Scale-Space Techniques for Fiducial Points Extraction from 3D Faces

Nikolas De Giorgis(✉), 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. We propose a method for extracting fiducial points from human faces that uses 3D information only and is based on two key steps: multi-scale curvature analysis, and the reliable tracking of features in a scale-space based on curvature. Our scale-space 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 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 images depicting human faces, which can be used to discriminate if two such images are obtained from the same face.

Many works have been presented dealing with color and gray-scale images, among which the most famous are *PCA* [9], *LDA* [7] and *EBGM* [22]. Recognition from 2D images, though, suffers from several 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 were needed for the extraction of 3D data and because of the consequent lack of publicly available datasets with good enough resolution. In the last few years, 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

V. Murino and E. Puppo (Eds.): ICIAP 2015, Part I, LNCS 9279, pp. 421–431, 2015.
DOI: 10.1007/978-3-319-23231-7_38

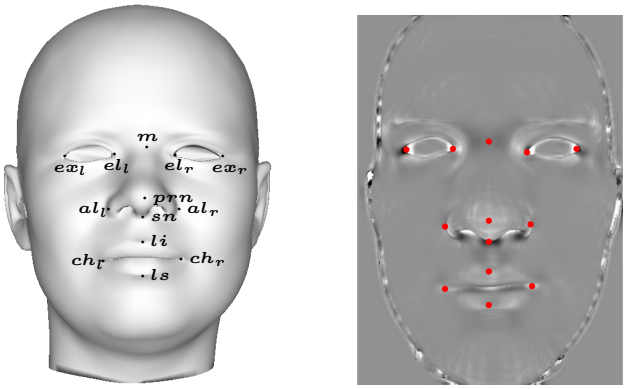


Fig. 1. On the left: names and positions of fiducial points; on the right: the same points, as extracted from a range image with our scale-space method.

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 [11] 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. [12] use shape index to detect a similar set of features. Gupta et al. [8] developed a method which detects a set of 10 points combining curvature information, *a priori* information and 2D techniques. Perakis et al. [18] 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. [5] developed a 3D method based on spin images which obtains a high accuracy but gets only three points on the faces. Segundo et al. [16] 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 [20] use ten facial landmarks for face recognition, but they do not give details about how these points are extracted. Sukno et al. [21] detect a set of fiducial landmarks using spin images as described in [10] but then use statistical models to filter out outliers and infer missing features. Bockeler and Zhou [3] detect a set of ten points with strong 2D information and antrophometric constraints. A work by Berretti et al. [1] computes DoG of a mean curvature scalar field and extracts a variable number of keypoints that are not necessarily located in meaningful parts of the face. Some works by Novatnack et al. [14], [13] [15] use mesh parameterization with a distortion-adapted Gaussian scale-space to extract features using image analysis techniques (edge and corner detection) on the 2D plane.

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 does not need any kind of learning or training phases. The input of the extraction algorithm is a range image; the core ingredient employed is a sequence of curvature fields, computed at different scales, which sets up a scale-space of differential properties of the original surface. We select fiducial points among curvature's critical points, using information computed from the scale-space as a guide. The 13 fiducial points we identify (shown in figure 1) are a subset of the 25 points presented in [6], and more precisely the 10 points that are found by the method developed in [8], plus the points named *sn*, *ls* and *li*.

2 Extraction of Morphological Features

Our goal is to extract all interesting morphological features from a 3D surface representing a human face. It seems intuitive that those features should occur where the surface varies the most, thus making the computation of curvature a very useful tool in this endeavor. Most methods for fiducial point extraction 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 rather adopt a multi-scale curvature analysis method based on surface fitting [17]. The scale parameter is the size of the local surface around a vertex that contributes to the computation of curvature at the vertex itself, with the size measured as the radius of a sphere.

For the purposes of our scale-space analysis we use the *Gaussian curvature,* a scalar field which provides a good characterization of surface features. Our claim is that fiducial points occur at "important" maxima and minima of Gaussian curvature. Therefore, reliable criteria are needed, which can discriminate critical points of Gaussian curvature worth keeping from others caused by noise or depicting irrelevant features.

In order to measure the importance of critical points, we employ a scale-space based approach. 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. 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*, i.e., the tracking of critical points of the signal as they change across the scale-space. Classic approaches to deep structure computation are prone to noise and tracking errors; we adopt a virtually continuous scale-space technique, introduced in [19], which solves many of those problems. This method, which is filter agnostic and relies on piece-wise linear interpolation across scales, provides a fine-grained and reliable tracking of critical points of two dimensional signals. After this last preprocessing step, the main identification phase starts: fiducial points are chosen among critical points using the importance criteria computed during the scale-space analysis and prior knowledge based on anthropometric constraints.

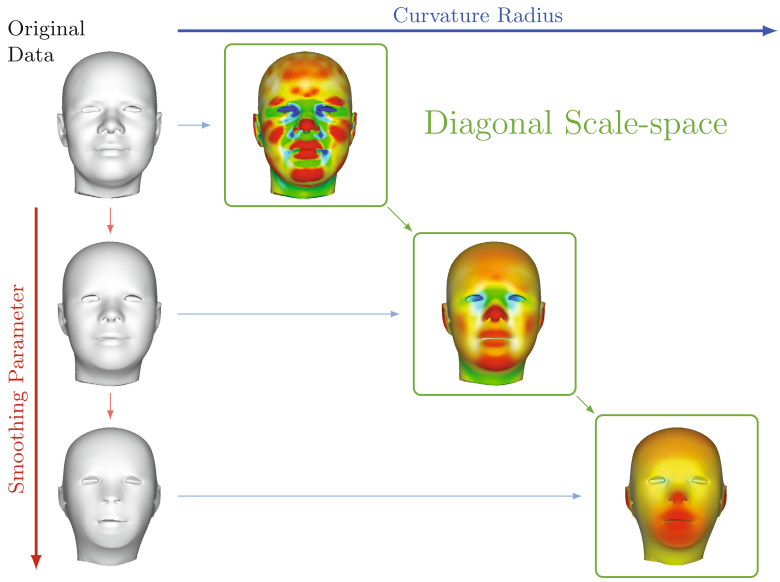


Fig. 2. The diagonal scale-space is composed by a sequence of curvature fields, obtained by computing curvature at increasing scales on increasingly smoothed surfaces.

2.1 Precomputation Phase

The main preprocessing steps consist in the computation of a scale-space which effectively encodes curvature information, and in the computation of the importance measures of critical points.

Diagonal Scale-Space. Computation of differential properties is severely affected by the presence of noise. This makes the most straightforward combination of the concepts outlined in Section 2 – a scale-space of curvature where the discrete levels were computed with radii of increasing size – a inadequate solution, because the number of critical points does not decrease fast enough as the curvature method's scale parameter increases, and, as a consequence, tracking them does not provide meaningful information. We thus propose a new type of scale-space that combines multi-scale curvature with a Gaussian scale-space, called *diagonal scale-space*.

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. 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

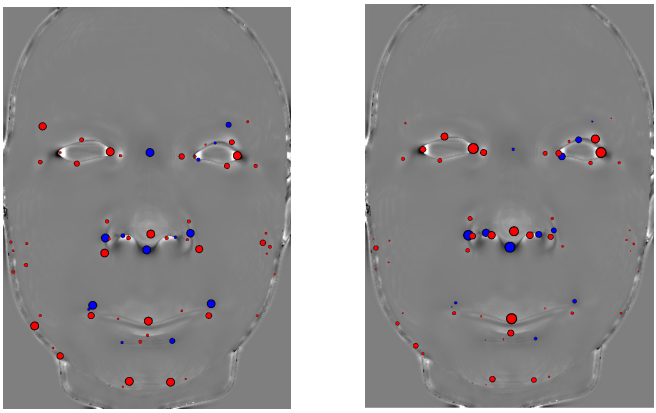


Fig. 3. Maxima (red) and minima (blue) of Gaussian curvature scaled by *life* (on the left) and *strength* (on the right).

on each level, with a correspondingly larger radius. A graphical account of this arrangement is depicted in Figure 2.

In this work, we process range images through a Gaussian smoothing filter, but the general idea of increasing two different scale parameters together, one for the amount of surface smoothing and the other for curvature computation, should be equally effective when processing full 3D data in the form of triangle meshes through a Laplacian smoothing filter. The end result is that noise is discarded in a more effective way, and the number of features decreases faster through scale. We are therefore able to achieve a meaningful tracking of the critical points of the Gaussian curvature through the scales.

Importance Measures: Life and Strength. After generation of the diagonal scale-space, we extract all the critical points in the original signal (which, in our case, is the curvature at the smallest radius computed on the original surface) and we track them through scales, using the virtually continuous scale-space method described in [19]. The output of the tracking algorithm is a data structure which encodes every critical point present in the scale-space, along with detailed information about their changes as scale grows. In particular, the data structure memorizes the moment each critical point disappears, because the feature it describes has been smoothed out and does not exist anymore. This death event marks the lifetime of a critical point in the scale-space, and we use this *life* value as our main importance measure. The life value of a critical point effectively measures the frequency of the signal that point corresponds to; critical points associated to information at higher frequencies will disappear faster than others.

Life is not the only importance measure that the proposed method employs; there is also a *strength* value that is used as a secondary criterion. Its aim is to assess the relative strength of the scalar field's maxima and minima, compared to the local trend on the surrounding surface. For each maximum, we compute the

average of the curvature field at the pixels that are below its value in a growing area around it, and return the highest difference between its value and that average; the same algorithm is applied to minima by taking into account only the surface values above the minimum. The radius of the local area is capped at a value related to the scale of its life in the scale-space. The resulting value corresponds to a sort of variable-scale Laplacian of the surface at a given point. Critical points scaled according to their life and strength values for one of the faces in our test bed are shown in Figure 3.

2.2 Identification of Fiducial Points

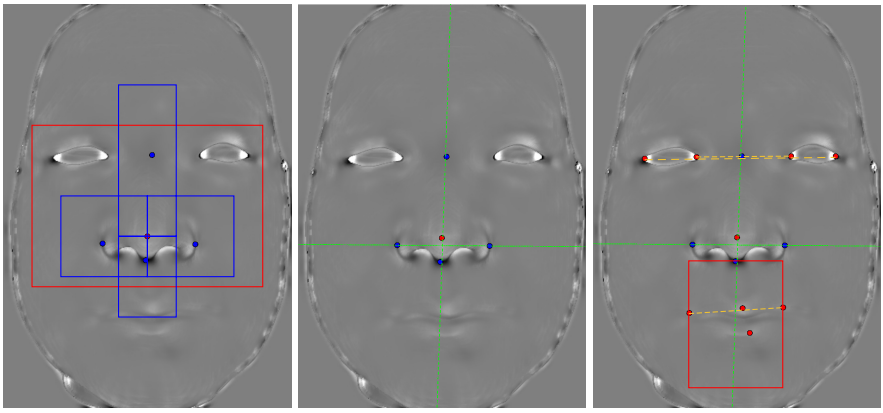
Fiducial points are selected among maxima and minima of the Gaussian curvature scalar field. Our strategy, which relies only on prior knowledge and on the *life* and *strength* measures, is based on a hierarchical search. We start by identifying the most prominent features and then seek out other features 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 (see Figure 4a):

- 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 simple: search is restricted to a wide rectangular area around the center of the range image, and the maximum of Gaussian curvature with the highest vertical value is selected.
- 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 the areas to the left and to the right of the nose tip, are considered, and the minima (one on the left and one on the right) with the highest life value in those areas are selected.
- The upper nose saddle, *m*. This is one of the most prominent saddles on a face's surface. In order to locate a rectangle located high above the nose tip is scanned, and the minimum which survives the longest is selected.
- 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 the search is narrowed down to an area located below the tip of the nose, and the point with the highest strength is selected.

Symmetry Axis and the Eyes. After points around the nose are identified, we use them to compute a vertical symmetry axis. The 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 that still need to be detected are symmetric pairs with respect to this axis. From now on, when we seek a pair of left and right points, their fitness is evaluated 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, the line connecting al_l and al_r is considered as the dividing line between the upper half and the lower half of the face. An example is shown in Figure 4b.



(a) (b) (c)

Fig. 4. (a): The first five points and the bounding boxes used to find them. (b): The horizontal line across al_l and al_r divides the face in an upper half and a lower half; the vertical line represents the symmetry axis computed on the given face. (c): Remaining points located through symmetric search, connected by a yellow dashed line, plus bounding boxes for points ls and li.

We employ the aforementioned strategy in order to find the pairs of fiducial points that characterize the eyes, as shown in the upper half of Figure 4c: 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); the two pairs that have the highest $s_a \cdot s_b$ value are selected.

The Mouth. This area contains four fiducial points (see lower half of Figure 4c): 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 two maxima of Gaussian curvature with the highest life in the area below sn delimited by ch_l and ch_r.

3 Experiments and Results

Experiments were run on Face Warehouse dataset [4], using meshes representing faces with neutral expressions, frontally projected in order to extract range images, 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 13 points shown in Figure 1.

The method's results are 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. Plots in figure 5 shows the percentage of meshes (Y axis) on which the distance is less than the given millimeters (X axis), for each fiducial point.

- Figure 5a 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. As far as we know, this work is the first to achieve 3D 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 achieves 90% at 11mm. It should be considered that this point is difficult to manually place, because the nose saddle is wide.
- Figure 5b shows results for the eyes' corners. Our method performs with good accuracy for these fiducial points. All four point already attain a detection rate above 90% within a 3mm distance.
- Figure 5c 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 ([8], [3], [2], [18] and [21]) 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 our knowledge, this work is the first one that performs 3D detection of fiducial points on the upper and lower lip, ls and li. Detection of this points achieves a 90% rate at 8mm.

Our prototype software was designed as a proof-of-concept to test the approach, by patching together previously existing packages that compute curvature and the scale-space. The resulting software is currently slow, especially in the preprocessing phase, because such packages were not optimized and also because they compute much more information than needed by our method. For a single face, building the diagonal scale-space and performing tracking on a commodity PC takes on average 50 seconds; while identifying the 13 fiducial points takes 3.75 seconds. We believe that an optimized implementation, also exploiting parallel computing, can easily achieve a speedup of two orders of magnitude.

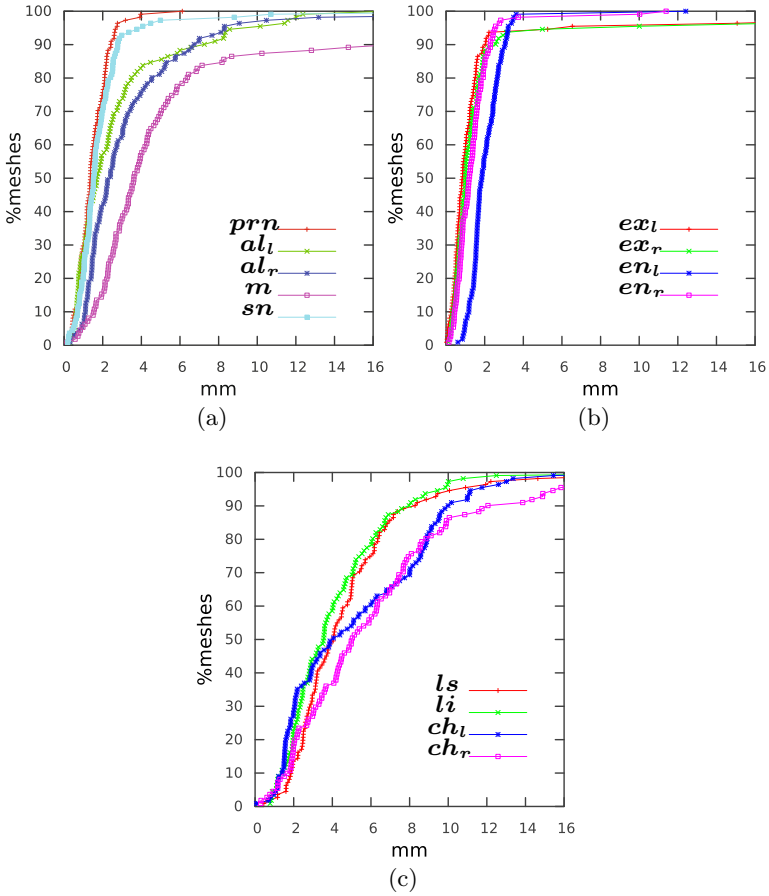


Fig. 5. Localization accuracy for fiducial points: (a) points in the nose area; (b) the corners of the eyes; (c) points around the mouth.

4 Conclusions

We presented a novel technique for extraction of fiducial points on human faces which makes use of 3D data only. Since the proposed method relies on the surface's morphological information only, no training is needed. 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 identification of three new, previously undetected, fiducial points is achieved. 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. In order to overcome the problems posed by non-frontal images, we plan to use critical points of the Gaussian curvature (a property not affected by the image's point of view) as input to an iterative refining process to detect the plane of symmetry of the face, followed by an appropriate transformation to have the mesh in the canonical view. We also plan to optimize execution times. The largest time is spent in computing curvature data; this task is suitable for parallelization and is a good candidate for GPGPU computation, because curvature on each vertex can be computed independently from other vertices. Moreover, additional research work could open the way to a curvature scale-space directly built from raw 3D data (e.g., point clouds) instead of meshes or range images, which would have even more dramatic advantages.

References

1. Berretti, S., Werghi, N., del Bimbo, A., Pala, P.: Matching 3d face scans using interest points and local histogram descriptors. Computers & Graphics **37**(5), 509–525 (2013). http://www.sciencedirect.com/science/article/pii/S0097849313000447
2. Beumier, C., Acheroy, M.: Automatic face verification from 3d and grey level clues. In: 11th Portuguese Conference on Pattern Recognition, pp. 95–101 (2000)
3. Bockeler, M., Zhou, X.: An efficient 3d facial landmark detection algorithm with haar-like features and anthropometric constraints. In: 2013 International Conference of the Biometrics Special Interest Group (BIOSIG), pp. 1–8, September 2013
4. 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)
5. 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)
6. Farkas, L., Munro, I.: Anthropometric facial proportions in medicine. Thomas (1987)
7. Friedman, J.H.: Regularized discriminant analysis. Journal of the American Statistical Association **84**(405), 165–175 (1989)
8. Gupta, S., Markey, M.K., Bovik, A.C.: Anthropometric 3d face recognition. Int. J. Comput. Vision **90**(3), 331–349 (2010)
9. Hesher, C., Srivastava, A., Erlebacher, G.: A novel technique for face recognition using range imaging. In: Proceedings of the Seventh International Symposium on Signal Processing and Its Applications, 2003, vol. 2, pp. 201–204 (2003)
10. Johnson, A., Hebert, M.: Using spin images for efficient object recognition in cluttered 3d scenes. IEEE Transactions on Pattern Analysis and Machine Intelligence **21**(5), 433–449 (1999)
11. 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, FGR 2006, pp. 585–590. IEEE, Washington, DC (2006)
12. 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 (2006)
13. Novatnack, J., Nishino, K.: Scale-dependent 3d geometric features. In: IEEE 11th International Conference on Computer Vision, ICCV 2007, pp. 1–8. IEEE (2007)

14. Novatnack, J., Nishino, K.: Scale-dependent/invariant local 3d shape descriptors for fully automatic registration of multiple sets of range images. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part III. LNCS, vol. 5304, pp. 440–453. Springer, Heidelberg (2008)
15. Novatnack, J., Nishino, K., Shokoufandeh, A.: Extracting 3d shape features in discrete scale-space. In: Third International Symposium on 3D Data Processing, Visualization, and Transmission, pp. 946–953. IEEE (2006)
16. Segundo, M.P., Silva, L., Bellon, O., Queirolo, C.: Automatic face segmentation and facial landmark detection in range images. IEEE Transactions on Systems, Man, and Cybernetics **40**, 1319–1330 (2010)
17. Panozzo, D., Puppo, E., Rocca, L.: Efficient multi-scale curvature and crease estimation. In: Proceedings Workshop on Computer Graphics, Computer Vision and Mathematics, September 2010
18. 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)
19. Rocca, L., Puppo, E.: A virtually continuous representation of the deep structure of scale-space. In: Petrosino, A. (ed.) ICIAP 2013, Part II. LNCS, vol. 8157, pp. 522–531. Springer, Heidelberg (2013)
20. Shin, H., Sohn, K.: 3d face recognition with geometrically localized surface shape indexes. In: 9th International Conference on Control, Automation, Robotics and Vision, ICARCV 2006, December 2006
21. Sukno, F.M., Waddington, J.L., Whelan, P.F.: 3d facial landmark localization using combinatorial search and shape regression. In: Fusiello, A., Murino, V., Cucchiara, R. (eds.) ECCV 2012 Ws/Demos, Part I. LNCS, vol. 7583, pp. 32–41. Springer, Heidelberg (2012)
22. Uchida, S., Sakoe, H.: A survey of elastic matching techniques for handwritten character recognition. IEICE - Trans. Inf. Syst. **E88-D**(8), August 2005