3D Face Sketch Modeling and Assessment for Component Based Face Recognition

Shaun Canavan[1], Xing Zhang[1], Lijun Yin[1], and Yong Zhang[2]

[1]State University of New York at Binghamton, Binghamton, NY.
[2]Youngstown State University, Youngstown, OH.

Abstract

3D facial representations have been widely used for face recognition. There has been intensive research on geometric matching and similarity measurement on 3D range data and 3D geometric meshes of individual faces. However, little investigation has been done on geometric measurement for 3D sketch models. In this paper, we study the 3D face recognition from 3D face sketches which are derived from hand-drawn sketches and machine generated sketches. First, we have developed a 3D sketch modeling approach to create 3D facial sketch models from 2D facial sketch images. Second, we compared the 3D sketches to the existing 3D scans. Third, the 3D face similarity is measured between 3D sketches versus 3D scans, and 3D sketches versus 3D sketches based on the spatial Hidden Markov Model (HMM) classification. Experiments are conducted on both the BU-4DFE database and YSU face sketch database, resulting in a recognition rate at around 92% on average.

1. Introduction

Face sketches can be drawn either by a trained police artist or using a composite software kit [1, 2]. Both types of sketches have been studied in the context of searching or matching a sketch to a subject's face in a database of photos or mug-shots [3,4,5,6,7,8]. Since all existing works were based on 2D sketches, issues of pose variations are still challenging. Recently, 3D face recognition has attracted much attention [9,10,13,14]. Along the same vein, 3D sketch models reconstructed from 2D sketches may improve sketch recognition performance. In order to increase the accuracy of geometric surface matching and efficiency of similarity measurement between 3D faces and probe sketch data, it is highly demanded to have 3D sketches matched up with the 3D scan models. Nevertheless, there is little investigation reported on 3D sketch modeling and 3D sketch recognition in the past.

In this paper, we address the issue of 3D sketch model construction from 2D sketches, and compare the 3D sketch models with the corresponding 3D facial scans. We further validate the models by conducting 3D face sketch identification on two 3D face databases. Note that there is no existing graphic tool for 3D sketch model construction from witness' description directly. One solution is to create 3D sketch models based on 2D sketches from hand-drawings by artists or conversion from 2D images [11, 12].

To build 3D sketch models, we applied a scale-space topographic feature representation approach to model the facial sketch appearance explicitly. We initially tracked 92 key facial landmarks using an active appearance model (AAM) [15], and then interpolated to 459. From the interpolated landmarks, we used a 3D geometric reference model to create individual faces. Based on the topographic features obtained from the sketch images, we applied a mesh adaptation method to instantiate the model.

In order to assess the quality of created 3D sketch models, we conducted a comparison study between the created 3D sketch models and their corresponding ground-true 3D scans. We show the difference between two data sets as well as the difference between the 3D sketch models created from hand-drawn sketches and the 3D sketch models created from machine-derived sketches. Moreover, in order to validate the utility of the 3D sketch models, we propose a new approach to decompose the 3D model into 6 independent component regions, and apply a spatial HMM model for sketch model recognition. The 3D sketch face recognition experiment is conducted on two databases: BU-4DFE [19] and YSU sketch database [20].

2. Source Data

Face database BU-4DFE [19] and YSU sketch database [20] have been used as our data source. Sample 3D scans of BU-4DFE are shown in Figure 1. The corresponding 2D textures are shown in Figure 2 (first row).

Based on six subjects of 4DFE, two forensic artists from Youngstown State University drew the corresponding 2D sketches of the same subjects. Thus we obtained Hand-Drawn (*HD*) sketch images of six subjects. Figure 2 (row-2) shows several samples of HD 2D-sketches.

Due to the time-consuming and intensive work of artist drawing, we created 2D sketches from 2D texture images of the 4DFE database. Our method can simulate the pencil sketch effect. The texture to sketch conversion follows a

978-1-4577-1359-0/11/$26.00 ©2011 IEEE

three-step image processing procedure: First, the image is processed by a de-saturation process, then the image is inverted. After applying a color dodge, the image is blurred with a Gaussian filter. Finally, the radius of pixels is adjusted to get an ideal sketch effect. Figure 2 (row-3) shows examples of MD 2D-sketch images.

We have also used YSU hand-drawn 2D facial sketches with 250 sketches. Figure 3 (row-1) shows examples of YSU HD 2D-sketches.

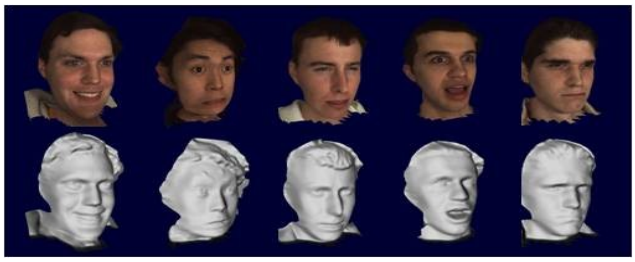


Figure 1: Examples of 3D scans from 4DFE: textured models shown in upper row and shaded models in bottom row.

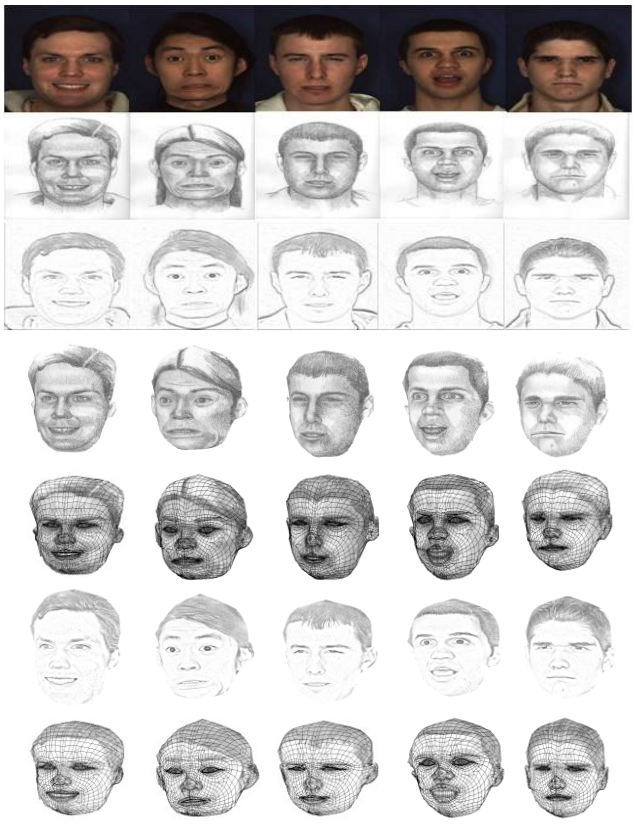


Figure 2: Examples of 2D images of 4DFE from top to bottom: Original textures (*row-1*); Hand-drawn (HD) 2D sketches (*row-2*), and Machine-derived (MD) 2D sketches (*row-3*). *Rows 4-5:* Created 3D sketches from HD sketches with textures and mesh models in different views. *Row 6-7*: Created 3D sketches from MD sketches with textures and mesh models in different views.

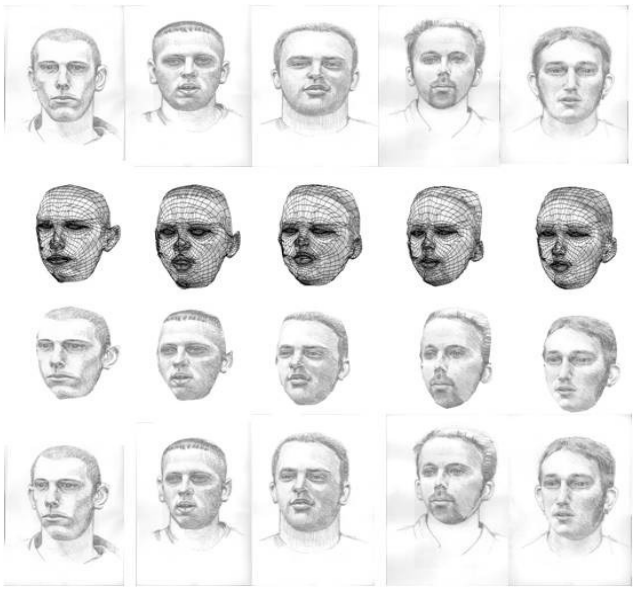
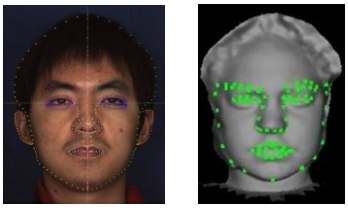


Figure 3: Samples of YSU 2D sketch database (*Row 1*) and the reconstructed 3D sketches (*Row 2-3*). *Row-4* shows synthesized 3D sketches with rotated heads on the corresponding shoulders.

3. 3D Sketches Creation from 2D Sketches

3.1. 3D sketch reconstruction

To build 3D sketch models, we developed a scale-space topographic feature representation approach to model the facial sketch appearance explicitly. Using an AAM we initially tracked 92 key facial landmarks, and then interpolated them to 459 using a Catmull-Rom spline [16] (as shown in Figure 4 (a)).



(a) (b)

Figure 4: (a) Illustration of 459 points on a sample face; (b) 83 feature points for evaluation.

From the interpolated landmarks, we used a 3D geometric reference model to create individual faces. The reference model consists of 3,000 vertices. Based on the topographic labels [17] and curvatures obtained from the sketch images, we then applied a spring-mass motion equation [18] to converge the reference model to the sketch topographic surfaces in both horizontal and depth directions. Existing topographic labeling approaches can create different levels of feature detail, depending on the variance of the Gaussian smoothing function (σ) and the fitting polynomial patch size (N) (both σ and N are called *scales*). The existing applications of topographic analysis

are limited in a "still" topographic map with a selected scale. Every label may represent various features in a specific image, various features (e.g., on the organs in the human face) may be "screened out" with various "optimal" scales. The idea of scale is critical for a symbolic description of the significant changes in images. A small scale could produce too much noise or fake features. A large scale may cause the loss of important features. Too many fake features could cause the model adaptation be distracted. More seriously, it could cause the adaptation to be unstable, (e.g., even not converge). Too few features will not attract the generic model into the local facial region with expected accuracy. Due to the difficulty to select an "optimal" scale, here we use a multi-scale analysis approach to represent the topographic features from a coarse level to a fine level as the scale varies. Applying the topographic labeling algorithm with different scales, we generated the topographic label maps of facial images at different levels of detail. Different scales will be applied to different levels of details of sketch images (e.g., hand-drawn (fine details) or machine derived (coarse details).

In order to deform the face model into the non-rigid facial area, we applied the adaptive mesh [18] to the facial areas in the *multi-scale* topographic domain. Such dynamic meshes are moved by not only the 2-D external force (*e.g.* topographic gradient) but also the depth force (*e.g.* topographic curvature) for model deformation in *multiple scales*. We take the model as a dynamic structure, in which the elastic meshes are constructed from nodes connected by springs. The 3D external force is decomposed into two components: the gradients of the topographic surface are applied to the image plane, and the curvatures of the topographic surface are applied as a force to pull or deflect meshes in the direction perpendicular to the image plane. As a result, the 3D shape of mesh becomes consistent with the face surface. This procedure was performed based on a series of numerical iterations until the node velocity and acceleration were close to zero. Such a mesh adaptation method was applied to sketch regions to instantiate the model. Figure 2 and Figure 3 shows examples of 3D sketch models reconstructed from 2D sketches for both HD and MD data.

3.2. 3D sketch accuracy evaluation

(1) Comparison: 3D HD sketches versus 3D scans

We also conducted an objective evaluation, by which we calculated the error between the feature points on the individualized sketch models and the corresponding manually picked points on the face scans. We selected 83 key points as the ground truth for assessment (see Figure 4 (b)). After creating sketch models from 4DFE, we conduct a quantitative measurement as follows: First, we normalize all the models into a range of [-50, 50] in three coordinates of x, y, and z. We then calculate the mean square error (MSE) between the feature points of the 3D sketch and the ground truth of a set of models. We define the one-point spacing as a closest pair of points on the 3D scans, which is approximately 0.5mm on the geometric surface of the 4DFE models. The mean error of two models can be computed by the average of point differences between two models. Figure 5 shows the error statistics, which is the average error and standard deviation on each of the 83 key points. The result shows that the MSE of the examined points is 8.71 point spacings. The average error ranges from 1 to 16 point spacings, with the most of points being less than 10 point spacings. The errors mainly lie in the left side and right side of the face contour and chin area, which are points 69-83.

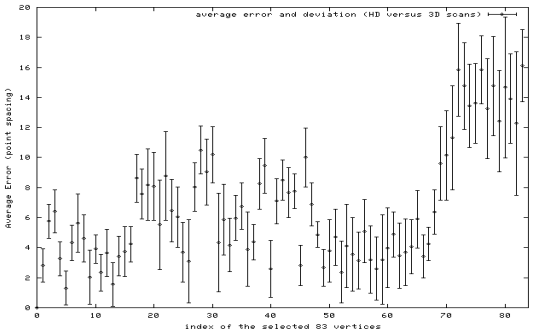


Figure 5: Error statistics of selected 83 testing vertices of a set of models (3D HD-sketch models and 3D scans). Mid-point of each line represents an average error of the vertex (MSE). The standard deviation is shown by the length of the line.

(2) Comparison: 3D MD sketches versus 3D scans

Similar to the above assessment, we also compare the difference between the 3D scans and 3D sketch models created from the MD sketches. Figure 6 shows the error statistics of the 83 points of 100 models. The MSE of the examined points is 6.84 point spacings. The MD sketch models show more accuracy than the HD sketches. The reason is that the MD sketches are derived from the 2D textures of 4DFE, thus, a better alignment can be obtained between the 3D sketches and 3D scans.

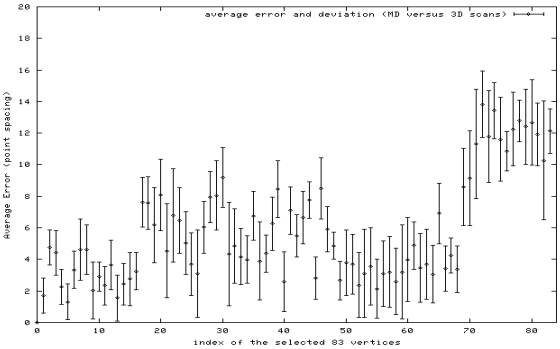


Figure 6: Error statistics of selected 83 testing vertices of a set of models (3D MD-sketch models and 3D scans).

(3) Comparison: 3D HD versus MD

In order to examine the similarity of the HD sketches and MD sketches. We compare the 3D sketch models created from hand drawn sketch image to the 3D sketch models created from machine generated sketch images. The MSE of the 83 points among those models is 3.78 point-spacings, which are very similar to each other. Figure 7 shows the error statistics. The results justify the approximate equivalence of both MD models and HD models, which can be used by subsequence study for face recognition.

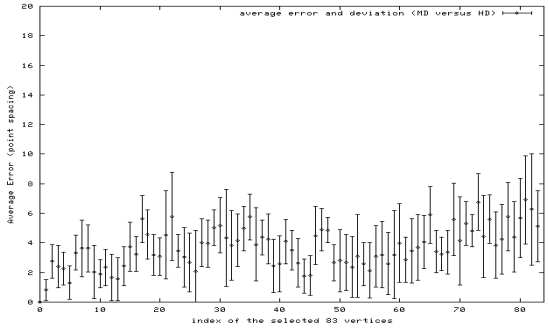


Figure 7: Error statistics of selected 83 testing vertices of a set of models (3D MD-sketch models and 3D HD-sketch models).

4. 3D Sketch Face Recognition

In order to validate the utility of the created 3D sketch models, we conducted experiments of 3D sketch model identification. To do so, we segment each sketch model and each scan model into six component regions. A conventional set of surface label features are used for the spatial HMM classification.

4.1. Component region segmentation

Given a 3D sketch model and the tracked feature points, we can easily segment the facial model into several component regions, such as the eyes, nose and mouth. However, without any assumption of feature points detected on the 3D scans, it is needed to automatically segment facial regions by a more general approach. We developed a simple yet effective approach for 3D facial component segmentation. This approach is general enough to be applicable to other kinds of mesh models, including 3D sketch models. The component segmentation works on the geometric surface directly. It includes mainly two steps:

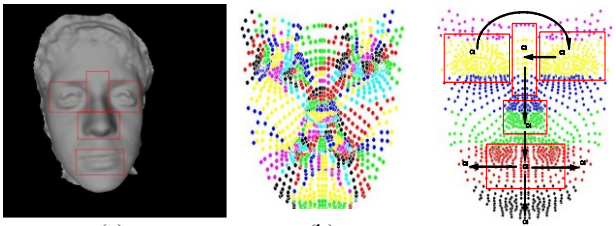
(1) Edge Vertices (EV) determination: Since the edge-feature-rich regions of a 3D facial model lie in regions of eyes, mouth, and nose, we detect edge vertices based on their vertex normals. To do so, a normal mapping scheme is used, where each vertex is assigned by a pseudo-color $\boldsymbol{p}_c = (r, g, b)$. p_c is assigned by the corresponding vertex normal n, i.e., $\boldsymbol{p}_c = \boldsymbol{n} = (n_x, n_y, n_z)$. Emulating the color to grayscale conversion, each vertex is assigned by an attribute value v_a:

$$v_a = 0.299 | n_x | + 0.587 | n_y | + 0.114 | n_z | \quad (1)$$

Initially, an edge vertex can be calculated by iterating through the neighbors of each vertex and calculating the difference (d_a) between v_a of the vertex and the average v_a of its neighbors. Thresholding on d_a values could indicate the edge vertices, however, it may not generate a reliable result. Rather than using a threshold, we apply a clustering method to get two groups of vertices: *edge* and *non-edge vertices*. To do so, a *k*-means clustering algorithm is applied, where $k=2$, to determine the two groups according to the d_a values. Whichever cluster a vertex is closer to (*edge* or *non-edge*), that vertex will be added to the corresponding cluster. This procedure is iterated until the centroids of the clusters remain unchanged.

Once we have obtained these edge vertices, a rectangular bounding box of the face model can be determined by a convex-hull of the edge vertices. We can also find the vertex with the highest z value of those edges, which is the vertex closest to the nose tip. Note that the top one-fourth of the face model is ignored to avoid noise from hair.

(2) Component regions determination: Within the bounding box of a facial model, we start to use four edge vertices as the initial centroids to cluster the edge vertices into four component regions, which are left eye, right eye, nose, and mouth. The initial centroids are determined simply by four edge vertices within the bounding box, which are *top-left*, *top-right*, *mid-bottom*, and *vertex close to nose tip*, respectively. The *k*-means clustering method ($k=4$) is applied using Euclidean distances of edge vertices to the four centroids. The centroids of four regions are updated iteratively until they remain unchanged. As a result, four component regions are detected. Furthermore, the nose bridge region can be determined by the eye and nose boundaries, and the top of the convex hull. The complementary region of the five component regions within the face convex-hull forms the sixth component region. Figure 8(a) shows an example of the resulting segmentation.



(a) (b) (c)

Figure 8: (a) Sample of component regions; (b) Sample of labeled surface of a sketch model, and (c) a component-based HMM based on six component regions.

4.2. 3D component feature representation

3D facial models of both scans and sketches can be characterized by their surface primitive features. This

spatial feature can be classified by eight types: convex peak, convex cylinder, convex saddle, minimal surface, concave saddle, concave cylinder, concave pit, and planar. Such a local shape descriptor provides a robust facial surface representation [17,22]. To label the model surface, we select the vertices of the component regions, and then classify them into one of the primitive labels. The classification of surface vertices is based on the surface curvature computation [22]. After calculating the curvature values of each vertex, we use the categorization method [21] to label each vertex on the range model. As a result, each range model is represented by a group of labels, which construct a feature vector: $G = (g_1, g_2, ..., g_n)$, where g_i represents one of the primitive shape labels, n equals the number of vertices in the component region. An example of the labeled surface is shown in Figure 8 (b).

Due to the high dimensionality of the feature vector G, where each of six component-regions contains vertices ranging from 300 to 700, we use a Linear Discriminative Analysis (LDA) based method to reduce the feature space of each region. The LDA transformation is to map the feature space into an optimal space that is easy to differentiate different subjects. Then, it will transform the n-dimensional feature G to the d-dimensional optimized feature O_G ($d<n$).

4.3. Spatial HMM model classification

In each frame, the 3D facial model is subdivided into six components (sub-regions) $C1$, $C2$, $C3$, $C4$, $C5$, and $C6$, as shown in Figure 8 (c), including regions of the eyes, nose, nose bridge, mouth, and the remaining face. From $C1$ to $C6$, we construct a 1-dimensional HMM which consists of the six states ($N = 6$), corresponding to six regions.

As aforementioned, we transform the labeled surface to the optimized feature space using LDA transformation. Given such an observation of each sub-region, we can train the HMM for each subject. Given a query sketch face model sequence of a length k, we compute the likelihood score for each frame, and use the Bayesian decision rule to decide which subject each frame is classified to. Since we obtain the k results for k frames, we take a majority voting strategy to make a final decision. As such, the query model sequence is recognized as subject Y if Y is the majority result among k frames. This method tracks spatial dynamics of 3D facial sketches, the spatial components of a face gives rise to the spatial HMM to infer the likelihood of each query model. Note that if k is equal to 1, the query sketch model sequence becomes a single sketch model for classification.

5. Experiments of Face Recognition

5.1. 4DFE: 3D sketch (training) vs. 3D sketch (testing)

The 3D sketch models include 3D models created from both HD sketch images and MD sketch images. For each subject, we randomly select 50% of the model frames for training, the remaining 50% of the data for test. For subjects with HD models, we also include half of the data in the training set, and the rest are included in the test set.

For each training sequence of 4DFE, 20 sets of three consecutive frames were randomly chosen for training. Following the HMM training procedure ($k=3$), we generated an HMM for each subject. The recognition procedure is then applied to classify the identity of each input sketch sequence ($k=3$) as the previous section described. Based on the 10-fold cross validation approach, the correct recognition rate is about 95.5%. The ROC curve is shown in Figure 9.

5.2. 4DFE: 3D scans (training) vs. 3D sketches (testing)

In order to validate the utility of the 3D sketches with respect to the 3D scans, we conducted the 3D sketch classification against the corresponding 3D scans. Similar to the above approach, for each subject, we randomly select 20 sets of three consecutive 3D scans for training. Following the HMM training procedure, we generated an HMM for each subject. The recognition procedure is then applied to classify the identity of each input 3D sketch sequence ($k=3$). Based on the 10-fold cross validation approach, the correct recognition rate is about 89.4%. The ROC curve is shown in Figure 9.

5.3. YSU: 3D sketch (training) vs. 3D sketch (testing)

The validation has also been conducted on the 3D sketch models created from YSU sketch database, where sketches from 50 subjects are created. Each subject has five sketches drawn by five artists separately. There are 250 sketches in total. For each subject, we randomly select 4 sketches for training, the remaining one for test. Following the HMM training procedure ($k=1$), we generated an HMM for each subject. The recognition procedure is then applied to classify the identity of each input sketch model ($k=1$). Based on the 10-fold cross validation approach, the correct recognition rate is about 92.6% (see ROC curve of Fig. 9).

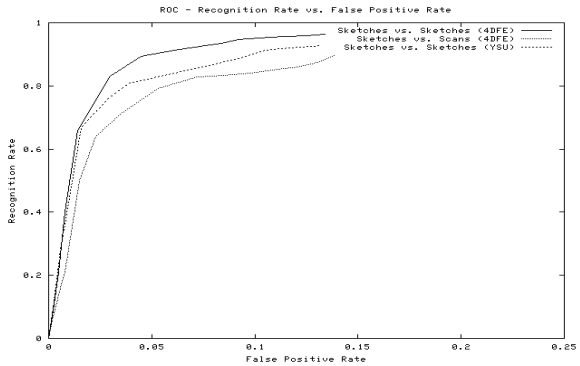


Figure 9: ROC curves of 3D sketch face recognition.

Due to the sketches drawn from different artists in the YSU database, the variation of the sketch styles and the single model query plus single model training of HMM degrades the recognition performance as compared to the 4DFE case (sketches-to-sketches). However, the cross modality matching between 3D scans (training) and 3D sketches (testing) shows the challenge for classification as the 3D sketches created from 2D images may not match well to the ground true data (3D scans). A further study using a more advanced classifier will be investigated in future work.

6. Conclusion and Future Work

This paper addresses the issues of 3D sketch modeling and its validation through 3D sketch recognition using a component based spatial HMM. The quality of 3D sketch models is evaluated by comparing to the corresponding ground-truth 3D scans. We have also shown the approximate equivalence of models between the 3D sketches from HD and MD. Among the test data (both 4DFE database and YSU database), on average a 92% correct recognition rate has been achieved for 3D sketch model identification.

Our future work consists of developing more robust algorithms to detect features in 3D space and improving the current approach for handling noisy models with more expression variations. More advanced 3D geometric surface measurement, representation, matching, and classification will also be investigated [23, 24]. We will also validate our approaches by testing on the larger volume of datasets (e.g., FRGC 2.0 dataset, etc.), and evaluate the performance by comparing it to the 2D sketch recognition approaches [5-8]. In addition, we will further investigate the issues of sketch variations and pose variations from difference sources or different artists, and seek to integrate 3D sketches and 2D sketches in order to improve the face recognition performance.

7. Acknowledgement

This work was supported in part by NYSTAR James Investigator Program, and the NSF under grants IIS-0414029 and IIS-1051103 at Binghamton University, and YSU Research Council Grant No. 08-8.

8. References

[1] L. Gibson, *Forensic Art Essentials: A Manual for Law Enforcement Artists*, Academic Press, 2007.

[2] C. D. Frowd, V. Bruce, A. McIntyre, D. Ross, and *el al.*, "Implementing holistic dimensions for a facial composite system", *Journal of Multimedia*, 1(3), pp. 42-51, 2006.

[3] P. Yuen, and C. Man, "Human face image searching system using sketches", *IEEE Trans. on SMC-A,* 37(4), 2007.

[4] C. Frowd, D. McQuiston-Surrett, S. Anandaciva, and *et al.* "An evaluation of US systems for facial composite production", *Ergonomics*, 50:562-585, 2007.

[5] B. Flare, Z. Li, and A. Jain, Matching forensic sketches to mug shot photos, *IEEE Trans. on PAMI*, 33(3):639-646, March, 2011

[6] X. Wang and X. Tang, "Face Photo-Sketch Synthesis and Recognition," *IEEE Trans. on PAMI*, 31(11), Nov. 2009.

[7] X. Gao, J. Zhong, J. Li, and C. Tian, "Face Sketch Synthesis Algorithm Based on E-HMM and Selective Ensemble," *IEEE Trans. Circuits and Systems for Video Technology*, 18(4):487-496, 2008.

[8] B. Xiao, X. Gao, D. Tao, and X. Li, "A New Approach for Face Recognition by Sketches in Photos," *Signal Processing*, 89(8):1576-1588, 2009.

[9] K. W. Bowyer, K. Chang, and P. J. Flynn. "A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition," *CVIU*, 101(1):1-15, 2006.

[10] P. J Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, and *et al.*, "Overview of the face recognition grand challenge," *IEEE CVPR05*, pp. 947-954, Washington DC, 2005.

[11] Z. Xu, H. Chen, S. Zhu, and J. Luo, A hierarchical compositional model for face representation and sketching, *IEEE Trans. on PAMI*, 30(6):955-969, June, 2008

[12] L. Clarke, M. Chen, and B. Mora, Automatic generation of 3D caricatures based on artistic deformation styles, IEEE Trans. on Visualization & Computer Graphics, 17(6), 2011.

[13] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Surveys*, 35(4), pp. 399-458, 2003.

[14] S. Z. Li and A. K. Jain (editors), *Handbook of Face Recognition*, Springer, 2005.

[15] T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Active appearance models", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23 no. 6. pp. 681-685, June 2001.

[16] L. Yin and K. Weiss. "Generating 3d views of facial expressions from frontal face video based on topographic analysis," *ACM Multimedia,* pp. 360-363, 2004.

[17] P. Flynn and A. Jain, "Surface classification: Hypothesis testing and parameter estimation," *IEEE CVPR* 1988.

[18] D. Terzopoulos and K.Waters. Analysis-synthesis of face image seq. using physical-anatomical models. *IEEE PAMI*, Vol.15, No.6, 1993.

[19] L. Yin, X. Chen, Y. Sun, T. Worm, and M. Reale, A high-resolution 3D dynamic facial expression database, in *IEEE Inter. Conf. on Face and Gesture Recognition,* 2008.

[20] H. Nizami, J. Adkins-Hill, Y. Zhang, *et al*, A biometric database with rotating head videos and hand-drawn face sketches*, IEEE 3rd Inter. Conf. on BTAS*, 2009.

[21] J. Wang, X. Wei, L. Yin, and Y. Sun, 3D facial expression recognition based on primitive surface feature distribution. In *IEEE CVPR 2006*.

[22] Y. Sun, X. Chen, M. Rosato, and L. Yin, "Tracking vertex flow and model adaptation for 3D spatio-temporal face analysis", *IEEE Trans. on SMC – Part A*. 40(3), 2010.

[23] J. Kittler, A. Hilton, *et al.* 3D assisted face recognition: A survey of 3D imaging, modeling and recognition approaches. In *CVPR05 Workshop on A3DISS*, 2005.

[24] X. Lu, A. Jain, and D. Colbry. Matching 2.5D face scans to 3D models. 28(1):31–43, Jan. 2006.