A coarse-to-fine curvature analysis-based rotation invariant 3D face landmarking

Przemyslaw Szeptycki, Mohsen Ardabilian and Liming Chen

Abstract— **Automatic 2.5D face landmarking aims at locating facial feature points on 2.5D face models, such as eye corners, nose tip, etc. and has many applications ranging from face registration to facial expression recognition. In this paper, we propose a rotation invariant 2.5D face landmarking solution based on facial curvature analysis combined with a generic 2.5D face model and make use of a coarse-to-fine strategy for more accurate facial feature points localization. Experimented on more than 1600 face models randomly selected from the FRGC dataset, our technique displays, compared to a ground truth from a manual 3D face landmarking, a 100% of good nose tip localization in 8 mm precision and 100% of good localization for the eye inner corner in 12 mm precision.**

I. Introduction

The use of 3D face models has emerged as a major face recognition solution in the last years to deal with unsolved issues, e.g. lighting and pose variations, in 2D face recognition solutions [8]. While the use of 2.5D or 3D face models instead of 2D texture face images can theoretically overcome the difficulty of lighting conditions, the head pose is still a challenge in 3D face recognition solutions [2], [6]. In such a context, 3D face landmarks, in particular nose tip and eye corners, are often used in the existing approaches for 3D face models normalization and registration. They can also be used for other applications including for instance face tracking, face expression recognition, etc.

The most of the existing works in the literature for 2.5D or 3D face landmarking are based on facial geometry analysis, making use of some a priori knowledge of face configuration. These solutions are unfortunately mostly 2.5D face rotation dependent. In this paper, we propose a rotation invariant 2.5D face landmarking solution based on facial curvature analysis combined with a generic 2.5D face model and make use of a coarse-to-fine strategy for more accurate facial feature points localization. An alternative approach also developed within our team is based on a statistical model learned from a dataset. However, this approach assumes that 3D face models are in a normalized frontal pose. In order to objectively assess the quality of an automatic feature point localization solution, we have manually landmarked feature points, including nose tip, inner eye corners, extern ones, etc., on the whole FRGC 1.0 and FRGC 2.0 datasets. Experiments on a significant subset of the FRGC dataset and compared with the manually labeled landmarks, our automatic face landmarking solution displays a 100% of good localization for the nose tip in 8 mm precision and 100% of good localization for the eye inner corner in 12 mm precision.

The rest of this paper is organized as follows. Section II overviews the related work. Section III describes our curvature analysis-based algorithm. Section IV presents the ground truth of manually labeled feature points on the FRGC datasets [14] and experimental results. Section V concludes the paper and gives some hints on future work.

Przemyslaw Szeptycki is with MI Department, LIRIS Laboratory, Ecole Centrale de Lyon, 69134 Lyon, France przemyslaw.szeptycki@ec-lyon.fr

Mohsen Ardabilian is with MI Department, LIRIS Laboratory, Ecole Centrale de Lyon, 69134 Lyon, France mohsen.ardabilian@ec-lyon.fr

Liming Chen is with MI Department, LIRIS Laboratory, Ecole Centrale de Lyon, 69134 Lyon, France liming.chen@ec-lyon.fr

II. Related work

2.5D face landmarking has gained increasing interest in the last years thanks to its diverse applications in face normalization, face registration, etc. Most of the existing works embed some a priori knowledge on face configuration into the algorithm and make use of facial geometry-based analysis to localize geometrically salient feature points such as nose tip, eye corners, etc. All these works can be categorized by their way of using curvature analysis, resulting in a rotation sensitive or not solutions.

Many works make use of HK classification. Chang et al. in [3] used method based on Mean and Gaussian curvatures. While HK-Classification is rotation invariant, the way they analyze it with assumption that the face model is in a frontal position to localize the nose tip and eye inner corners, makes their technique sensitive to roll and yaw rotation. HK-Classification was also used by Colombo et al. in [5] for locating the same three main face points. However, their validation step is based on correlation with a template of a cropped face around the three main points, makes their solution very sensitive to yaw and pitch rotations. Moreover, a depth map used as template is very sensitive to spikes which may significantly change depth map appearance. Another work proposed by Sun et al. in [15] also relies on HK classification for automatic facial pose estimation. Pose is estimated by locating the three main points as well. They proposed to localize correct eye regions using statistical distances between the eyes and to remove clusters which have less number of points. Removing concave regions in such a way may simply lead to the removal of the eye regions.

Other works are based on the shape index, for example: Colbry et al. [4] and Lu et al. [10]. They analyzed the shape index values to find feature points of the face. Nevertheless, they assume that the eyes are above the nose in the face scan and therefore the major axis of the scanned image must be vertical.

There also exist many other works making use of particular curvature signature to embed into the algorithm some a priori knowledge on facial geometric configuration. Faltemier et al. proposed in [8] ”Rotated Profile Signatures” for nose tip detection. In their approach, the face model is rotated around the y axis, the intersection of the 3D model and (x,y) plane in the right gives them a profile signature. This profile is then compared with two known templates to localize the nose tip. Despite of reported results (100% nose tip localization in 10mm precision for frontal faces), their algorithm is not invariant to the roll rotation (along z-axis). Also, the use of skin color detection to remove non skin color vertexes could also be source of errors due to the illumination variability. A very similar work was proposed by Lu et al. [11]. Their approach is based on the simple assumption that the nose tip should have the highest z value in a frontal position. The face model is then rotated around y axis at a fix step and the point with the largest value in the z-axis in each rotation is considered as the nose tip candidate. Verification stage based on vertical profile in the nose tip point gives them correct result, nevertheless verification based on the vertical line is not invariant to the roll rotation.

Alternative method for the nose tip localization was proposed by Mian et al. [12]. They slice horizontally the face and consider a moving circle centered on each slice point. The maximum of the triangle altitude formed by the center of the circle and its intersections with the slice indicate the nose tip candidates. The point having maximum triangle altitude is considered as the nose tip. Horizontal slicing used in this approach makes it sensitive to roll rotation.

Jahanbin et al. [9] proposed to locate facial feature points using ”gabor jets”. The general appearance of each fiducial point is modeled by ”gabor jets” extracted from several manually marked examples. At localization stage which is based on ”bunch graph matching”, the search area of each feature point is constrained by penalizing the deformation of a graph connecting the fiducials through an optimization algorithm. However, they imposed the constraints that the nose is expected at the center, inner corner of left eye located above and to the left of the nose tip, thus discarding this method from rotation invariant. A significant work was proposed by D'Hose et al. [6]. 3D face landmarks in their work are located by using the Gabor decomposition to filter and amplify curvature information. They perform Gabor filtering in vertical and horizontal directions and localize the nose tip candidates based on these two Gabor filters' responses. Finally, an ICP matching is performed between the template of a nose region and nose tip candidate and the best matching delivers nose tip location on the face. They report a 99.89% and 99.37% of correct nose tip localization in respectively 20 mm and 10 mm precision.

As we can see from this overview, most of these works are face rotation sensitive, thus considerably limiting their use in real-life applications which generally require the least possible constraint. In our work, we make use of facial curvature analysis and embed some a priori knowledge on face configuration into the algorithm to localize facial feature points based on their salient geometrical properties. Meanwhile, in order to have a rotation invariant solution, we further make use of a generic face model generated from a set of 2.5D face models in IV2 dataset. Moreover, a coarse-to-fine strategy is also applied for more accurate feature point localization.

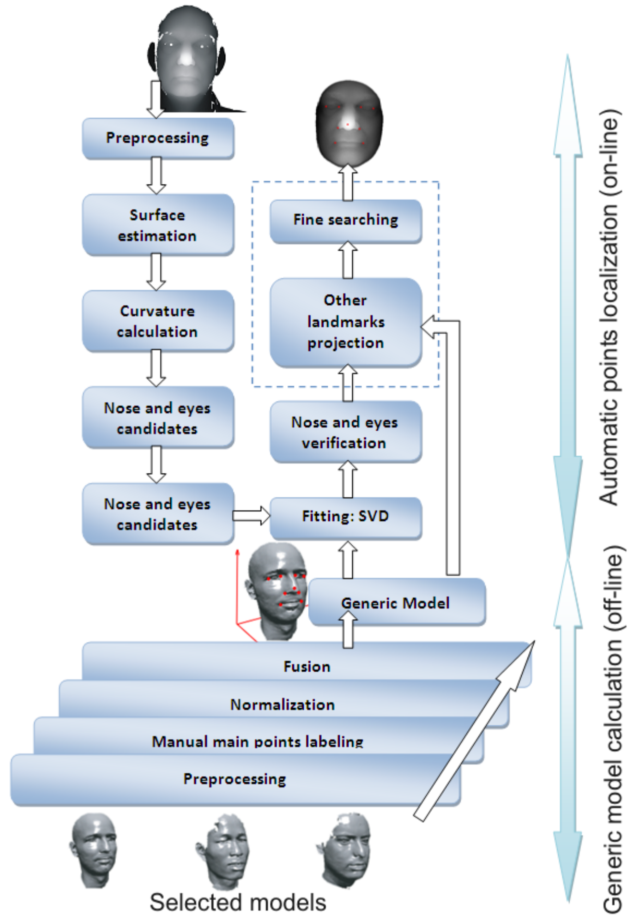

Fig. 1. Schema of our algorithm for automatic main points localization.

III. Face main points localization algorithm

Our algorithm for automatic feature point localization is based on facial curvature analysis and makes use of a coarse-to-fine search strategy and consists of two main steps. At a coarse search step, candidate points for the three most salient facial feature points (nose tip and the two inner eye corners) are first identified based on curvature analysis and generic model fitting. At a fine search step, a generic face model is used to locate other feature points within the neighborhood of the projection of the ones from the generic face model. But first of all, 2.5 face scans are usually noisy and thus need to be cleaned up in a preprocessing step. Figure 1 sketches these steps which are detailed in the subsequent subsections.

A. 2.5D models pre-processing

2.5D face models delivered so far by various scanners usually are corrupted by impulse noise and holes. To reduce influence of the underlying face model quality on face landmarking algorithm, holes and spikes need first to be removed.

The most popular technique to remove impulse noise (spikes) is the median filter. This method removes noise but also tends to affect fine details in many cases. To avoid changes in correct vertexes, we have used in our work a decision-based median filtering technique which applies the median filter only on vertexes classified as potential spikes after a thresholding operation. This method can efficiently remove all the spikes without touching properly scanned points.

Once all the spikes are removed, we also need to fill holes which often occur in 3D face scans. Our method for removing this kind of discontinuity consists of fitting a mean square surface to the border of the hole, the hole border being located by searching vertexes having less than 8 neighbors. In comparison with linear or cubic interpolation, our method takes into account all the directions of the surface changes and it is more convenient for 3D models.

B. Curvature analysis-based coarse search

The aim of this coarse search step is to localize feature point candidates on a 2.5D face model. In order to achieve rotation invariant feature point localization, we make use of Mean and Gaussian curvature analysis and classification [1] and we are targeting the three most salient feature points from the geometric perspective, namely the nose tip and the two inner eye corners. Indeed, the nose tip appears as a convex point on a facial surface while the two inner eye corners as concave points.

The range data of a 2.5D face model is thus first segmented into regions of homogeneous shapes according to HK classification (tab. I). The HK classification labels each vertex into basic geometric shape class (tab. I), using the sign of the mean (eq. 1) and the Gaussian (eq. 2) curvature [16] which can be respectively computed by the following equations on a 2.5D surface:

$$H(x,y) = \frac{(1+f_y^2)f_{xx} - 2f_xf_yf_{xy} + (1+f_x^2)f_{yy}}{2(1+f_x^2+f_y^2)^{\frac{2}{3}}}, \quad (1)$$

$$K(x,y) = \frac{f_{xx}f_{yy} - f_{xy}^2}{(1+f_x^2+f_y^2)^2}, \quad (2)$$

where $f_x, f_y, f_{xx}, f_{yy}, f_{xy}$ are the first and second derivatives of a surface f in (x,y) [16], [5].

Since our surface representation is discrete, the partial derivatives need to be estimated based on a surface equation derived from the neighborhood of a point. The problem of finding the surface equation can be represented as:

$$AW = Y, \quad (3)$$

	K < 0	K = 0	K > 0
H < 0	Hyperbolic	Cyl. convex	Ellip. convex
H = 0	Hyperbolic	Planar	Impossible
H > 0	Hyperbolic	Cyl. concave	Ellip. concave

TABLE I

HK-CLASSIFICATION [16].

where matrix A contains x and y values, W is a matrix of the coefficients from the estimated function and matrix Y contains function results.

However, this curvatures approximation is very sensitive to the noise because of the second derivatives. Such noise sensitiveness can be reduced by varying the extent of the neighborhood size. Figure 3 shows the HK-Classifications achieved in varying the neighborhood size. As we can see, large neighborhood used in surface equation estimation hides noise and also helps to find most marked out points.

In our experiments we fix the neighborhood size to 25 mm in geodesic distance (based on observations of the curvatures decomposition) which is more appropriate in such computations because it takes into account the facial surface shape. Recall that geodesic distance between point A and B is defined as sum of Euclidean distances between points in the shortest path between points A and B on the underlying surface.

To localize the three most salient facial feature points, namely the nose tip and the two eye inner corners, the concave and the convex regions of the nose and the eyes respectively are first searched. Each region can be localized according to HK-Classification. To reduce number of regions to be considered, the Gaussian curvature is also thresholded with $K > 0.001$ for the nose and $K > 0.00005$ for the eyes regions (figure 2b). The thresholding process and curvatures calculation is derived from articles [5], [3], [13] with small changes in thresholds values. The main difference between our method and other methods is in main points extraction from convex and concave regions and face validation based on geometrical generic face model.

In each located region (figure 2b), the most representative point in term of concavity or convexity is then identified. As can be seen in table I, changes in Gaussian curvature result in shape changes and maximum Gaussian curvature in each region gives maximum convex or concave point. Such point will be labeled as landmark candidate in the convex regions for the nose tip and in the concave regions for the eye inner corners.

C. Generic face model-based fine search step

Once the landmark candidates were generated for the three most salient feature points (nose tip and the two eye inner corners), we proceed to a fine search step in order to sort out the true three main feature points and to localize other feature points by making use of a generic face model. In the following, we describe our generic face model and its use for locating first the true three main points and the other ones from a model projection.

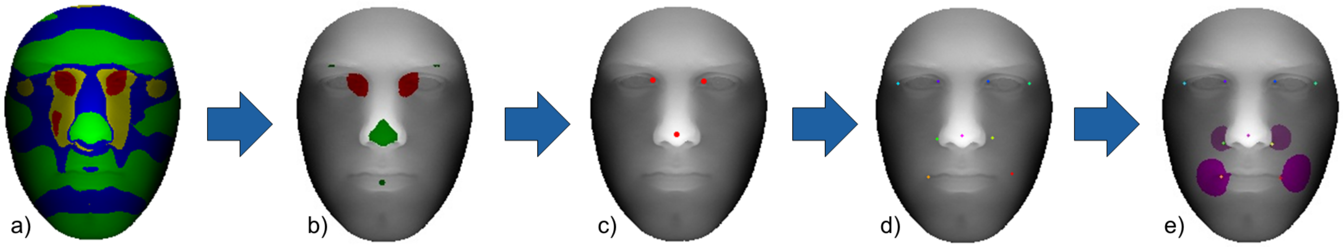

Fig. 2. Main points localization algorithm: a) HK-Classification, b) nose and eyes regions, c) (coarse localization) the nose tip point and the inner corners of eyes points, d) generic model aligment, e) fine adjusting of points

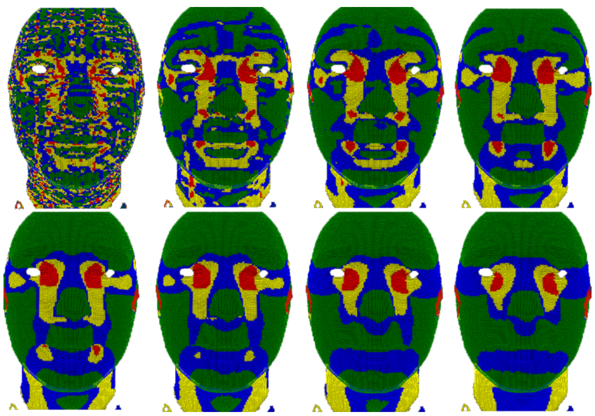

Fig. 3. The HK-Classification with different neighborhood in the curvature calculation between 5 mm(left top) and 40 mm(right bottom) (elliptical concave: red, elliptical convex: green, hyperbolic concave: yellow, hyperbolic convex: blue).

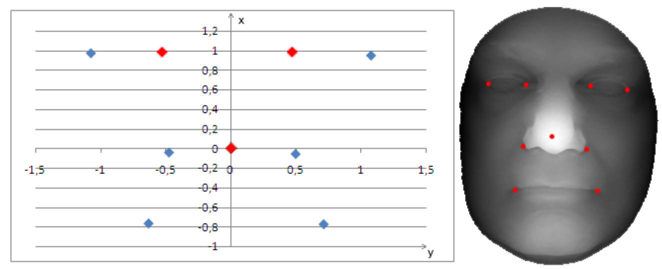

Fig. 4. Generic model made based on 40 models from IV2 data set (x,y projection, red points - main three points - inner corners of the eyes and the nose tip).

1) Generic face model building: Our generic face model (figure 4) is built on 40 models randomly selected from the IV2[1] data set. The generic model is composed from 9 main face points which positions have been calculated based on selected 2.5D facial models. These models were first manually landmarked for 9 feature points. Next, all models were translated and rotated to a frontal position having the nose tip as the origin. Fusion of all models relay on mean main point position calculation in 3D space. The generic model is further normalized so that the distance between the two eye inner corners is 1 mm. Model was made from 9 points based on our observations of curvatures decomposition on a face. We choose those points which can be described in curvature space (all of them belongs to convex or concave regions).

2) The three main feature points identification: The calculated generic face model is now used to sort out the true three main feature points (the nose tip and the two inner eye corners) from the set of candidate landmarks resulted from the curvature analysis (section III-B). As our 2.5D face model can be in arbitrary position and we do not have any priory information about it, the basic idea is to consider all combinations of any three landmark candidates (the nose tip candidate and the inner corners of the eyes candidates).

To select true main three points from points, candidates selection step needs to be performed. Selection step is based on error calculation between generic model in certain position and face surface. To calculate error between generic model and facial surface, for each combination of landmarks candidates (we are taking under consideration, with one assumption that one of eyes have to be above the nose, always two points from the eyes candidates and one from the nose candidates) we are moving whole generic model above the face surface and calculating sum of distances for all generic model points to the closest points on the face surface. The movement is based on the rotation and the translation founded by SVD algorithm based on the three main points from the generic model (red points on fig. 4) and considered landmarks candidates. Singular Value Decomposition algorithm [17], [7], [18] is a matrix decomposition algorithm which has been used iteratively in the ICP (Iterative Closest Point) algorithm. Algorithm let us to find fine translation and rotation between objects in correspondence based on their covariance matrix.

To be invariant to the scale, the generic model has been scaled based on the distance between concave points (the eyes candidates) under the consideration.

The smallest error between generic model and face surface under specific position identifies the true main feature points thanks to the associated manually labeled landmarks on the generic model (fig. 2c).

3) Other feature points localization: The aim of our work is to locate nine feature points on a 2.5D face model (fig. 4). Once the three main feature points have been located, we

[1]IV2 - French biometric data base created in cooperation of few laboratories.

proceed to localize the other feature points. For this purpose, we project these manually labeled feature points from the generic face model onto the 2.5D face model, using the same rotation and translation computed previously by SVD. The closest points on the 2.5D face model to the generic model point will became the landmark candidates and succeed their labels (figure 2d).

A coarse-to-fine search strategy is again applied here by a local search to deliver the better accurate location of the projected feature points. Indeed, the two corners of the lips, the two outer nose corners and the inner eye corners can be characterized as concave points within a certain neighborhood resolution. As we can see on figure 3, that smaller neighborhood uncovers details on the surface like the lips corners or the nose corners. To localize them precisely, we calculate curvatures using a smaller neighborhood around these points in the surface approximation. In our work, 15 mm neighborhood size is chosen for the lips while 10 mm neighborhood size is chosen for the nose, figure 2e based on observations of HK-Classification (figure 3) decomposition in different neighborhood size of investigated point. The vertex having its maximum Gaussian curvature gives us the most concave point in the concave region and is labeled as final anchor point.

IV. Experiments and discussion

Our 3D face landmarking solution was benchmarked on a significant subset from FRGC datasets 1.0 and 2.0 [14]. To set out the ground truth, the whole FRGC datasets were manually marked out by our team. As we can see in figure 5, these manually labeled anchor points include the eye and lips corners, the nose corners and its tip, and also upper and lower points at the eyelid and lips middle for future investigation. **These manually labeled landmarks are available to the public for research purpose**. The quality of these manual landmarks was also assessed using randomly selected 3D models on which 10 people were asked to manually label the previously defined anchor points. We then computed mean error and standard deviation for each landmark which is summarized in Table II. As we can see from the table, the biggest manual errors as expected were made on landmarks not precisely defined such as right and left corners of nose while nose tip was among the anchor points labeled with the least errors. This experiment shows that each feature point for different person does not locate accurately at the same place, therefore anchor point on a 3D face model should be considered more as a region than an exact point.

In this work we have chosen 9 prominent feature points from the curvature viewpoint to assess our automatic 3D face landmarking solution. For this purpose, more than 1600 of models from FRGC datasets were randomly selected. While many of these face models have facial expressions (FRGCv2), all of them are in roughly frontal position. In order to test robustness of our solution as compared to rotation, each selected 3D face model was rotated randomly in yaw (from -90 to 90 degrees), pitch (from -45 to 45

Anchor Point	Mean Error	Standard Deviation
Left Eye Left Corner	2.9531	1.4849
Left Eye Right Corner	2.4194	1.0597
Left Eye Upper Eyelid	2.0387	1.3755
Left Eye Bottom Eyelid	1.9424	0.8507
Right Eye Left Corner	2.0473	1.0770
Right Eye Right Corner	2.7559	1.5802
Right Eye Upper Eyelid	2.1080	1.6449
Right Eye Bottom Eyelid	1.8362	0.8105
Left Corner Of Nose	3.8023	1.9839
Nose Tip	1.9014	1.0474
Right Corner Of Nose	4.4974	2.1489
Left Corner Of Lips	1.9804	1.1045
Right Corner Of Lips	1.9891	1.1905
Upper Lip	3.0414	1.5292
Bottom Lip	2.0628	1.3052

TABLE II

Mean error and standard deviation of manual anchor points based on 10 models and 10 samples per model.

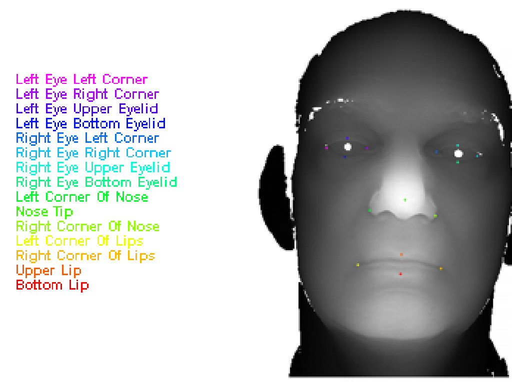

Fig. 5. Manual main points in FRGC dataset.

degrees) and roll (from -30 to 30 degrees) before applying our automatic 3D face landmarking.

Figure 6 shows the localization results by our algorithm. Accumulative precision is displayed together with localization accuracy rate given in mm precision.

As we can see, the best result was achieved for the nose tip localization with a 100% accuracy in 8 mm precision while the eye inner corners were localized with a 100% accuracy for left inner eye corner in 12 mm precision and 13 mm precision for the right inner eye corner. Therefore, the inner eye corners were located with more than 99% accuracy in a 10 mm precision.

With an accuracy of 88.75% for left lips corner and 87.45% for right lips corner in 20m precision, our algorithm displays the worst localization results which are mainly caused by facial expressions, mouth movement, etc. Fair localization is achieved for other feature points with respectively 99.62% and 99.87% accuracies for the left and right eye outer corners in 20 mm precision, and 98.2% and 99.0% accuracies for the left and right nose corners. The whole results curves can be seen on figure 6.

A major application of 3D face landmarking is 3D face registration and normalization. Thus rotation robustness of

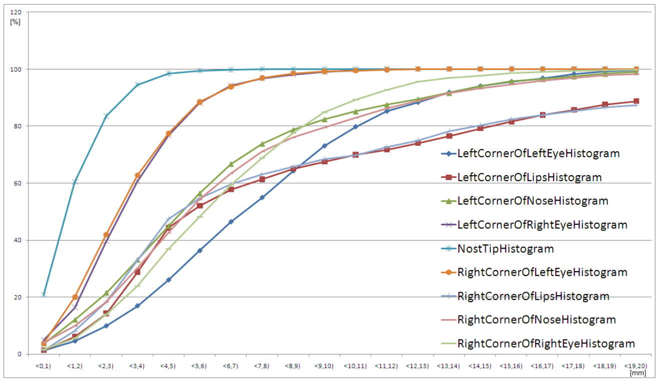

Fig. 6. Precision curves for all points (precision in mm).

a 3D face landmarking solution is important as it relaxes constraints on the input 2.5D or 3D face model, making 3D-based face processing techniques closer to realistic application requirements. In our work, such a rotation invariance was made possible thanks to curvature-based analysis and the use of a generic 3D face model. As compared to [8], [6], our approach achieves higher precision in nose tip and eye inner corners localization while automatically providing up to nine 3D face landmarks.

V. Conclusion and future work

In this paper we presented a rotation invariant 2.5D face landmarking algorithm which makes use of curvature analysis combined with a generic face model and applies a coarse-to-fine strategy for accurate localization of nine feature points, including the nose tip and eye inner and outer corners. The experiments carried out on a randomly selected models from FRGC dataset showed the effectiveness of our technique assessed in terms of locating precision in mm as compared to manually labeled landmarks.

As many other curvature analysis-based techniques in the literature, our algorithm embeds some a priori knowledge on 2.5D face geometry, especially the most salient feature points such as the nose tip and the inner eye corners. While our approach is made rotation invariant thanks to curvature based analysis and the use of a generic model, accurate localization of a landmark is sensitive to its curvature saliency, making difficult precise locating of less salient face landmarks. Alternatively, a learning-based approach based on a statistical model, also developed within our team, can be used. However, such an approach assumes that the face models for learning and testing are in a frontal position. These two approaches, curvature analysis-based and learning based, are thus complementary.

In our future work, we are moving toward 2.5D face pose estimation and face landmarking on partial 2.5D face models.

VI. Acknowledgment

This work has been partially supported by the ANR under the FAR 3D project with the reference ANR-07-SESU-004.

References

[1] P. Besl and R. Jain. Invariant surface characteristics for 3-d object recognition in range images. *Comput. Vision, Graphics Image Proc.*, 33:33–80, 1986.

[2] K. W. Bowyer, K. Chang, and P. Flynn. A survey of approaches and challenges in 3d and multi-modal 3d + 2d face recognition. *Computer Vision and Image Understanding*, 101(1):1–15, 2006.

[3] K. I. Chang, K. W. Bowyer, and P. J. Flynn. Multiple nose region matching for 3d face recognition under varying facial expression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(10):1695–1700, 2006.

[4] D. Colbry, G. Stockman, and J. Anil. Detection of anchor points for 3d face verification. *Computer Vision and Pattern Recognition*, pages 118–118, 2005.

[5] A. Colombo, C. Cusano, and R. Schettini. 3d face detection using curvature analysis. *Pattern Recognition*, 39(3):444–455, 2006.

[6] J. D'Hose, J. Colineau, C. Bichon, and B. Dorizzi. Precise localization of landmarks on 3d faces using gabor wavelets. *Biometrics: Theory, Applications, and Systems*, pages 1–6, 2007.

[7] D. Eggert, A. Lorusso, and R. Fisher. Estimating 3-d rigid body transformations: a comparison of four major algorithms. *Machine Vision and Applications*, 9:272–290, 1997.

[8] T. C. Faltemier, K. W. Bowyer, and P. J. Flynn. Rotated profile signatures for robust 3d feature detection. *IEEE Int. Conf. on Automatic Face and Gesture Recognition*, 2008.

[9] S. Jahanbin, A. C. Bovik, and H. Choi. Automated facial feature detection from portrait and range images. *IEEE Southwest Symposium on Image Analysis and Interpretation*, pages 25–28, 2008.

[10] X. Lu, D. Colbry, and A. K. Jain. Three-dimensional model based face recognition. *Int. Conf. on Pattern Recognition*, 1:362–366, 2004.

[11] X. Lu and A. K. Jain. Automatic feature extraction for multiview 3d face recognition. *Automatic Face and Gesture Recognition*, pages 585–590, 2006.

[12] A. Mian, M. Bennamoun, and R. Owens. An efficient multimodal 2d-3d hybrid approach to automatic face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(11):1927–1943, 2007.

[13] A. Moreno, A. Sanchez, J. Velez, and F. Diaz. Face recognition using 3d surface-extracted descriptors. *Irish Machine Vision and Image*, 2003.

[14] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. *Computer Vision and Pattern Recognition*, 1:947–954, 2005.

[15] Y. Sun and L. Yin. Automatic pose estimation of 3d facial models. *Int. Conf. Pattern Recognition*, 2008.

[16] E. Trucco and A. Verri. *Introductory Techniques for 3-D Computer Vision*. Prentice Hall PTR, Upper Saddle River, NJ, USA, March 1998.

[17] S. Umeyama. Least-squares estimation of transformation parameters between two point patterns. *Transactions on Pattern Analysis and Machine Intelligence*, 13:376–380, 1991.

[18] G. Wen, Z. Wang, S. Xia, and D. Zhu. Least-squares fitting of multiple m-dimensional point sets. *The Visual Computer*, 22(6):387–398, 2006.