariants depend on the lower the differential order. For instance, invariants that are functions of only one point varying on the surface ($I = I(p)$, $p \in F$) have differential order equal to 2. These are the classical Gaussian and Mean curvatures. In order to trade the computational time with robustness to noise the invariants are built depending on three points at one time. The result is a set of nine invariants, three of differential order zero, and six of order one.

3.1 3-Points Invariants

Let p_1, p_2, $p_3 \in F$ and ν_i be the normal vector at p_i. The directional vector v of the line between p_1 and p_2 and the normal vector ν_t to the plane through p_1, p_2, p_3, are defined as:

$$v = \frac{p_2 - p_1}{\|p_2 - p_1\|} \qquad \text{and} \qquad \nu_t = \frac{(p_2 - p_1) \wedge (p_3 - p_1)}{\|(p_2 - p_1) \wedge (p_3 - p_1)\|}.$$

The zero order invariants are the inter-point distances $I_1 = \|p_2 - p_1\|$, $I_2 = \|p_3 - p_2\|$ and $I_3 = \|p_3 - p_1\|$ whereas the first order invariants are

$$J_k(p_1, p_2, p_3) = \frac{(\nu_t \wedge v) \cdot \nu_k}{\nu_t \cdot \nu_k} \quad \text{and} \quad \tilde{J}_k(p_1, p_2, p_3) = \frac{v \cdot \nu_k}{\nu_t \cdot \nu_k} \quad \text{for } k = 1, 2, 3.$$

Each triplet (p_1, p_2, p_3) on the surface can now be linked with a point of the signature in 9-dimensional space whose coordinates are given by
$(I_1, I_2, I_3, J_1, J_2, J_3, \tilde{J}_1, \tilde{J}_2, \tilde{J}_3)$.

3.2 Matching 3D Face Scans

For each triplet of feature points extracted from a sample face scan F the invariants are computed and stored into a signature S that characterizes F. Two face scans F and F' can be compared by computing the intersection between the two signatures S and S'. If the intersection between S and S' is not null, then exists a subset of feature points belonging to the two scans holding the same properties, i.e. the same inter-point distances and normal vectors (up to Euclidean motion). The signature points are compared by computing the Euclidean distance: given a threshold ϵ, if $s \in S$, $s' \in S'$ and $|s - s'| \leq \epsilon$, then the triplets that generated the signature points are matched. From the triplets the roto-translation $(\mathbf{R}, \mathbf{t})$ that takes the second into the first can be computed. Given $\{t_1, \ldots, t_m\}$ the set of triplets of the face scan F that are matched to the triplets in S', each matched triplet generates a roto-translation $(\mathbf{R}_i, \mathbf{t}_i)$. To select the best registration parameters among those computed, each $(\mathbf{R}_i, \mathbf{t}_i)$ is applied to F', so that $F'' = \mathbf{R}F' + \mathbf{t}$ and the registration error is computed according to the following procedure.

For each point $q_i \in F'$ the closest point p_i in F is computed together with the corresponding Euclidean distance $d_i = \|q_i - p_i\|$. A set of distances $D = \{d_i\}_{i \in I}$ is obtained where I is the cardinality of F'. The registration error is defined to be the median of $D = \{d_i\}_{i \in I}$. The pair $(\mathbf{R}_m, \mathbf{t}_m)$ corresponding to the minimum registration error d_m is chosen as the best registration between the two faces. Whenever the registration step fails (there are no matching points in the signature space and so triplets) the result is accounted as a negative match.

3.3 3D Face Identification

The registration error is used as matching score between two faces F and F'. Due to noise and occlusions, the computed registration score can be still inaccurate. Another algorithm, such as ICP, is applied to refine the registration. In the first iteration, ICP takes as input the two scans aligned through the invariant matching. The registration error after the last iteration is the final matching score. After registration, two scans will be considered a match, i.e. belonging to the same individual, id the matching score is below a fixed threshold σ. After a successful registration of two fairly neutral scans of the same subject, the median distance d_m can be assumed to be $\delta/2 < d_m < \delta$ where δ is the average resolution of the scans.

4 Experimental Results

The proposed algorithm was tested on the Bosphorus database [2]. The database contains about 50 scans of 105 individuals, with 61 male and 44 female subjects. 31 out of the total male subjects have a beard and mustaches. Each scan either presents a different facial expression (anger, happiness, disgust), corresponding to a "Face Action Unit", or a head rotation along different axes. Since the subjects to be identified can be assumed to be cooperative, we will simulate an authentication scenario using the sets of faces that are fairly neutral and only slightly rotated sideways, upwards and downwards. Examples of the scans for two subjects are shown in figure 2. The picture shows (in a clockwise direction): a neutral pose, a slight downwards rotation, a slight upwards rotation, a 10° head rotation on the right.

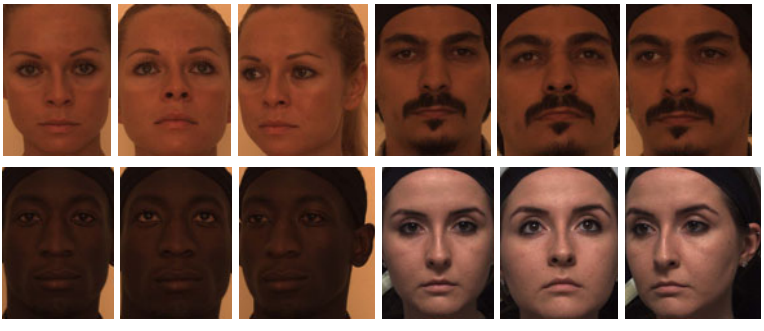


Fig. 2. Sample 3D scans of four subjects in the Bosphorus database

This database has been chosen because it contains a large number of subjects and an excellent variety of poses. Furthermore, despite only geometric information (3D points) is used for identificaiton, the availability of landmark points constitutes a ground truth which makes it possible to compare the methodology with a baseline algorithm.

The database was divided into a gallery set G and two probe sets P_1, P_2. The gallery G consists of one neutral face scan for each individual (named N-N in the database). The neutral scan could be stored in a smart card or ID card of an individual in the form of text file, whereas the poses in P_i, $i = 1, 2$ can be assumed to be the scans taken from the acquisition device when the subject undergoes authentication.

Three authentication tests were run. In all of them, the gallery consisted of the neutral poses, such as the first sample of each subject in figure 2.

1. **P_1 vs G**. The probe set P_1 consists of the scans labeled PR-SU in the database (105 scans in total, one for each subject). The pose is a slight rotation of the face upwards as shown in the second image of each subject in figure 2. Each scan of P_1 was compared to all scans of the neutral gallery G using the methodology described in section 3.3.

2. **P_2 vs G**. The probe set P_2 consists of the scans labeled YR-R10 in the database (105 scans in total, one for each subject). The pose is a rotation of the face of about 10 deg on one side, as shown in the third image of each subject in 2. Again, each scan of P_2 was compared to all scans of the neutral gallery G as in 3.3.
3. **Manual P_1 vs G**. This is the baseline algorithm. Each scan in P_1 was roughly aligned with each scan of G using three of the manually selected landmarks provided by the database (the two inner eye corners and nose tip) and the alignment refined with ICP.

All algorithms were implemented in MatLab. On a consumer PC, the computational time to extract the features from a face scan of 30.000 points was on average 2 min. The signature generation took about 3 sec. For the registration of two scans, times varied from 2 sec for scans of different subjects to 20 sec for those of the same subject. The results of the tests are summarized in table 1. The failed registrations (*F.R.*) is the number of subjects for which no triples were matched in the signature space. These numbers are indicative of the robustness of the method. In fact, if a registration fails there is no later chance of refinement. No registration failures occurred in experiment 1 and 2.

Table 1. Matching scores

Experiment	F.R.	A.R.	T.P.	F.P.	F.N.	T.N.	Acc
1	0	0.981	103	0	2	10920	0.9998
2	0	0.924	97	0	8	10920	0.9992
3	0	0.99	104	0	1	10920	0.9999

In the third column of table 1, *A.R.* indicates the authentication rate (number of correctly identified subjects over the total of 105) obtained using as matching score the registration error that follows from the automatic feature extraction and the registration through invariants refined by ICP. *T.P.* is the number of true positives, *F.P.* the number of false positives, *F.N.* that of false negatives, and *T.N.* that of true negatives. In the last column, *Acc* stands for accuracy and it is defined by $Acc = \dfrac{TP + TN}{P + N}$, where $P = 105$ is the number of positives and $N = 10920$ is the number of negatives. The noise associated to some of the scans accounts for the false negative, therefore preprocessing the data or acquiring them with a lower noise system would reduce the number significatively. In figure 3, the matching scores after registration of the probe scans to the gallery scans are shown. The cases of registration failure between different subjects are omitted. Each number from 1 to 105 along the x-axis refers to the gallery scan of a subject. For each subject i, the matching scores after registration of the gallery scan with all probe scans are represented along the column (i, y). The scores are marked by gray circles if the probe subject is different from the i subject and with a black star if the probe scan is of subject i. As we can see from figure 3, the threshold (horizontal line, set to be equal to 0.65 for this database), separates

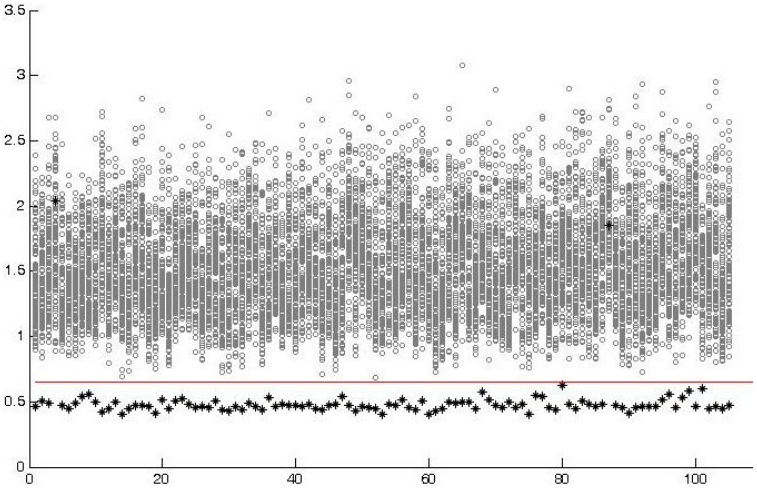


Fig. 3. Distribution of scores from experiment 1

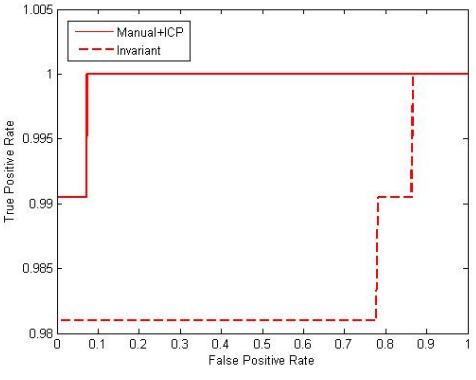


Fig. 4. ROC curves for experiments 1 (red) and 3 (blue)

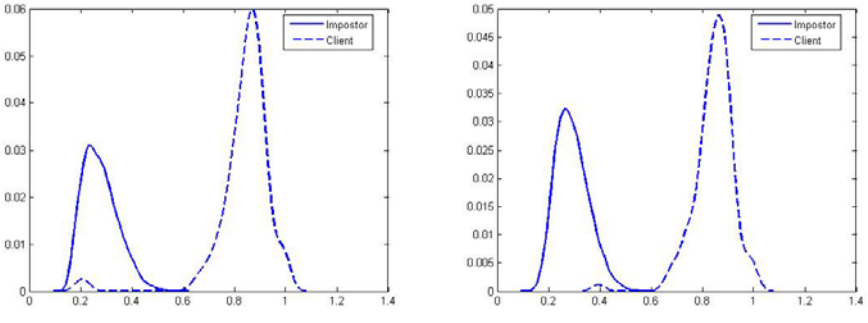


Fig. 5. Impostor and client distribution for experiment 1 (left), and 3 (right)

the two classes client and impostor very well. The performance as the threshold varies is shown by the two ROC curves in figure 4. The images in figure 5 show the separation of the client and impostor classes in experiment 1 and 3. On the x-axis a similarity measure of two faces is given as the inverse of the registration error. It can be seen that the baseline algorithm does not significantly improve the separation of classes obtained with the automatic one, although it manages to identify one of the two subjects on which the proposed method fails.

5 Conclusions

The identification of individuals on the basis of 3D shape information only has been addressed. This is a very promising and challenging biometric technology at the same time, because of the difficulties in processing three-dimensional data and of the advantages such as the relative insensitivity to illumination changes. The proposed method, based on the scale-space theory for the extraction of stable 3D feature points and on the generation of an invariant signature to characterize the face shape, proved to be very robust at identifying subjects, providing very good performances in terms of matching accuracy avoiding any data pre-processing to either fill-in holes or smooth the face surface to remove spikes within the points cloud. Moreover, the procedure is highly flexible regarding the storage and on-line processing: by storing the 3D points only more on-line processing is required, whereas by storing the feature points or the signature of the face shape, on-line processing is increasingly reduced.

Even though changes in facial expression were not addressed, these can be very important in building a proper man-machine interface [11]. Facial expressions play a central role in understanding the mood, emotions and even intentions of the counter-part. As such, further work will be devoted to properly understand and model the deformation of key face areas. The scale-space analysis presented in this paper can be further enhanced to extract and build 3D invariants which are robust to such deformations, but also to define a dynamic facial signature which encompasses the identity as well as the emotional state of the subject.

Further performance improvements are expected with a light data pre-processing, e.g. cropping the central part of the face to remove spikes due to hair or acquisition artifacts, or by consolidating the extraction the feature points of the gallery image with the aid of texture information.

References

1. Besl, P.J., McKay, N.D.: A Method for Registration of 3-D Shapes. IEEE Trans. on Pattern Analysis and Machine Intelligence 14, 239–256 (1992)
2. Savran, A., Alyüz, N., Dibeklioğlu, H., Çeliktutan, O., Gökberk, B., Sankur, B., Akarun, L.: Bosphorus database for 3D face analysis. In: Schouten, B., Juul, N.C., Drygajlo, A., Tistarelli, M. (eds.) BIOID 2008. LNCS, vol. 5372, pp. 47–56. Springer, Heidelberg (2008)

3. Çeliktutan, O., Çinar, H., Sankur, B.: Automatic Facial Feature Extraction Robust Against Facial Expressions and Pose Variations. In: IEEE Int. Conf. on Automatic Face and Gesture Recognition. Holland, Amsterdam (2008)
4. Dibeklioğlu, H., Salah, A., Akarun, L.: 3D Facial Landmarking Under Expression, Pose, and Occlusion Variations. In: IEEE 2nd International Conferance on Biometrics: Theory, Applications, and Systems (IEEE BTAS), Washington, DC, USA (September 2008)
5. Gorkberk, B., Savran, A., Ali, A., Akarun, L., Sankur, B.: 3D face recognition benchmarks on the bosphorus database with focus on facial expressions. In: Schouten, B., Juul, N.C., Drygajlo, A., Tistarelli, M. (eds.) BIOID 2008. LNCS, vol. 5372, pp. 57–66. Springer, Heidelberg (2008)
6. Lindeberg, T.: Feature Detection with Automatic Scale Selection. International Journal of Computer Vision 30(2), 77–116 (1998)
7. Pauly, M., Keiser, R., Gross, M.: Multi-scale Feature Extraction on Point-sampled Surfaces. In: Proceedings of Eurographics 2003, vol. 22(3) (2003)
8. Witkin, A.: A Scale Space Filtering. In: Proc. 8th Int. Joint Conference on Artificial Intelligence (1983)
9. Olver, P.J.: Joint Invariants Signatures. Found. Comput. Math. 1, 3–67 (2001)
10. Cadoni, M.I., Bicego, M., Grosso, E.: 3D Face Recognition Using Joint Differential Invariants. In: Tistarelli, M., Nixon, M.S. (eds.) ICB 2009. LNCS, vol. 5558, pp. 11–25. Springer, Heidelberg (2009)
11. Tistarelli, M., Schouten, B.: Biometrics in ambient intelligence. Journal of Ambient Intelligence and Humanized Computing, 1–14 (November 2010), http://dx.doi.org/10.1007/s12652-010-0033-z