# Scale-space Techniques for Fiducial Points Extraction from 3D Faces

Nikolas De Giorgis and Luigi Rocca and Enrico Puppo

Department of Informatics, Bio-engineering, Robotics and System Engineering
University of Genova - Via Dodecaneso 35 - 16146
Genova - ITALY {nikolas.degiorgis|rocca|puppo}@dibris.unige.it

**Abstract.** The quest for fast and reliable extraction of fiducial points from human faces using three dimensional surface information is still a work in progress in current literature, despite its many potential advantages compared to image-based extraction, such as less sensitiveness to variations in pose and illumination. We propose an extraction method that uses 3D information only and is based on two key steps: multi-scale curvature analysis, and the reliable tracking of the deep structure of a virtually continuous scale-space based on curvature. Our scale based analysis, coupled to careful use of prior information based on variability boundaries of anthropometric facial proportions, does not require a training step, because it makes direct use of morphological characteristics of the analyzed surface. The proposed method precisely identifies important fiducial points with high precision and is able to extract new fiducial points that were previously unrecognized, thus paving the way to more effective recognition algorithms.

**Keywords:** scale-space, multi-scale, curvature, 3D computer vision, shape analysis, fiducial points extraction

## 1 Introduction

Face recognition has been widely studied and addressed in the literature, mainly in the image processing field. It can be described as the task of extracting descriptors, from an image depicting a human face, which can be used to discriminate if two of those are obtained from images of the same face.

Many works have been presented dealing with color and gray-scale images, among which the most famous are *PCA* [8], *LDA* [6] and *EBGM* [18]. Recognition from 2D images, though, suffers from a lot of known problems, such as a strong dependency on consistent illumination and pose. Moreover, it is straightforward to see that images cannot carry all the original information about a face's structure. Despite these shortcomings, work on 3D face recognition has been less investigated in the past, because complex and exotic hardware is needed for the extraction of 3D data and because of the consequent lack of publicly available datasets with good enough resolution. All this has started to change in the last few years, as the hardware landscape improved and the growth in available computational power not only unlocked usage of more complex software techniques

during processing steps, such as surface reconstruction and meshing, cleaning and smoothing, but also enabled novel extraction techniques of 3D raw data, such as photogrammetry.

Existent methods that extract fiducial points from 3D data can be roughly subdivided into *appearance based* and *feature based*; the first class is made of methods that are typically modified versions of 2D algorithms extended to work with range images. The second class contains methods that work by extracting local relevant features. The method we propose falls into the latter category, of which we are going to give a brief overview. Lu and Jain [10] proposed a method that combines 2D and 3D techniques to extract a small set of facial features: they use *a priori* knowledge to detect the tip of the nose, and then detect the mouth and eye corners using the shape index from the range image. Colbry et al. [11] use shape index to detect a similar set of features. Gupta et al [7] developed a method which detects a set of 10 points combining curvature information, *a priori* information and 2D techniques. Perakis et al [14] developed a method which aims at detection of facials landmarks in presence of large yaw and expression variations using shape index and spin images. Conde et al [4] developed a 3D method based on spin images which obtains a high accuracy but gets only three points on the faces. Segundo et al [12] use curvature information combined with depth values from range images to detect a small set of points (nose tip, nose and eye inner corners). Shin and Sohn [16] use ten facial landmarks for face recognition, but they do not give details about how these points are extracted. Sukno et al. [17] detect a set of fiducial landmarks using spin images as described in [9] but then use statistical models to filter out outliers and infer missing features. Bockeler and Zhou [2] detect a set of ten points with strong 2D information and antrophometric constraints.

Our method extends the family of techniques based on curvature and on prior knowledge of anthropometric features' locations. However, it makes use of the 3D surface only, without need for color or light intensity information, and it doesn't need any kind of learning or training phases. First of all, we compute curvature at different scales of increasing sizes, from finer to coarser. This allows us to build a scale-space based on a curvature derived scalar field, and to perform a tracking of its extremalities through scales. These processing steps give us a selection of points, representing notable features of the surface, ranked by an importance measure that corresponds to their lifetime in the scale-space. We can thus take into account both the actual strength and the importance in terms of scale of feature points when fiducial points are searched among them, achieving reliable recognition of all fiducial points identified in previous literature and others that were, to the best of our knowledge, previously unrecognized.

## 2 Multi-scale Curvature

Curvature is a second order differential property of surfaces that is represented by a *tensor field*. at each point two principal directions of curvature exist, along which the curvature is respectively maximum and minimum; the values $k_1, k_2$ of

the curvature along these directions are called *principal curvatures*. In geometry processing literature, two main families of curvatures estimation methods for digital surfaces can be identified: discrete methods, which directly derive differential properties from the discrete geometry of the mesh, and fitting methods, which fit an analytic surface to input data and derive differential properties analytically. Most fiducial point extraction works that use curvature compute it using discrete methods, which tend to highlight features at the finest scales and to be prone to noise. Moreover, it is accepted knowledge in the geometry processing field that these disadvantages tend to be exacerbated, instead of being reduced, as the resolution and size of datasets grow. We suggest the adoption of multi-scale curvature analysis methods; in particular, we adopt a method based on fitting[13]. The scale parameter is the size of the local surface around a vertex that contributes to the computation of the curvature value of the vertex itself, with the size measured as the radius of a sphere.

For the purposes of our scale-space analysis we need a scalar field based on curvature. We therefore derive from the curvature tensor field the *Gaussian curvature* $K = k_1 * k_2$, a scalar field which provides a good characterization of a surface's features; in particular:

- Points in which $K < 0$ are hyperbolic points (saddles)
- Points with $K > 0$ are elliptic point (either pits or peaks)
- Points with $K = 0$ are parabolic points.

We are interested in points that are either maxima or minima of Gaussian curvature. As the scale parameter increases, extremal points tend to decrease and to identify bigger, coarser features.

## 3 Virtually Continuous Scale-space

Since their introduction, scale-space methods have been widely used in computer vision and image processing; the general idea is to build a one parameter family of images from an input signal. This is usually done by applying a filter repeatedly, thus building a discrete sequence. Among the many scale-spaces that have been proposed, the linear scale-space, which adopts Gaussian smoothing as a filter, is one of the most studied. An example of an alternative scale-space could be a curvature scale-space, built by repeated computation of curvature using the method we adopt in Section 2, as the radius parameter, and thus the size of the local surface, increase at every step.

The main goal of scale-space methods is to highlight features at different levels of detail and importance. One of the classic approaches to this end is the computation of the *deep structure*, which is the tracking of the extremalities of the signal as they change across the scale space. Tracking, though, is traditionally performed by neighborhood search across discrete levels of the scale-space. This approach is prone to errors and mismatches, and has coarse results. To overcome this difficulties we employ a novel method, which exploits linear interpolation between consecutive levels in order to approximate the continuous process that a

scale-space represents with discrete samples. This virtually continuous approach has been introduced in [15]. The most interesting feature of this method is that each critical point that is present in the original signal can be tracked across scales at the finest level of detail possible, and its lifetime across scales can be reliably reconstructed. In particular, we can consider the filtering process as a time-dependent function and associate each point to the moment in time of their disappearance in the scale-space. This lifetime value is well suited in the role of an importance score, associated to each critical point and to the feature in the signal that the critical point represents. It is also worth mentioning that this tracking method is filter-agnostic and it is not limited to Gaussian smoothing.

## 4 Extraction Method

Our extraction method is based on two main phases: a precomputation phase, in which we employ the multi-scale techniques previously described in order to select a set of candidate points and measure their importance; and an identification phase, in which we use prior knowledge in order to identify fiducial points among candidates in the previously selected set.

### 4.1 Selection and Ranking of Candidate Points

Our goal is to extract all interesting morphological features on a 3D surface representing a human face and measure them according to criteria that will enable us to discriminate between different characteristics that they may have. In particular, our main criterion is their lifetime in the scale-space, but we also developed a secondary criterion which takes into account the relative strength of a feature in comparison to the local surface surrounding it.

**Diagonal scale-space.** Computation of differential properties is severely affected by the presence noise. This makes the most straightforward combination of the concepts that we delineated in previous sections—a scale-space of curvature where the discrete levels were computed with radii of increasing size—an inadequate solution, because the number of extremalities does not decrease fast enough as the curvature method's scale parameter increase, and, as a consequence, tracking them does not provide meaningful information. We propose a new type of scale-space that combines multi-scale curvature with a Gaussian scale-space, called *diagonal scale-space*, that offers a solution to this problem.

The diagonal scale-space is made up by scalar fields representing Gaussian curvature at increasing scales, but it is generated by employing both a smoothing filter on the original 3D surface and by varying at the same time the parameter of the multi-scale curvature computation method. In fact, we initially compute a linear scale-space of the original surface, with consecutive samples generated by repeated smoothing with variances of increasing size. We then compute curvature on each level, with a correspondingly bigger radius the smoother that level is

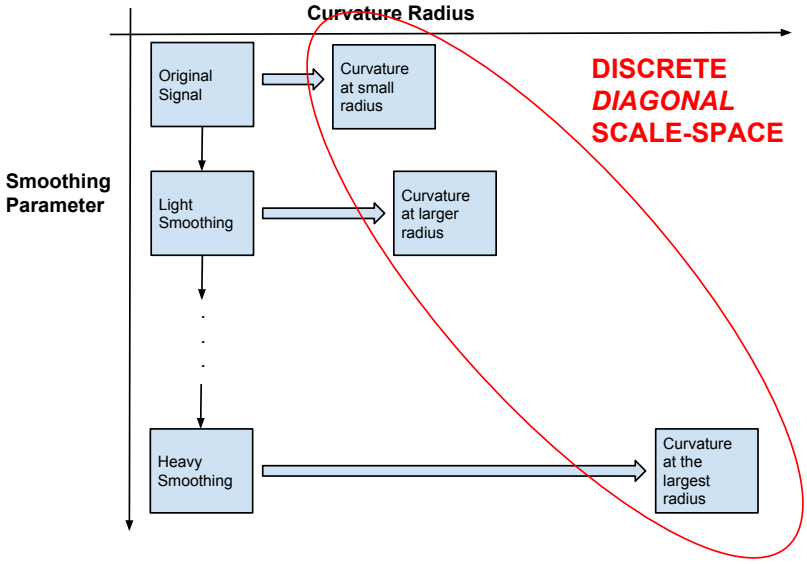


**Fig. 1.** Diagonal Scale-Space

in the linear scale-space. A graphical account of this arrangement is depicted in Figure 1.

We used range images and a Gaussian smoothing filter, but the general idea of increasing two different scale parameters together, one the amount of surface smoothing and the other the local surface's size applied during curvature computation, should be equally effective with triangle meshes and a Laplacian smoothing filter. The end result is that noise is discarded in a more effective way, and that the number of features decreases faster. We are therefore able to achieve a meaningful tracking of the extremalities of the Gaussian curvature scalar field.

**Importance measures: life and strength.** In our experiments we generated depth images of the faces at our disposal by rendering them and getting depth information from the *z buffer*, because the tracking algorithm supports range images only. We then generate the diagonal scale-space, extract all the extremalities in the original signal (which, in our case, is the curvature at the smallest radius computed on the unsmoothed surface) and track them through scales. We do this by identifying which pairs of neighboring pixels invert the relation between their values in consecutive levels of the scale-space. This events, called *flips*, can be sorted in time using linear interpolation. The result is an ordered sequence of all the events that may have an influence on the position or existence of extremalities in the scale-space, which makes it possible to easily track features

across the scale-space with high granularity and precision; the result is a detailed log of their movements, disappearances, and appearances. A detailed account of the tracking method can be found in [15]. Extremal points always appear and disappear in pairs (a saddle coupled to either a maximum or a minimum); we call this events *birth* and *death* events. The moment in time of death events is their lifetime in the scale-space. Birth events are disruptive, because these new features do not correspond to original features in the scale-space and sometimes have long lives in the rest of the scale-space. We recover the lifetime of original critical points by following the chain of their death events as long as it is made of newborn points. Their lifetime value becomes the time of the first death that happens with an original point; we consider this recovered measure to be the *life* value, our main criterion in the identification phase.

From here on, we are interested only in maxima and minima of Gaussian curvature; saddles can be discarded. After tracking in the diagonal scale-space has been performed, we compute the *strength* value, our secondary criterion. Our goal is to assess the relative strength of the scalar field's extremalities, compared to the local trend on the surrounding surface. For each maximum, we compute the average of the pixels that are below its value in a growing area around it, and return the highest difference between its value and the average; the same algorithm is applied to minima by taking into account only the surface values above the minimum. The growth of the local area is capped at a value related to its life in the scale-space.

### 4.2 Identification of Fiducial Points

The fiducial points we identify are a subset of the 25 points presented in [5], and more precisely the 10 points that are found by the method developed in [7], plus the points named *sn*, *ls* and *li*, as shown in figure 2. Fiducial points are selected among extremalities of the Gaussian curvature scalar field. Our strategy, which relies only on prior knowledge and on the *life* and *strength* measures, is based on a hierarchical search. We start by identifying the most prominent features and then seek out other features in narrowed down areas, found by displacements relative to previously found ones. In particular, we identify fiducial points that characterize the nose and compute a symmetry axis that separates the left and right parts of the face. We then proceed to identification of the eyes' corners and of peculiar points on the mouth.

**The nose.** Five fiducial points characterize this area:

- The nose tip, *prn*. This is a very prominent feature which is characterized by a high Gaussian curvature, a long life in the scale-space and by having the highest vertical value. The best strategy is very simple: we restrict the search to an area centered around the center of the mesh, and we extract the maximum of Gaussian curvature with the highest vertical value.
- The sides of the nose, $al_l$ and $al_r$. Those two points are saddles on the surface, which means they are minima of Gaussian curvature. To detect them we look

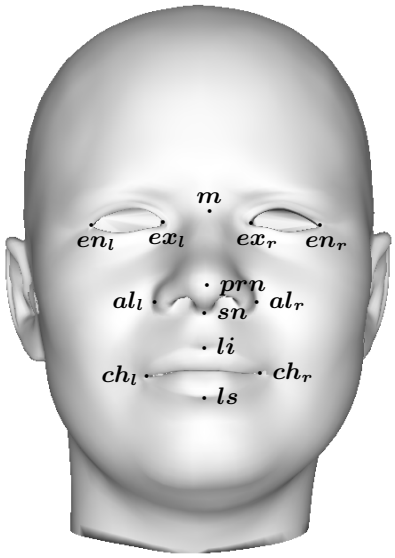


**Fig. 2.** Names and positions of fiducial points extracted by the proposed method.

to the left and to the right of the nose tip and we extract the minima (one on the left and one on the right) with the highest life value in those areas.

- The upper nose saddle, $m$. This is one of the most prominent saddles on a face's surface. In order to locate it we scan an area directly above the nose tip and select the minimum which survives the longest.
- The lower limit of the nose, $sn$. This point, located on the saddle where the nose ends, is a minimum of Gaussian curvature. We have discovered that this point is more reliably characterized by strength; in order to find it we narrow down our research to an area below the tip of the nose and we look for the point with the highest strength.

**Symmetry axis and the eyes.** After points around the nose are identified, we use them to compute a vertical symmetry axis. Our goal is to take advantage the intrinsic symmetry of the human face during the next phases. The axis is computed as the average of the line that fits the points $m$, $prn$, $sn$ and the line orthogonal to the one connecting the two points identified as $al$. Most of the other fiducial points we still need to detect are pairs of symmetric points, specular in relation to the axis. From now on, when we seek a pair of left and right points, we evaluate their fitness together by requiring them to be almost symmetric, within a given tolerance, on top of any other criteria that may be necessary in order to identify them. Moreover, we consider the line connecting $al_l$ and $al_r$ as the dividing line between the upper half and the lower half of the face.

We employ the aforementioned strategy in order to find the pairs of fiducial points that characterize the eyes: the external corners, $ex_l$ and $ex_r$, and the internal corners, $en_l$ and $en_r$. These points are in pit regions, which means they have high Gaussian curvature. We wish to extract the two symmetric pairs in the upper half of the image that have the highest strength value. We perform this by selecting all possible symmetric pairs of maxima $(a, b)$, with strength values $(s_a, s_b)$; we keep the two pairs that have the highest $s_a \cdot s_b$ value.

**The mouth.** This area contains four fiducial points: the pair that represents the corner of the mouth, and the two points representing the tip of the higher lip and the tip of the lower lip. The corners of the mouth, $ch_l$ and $ch_r$, are identified with the same strategy employed for the eyes' corners, applied to the lower half of the face. The upper lip, $ls$, and the lower lip, $li$, are identified as the maxima of Gaussian curvature with the highest life in the area below $sn$ delimited by $ch_l$ and $ch_r$.

## 5 Experiments and Results

Experiments were run on Face Warehouse dataset [3], using meshes representing faces with neutral expressions, for a total of 111 different faces. The dataset does not provide a ground truth for fiducial points, so we created one by manually selecting on every face the thirteen points shown in Figure 2. The method's performance is evaluated by measuring the distance in millimeters between each fiducial point we extract and the corresponding ground truth, for every mesh in the test set. Figures 3, 4, and 5 show the percentage of meshes (Y axis) on which the distance is less than the given millimeters (X axis), for each fiducial point.

- Figure 3 shows results for fiducial points depicting features on the nose. The localization accuracy in this area is high: when the distance from the ground truth for fiducial point reach 4mm, $prn$ is localized on 99% of the dataset, and $sn$ is localized on 94% of the dataset. It should ne noted that, as far as we know, this work is the first to achieve detection of this particular fiducial point. At 7mm, $a_l$ and $a_r$ reach a detection rate of 90%. The point with worst performance in this area is $m$, the nose saddle, which reaches 90% at 11mm. It should be noted that this point is difficult to manually place, because the nose saddle is wide.
- Figure 4 shows results for the eyes' corners. Our method performs with good accuracy for these fiducial points. All four point already reach a detection rate above 90% within a 3mm distance.
- Figure 5 shows results for fiducial points located around the mouth. Features depicted by these points are subtle, and extraction is easily affected by noise. In fact, only a few works have tried to detect the mouth corner, $ch_l$ and $ch_r$ ([7], [2], [1], [14] and [17]) and they always use also 2D information. In our case, extraction suffers because a lot of points along the mouth tend to have similar curvature values. 90% accuracy is reached at 14mm. To the best of

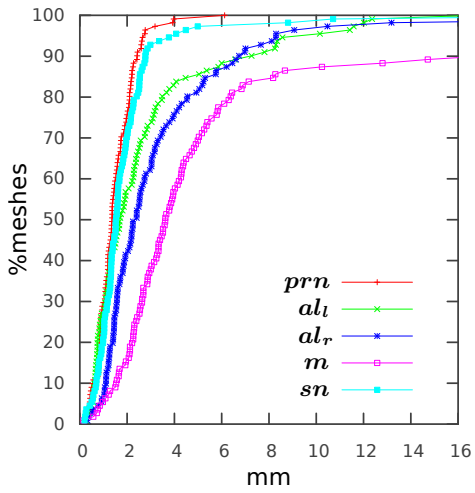


**Fig. 3.** Localization accuracy for fiducial points $prn$, $m$, $al_l$, $al_r$, and $sn$.

our knowledge, this work is the first one that performs detection of fiducial points on the upper and lower lip, $ls$ and $li$. Detection of this points reach a 90% rate at 8mm.

## 6 Conclusions

We presented a novel technique for extraction of fiducial points on human faces which makes use of 3D data. Since we rely on the surface's morphological information only, no training is needed for the method to be effective. Fiducial points that were already extracted using 2D+3D techniques in previous works are detected with a performance that is at least as good, and we show the capability of our methodology by extracting three new fiducial points that were previously undetected. Results are promising and we plan to extend the method and test it on a wider range of datasets. In particular, we are currently working on a version that uses triangle meshes and a Laplacian filter. We plan to experiment with meshes with different facial expressions, and with range images taken from a lateral point of view (or 3D meshes with occlusions and missing pieces), in order to test for robustness in unstaged settings, where non-neutral expressions and large variations in roll and yaw in the range data could occur. we also plan to optimize execution times for the preprocessing part of our algorithm. The largest time is spent in computing curvature data; this task is trivial to parallelize and is a good candidate for GPGPU computation, because curvature on each vertex can be computed independently from other vertices. More research work could open the way to a curvature scale-space directly built from 3D raw data, instead of meshes or range images, which could be computed in a dramatically faster way.

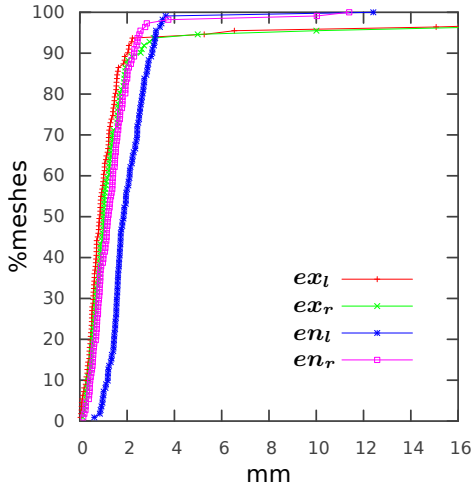


**Fig. 4.** Localization accuracy for fiducial points $en_l$, $en_r$, $ex_l$, and $ex_r$.

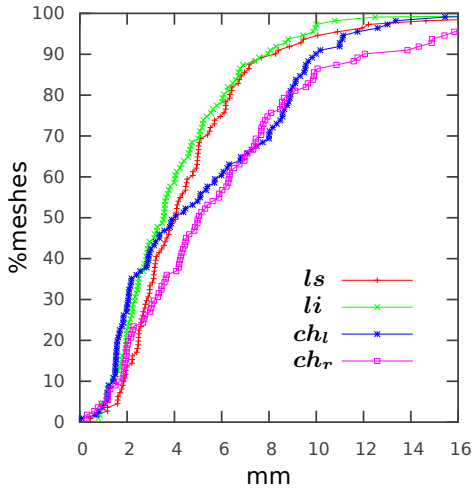


**Fig. 5.** Localization accuracy for fiducial points $ch_l$, $ch_r$, $ls$, and $li$.

## References


1. Beumier, C., Acheroy, M.: Automatic face verification from 3d and grey level clues. In: 11TH PORTUGUESE CONFERENCE ON PATTERN RECOGNITION. pp. 95–101 (2000)
2. Bockeler, M., Zhou, X.: An efficient 3d facial landmark detection algorithm with haar-like features and anthropometric constraints. In: Biometrics Special Interest Group (BIOSIG), 2013 International Conference of the. pp. 1–8 (Sept 2013)
3. Cao, C., Weng, Y., Zhou, S., Tong, Y., Zhou, K.: Facewarehouse: A 3d facial expression database for visual computing. IEEE Transactions on Visualization and Computer Graphics 20(3), 413–425 (2014)
4. Conde, C., Cipolla, R., Rodríguez-Aragón, L.J., Serrano, Á., Cabello, E.: 3d facial feature location with spin images. In: MVA. pp. 418–421 (2005)
5. Farkas, L., Munro, I.: Anthropometric facial proportions in medicine. Thomas (1987)
6. Friedman, J.H.: Regularized discriminant analysis. Journal of the American Statistical Association 84(405), 165–175 (1989)
7. Gupta, S., Markey, M.K., Bovik, A.C.: Anthropometric 3d face recognition. Int. J. Comput. Vision 90(3), 331–349 (Dec 2010)
8. Hesher, C., Srivastava, A., Erlebacher, G.: A novel technique for face recognition using range imaging. In: Signal Processing and Its Applications, 2003. Proceedings. Seventh International Symposium on. vol. 2, pp. 201–204 vol.2 (2003)
9. Johnson, A., Hebert, M.: Using spin images for efficient object recognition in cluttered 3d scenes. Pattern Analysis and Machine Intelligence, IEEE Transactions on 21(5), 433–449 (May 1999)
10. Lu, X., Jain, A.K.: Automatic feature extraction for multiview 3d face recognition. In: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition. pp. 585–590. FGR '06, IEEE, Washington, DC, USA (2006)
11. Lu, X., Jain, A.K., Colbry, D.: Matching 2.5d face scans to 3d models. IEEE Trans. Pattern Anal. Mach. Intell. 28(1), 31–43 (Jan 2006)
12. Pamplona Segundo, M., Silva, L., Bellon, O., Queirolo, C.: Automatic face segmentation and facial landmark detection in range images. Systems, Man, and Cybernetics, IEEE Transactions on 40, 1319–1330 (Oct 2010)
13. Panozzo, D., Puppo, E., Rocca, L.: Efficient multi-scale curvature and crease estimation. Proceedings Workshop on Computer Graphics, Computer Vision and Mathematics, (Sep 2010)
14. Perakis, P., Passalis, G., Theoharis, T., Kakadiaris, I.A.: 3d facial landmark detection under large yaw and expression variations. IEEE Transactions on Pattern Analysis and Machine Intelligence 35(7), 1552–1564 (2013)
15. Rocca, L., Puppo, E.: A virtually continuous representation of the deep structure of scale-space. Int. Conf. on Image Analysis and Processing (2013)
16. Shin, H., Sohn, K.: 3d face recognition with geometrically localized surface shape indexes. In: Control, Automation, Robotics and Vision, 2006. ICARCV '06. 9th International Conference on (Dec 2006)
17. Sukno, F., Waddington, J., Whelan, P.: 3d facial landmark localization using combinatorial search and shape regression. In: Fusiello, A., Murino, V., Cucchiara, R. (eds.) Computer Vision ECCV 2012. Workshops and Demonstrations, Lecture Notes in Computer Science, vol. 7583, pp. 32–41. Springer Berlin Heidelberg (2012)
18. Uchida, S., Sakoe, H.: A survey of elastic matching techniques for handwritten character recognition. IEICE - Trans. Inf. Syst. E88-D(8) (Aug 2005)