Digital Geometry Image Analysis for Medical Diagnosis

Jiandong Fang Shiaofen Fang Jeffrey Huang Mihran Tuceryan

Department of Computer and Information Science
Indiana University Purdue University Indianapolis
723 W. Michigan St., SL 280
Indianapolis, IN 46202, USA

1-317-274-9731

jfang[sfang,huang,tuceryan]@cs.iupui.edu

ABSTRACT

This paper describes a new medical image analysis technique for polygon mesh surfaces of human faces for a medical diagnosis application. The goal is to explore the natural patterns and 3D facial features to provide diagnostic information for Fetal Alcohol Syndrome (FAS). Our approach is based on a digital geometry analysis framework that applies pattern recognition techniques to digital geometry (polygon mesh) data from 3D laser scanners and other sources. Novel 3D geometric features are extracted and analyzed to determine the most discriminatory features that best represent FAS characteristics. As part of the NIH Consortium for FASD, the techniques developed here are being applied and tested on real patient datasets collected by the NIH Consortium both within and outside the US.

Categories and Subject Descriptors

I.5.2 [**Pattern Recognition**]: Design Methodology – *Feature evaluation and selection.* I.3.5 [**Computer Graphics**]: Computational Geometry and Object Modeling – *Geometric algorithms*.

General Terms

Algorithms, Measurement.

Keywords

Digital geometry, Pattern recognition, Classifier, 3D facial analysis, Medical diagnosis, Fetal alcohol syndrome.

1. INTRODUCTION

Technological advances in recent years have led to an era of information explosion. The enormous proliferations of data and information, in particular multimedia data, have created a great challenge and urgent need for new and more effective analysis techniques for data from multimedia sources. At the same time, biomedical sciences have been increasingly relying on information technology and multimedia data analysis for biomedical research, clinical studies and medical diagnosis.

While traditional 3D medical imaging technologies focus on volumetric images such as CT and MRI, there is a tremendous need in medical applications for the image analysis techniques of surface scans that capture detailed geometric and texture information. Many diseases, particularly neurological diseases, exhibit strong correlations between facial feature anomalies and neurological conditions. 3D data analysis techniques that capture such correlations can potentially provide effective diagnosis tools for medical and clinical studied, especially in the pre-screening process and in early diagnosis of children. This technology becomes more attractive recently as 3D laser scanners become cheaper, faster and more portable with higher resolution. More importantly, many of the later models of laser scanners have been made eye safe, which is essential for human subject studies.

This paper aims to develop advanced geometry analysis techniques for 3D facial images collected using a 3D laser scanner. We are specifically targeting the diagnosis problem of fetal alcohol syndrome disorder (FASD), mainly because of our participation in an NIH project that will allow us to test our theory and techniques with real clinical data. Fetal alcohol syndrome (FAS) is a neurological disorder due to alcohol exposure, and is the most common nonhereditary cause of mental retardation. A number of studies have examined different populations both within the United States and throughout the world to estimate the incidence and prevalence of this devastating syndrome. It is estimated that the prevalence in the general population of FAS is likely to be between 0.5 and 2.0 per 1,000 births. Importantly, studies outside the U.S. have found even higher rates of prevalence in particular geographic regions.

The basic features often associated with prenatal exposure to alcohol include growth deficiencies and neurodevelopmental abnormalities of the central nervous system, and a pattern of various facial anomalies [17]. A powerful FAS diagnosis method is the 4-Digit Diagnostic Code. It uses a numerical scale that measures 4 key diagnostic features of FAS: 1) growth deficiency, 2) FAS facial phenotype, 3) brain damage/dysfunction, 4) gestational alcohol exposure. Each was ranked independently on a 4-point Likert scale. While the combination of the four features constitutes a complete diagnosis criterion, facial anomalies is the only one that can be easily detected and potentially automated. Naturally, facial data analysis approach is becoming an important tool for the development of early diagnosis and treatment for general pre-screening.

Our approach is based on a digital geometry analysis framework that employs 3D geometry analysis techniques on the polygon mesh surfaces of the scanned facial images for feature detection and classification. Digital geometry refers to the digital representation of geometric information of 3D surfaces, and is commonly represented as dense polygon meshes. Such mesh data often comes from 3D surface scanning devices such as laser scanners, or from surface extraction in volumetric images (e.g. CT image). They are usually not in a regular range image format. For example, 3D laser scanning often requires the merging and stitching of multiple scans taken from different angles, at different positions, or simply using multi-pass repeat scans. Thus, traditional approaches such as range image processing and analysis are not feasible with digital geometry data.

We are currently working within an NIH-funded international Research Consortium to develop effective FAS diagnosis solutions. The Consortium has dispatched data collection teams in several parts of the world to collect 3D facial images for both FAS patients and normal people, in various age groups and races. These datasets will be a great source for system training in machine learning and for testing.

2. RELATED WORK

Digital Geometry Processing (DGP) has only recently been recognized as a separate field that dedicated to discrete geometric information processing and representation [16]. While DGP does not directly support analysis applications, it provides the basic algorithms and tools for analysis applications [19].

There has been a large number of literatures on feature identification and extraction in 2D images. In 1991, Gordon [8] proposed a 3D face recognition method using curvature calculation based on range image data obtained from a rotating laser scanner. Tanaka *et al*. [18] extended the concept of free-form curved surface in 3D shape recognition problem to 3D face recognition application. Based on Extended Gaussian Image (EGI) representation, he extracted face signature using principal curvatures and their directions. Beumier and Acheroy [6] used central and lateral facial profiles to generate curvature values as feature vector for face authentication.

Range image is a special case of digital geometry data. Many of the range image processing operations have been extended to digital geometry data by the digital geometry processing community. Segmentation of range images and extraction of 3D features has been intensely studied [4,22]. These segmentation results are then used to do object recognition [1,3,12]. Object recognition strategies based on surface properties such as surface areas and curvatures have also been used [3,4].

Machine learning and pattern recognition provides a computational framework for developing intelligent systems for analyzing geometric patterns. There are two general approaches to identifying optimal subsets of features: 1) strategies which evaluate subsets using abstract measures felt to be relevant to important properties of good feature sets, such as orthogonality, information content and low variance [20,17,7], and 2) strategies which involve actually building a classifier from the feature subset and evaluating its performance on actual classification tasks [11]. The first step in any classification process is to choose candidate discriminatory features and evaluate them for their usefulness. Principal Component Analysis (PCA) is first proposed to define a subspace whose basis vectors are called eigenfaces [13,20]. To use higher order statistics of the training data, Independent Component Analysis (ICA) has gained importance in recent times [2]. Although the details vary, these techniques intend to project the set of training images onto a lower dimensional subspace. Features are now selected from this lower dimensional space and used for the classification module.

3. FACE ALIGNMENT AND MAPPING

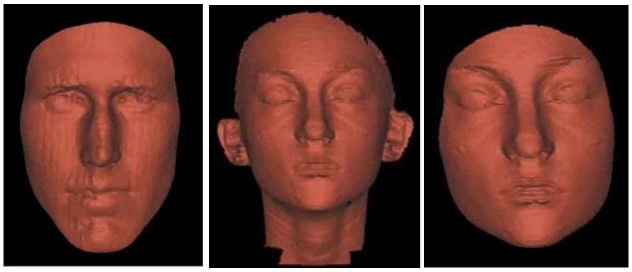
Facial datasets are collected by various groups under different conditions. In order to properly compare 3D facial datasets, all face scans need to be precisely aligned in a common coordinate system. This can be done by using a template face (a standard face dataset) and aligning each new dataset with the template face. The problem of 3D alignment has been studied extensively. The most effective solution is the Iterative Closest Point (ICP) algorithm [5], which computes the optimal transformation by iteratively finding a local minimum of a mean-square distance metric. In our ICP algorithm, the cost function is defined as the sum of the least square distances from the vertices of one face dataset to the other face surface. Powell’s direction set method is employed for optimization [15] as it provides a way to estimate the optimal direction in the *n*-dimensional space during iteration without computing the derivatives of the cost function.

A space encoding method is designed to allow fast minimum distance computation by localizing the search space. We use a simple uniform space subdivision in the volume space of the template dataset to encode the vertices of the template dataset first. A bounding box of the template face is subdivided into an $n \times m \times l$ 3D grid. The vertices of template face as well as the face to be compared with will be registered with the corresponding grids to facilitate fast search and cost function computation in the Powell's iteration algorithm. The establishment of the distance function allows us to define a *distance map* as a mapping from a vertex on the template face to its closest point on the surface of another face dataset that is aligned with the template. Let $\{P_i\}$ be the vertices of the template dataset and $\{Q_i\}$ be the vertices of an aligned face dataset. The *distance map* is defined as:

$$DM(P_i) = \min_{Q} dist(P_i, \{Q_i\})$$

This closest distance can be computed quickly using the neighborhood information encoded with spatial encoding. We can define certain position-related features, such as landmarks, regions and boundary lines, on the template dataset first, and then automatically map these features, by the distance map, to another face dataset that is aligned with the template face.

Distance map may also be used to generate a uniform *cut* for all datasets using the boundary of the template face, so that they all represent approximately the same area of the face. This is necessary in order to carry out some global comparisons among difference datasets, such as curvature histogram comparison, global measurements, and inter-face distances. The *cut* is done by mapping the points on the boundary of the template face to another dataset to form its new boundary, and then cutting the dataset along the boundary into a standard region. Figure 1 shows some alignment and cutting examples.



(a) (b) (c)

Figure 1. Face alignment and cutting. (a) Template face; (b) a new face before alignment; (c) after alignment and cutting.

4. FEATURE COMPUTATION AND EXTRACTION

The most critical component of this research is the detection of salient features that may be used as part of the diagnostic conditions, i.e. separating FAS faces from a non-FAS faces

4.1 Difference Map Visualization

Since there are infinite numbers of possible features that can be extracted from a face dataset, it is imperative to start from a set of features that have the best chance of being "salient". We are experimenting with a *visual data mining* approach which uses a visualization of the "difference" between two known groups to provide visual clues about possible salient features.

The template face is used as a parametric domain to correspond points between faces using a distance map. We are given two groups of datasets for training: one contains known FAS faces, and the other contains known non-FAS faces (*controlled* group). After alignments and standard cuts, The difference map can be obtained by first computing the average displacement vector of the FAS faces at each vertex of the template face. Similar displacement vectors are also computed using the non-FAS faces. The vector differences of the displacement vectors on the template face provide the difference map that can be visualized to provide an intuitive visual impression about the differences of the two groups. This can help identifying areas or relationships that are most "salient" to the FAS problem.

Displacement vectors represent the zero-order differences. 1st order and 2nd order difference maps can also be computed similarly. For instance, first-order difference vector can be computed using surface normal vectors. Second-order difference can be computed using one of the local surface curvatures. The combination of several such difference maps may provide much more information in detecting more accurate salient features.

4.2 Feature Computation

Discrete differential geometry algorithms and digital geometry operators are appllied to extract the potential salient features. The features computed include: curved distances, local curvatures, regional moments, and flatness measures.

Curved distances. Curved distances on a surface provide more accurate measures than Euclidean distances. We employed the Dijkstra's algorithm is used to compute an approximated shortest distance between landmark points over a polygon mesh.

Local Curvatures. Curvatures are local properties of a landmark point on the surface. Curvature computation requires the computation of the first and second fundamental forms of the surface. For digital geometry datasets, discrete operators will need to be applied to approximate the curvatures [14]. Typical curvatures include Principal curvatures κ_1 and κ_2 (maximum and minimum curvatures in the 2 principal directions), Gaussian curvature $\kappa_G = \kappa_1 \times \kappa_2$ (local shape classification) and Mean curvature $\kappa_H = (\kappa_1 + \kappa_2)/2$ (average over all directions).

Both the distance and curvature features are based on landmark points that are defined on the polygon mesh surfaces. These are either natural feature points (e.g. eyes, nose, etc.) or meaningful biological points defined by anthropologists or biologists.

Moment features. Moments of inertia are used to represent global shape information of a region [10]. The $(p+q)^{th}$ order moments of a density distribution function $\rho(x,y)$ are:

$$m_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^p y^q \rho(x,y)dxdy, \quad p,q = 0,1,2,\cdots.$$

The central moments μ_{pq} are defined as

$$\mu_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\bar{x})^p (y-\bar{y})^q \rho(x,y)dxdy, \quad \text{where}$$

$$\bar{x} = m_{10}/m_{00}, \quad \bar{y} = m_{01}/m_{00}$$

Order 0 and 1 central moments are constant for every standardized distribution function. Seven lower order (order 2, and 3) central moments are used in our application:

$$\mu_{20} = m_{20} - \bar{x}^2, \quad \mu_{11} = m_{11} - \overline{xy}, \quad \mu_{02} = \bar{y}^2$$

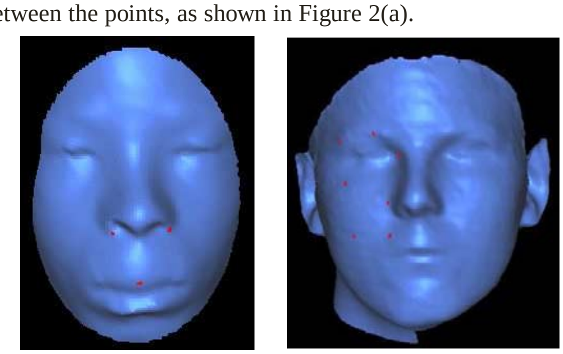
$$\mu_{30} = m_{30} - 3m_{02}\bar{x} + 2\bar{x}^3,$$

$$\mu_{21} = m_{21} - m_{20}\bar{y} - 2m_{11}\bar{x} + 2\bar{x}^2\bar{y},$$

$$\mu_{12} = m_{12} - m_{02}\bar{x} - 2m_{11}\bar{y} + 2\overline{xy}^2,$$

$$\mu_{03} = m_{03} - 3m_{02}\bar{y} + 2\bar{y}^3$$

In our application, a moment feature is extracted from the Philtrum region and the whole face region after cutting. The Philtrum is often considered an important region for FAS, and is defined as the groove between the nose and upper lip, and can be defined by three landmarks and the curved distance paths between the points, as shown in Figure 2(a).



(a) (b)

Figure 2. (a) the Philtrum region; (b) flatness test region.

Flatness features. Flatness feature can be computed by fitting a planar surface to the defined region using a least square fitting. Again, a region can be defined by a sequence of landmark points and the curved distance paths between them, as shown in Figure 2(b). The objective function is defined as the sum of the squares of the distances from the vertices to a parameterized plane.

5. FEATURE SELECTION AND CLASSIFICATION

5.1 Feature Selection

This is to determine an optimal subset of the initial features that has the best discriminatory (diagnostic) power. Various heuristic search strategies such as hill climbing and Best First may be applied. We use a Correlation-based Best First search approach in our implementation. The Best First search starts with an empty set of features and generates all possible single feature expansions. The subset with the highest evaluation is chosen and is expanded in the same manner by adding single features. If expanding a subset results in no improvement, the search drops back to the next best unexpanded subset and continues.

The Correlation-based Feature Selection (CFS) uses a search algorithm to evaluate the merit of the feature subsets. The heuristic by which CFS measures the "goodness" of features takes into account the usefulness of individual features for predicting the class label along with the level of intercorrelation among them. It is based on the hypothesis that "Good feature subsets contain features highly correlated with (predictive of) the class, yet uncorrelated with (not predictive of) each other".

5.2 Feature Classification

Multilayer Perceptron Networks. It is a classifier that uses back-propagation to classify instances. It is a network of simple neurons called perceptrons [9], which compute a single output from multiple real-valued inputs by forming a linear combination according to its input weights and then possibly putting the output through some nonlinear activation function, i.e.

$$y = \varphi(\sum_{i=1}^{n} w_i x_i + b) = \varphi(\mathrm{w}^{\mathrm{T}}\mathrm{x} + \mathrm{b})$$

where **w** denotes the vector of weights, **x** is the vector of inputs, **b** is the bias and φ is the activation function.

A typical multilayer perceptron network consists of a set of source nodes forming the input layer, one or more hidden layers of computation nodes, and an output layer of nodes. The input signal propagates through the network layer-by-layer. MLP networks are typically used in supervised learning problems, and can be solved by a back-propagation algorithm.

Support Vector Machines. It is a method for creating functions from a set of labeled training data [21]. For classification, SVMs operate by finding a hypersurface in the space of possible inputs. This hypersurface will attempt to split the positive examples from the negative examples. The split will be chosen to have the largest distance from the hypersurface to the nearest of the positive and negative examples. Intuitively, this makes the classification correct for testing data that is near, but not identical to the training data.

Sequential Minimal Optimization (SMO) is a fast method to train Support Vector Machine. Training an SVM requires the solution of a large quadratic programming optimization problem. SMO breaks the problem into a series of smallest possible QP problems, which can be solved analytically, and thus avoids a time-consuming numerical QP optimization as an inner loop.

6. RESULTS

We have implemented the techniques described in this paper, and have applied and tested them on some preliminary datasets collected by the NIH consortium: "Collaborative Initiative on Fetal Alcohol Spectrum Disorder (CIFASD)". Total 65 face datasets are collected using a Minolta Vivid 910 laser scanner. Among them, 44 are from patients with FAS, and the other 21 are from a controlled population (Non-FAS).

We used total 49 features as the initial feature vector. They include the mean curvature of the center of the philtrum, the flatness of the left cheek, the length of philtrum, the length of nose, and their ratio, the biocular breadth, the palpebral fissue length, moment features of the entire face, moment features of the philtrum, curvature moment features of the philtrum, etc. Age and gender information are also included.

We first applied Correlation-based Best First search method for feature selection. The algorithm produces a subset of 8 features out of the total 49 features. Two classification methods are then applied on these 8 features: Multilayer Perceptron Network classifier and Support Vector Machine classifier. For a two-way classification, four classification outcomes are possible, which can be displayed in a confusion matrix. Errors occur as either False Positive (FP) or False Negative (FN).

The results are validated using both Test-Set validation and Leave-One-Out cross validation. In Test-Set validation, one-third of the total datasets are randomly selected as the test set. The remaining datasets are used as a training set. Table 1 and Table 2 show the confusion matrices under test-Set validation. The success rate is 95.6522% with the Support Vector Machines classifier and 86.9565 with the Multilayer Perceptron Networks classifier. In Leave-One-Out cross validation, each dataset by itself is used as the test set at a time, and the remaining *n-1* datasets become the training set. This process will repeat *n* times, and the average performance can then be measured. Table 3 and Table 4 show the confusion matrices under Leave-One-Out cross validation. The success rate is 89.23% with both the Support Vector Machines classifier the Multilayer Perceptron Networks classifier. These results are very good with a relatively small training set. We are expecting several hundreds of new datasets to become available in the next few months. With a larger training set and more testing data, we expect the performance to become even better and more reliable.

Table 1. The confusion matrix using Support Vector Machines and Test-Set validation

	Ground Truth	
Prediction	FAS (+)	FAS (–)
FAS (+)	17	1
FAS (–)	0	5

Table 2. The confusion matrix using Multilayer Perceptron Networks and Test-Set validation

	Ground Truth	
Prediction	FAS (+)	FAS (–)
FAS (+)	15	1
FAS (–)	2	5

Table 3. The confusion matrix using Support Vector Machines and Leave-One-Out cross validation

	Ground Truth	
Prediction	FAS (+)	FAS (–)
FAS (+)	42	5
FAS (–)	2	16

Table 4. The confusion matrix using Multilayer Perceptron Networks and Leave-One-Out cross validation

	Ground Truth	
Prediction	FAS (+)	FAS (–)
FAS (+)	41	4
FAS (–)	3	17

7. CONCLUSIONS

A new 3D image analysis approach for a medical diagnosis application is presented. The 3D geometry-based data analysis approach, the automatic 3D feature generation techniques, and its application in medical diagnosis are novel. We believe that this type of 3D analysis problems will become increasingly important as new sensory technologies become more ubiquitous. It can have significant impacts to a wide range of applications both within and outside the biomedical fields.

In the future, we would like to investigate more sophisticated geometric features, including frequency domain features and higher-degree surface fitting features. We would also like to study a larger array of classifiers to determine the optimal classification methods for this type of 3D data analysis problems. A larger scale experiment with additional datasets will certainly be very desirable. We will report the results of this research in real clinical tests, whenever it becomes available.

8. ACKNOWLEDGMENTS

This research is supported in part by the NIH grant 5U24AA014809-02. We would like to thank Prof. Ed Riley (SDSU), Prof. Rick Ward (IUPUI), Prof. Tatiana Foroud and Dr. Elizabeth Moore (IU Medical Center) for their collaborations and help in making the datasets available. We also like to thank Jeff Rogers for providing excellent technical support to this project.

9. REFERENCES

[1]. Arman, F., and J. K. Aggarwal (1993). "Model-based object recognition in dense-range images—a review." *ACM Computing Surveys* 25(1): 5-43.

[2]. Barlett, M. and T. Sejnowski, "Independent components of face images: A representation for face recognition," *Proc. 4th Annual Joint Symposium on Neural Computation,* Pasadena, CA. May 17, 1997.

[3]. Besl, P. and R. C. Jain (1985). "Three-dimensional Object Recognition." ACM Computing Surveys 17(1): 75-145.

[4]. Besl, P. J. and A. Jain (1988). "Segmentation through variable-order surface fitting." *IEEE Transaction on Pattern Analysis and Machine Intelligence* 10(2): 167-192.

[5]. Besl, P. and N. McKay (1992), "A Method for Registration of 3D Shapes", *IEEE Trans. PAMI*, vol. 14, no. 2, 239-256.

[6]. Beumier, C. and M. Acheroy (2000), Automatic 3D face authentication, *Image & Vision Computing*, 18(4), 315-321.

[7]. Comon, P. (1994). "Independent Component Analysis: A New Concept?" *Signal Processing* 36: 287-314.

[8]. Gordon, G. (1991), Face recognition based on depth maps and surface curvature, in *SPIE proc. on Geometric Method in Computer Vision*, vol.1570, pp.234-247.

[9]. Haykin, S, Neural Networks - A Comprehensive Foundation, 2nd ed. Prentice-Hall, 1998.

[10]. Hu, M. K. (1962). "Visual Pattern Recognition by Moment Invariants." *IRE Trans. on Information Theory* 8: 179-187.

[11]. Huang, J. and H. Wechsler, Visual Routines for Eye Location Using Learning and Evolution, *IEEE Trans. on Evolutionary Computation (EC),* 4 (1), pp.73-82, 2000.

[12]. Jain, A. K. and R. Hoffman (1988). "Evidence based recognition of 3D objects." *IEEE Transaction on Pattern Analysis and Machine Intelligence* 10(6): 783-802.

[13]. Jolliffee, I.T., "Principal Component Analysis", New York, Springer Verlag, 1986.

[14]. Meyer, M., M. Desbrun, P. Schroder, and A. Barr, "Discrete Differential Geometry Operators for Triangulated 2-Manifolds", VisMath '02, Berlin (Germany), 2002.

[15]. Powell, M., "An efficient method for finding the minimum of a function of several variables without calculating derivatives", *Computer Journal*, vol. 7, 155-162, 1964.

[16]. Schroder, P. and W. Sweldens, "Digital Geometry Processing". In Sixth Annual Symposium on Frontiers of Engineering, 41–44, 2001.

[17]. Swets, D. L. and J. Weng (1996). "Using Discriminant Eigenfeatures for Image Retrieval." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18(8): 831-836.

[18]. Tanaka, H.T., M. Ikeda, and H. Chiaki (1998), Curvature-base face surface recognition using spherical correlation – principal directions for curved object recognition, *Proc. Int. Conf. on Automatic Face Gesture Recognition*, 372-377.

[19]. Taubin, G. (2000). "Geometric Signal Processing on Polygonal Meshes." *Eurographics 2000, State of the Art Report (STAR)*, INTERLAKEN, Switzerland.

[20]. Turk, M. and A. Pentland (1991). "Eigenfaces for Recognition." *J. of Cognitive Neuroscience* 3(1): 71-86.

[21]. Vladimir N. Vapnik, The Statisitcal Learning Theory. Springer, 1998.

[22]. Zhao, C., D. Zhao, et al. (1996). "Simplified Gaussian and Mean Curvatures to Range Image Segmentation." Proceedings of the 1996 International Conference on Pattern Recognition (ICPR' 96) IEEE.