THE INTERPLAY BETWEEN RECONSTRUCTION AND RECOGNITION

Shireen Y. Elhabian and Aly A. Farag

Computer Vision and Image Processing Laboratory
Department of Electrical and Computer Engineering
University of Louisville
Louisville, Kentucky

August 2009

ABSTRACT

This report is concerned with understanding the interplay between 3D face reconstruction and recognition. We provide analysis for the effect of adaptive sparse 3D reconstructions on recognition rate. This work illustrates that, for 3D face recognition, a proper mesh containing nearly 2% of the facial information and about 5 times the initial/root mesh provides similar recognition rate as one with full 3D information. The report describes a systematic approach to generate adaptive meshing staring from a root mesh of basic facial landmarks. This work is a crucial first step towards a successful incorporation of 3D face recognition approaches in near real time scenarios under effects of pose and illumination.

1. INTRODUCTION

3D reconstruction encapsulates the information content of an object, thus permitting recognition under the constraints of the reconstruction environment. In the 3D data domain, effects like pose, illumination, and occlusions can be quantified, and recognition is much easier and more robust than approaches based on 2D projections of the 3D objects. However, it is often the case that the prevailing features which determine the accuracy of the recognition are only a small subset of the 3D information in the object. In real world applications (e.g., video-based human identification), various regions in the object (e.g., the cheeks in the human face) are nearly featureless and do not significantly affect the recognition rate. Hence we do not have to reconstruct all 3D information for the object at hand; this motivates the subject of this report: how sparse a reconstruction can be and still provide recognition performance comparable to that based on the whole 3D information.

The major contribution of this study is developing an approach for feature sampling for 3D face recognition. The methodology is generic and would be applicable for various applications in 3D active computer vision. This work addresses the issue of reconstruction resolution, i.e. Do we have to reconstruct all facial details with high resolution in order to maintain high recognition rate? We believe that this question is worth answering especially with the absence of a perfect 3D facial reconstruction. Our course of action is divided into two stages, first start with a perfect reconstruction (3D facial scans) then move to the real world where there is no perfection, the first stage is introduced in this report, in the meantime we are working on the second one. We have used a benchmark 3D data (FRGC) in order to compare our work with the literature, we have used a standard recognition method (PCA-based) and standard mesh

subdivision algorithm (Loop's mesh subdivision) to isolate biasness of our results due to any new proposed approach, i.e. we are using existing techniques to defend our hypothesis. For recognition purposes, we use a minimum distance classifier in the subspace of facial scans produced by PCA.

The data of a 3D shape can be represented at different levels of abstraction. The first level is a set of points in 3D space. The second level is the boundary of the shape. Finally, the third level is the volume that the shape occupies. We provide a brief description of these representations [e.g., 9]:

Point-Based Representation: The boundary of the object is described by either a cloud of points or range images. Range images are similar to intensity images in the sense that they capture the shape from one point of view except that the color information of the pixel carries the depth information of the surface point from the camera. This representation is generally used in 3D model reconstruction from multiple range images of an object that is acquired using a laser scanner. In general, point-based representation lacks the structure information of a shape but it is enough for visualization purposes.

Boundary-Based Representation: A 3D object can be represented in terms of its boundary or surface, which is the common representation in a computer aided design (CAD) and computer graphics industries. The boundary data can be described by polygonal meshes, parametric forms, or implicit surfaces. Although local boundary features can be computed efficiently, global shape features such as moments are very difficult to compute. In addition, such methods do not capture the hierarchical structure of the shape, which is an essential requirement for several applications such as object recognition and classification.

Volume-Based Representation: A 3D object is represented by the volume it occupies. The volumetric data is described by voxels. This representation is commonly used in the field of computer vision and medical imaging due to the nature of the acquisition process. In Constructive Solid Geometry (CSG), a complex surface or object is composed of simple solid objects, which are glued to each other through Boolean operators on sets such as union, intersection, and difference. Often CSG presents a model or surface that appears visually complex, but is actually little more than cleverly combined or de-combined objects. The simplest solid objects used for the representation are called primitives. Typically they are the objects of simple shape: cuboids, cylinders, prisms, pyramids, spheres, and cones.

A compact object representation is essential to understand the effects of sparsity. In this study we do not consider the set of shapes that can be represented by closed surfaces or contours; object in this set may have a robust representation in terms of curve skeletons which produce enormous data reduction and can be used for accurate recognition [9]. The topologies considered in this study are more general, and a triangular mesh representation is adequate. This class may result from various reconstruction methods or sensors. The human face is a member of this class. The problem under consideration can be stated as follows: starting from basic landmarks in the object, construct a mesh that can be adaptively densified through a well-structured geometric approach. On each level of meshing, we compute a recognition measure; the goal is to understand which meshing level provides adequate recognition accuracy. From a practical standpoint, the results of this work will make the reconstruction algorithms much simpler and faster at almost no cost for the accuracy of recognition.

Starting from perfect 3D human face reconstructions, in particular facial laser scans, the prevailing features are extracted using psychological and geometric measures devised in the past

decade. A geometric surface mesh is generated using the statistics of these features over a very large sample of standard 3D facial scans. This mesh is used to adaptively evaluate the effect of reconstructions on recognition, authenticating a number of studies reported in the most recent literature. This mesh will constitute the feature mask (or automatic feature sampler) for 3D recognition. The mesh will lend itself to various related research work under active appearance modeling (AAM), statistical face modeling and recognition.

From a practical standpoint, perfect 3D information is a luxury that does not exist; 3D objects (e.g., 3D reconstruction of the human face) are generated by 2D sensors or data sources (e.g., 2D images) through shape-from-X algorithms. Stereo-based approaches reconstruct metric 3D information of an object using a stereo-pair of images. However, such approaches suffer from finding correspondences between the given pair of images. At the end of the report we discuss the application of the aforementioned mesh in automatic stereo-based reconstructions for recognition.

Interest in 3D face recognition has grown in the computer vision community in recent years. An informative survey regarding three-dimensional face recognition is available in [1]. There are three common approaches to extracting 3D information from faces, namely: (a) passive stereo, (b) structured light, and (c) hybrid combination of passive stereo and structured light [2]. From the perspective of passive stereo, generation of dense high-resolution 3D reconstructed scans involves high computational effort due to correspondence computations [3]. Instead of performing dense operations, Park and Jain [4] performed stereo reconstruction on face feature points related to AAM/ASM vertices [5]. Wright, et al. [6] examined the effects of occlusion and disguise on recognition via sparse signal representation from video. A similar study on sparse face recognition uses various descriptors (e.g., BHG, SIFT, and LBP descriptors)

to determine the face feature points for recognition [7]. Finally, in a study not related to face recognition, Ramnath [8] investigated the fitting performance of AAM. The study showed that denser AAM's can be fit more robustly than the sparser counterparts.

The rest of this report is organized as follows: Section 2 describes adaptive facial meshing used in this work; Section 3 details the basic relationship between 3D face reconstruction and recognition; Section 4 briefly outlines a stereo-based 3D face recognition approach; and Section 5 provides conclusions and ongoing related work.

2. ADAPTIVE FACIAL MESHING

In order to quantify the relationship between the sparsity of the face model and the recognition accuracy, we need to find a systematic way to down-sample the 3D object representation (i.e., generate less number of vertices/triangles). However, simple uniform down-sampling cannot serve our purpose, we need to have an adaptive/non-uniform down-sampling which enable us to have more samples in featured regions such (e.g., the eyes in human face) and less samples in smooth regions (e.g., the cheeks). Such metric will emphasize the most prevailing features for recognition and deemphasize the least important ones. This process may be accomplished by the so-called mesh simplification algorithms [e.g. 10]. However, moving in this way will not guarantee maintaining the salient features of a facial surface such as nose tip, eye centers … etc. The other way round is to start with an initial root mesh which contains the salient facial features, then utilize mesh subdivision approach to densify such mesh and hence we obtain more and more details through mesh upsampling while maintaining such prevalent features.

Definition 1: *A* **root mesh** *is the set of triangles describing a connected surface linking the predominant facial landmarks.*

In terms of recognition, the basic landmarks from psychological studies include: the eyes, nose, and lips. Various approaches can be used to reconstruct the root mesh. The number of vertices/triangles in the root mesh is arbitrary; however, it should contain one or more points in four main facial regions (left and right eyes, nose and mouth). Therefore, there is a large degree of flexibility in choosing the root mesh. The most important characteristic is that subsequent additional meshes include the root mesh.

Definition 2: *An* **adaptive mesh** *is the set of triangles describing a connected surface such that the root mesh can always be recovered by an iterative down sampling process.*

Fig. 1 illustrates a root mesh layout on the average face generated from the University of South Florida (USF) database [11].

Various root meshes could be generated with landmarks on the nose, eyes, lips and tip of the chin. Likewise, a mesh linking the basic morphometric measures may be used. In our subsequent analysis, we study the performance of the recognition rate starting from the root mesh and the subsequent densified meshes.

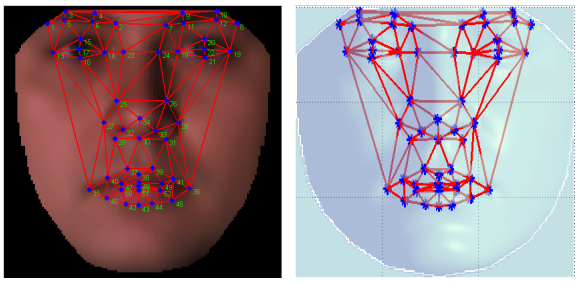


Figure 1 – A sparse facial mesh defined on the USF mean face

The main objective of a mesh subdivision algorithm is to produce a smooth and visually pleasing surface whose shape is controlled by the initial mesh [12, 13]. Subdivision produces a sequence of mesh refinements where new vertices are defined as local affine combinations of nearby vertices. Over the last decade, interpolating subdivision has been used for multi-resolution analysis of complex surface geometries. Mesh subdivision is a two phase process. The refinement phase creates new vertices and new triangles from the root mesh, and the smoothing phase computes new positions for some or all of the vertices in the new mesh. We exploit a loop subdivision scheme [10] where mesh refinement is achieved by refining each triangle into four triangles by splitting each edge and connecting new vertices, locations for new vertices are weighted average of original vertices in local neighborhood, see Fig. 2. (for details refer to Appendix A)

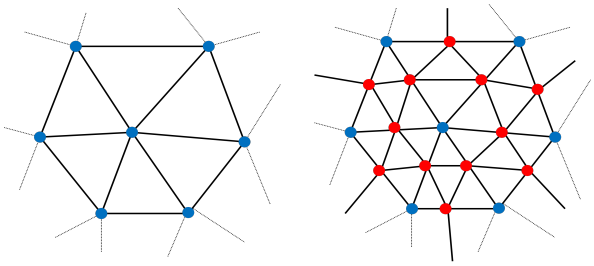


Figure 2 – Mesh refinement is accomplished by dividing each triangle into four triangles by splitting each edge and connecting new vertices (these points - appear in red in a color print - form an inner most triangle).

Fig. 3 shows different sampling meshes defined on the mean shape surface starting from our root mesh (down-sampling factor is 99.65%) till the full resolution facial surface (0% down-sampling).

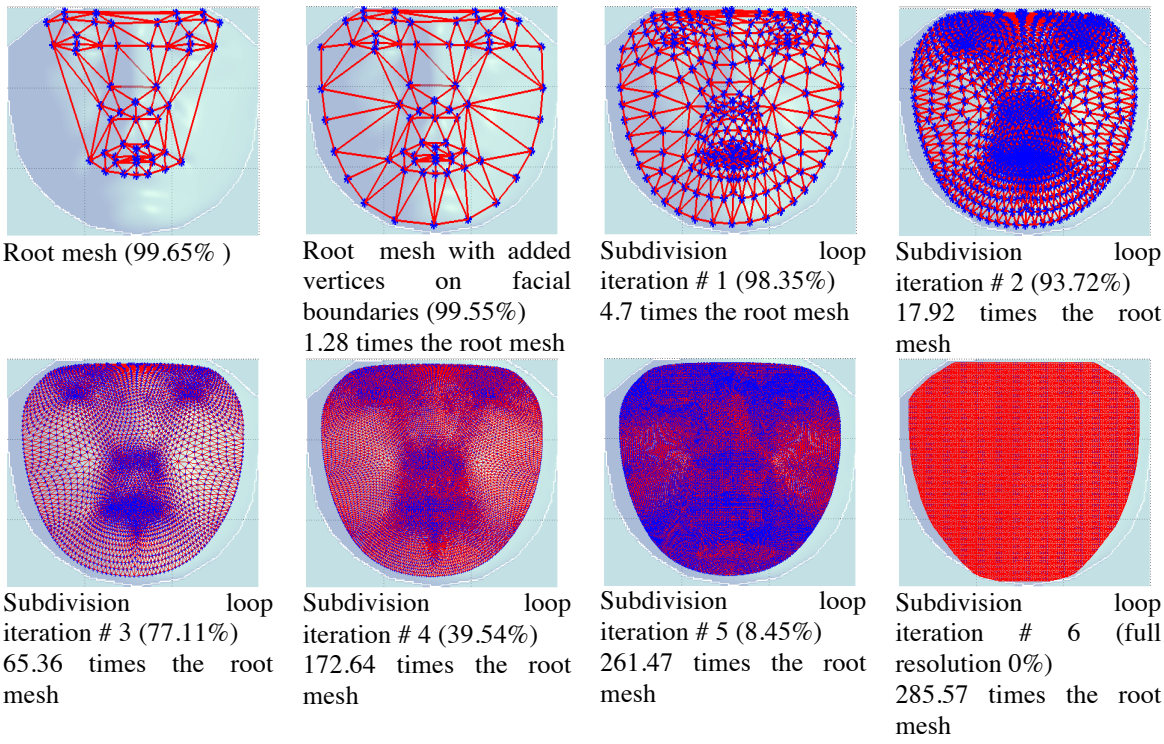


Figure 3 – Different facial surface sampling meshes, percentages indicate the down-sampling factor of each sampling mesh

It is important to note the difference between the adaptive mesh generated in this report and that used in the AAM-based recognition approaches [e.g., 8]. The mesh in the AAM is fitted over a given intensity image by minimizing an objective function that involves shape and appearance parameters, whereas in our case, we use the raw 3D data to generate the root mesh and the subsequent densifications.

3. 3D FACE RECOGNITION

The main purpose of this report is studying the effects of sparsity of the reconstruction on the recognition rate. We will use the standard FRGC database [14, 15] and the basic recognition

approaches reported elsewhere [eg. 16]. Of course, enhancements in the data acquisition and recognition methods may be easily incorporated in the approaches devised in this report. We describe the preprocessing steps and the recognition criterion. Then we perform the recognition on the surfaces generated by the adaptive meshing approach described in the previous section.

3.1 Data Processing

We used the FRGC data corpus [14,15] which contains 3D face images (laser scans) acquired under structured light sensor that generates 640 by 480 range sampling and a registered color image. We used two datasets, range images collected in spring 2003 (943 scans) and spring 2004 (2114 scans), total of 3057 facials scans for more than 480 subjects. However we excluded scans having expressions and other artifacts due to acquisition such as un-registered pair of range and texture images. A total of 1164 scans were excluded. The resulting database contains 436 subjects of which 89 subjects with only one scan. The number of scans per subject ranges from one to fifteen scans, with mean value of 4.2867 and standard deviation of 3.0884. Fig. 4 shows the histogram of number of scans per subject, while Table 1 shows number of subjects having at least N-scans where N ranges from 1 to 7.

Table 1 - Number of subjects having at least N-scans

N-Scans	1	2	3	4	5	6	7
# of subjects	436	347	281	223	171	125	85

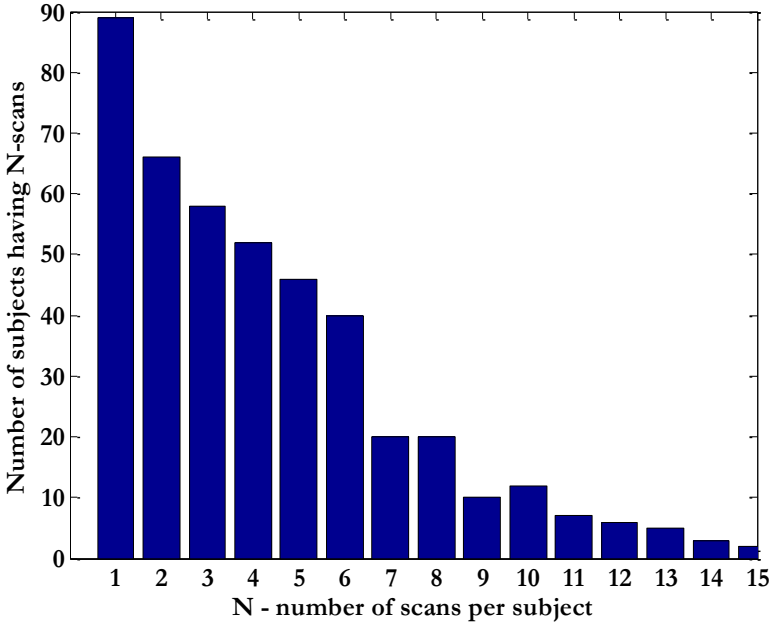


Figure 4 - Histogram of number of scans per subject, after excluding scans having expressions and other artifacts due to acquisition, a total of 1893 scans were included in this study.

Common problems with the raw 3D data include existence of holes, spikes, variable pose and location with respect to the acquisition device can be alleviated in the preprocessing stage. Guided by the preprocessing steps outlined by Chang *et al*. [15], we can summarize this stage as follows: i) masking the range image so that relevant face region will be maintained. This is accomplished through thresholding the height map taking into consideration that subjects stood or sat approximately 1.5 meters from the sensor [15]. Using connected component labeling on a mask of ones where the height map was thresholded, we maintain the region with the largest area. We fill the holes due to missing 3D data during sensing by interpolation using the information provided by the neighborhood region. We crop the facial region to remove the background. ii) The next step removes spike artifacts that can occur in the 3D image, where the variance in the Z component of the 3D data is computed over an 11x11 neighborhood, a point is considered a part of a spike if its variance exceed a predefined threshold, spike points are

removed by linear interpolation using the neighborhood/non-spike points. Finally median filtration is applied for smoothing. Fig. 5 shows the main steps of scan preprocessing along with intermediate results on a sample scan.

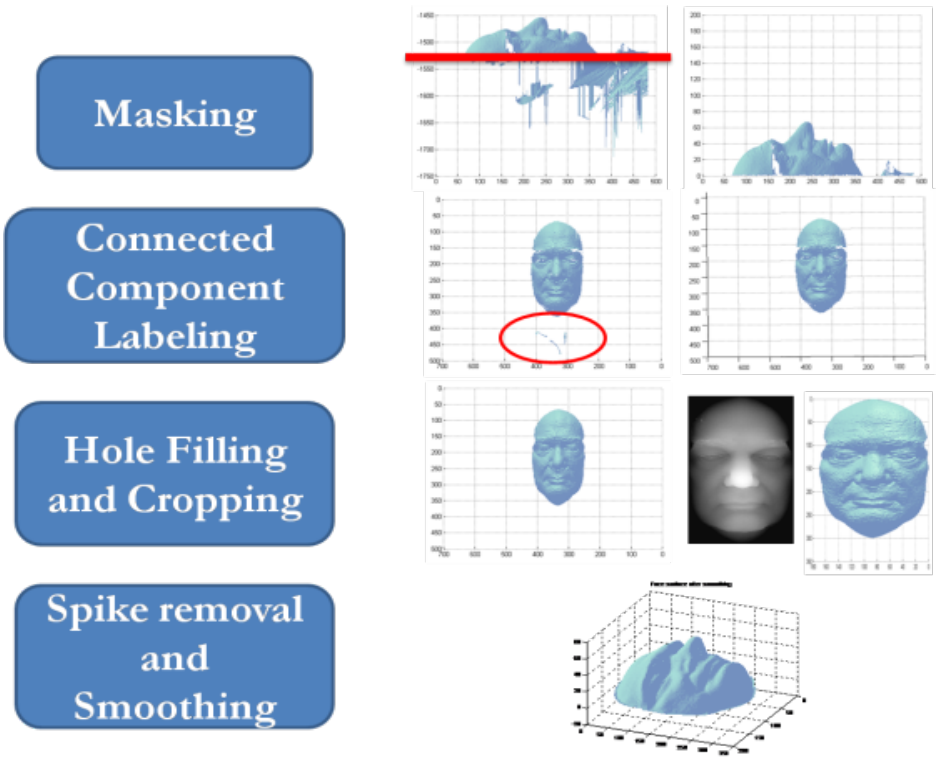


Figure 5 - Main steps of scan preprocessing. The facial region is masked and geometric data processing is performed to fill in the missing holes in the facial scans and spurious details/spikes are removed.

3.2 Alignment/Rigid Registration

The main characteristic that distinguishes a point in 3D space in a range image is its depth value along the Z-axis which defines the distance between the sensing device and each point on the object being scanned. Hence, all facial scans share the same spatial support (xy-

plane), however they do not share the same pose and location with respect to the sensor. To correct pose variation, geometric normalization is needed, where four landmark points are manually selected; the left and right outer eye corners, nose tip and chin center.

As noted by Chang *et al* [15]; "The outer eye corners rather than eye centers are used as landmark points because the eyeball is an artifact-prone region for the range sensor, whereas the eye corners marked on the skin are more reliable." We chose to manually mark these landmarks in order to isolate sources of error from the automatic detection algorithm; however, we refined the location of the marked nose tip over a neighborhood of 21x21, from the geometric aspect, a nose tip can be considered as a spherical cup or a dome. It can be defined using differential geometry-based concepts such as shape index [17] which describes the local behavior of surfaces in a small neighborhood; hence we considered the refined nose tip to be the local maxima of a homogenous region having a shape index that ranges from 0.8125 to 1. Fig. 6 outlines the nose tip refinement procedure.

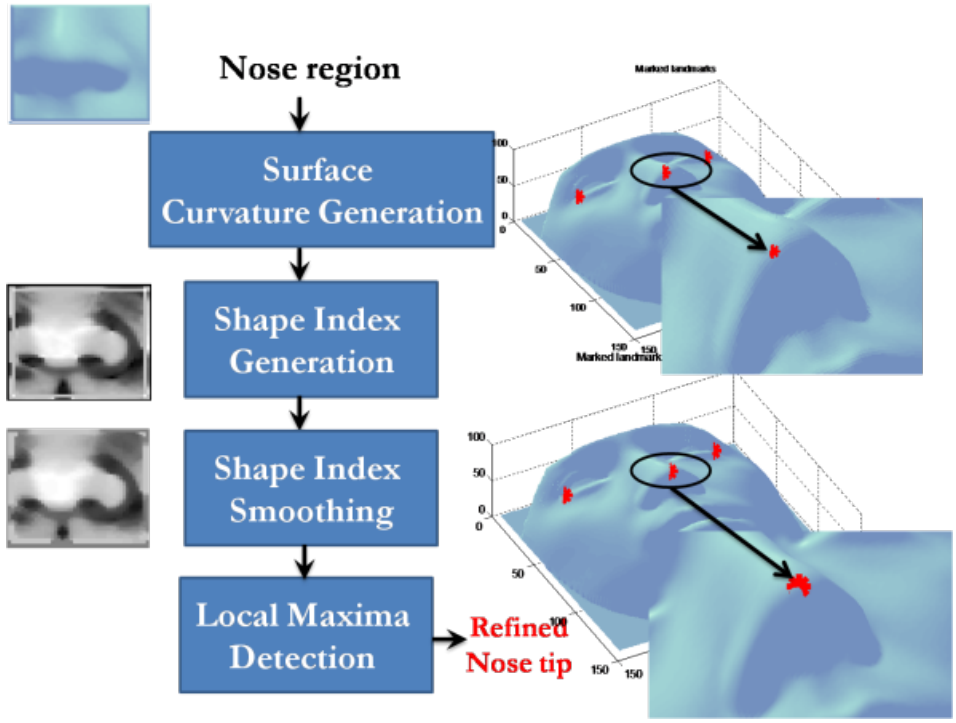


Figure 6 - Nose-tip refinement procedure

The marked landmark points are used to place the cropped 3D image in a standard pose. We roughly align the range and texture image to predefined reference points of a standard face pose and location defined by the mean shape/face of the University of South Florida (USF) database [11]. The alignment is achieved by affine transformation estimated using Procrustes analysis [18]. Fig. 7 shows the reference points of the USF mean face used to align FRGC facial scans. A sample scan along with its marked landmark points is shown, after geometric normalization, the scan (range and texture images) share the same spatial support and same landmarks location with the mean face. This process resulted in a spatial support of dimensions 142x145.

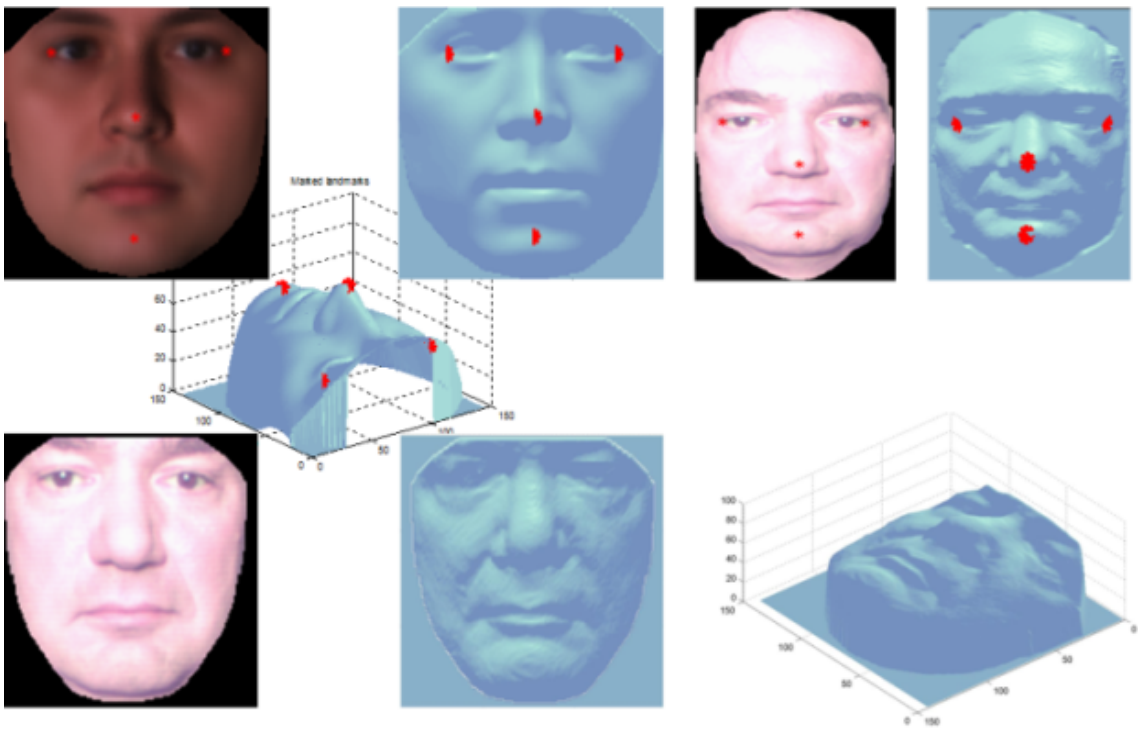


Figure 7 - The reference points of the USF mean face used to roughly align FRGC facial scans (a sample scan is shown), procrustes analysis is used to estimate the affine transformation parameters which causes the range and texture images of the scan to share the same spatial support and landmark points location with the mean shape.

3.3 Baseline Recognition

A PCA-based face recognition algorithm was selected as the baseline algorithm [14]. Figs. 8 and 9 show the outline of the training and test stages where the first scan of each subject was taken to probe a gallery containing a maximum of 7 scans for each subject. The face space (eigenFaces) was trained using the scans that define the gallery. Results in this subsection are reported using different similarity measures, including L1 norm, L2 norm, Cosine, Chebychev and Canberra similarity measures which are defined as follows;

Definition 3: *Let* **x** *and* **y** *be two feature vectors*, *Theirsimilarity measures can be defined as*;

$d_{L1norm} = \sum_{i=1}^{n} |x_i - y_i|$, $d_{L2norm} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$, $d_{cosine} = \frac{\boldsymbol{x}.\boldsymbol{y}}{||\boldsymbol{x}||||\boldsymbol{y}||}$, $d_{\text{chebychev}} = \max_i |x_i - y_i|$ and $d_{canberra} = \sum_{i=1}^{n} \frac{|x_i - y_i|}{|x_i| + |y_i|}$.

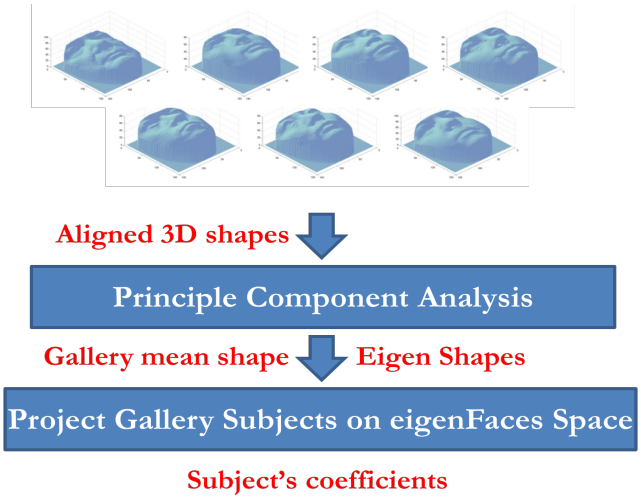


Figure 8 – Training (gallery building) stage, gallery contains lasers scans after preprocessing and alignment, where PCA is trained using the Z-component of each scan since all scans share the same spatial support.

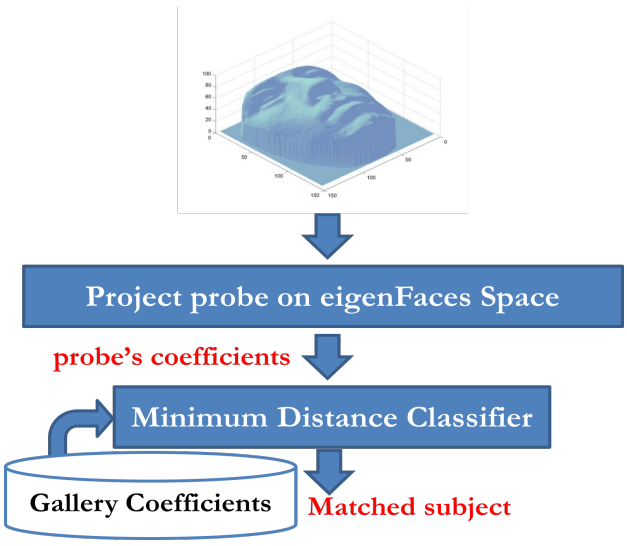


Figure 9 - Testing (probe matching) stage, a probe scan is used to query the gallery where the scan is projected on the eigenFaces space and probe matching is achieved using minimum distance classifier

Having multiple scans per subject in the gallery with maximum seven scans has lead to the best baseline performance (Fig. 10) when compared to having a single scan per subject in the gallery (Fig. 11). We maintained the eigen vectors corresponding to eigen values which maintain 98% of data variability.

3.4 Effect of Down Sampling

Based on extensive experimentations, we have shown that the L1 norm outperforms other similarity measures regardless of the down-sampling factor; on the other hand, Canberra distance provides the poorest recognition accuracy. Fig. 12 shows the CMS curves using meshes of different down-sampling factors.

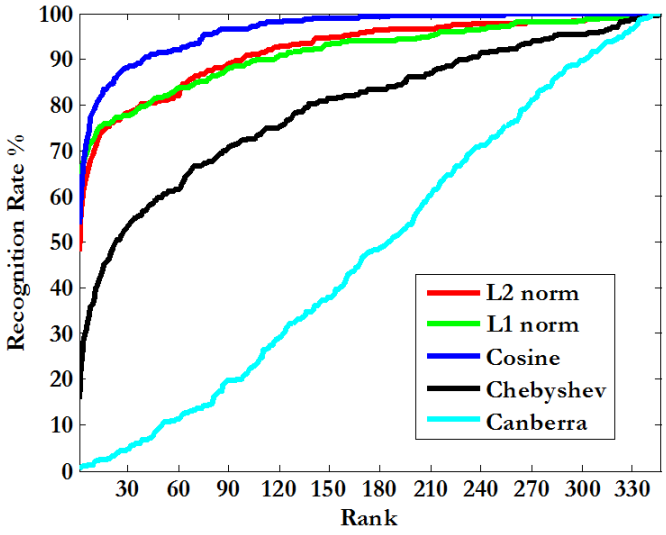


Figure 10 - CMS curve of PCA-based face recognition algorithm when using one scan per subject in the gallery, cosine measure outperform other similarity measures with rank one = 53.89%

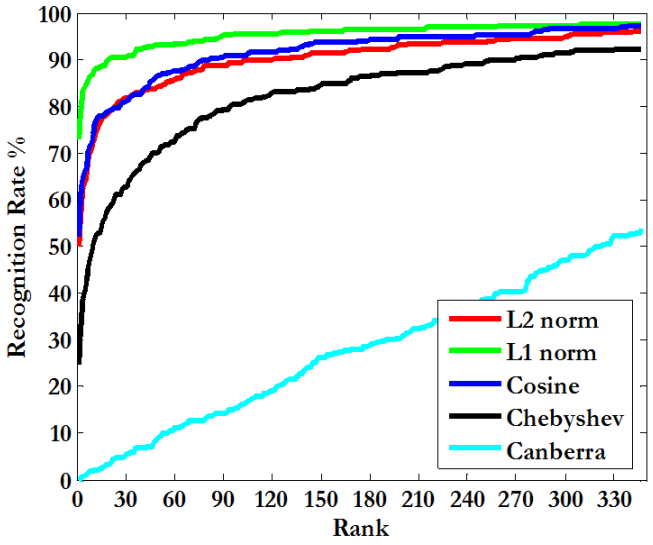
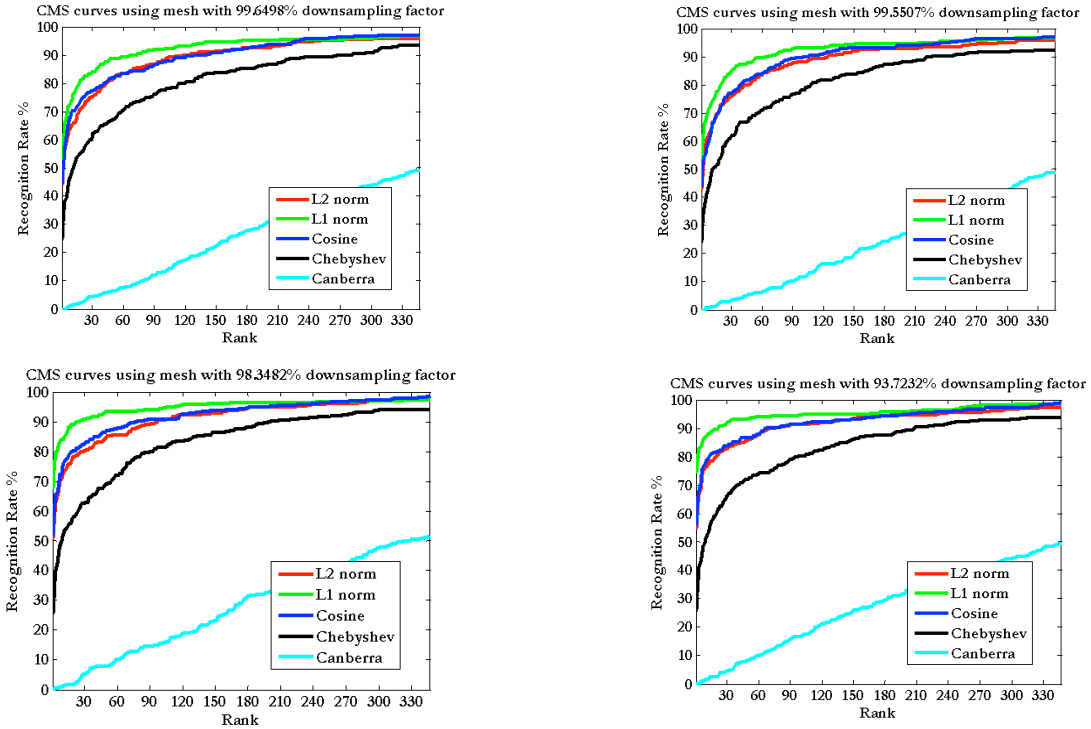


Figure 11 - CMS curve of PCA-based face recognition algorithm when using multiple scans per subject in the gallery with maximum 7 scans, L1 norm outperform other similarity measures with rank one = 73.2%



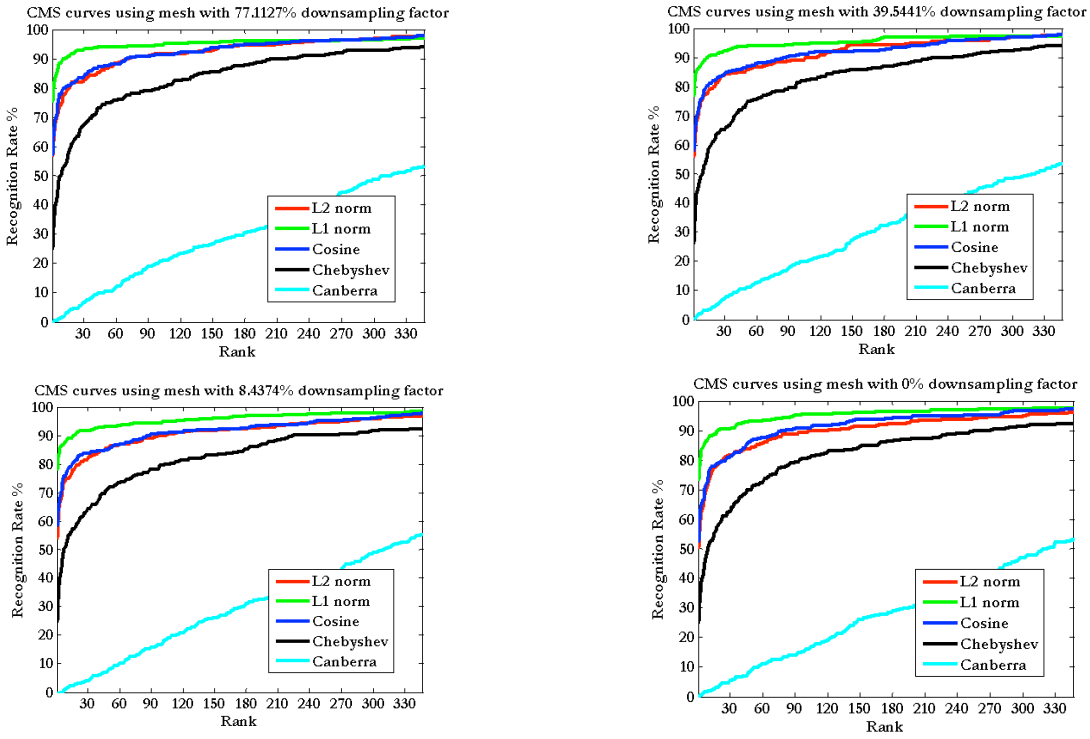


Figure 12 – CMS curves using meshes of different down-sampling factors, different distance measured behave similarly regardless of scan's resolution

Fig. 13 shows the CMS curve using L1 norm similarity measure for meshes of different down-sampling factors. We observe that meshes with down-sampling factors ranging from 8.43% to 98.34% provided similar recognition accuracy as full resolution scans. To the best of our knowledge this conclusion, which is based on extensive empirical results, documents for the first time in the literature the effects of proper meshing on the recognition.

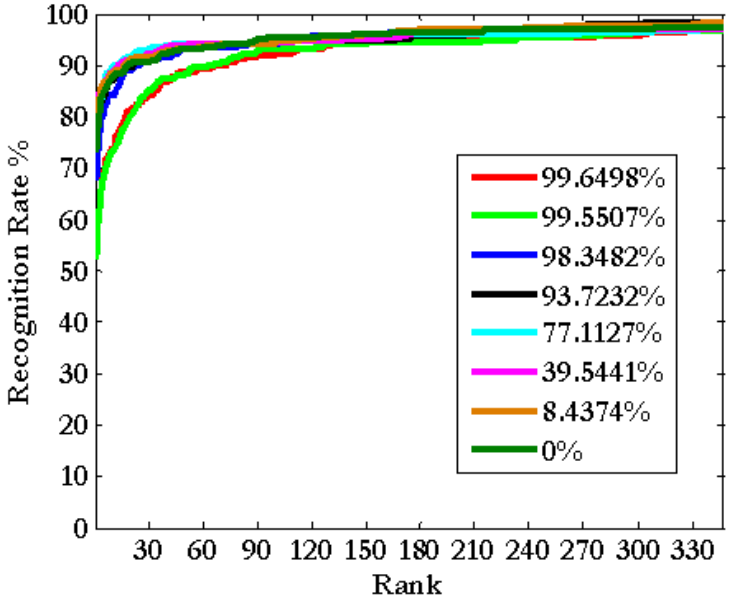


Figure 13 – CMS curves using L1 norm for meshes of different down-sampling factors, it can be observed that meshes with down-sampling factors starting from 8.43% to 98.34% give us similar recognition accuracy as if we are using the facial mesh with the full resolution.

Fig. 14 shows the behavior of rank one recognition results for different down-sampling factors. It can be observed that recognition rate degrades after 98.35% down-sampling factor, however, other factors lead to similar performance to the full resolution case, moreover, a little bit down-sampling (8.44%) might enhance the recognition accuracy, hence it can be concluded that full resolution might have noisy data.

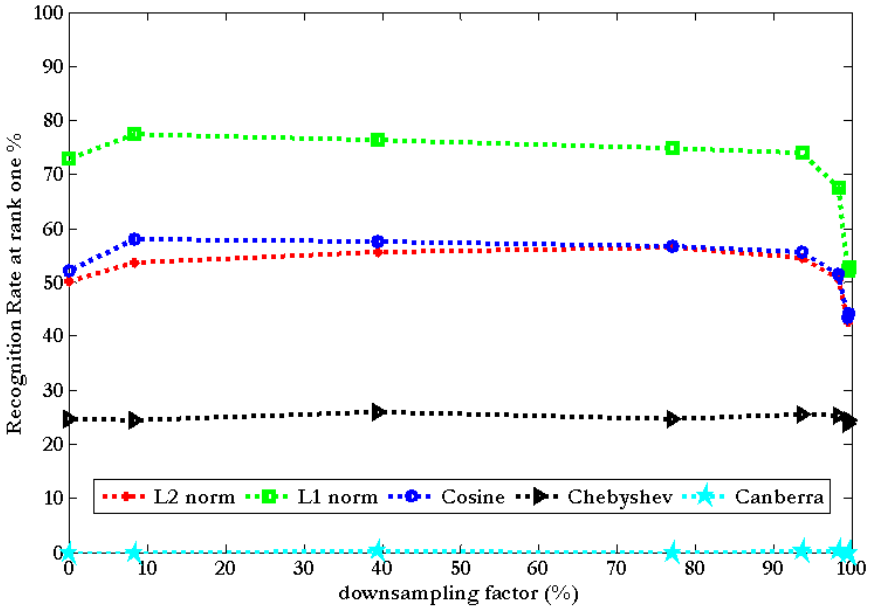


Figure 14 – rank one recognition results for different down-sampling factors.

The results suggest that a mesh with size of almost 2% of the original data and about 5 times the root mesh provided nearly same recognition as having full scanning information. This observation has been confirmed through various repeated experimentations.

4. APPLICATION TO STEREO-BASED RECOGNITION

Recent focus of face recognition research has been directed towards 3D-based approaches where a 3D image may be constructed directly from 3D scanner or inferred from stereo approaches and model-based approaches such as statistical shape from shading [19, 20]. We use an adaptive mesh, with higher resolution around facial features, as motivated by psychological studies. However, for a fixed number of mesh points, an adaptive mesh is arguably harder to obtain than a regular mesh, which raises questions in terms of applicability to real-time stereo based on such adaptive mesh; this root mesh can be generated using Active Appearance Models fitted on the stereo-pair. Our group has constructed a fully operational stereo-based 3D

recognition system. The sparsity analysis reported in this paper, using large FRGC database of objects, provided us with important clues in reconstructions from stereo. In fact, the root mesh and its 2D projection serve multiple purposes: constricting the reconstruction region, and an initial solution of the correspondence problem in stereo matching. Due to space limitations we will not be able to describe the stereo-based approach. Indeed, extensive results based on a stereo database of 30 subjects taken at three different distances (3, 12 and 33 meters) confirm that robust near real-time 3D face recognition can be obtained based on sparse facial reconstruction. We intend to make this database public pending approval of the sponsor. This stereo database will be a vehicle of various algorithmic and validation studies in 3D face recognition.

5. CONCLUSION

This report describes an approach to test the effect of sparsity of reconstruction on the recognition rate. Our results show that a small sample of data points forming *a root mesh* may be adequate vis-à-vis recognition. A mesh of nearly 2% of the original data points provided almost the same recognition accuracy, using the L1 norm, as the full information. This finding suggests that most of the information for recognition comes from the low-frequency components, corresponding to a sparse mesh in our case. This conclusion supports the findings by psychologists, for example [22]. A figure of merit on the effectiveness of sparse reconstruction on recognition may pave the way to real-time 3D face recognition based on sparse meshes from stereo.

Ongoing efforts are focused on constructing the *root mesh* based on well-established psychological studies and morphometric analysis. Theoretical work is ongoing to quantify the

topological characteristics of the root mesh and studying invariance of geometric sampling. Preliminary results on the meshing approach shows that the root mesh lends credence to studies in concentration factors used as measures of facial information (e.g., [19]) and other human-based recognition.

REFERENCES

[1] Bowyer, K.W., Chang, K., Flynn, P.J.: A survey of approaches to three-dimensional face recognition. International Conference of Pattern Recognition, Vol I. (2004) 358–361

[2] Kevin W. Bowyer, Kyong Chang, and Patrick J. Flynn, A Survey of Approaches and Challenges in 3D and Multi-modal 3D+2D Face Recognition, Computer Vision and Image Understanding 101 (1), January 2006, 1-15.

[3] Takuma Shibahara, Takafumi Aoki , Hiroshi Nakajima, Koji Kobayashi , A sub-pixel stereo correspondence technique based on 1D phase-only correlation. International conference of image processing (ICIP) 2007

[4] Unsang Park and Anil K. Jain, "3D Face Reconstruction from Stereo Video," First International Workshop on Video Processing for Security (VP4S-06), June 7-9, Quebec City, Canada, 2006.

[5] T.F. Cootes, C.J. Taylor, Statistical Models of Appearance for Computer Vision, Technical Report, University of Manchester.

[6] Wright, J. Yang, A.Y. Ganesh, A. Sastry, S.S. Yi Ma , Robust Face Recognition via Sparse Representation IEEE Transactions on **Pattern Analysis and Machine Intelligence,** Feb 2009, Vol 31, issue: 2, pp: 210-227.

[7] Johan W.H. Tangelder Ben A.M. Schouten, "Learning a Sparse Representation from Multiple Still Images for On-Line Face Recognition in an Unconstrained Environment," ICPR 2006

[8] Ramnath, K.; Baker, S.; Matthews, I.; Ramanan, D., "Increasing the density of Active Appearance Models," CVPR 2008

[9] M. Sabry Hassouna and A. A. Farag, TPAMI-0845-1206 - Variational Curve Skeletons Using Gradient Vector Flow," IEEE Transactions on Pattern Analysis and Machine Intelligence - (Accepted October 2008).

[10] Jan Philipp Hakenberg, Smooth Subdivision for Mixed Volumetric Meshes, Master thesis, Rice university, Houston, Texas, August 2004.

[11] K. Bowyer, S. Sarkar, USF 3D Face Database. http://www.csee.usf.edu/~sarkar/index_files/DataAndCode.htm

[12] Denis Zorin, Peter Schröder, and Wim Sweldens. Interpolating subdivision for meshes with arbitrary topology. *ACM SIGGRAPH 1996 Conference Proceedings*, p.189 - p.192, 1996.

[13] D. Zorin, P. Schröder, A. DeRose, L. Kobbelt, A. Levin, W. Sweldens, Subdivision for modeling and animation*, SIGGRAPH 2000 Course Notes*.

[14] P. Jonathon Phillips, Patrick J. Flynn, Todd Scruggs, Kevin W. Bowyer, Jin Chang, Kevin Hoffman, Joe Marques, Jaesik Min, and William Worek, *Computer Vision and Pattern Recognition (CVPR 2005)*, San Diego, June 2005, I:947-954.

[15] Kyong I. Chang, Kevin W. Bowyer, and Partrick J. Flynn, An Evaluation of Multimodal 2D+3D Face Biometrics, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 4, April 2005.

[16] M.A. Turk, A.P. Pentland, Face Recognition Using Eigenfaces, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3-6 June 1991, Maui, Hawaii, USA, pp. 586-591

[17] Dorai, C.; Jain, A.K., "COSMOS-A representation scheme for 3D free-form objects," *Transactions on Pattern Analysis and Machine Intelligence* , vol.19, no.10, pp.1115-1130, Oct 1997

[18] F.L. Bookstein, *Morphometric tools for landmark data*, Cambridge University Press, (1991).

[19] W.A.P. Smith and E.R.Hancock, "Facial Shape-from-shading and Recognition Using Principal Geodesic Analysis and Robust Statistics", *International Journal of Computer Vision,* 76, 71—91, 2008.

[20] Abdelrehim Ahmed, Aly Farag and Thomas Starr, A new symmetric shape from shading algorithm with an application to 3-D face reconstruction. Proc. of IEEE International Conference on Image Processing (ICIP'08), San Diego, California, pp. 201-204, October 12-15, 2008.

[21] P Sinha, B Balas, Y Ostrovsky, R Russell , Face recognition by humans: Nineteen results all computer vision researchers should know about. Proceedings of the IEEE, 2006.

APPENDIX A: LOOP'S SUBDIVISION

Loop's subdivision scheme works on triangular meshes. The surface is an approximating surface, that is it doesn't interpolate (passes through) its control points. A new vertex is added for each edge, at each subdivision step.

Definition A.1*: The* **valence** *of a vertex* **x** *on the subdivision surface is the number of neighboring vertices, that is the number of vertices connected to* **x** *by an edge*.

Definition A.2*: A vertex with valence 6 is known as* **regular** *or* **ordinary** *vertex, otherwise it is known as* **irregular** *or* **extraordinary** *one*.

Loop's subdivision scheme can be summarized as follows;

(1) At each subdivision step k, the existing point $\mathbf{x}^k$ is updated with the scheme,

$$\mathbf{x}^{k+1} = (1 - n\beta)\mathbf{x}^k + \beta \sum_{i=1}^{n} \mathbf{y}_i^k$$

Where $\mathbf{y}_i^k$ are the n neighboring vertices and β is a constant determined by n. Loop suggests the function,

$$\beta(n) = \frac{1}{n}\left(\frac{5}{8} - \frac{\left(3 + 2\cos\left(\frac{2\pi}{n}\right)\right)^2}{64}\right)$$

While there is another alternative definition,

$$\beta(n) = \frac{3}{n(n+2)}$$

In both cases the surface is C^2 at vertices of regular valence and C^1 otherwise.

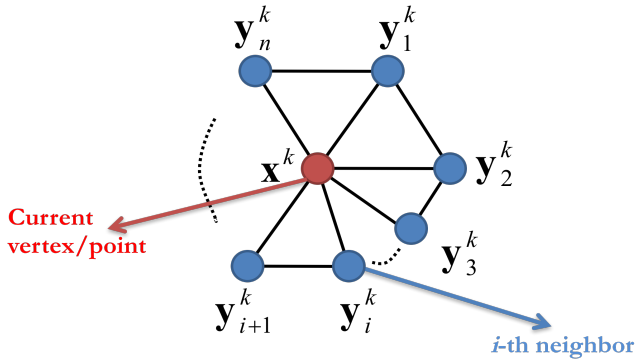


Figure A.1 – Neighborhood/valence of the current point

(2) For each edge connecting $\mathbf{x}^k$ to a neighbor, a new vertex is created via,

$$\mathbf{x}_i^{k+1} = \frac{3\mathbf{x}^k + 3\mathbf{y}_i^k + \mathbf{y}_{i-1}^k + \mathbf{y}_{i+1}^k}{8} \quad , \qquad i = 1,2,\dots,n$$

The subscripts i are calculated modulo n.

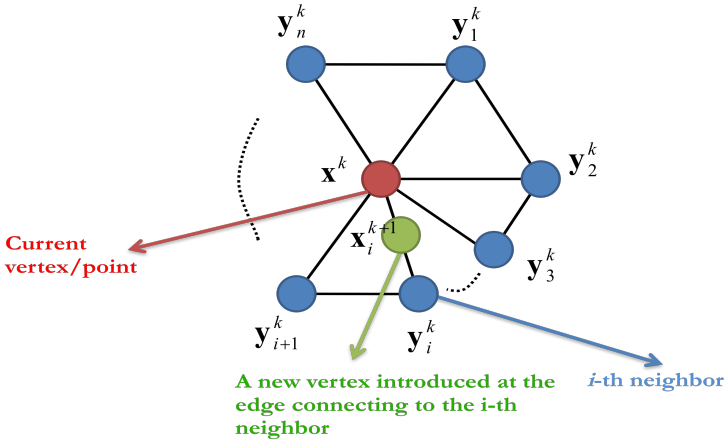


Figure A.2 – New vertex introduced on the edges connecting neighboring vertices

(3) Once the new vertices have been determined, each triangle is subdivided into four triangles.

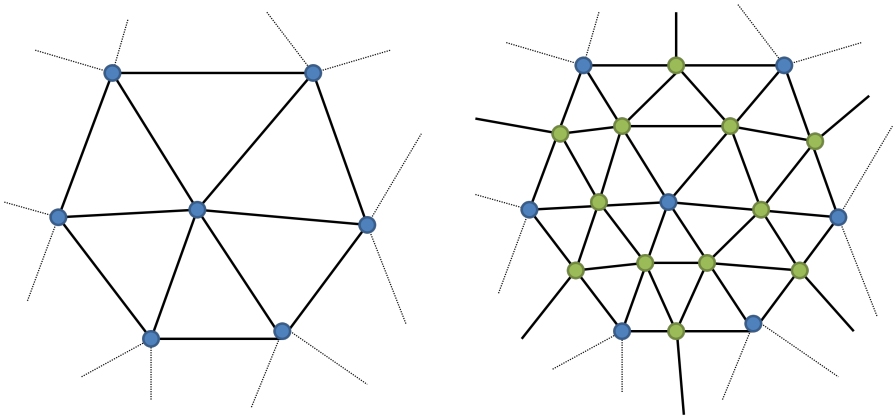


Figure 4 – On each edge a vertex will be added, then each triangle is subdivided into four triangles

Often a mask or stencil is used to visualize the subdivision, the entries in the mask are the weights for the contributing points. Noting that the sum of the weights is one. There are two types of masks, the first one is called the vertex mask which holds the weights of the current point $\mathbf{x}^k$ and its neighboring vertices $\mathbf{y}_i^k$ used to update the current point to obtain $\mathbf{x}^{k+1}$.

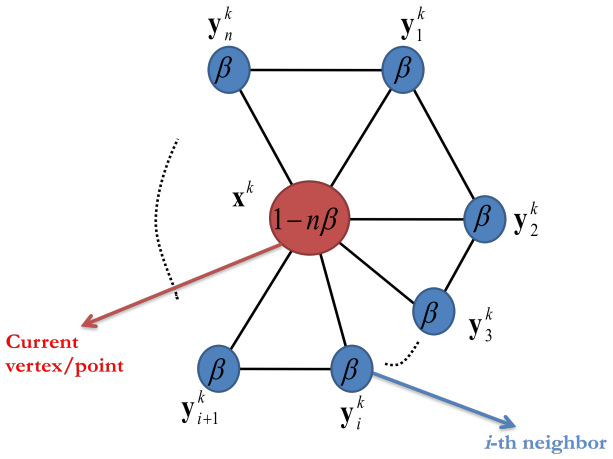


Figure A.4 –Vertex mask for Loop subdivision

The second type of masks is the edge mask, which defines the weights of the four neighboring vertices used to generate a new vertex $\mathbf{x}_i^{k+1}$ on the edge connecting the current point $\mathbf{x}^k$ and its i-th neighbor $\mathbf{y}_i^k$.

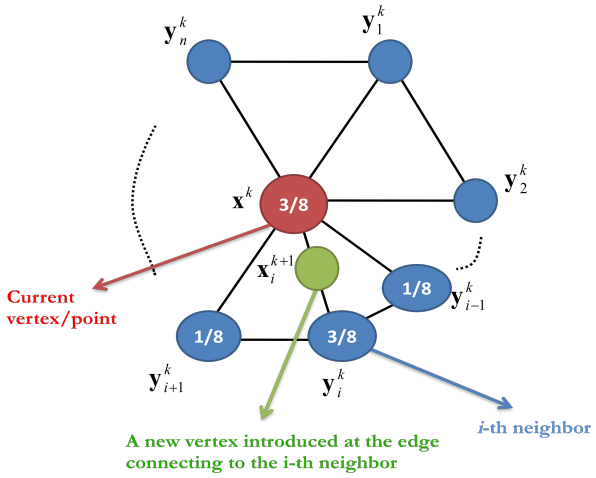


Figure A.5 - Edge mask for Loop subdivision

The scheme listed so far can work on closed surfaces however it needs some adjustments to handle open surfaces. Edges and vertices on the boundary of an open surface have different masks shown in Fig. A.6.

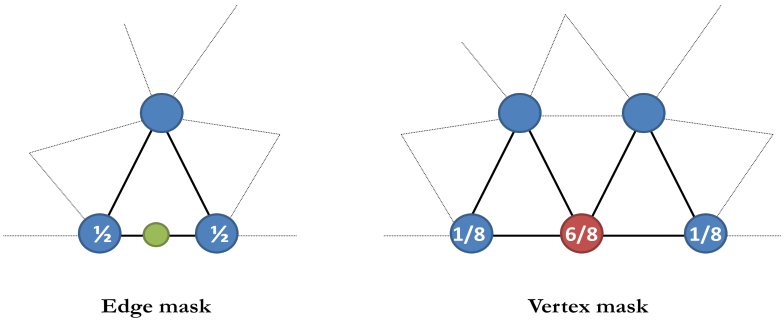


Figure A.6 – Subdivision masks for boundary conditions

It is important to note that the generated surface is affine invariant and lies in the convex hull of its control points. Two limit tangents can be computed by a weighted sum of the immediate control points or 1-*ring* or 1-*neighborhood* of a point $\mathbf{x}^k$. The tangents are,

$$\mathbf{t}_u = \sum_{i=0}^{n-1} \cos\left(\frac{2\pi i}{n}\right) \mathbf{y}_i^k$$

$$\mathbf{t}_v = \sum_{i=0}^{n-1} \sin\left(\frac{2\pi i}{n}\right) \mathbf{y}_i^k$$

The *normal vector* is then given by $\mathbf{n} = \mathbf{t}_u \times \mathbf{t}_v$. The resulting surface is quite *fair* that is it bends quite smoothly.