Published in final edited form as:
Proc IAPR Int Conf Pattern Recogn. 2010 ; 2010: 2444–2447. doi:10.1109/ICPR.2010.598.

The Use of Genetic Programming for Learning 3D Craniofacial Shape Quantifications

Indriyati Atmosukarto,
Department of Computer Science and Engineering, University of Washington, Seattle, WA, USA, 98105

Linda G. Shapiro, and
Department of Computer Science and Engineering, University of Washington, Seattle, WA, USA, 98105

Carrie Heike
Seattle Children's Hospital, Craniofacial Center

Indriyati Atmosukarto: indria@cs.washington.edu; Linda G. Shapiro: shapiro@cs.washington.edu; Carrie Heike: carrie.heike@seattlechildrens.org

Abstract

Craniofacial disorders commonly result in various head shape dysmorphologies. The goal of this work is to quantify the various 3D shape variations that manifest in the different facial abnormalities in individuals with a craniofacial disorder called 22q11.2 Deletion Syndrome. Genetic programming (GP) is used to learn the different 3D shape quantifications. Experimental results show that the GP method achieves a higher classification rate than those of human experts and existing computer algorithms [1], [2].

Keywords

3D shape quantification; genetic programming

I. Introduction

Evaluation of the facial variation observed in genetic conditions has traditionally relied on clinical description and select measurements taken directly with calipers. Three-dimensional surface imaging systems now allow for craniofacial assessment through evaluation of 3D shape. Quantification of the craniofacial variation may aid investigations into the genotype-phenotype associations in conditions that affect the head and neck.

This paper describes the use of genetic programming for learning 3D shape quantification and investigates its application in quantifying the dysmorphologies associated with 22q11.2 deletion syndrome (22q11.2DS), a disorder associated with a 1.5–3MB deletion on chromosome 22 that occurs in 1:4000 individuals. Phenotypic variability is a hallmark of this condition, and over 180 features have been associated [3]. The craniofacial features in this condition have been well-described, and include a bulbous nasal tip, tubular appearance to the nose, retrusive chin, among others. A few studies have incorporated direct anthropometric measurements [4], while two automated methods have been proposed.

Boehringer et al. [5] applied PCA on the Gabor wavelet transformation of 2D photographs of individuals, while Hammond et al. [6] used the dense surface model and average face approach to classify individuals with 22q11.2DS.

The remainder of this paper is organized as follows. First, the dataset used to develop and test the methodology is described. The overall methodology is described next. In the experimental results section, the classification and quantification experiments for nine facial abnormalities associated with 22q11.2DS are described and analyzed. Finally, conclusions are provided.

II. Datasets

Our 22q11.2DS dataset consists of the 3D craniofacial surface meshes obtained from the 3dMD imaging system [7]. The obtained 3D face models were pre-processed and pose-normalized using an automated system [8]. The dataset contains 86 individuals: 43 cases and 43 controls. We asked three dysmorphologists to rate the facial features observed on a 4-point scale. This study focuses on nine facial features: midface hypoplasia, tubular nose, bulbous nasal tip, prominent nasal root, small nasal alae, retrusive chin, small mouth, open mouth, and downturned mouth. The groundtruth for each facial feature of an individual in the dataset is the logical OR of the labels from each of three craniofacial experts.

III. Methodology

Our quantification learning framework begins with the selection of the facial region that is most pertinent to a given facial abnormality. Features in the form of a 2D histogram of azimuth and elevation angles of surface normals are then extracted from the selected facial region. The framework continues by selecting features from this region using Adaboost [9]. Genetic programming is then used to combine the selected features and produce the quantification of a given facial abnormality.

A. Facial Region Selection

To study the different facial abnormalities, we focus on the facial region that is most pertinent to that abnormality. The nine facial abnormalities cover three different areas of the face: midface, nose, and mouth. The nose region is extracted using a trapezium bounding box that covers the nose area of the face (Figure 1(a)). The mouth region is extracted using a rectangular bounding box that covers the mouth area of the face (Figure 1(b)), while the midface region is extracted using a rectangular bounding box that covers the middle portion of the face (Figure 1(c)).

B. 2D Histogram of Azimuth-Elevation Angles

Once the facial region pertinent to a given facial abnormality has been selected, features are extracted from the selected region. Our methodology for representing 3D facial shape uses 2D histograms of the azimuth and elevation angles of surface normal vectors of the 3D points in the region [2]. Figure 2(a) shows the selected midface region of an individual in the dataset, while Figure 2(b) shows the 8×8 2D histogram of the region displayed on a color map. High histogram bin values are represented by warm colors (red, orange, yellow), while

low bin values correspond to cool colors (blue, cyan, green). The histogram size was empirically selected.

The vector of histogram bin values is treated as a feature vector and used for classification. Rather than using a linear combination of all the histogram bins, our methodology first determines the bins that are most important and most discriminative in classifying the different facial abnormalities. It then applies genetic programming to find the best way to combine the discriminative histogram bin values in order to generate a quantification for each of the different facial abnormalities. Genetic programming (GP) is a method that follows the theory of evolution by evolving individual computer programs following the survival of the fittest approach. GP has been used in computer vision by Perez et al. [10], Torres et al. [11], and Zhang et al.[12].

C. Feature Selection

To determine the histogram bins that are most discriminative in classifying and quantifying given facial abnormalities, Adaboost learning was used to select the bins that optimized classification performance. The Adaboost algorithm obtains a strong classifier by combining a set of weak classifiers with different weights to minimize the classification error. In our experiments, we used the WEKA [9] implementation of Adaboost learning and selected the decision stump as the weak classifier, because of its high classification rate in our preliminary experiments. In addition, we selected a maximum of ten most discriminative histogram bins for each of the different facial abnormalities. The values of the selected bins were concatenated into a feature vector for use in both classification and disease quantification. Figure 3(a) shows the selected bin values of the 2D histogram in Figure 2(b) highlighted in red. Note that both low and high-valued bins were selected. Figure 3(b) shows the projection of all of the selected bins in Figure 3(a) back onto the face. Interestingly, though midface hypoplasia occurs on both sides of the face, the algorithm selected bins mostly from the right side of the face.

D. Feature Combination

The goal of our work is to quantify the different shape variation that manifests in the different facial abnormalities. The Genetic Programming (GP) methodology combines and evolves the values of the selected discriminative histogram bins to produce mathematical expressions that quantify the shape variation of the different facial abnormalities. The GP approach starts with assigning the measure of performance, commonly called the “fitness test” to an individual from a population. In this work, the F-measure is used as the fitness function to measure the presence of a given facial abnormality. The F-measure is defined as

$$F(prec, rec) = \frac{2 \times (prec \times rec)}{prec + rec}$$

where *prec* and *rec* are the precision and recall metrics at a given threshold. The final precision and recall metrics are calculated at the threshold that maximizes the area under the Receiver Operating Characteristics (ROC) curve. The best individual in the population is the individual with the maximum F-measure value.

In genetic programming, the genes of an individual program form a tree-like structure with two different types of genes: functions and terminals. In our approach, the terminal sets are

the selected histogram bins and the branch nodes are the functions used to combine their values. The genetic programming method evolves the individuals through a number of set iterations and selects the individual with the maximum F-Measure. The approach produces the tree-structure of the best individual, which can be translated into the best mathematical expression.

IV. Experimental Results

For our experiments, we used the MATLAB implementation of the GP approach called GPLAB [13]. We empirically tested four different GP function sets and analyzed their classification performance for each of the facial abnormalities. The four function sets were: (1) {+, −, *, *min, max*}, (2) {+, −, *, *min, max, sqrt*, log_2, log_{10}}, (3) {+, −, *, *min, max*, 2*x*, 5*x*, 10*x*, 20*x*, 50*x*, 100*x*}, and (4) {+, −, *, *min, max, sqrt*, log_2, log_{10}, 2*x*, 5*x*, 10*x*, 20*x*, 50*x*, 100}. The last two sets were chosen to introduce weighting, which is not an explicit part of GP. Table I shows the best and second best performing GP function sets for each of the facial abnormalities and their respective F-measures. The simplest function sets, (1) and (2), were selected as the best or second best performing function sets for all nine conditions. Figure 4 shows the tree structure of the best performing GP function set for quantifying midface hypoplasia. Preliminary experiments using random subsets of the dataset for cross validation testing showed that the GP approach was not overfitting the training data, hence the rest of the experiments were conducted using the whole dataset for generating the tree structures.

The second experiment was designed to measure the performance of the different facial region descriptors in classifying individuals to a given facial abnormality. The goal of the experiments is to classify each individual in the dataset as either affected or unaffected by a given facial abnormality. We tested two different classifiers, Adaboost and SVM, and are reporting on the best performing classifier for each facial abnormality. Table II shows the classification performance for the nine different facial abnormalities for full region histograms, selected bins only, and the GP expressions over selected bins. For all facial abnormalities, using the genetic programming approach to quantify and then classify the facial abnormalities performed the best.

In our third experiment, we compared our region-based results to two global approaches to representing the face and the various facial features. The first comparison is to a previous work of representing the whole face using a global saliency map [2]. The second comparison is to a global approach of representing the whole face, instead of only a specific facial region, using a 2D histogram of azimuth-elevation angles [1]. Table III shows the comparison results. It can be seen that using genetic programming quantification to represent the facial regions achieves a higher classification performance than the global approaches.

The last experiment was to analyze how the learned genetic programming quantification performed in predicting 22q11.2 deletion syndrome for individuals in the dataset. In this experiment, the best performing mathematical expression obtained by the genetic programming quantification for each of the nine facial abnormalities was evaluated, and the

resulting values concatenated into a feature vector of dimension nine. The resulting feature vector was then used to classify the individuals as either affected or unaffected by 22q11.2DS. The F-measure using this quantification feature vector was 0.709 with SVM and 0.721 with Adaboost respectively. However, evolving the resulting concatenated feature vector with dimension nine using the genetic programming approach obtained a much higher classification performance of 0.821. Table IV compares the F-measures for predicting 22q11.2 deletion syndrome. The top three rows use the 9-dimensional vector containing the quantifications of the nine separate abnormalities. Results for the global saliency map [2], which uses curvature as its low-level feature, as a whole and with Adaboost-learning selected bins, are shown next. Results for the global 2D histogram of azimuth and elevation angles (with dimensions 24×24) [1] as a whole and with Adaboost-learning-selected bins, are given next. Using genetic programming to evolve the Adaboost-learning-selected bins of both the global saliency map and the global 2D histogram of azimuth and elevation angles further improved the F-measures. Finally, the median score obtained by our three human experts is given for comparison. All of the automatic results are improvements over the median of human experts.

V. Conclusion

This paper has discussed a new methodology for learning 3D shape quantification using a genetic programming approach and investigated its application in analyzing facial shape variations in individuals with 22q11.2DS. Experimental results show that using genetic programming to quantify the different facial abnormalities has the highest classification rate of any tested approach. Although the focus of this work is on individuals with 22q11.2DS, we are investigating its application in quantifying other shape conditions associated with craniofacial dysmorphologies such as deformational plagiocephaly and cleft lip/palate.

Acknowledgments

This research was supported by the National Science Foundation Grant DBI-0543631, NIH/NIDCR Grant U01DE020050-01, and NIH Grant K23-DE017741.

References

1. Atmosukarto I, Shapiro L, Starr JR, Heike CL, Collett B, Cunningham ML, Speltz M. 3d head shape quantification for infants with and without deformational plagiocephaly. CPCJ. 2009
2. Atmosukarto I, Wilamowska K, Heike C, Shapiro LG. 3d object classification using salient point patterns with application to craniofacial research. Pattern Recognition. 2009
3. Robin N, Shprintzen R. Defining the clinical spectrum of deletion 22q11.2. J Pediatr. 2005; 147:90–96. [PubMed: 16027702]
4. Guyot L, Dubuc M, Pujol J, Dutour O, Philip N. Craniofacial anthropometric analysis in patient with 22q11 microdeletion. Am J Med Genet. 2001; 100:1–8. [PubMed: 11337741]
5. Boehringer S, Vollmar T, Tasse C, Wurtz RP, Gillessen-Kaesbach G, Horsthemke B, Wieczorek D. Syndrome identification based on 2d analysis software. Eur J Hum Gen. 2006; 14:1082–1089.
6. Hammond P. The use of 3D face shape modeling in dysmorphology. Arch Dis Child. 2007; 92:1120–1126. [PubMed: 18032641]
7. 3dMD, “http://www.3dmd.com.”
8. Wilamowska K, Shapiro LG, Heike CL. Classification of 3D face shape in 22q11.2 deletion syndrome. IEEE ISBI. 2009

9. Witten, IH.; Frank, E. Data Mining: Practical machine learning tools and techniques. 2. Morgan Kaufmann; San Fransisco: 2005.

10. Perez CB, Olague G. Evolutionary learning of local descriptor operators for object recognition. GECCO. 2009

11. da R, Torres S, Falcao AX, Goncalves MA, Papa JP, Zhang B, Fan W, Fox EA. A genetic programming framework for content-based image retrieval. Pattern Recognition. 2009; 42:283–292.

12. Zhang M, Bhowan U, Ny B. Genetic programming for object detection: A two-phase approach with an improved fitness function. Electronic Letters on Computer Vision and Image Analysis. 2007; 6(1):27–43.

13. GPLAB. http://gplab.sourceforge.net/

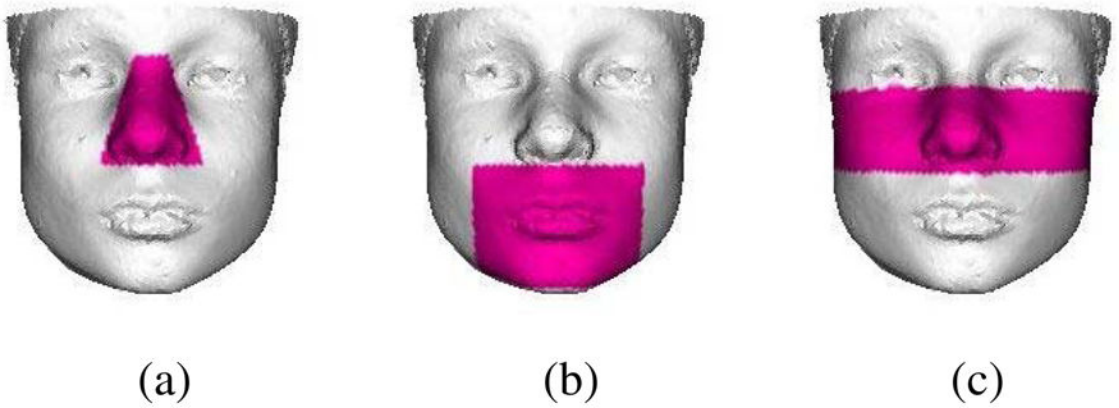

Figure 1.
Different facial regions are selected depending on the facial abnormality being studied: (a) nose, (b) mouth, and (c) midface.

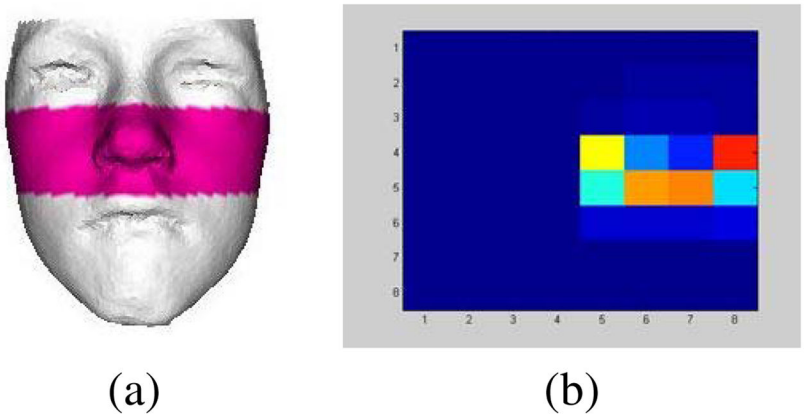

Figure 2.

(a) Selected midface facial region of an individual in the dataset. (b) The constructed 2D histogram of the azimuth and elevation angles of the surface normals of the points in the selected facial region.

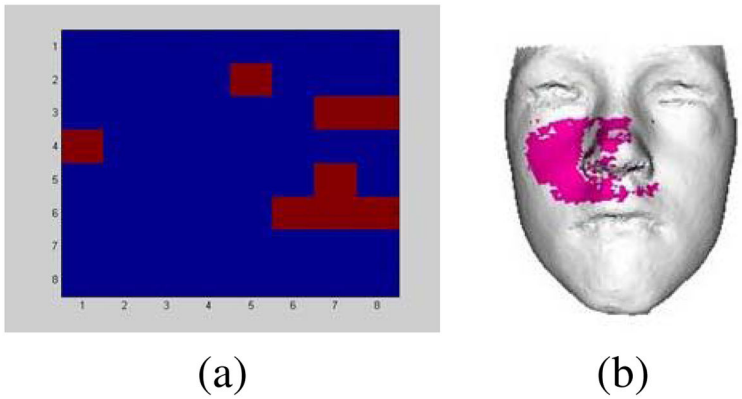

Figure 3.

(a) Adaboost learning selects the most discriminative histogram bins for classifying midface hypoplasia (highlighted in red). (b) Positional information about which points in the selected region contribute to the selected bins for classifying midface hypoplasia (highlighted in pink).

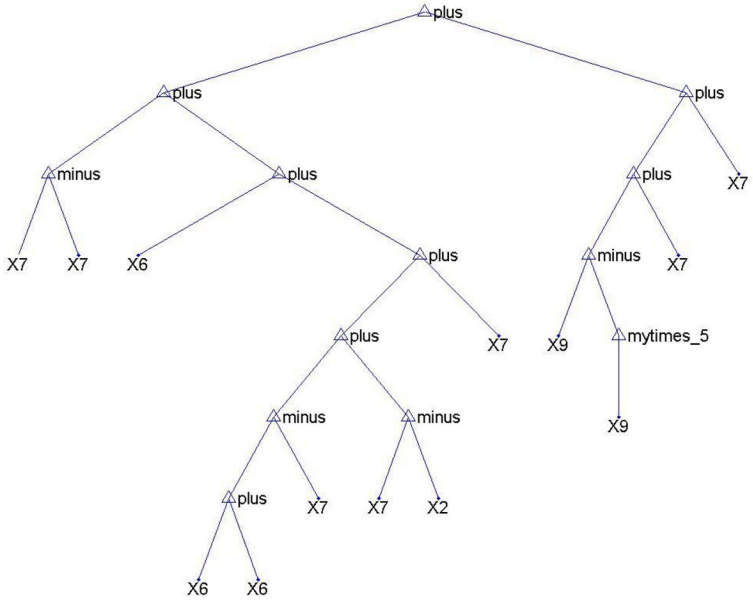

Figure 4.

The output of the GP Tree structure of the best performing function set for quantifying midface hypoplasia. The equivalent mathematical expression can be written as $(X7 - X7) + (X6 - (((X6 + X6) - X7) + (X7 - X2)) + X7) + (X9 - 5X9) + X7 + X7)$

Table I

Best (F_1) and second best (F_2) performing GP function set for each facial abnormality and their respective F-measures.

Facial anomaly	F_1	F-meas.	F_2	F.meas
Midface Hypoplasia	3	0.8527	1	0.8393
Tubular Nose	4	0.8813	2	0.875
Bulbous Nasal Tip	1	0.8545	2	0.875
Prominent Nasal Root	1	0.8667	3	0.8103
Small Nasal Alae	1	0.8846	3	0.8571
Retrusive Chin	2	0.8000	1	0.7952
Open Mouth	2	0.9714	1	0.9444
Small Mouth	4	0.7750	2	0.7568
Downturned Mouth	1,3,4	0.8000	2	0.7797

Table II

Classification performance using various facial shape descriptors to classify the nine different facial abnormalities.

Facial abnormality	Rgn Hist	Slct Bins	GP
Midface hypoplasia	0.697	0.721	**0.853**
Tubular nose	0.701	0.776	**0.875**
Bulbous nasal tip	0.617	0.641	**0.855**
Prominent nasal root	0.704	0.748	**0.867**
Small nasal alae	0.733	0.801	**0.8846**
Retrusive chin	0.658	0.713	**0.8000**
Open mouth	0.875	0.889	**0.9714**
Small mouth	0.694	0.725	**0.7750**
Downturned mouth	0.506	0.613	**0.8000**

Table III

Comparing the Genetic Programming quantification approach to the global approaches.

Facial abnormality	GP	Sa. Map	Glbl 2D Hist
Midface hypoplasia	**0.853**	0.674	0.744
Tubular nose	**0.875**	0.628	0.709
Bulbous nasal tip	**0.855**	0.616	0.639
Prominent nasal root	**0.867**	0.663	0.658
Small nasal alae	**0.8846**	0.779	0.675
Retrusive chin	**0.8000**	0.628	0.674
Open mouth	**0.9714**	0.707	0.875
Small mouth	**0.7750**	0.581	0.752
Downturned mouth	**0.8000**	0.566	0.630

Table IV

Classification performance in predicting 22q11.2 Deletion Syndrome.

Method	F-measure
Quantification vector with SVM	0.709
Quantification vector with Adaboost	0.721
Quantification vector with GP	0.821
Global saliency map [2]	0.764
Selected bins of global saliency map	0.9
Global 2D histogram [1]	0.79
Selected bins of global 2D histogram	0.9
Selected bins of global saliency map with GP	0.96
Selected bins of global 2D histogram with GP	0.92
Expert's median	0.68