

QUANTIFICATION OF 3D FACE SHAPE IN 22Q11.2 DELETION SYNDROME

Katarzyna Wilamowska[†], Linda Shapiro[†] Fellow, IEEE, Carrie Heike[‡]

[†]Department of Computer Science and Engineering, University of Washington;

[‡]Department of Pediatrics, University of Washington; Craniofacial Center, Seattle Children's Hospital

ABSTRACT

Given a set of labeled 3D meshes acquired from stereo imaging of heads, the goal of this research is to develop a successful methodology for discriminating between 22q11.2 Deletion Syndrome affected individuals and the general population and for quantifying the degree of dysmorphology of facial features. Although many approaches for such discrimination exist in the medical and computer vision literature, the goal is to develop methods that focus on shape-based morphological differences of facial features.

Index Terms— biomedical image processing, biomedical measurements, medical expert systems, medical information systems

1. INTRODUCTION

22q11.2 deletion syndrome (22q11.2DS) is a common genetic condition with an estimated prevalence between 1:2000 and 1:6000 live births in the United States [1]. It is associated with more than 180 clinical features, including over 25 dysmorphic craniofacial features. No single feature occurs in every individual with 22q11.2DS, and there are no individuals who have most or all of the clinical features. Additionally, the expression of a specific feature may be quite varied. For example, palatal anomalies can range from an obvious cleft palate to dysfunction of the palatal muscles in a normal-appearing palate [2]. Finally, while many individuals with 22q11.2DS have a characteristic facial appearance, even experts have difficulty in diagnosing 22q11.2DS from frontal facial photographs alone [3].

Quantification of the facial dysmorphology observed in 22q11.2DS provides an objective way of characterizing subtle anomalies. Demonstration of an association between specific facial features and genetic variants will allow for better understanding of the etiology of craniofacial malformations and pathogenesis of 22q11.2DS, provide insight into the genetic contribution to typical craniofacial development, and facili-

tate discovery of an association between facial features and other physical anomalies (such as cardiac defects).

In this paper we describe a new methodology for quantifying the phenotypic variations associated with 22q11.2DS. Our method takes a 3D surface image of a head, and applies noise cleaning and pose normalization and then extracts three different representations of the 3D data: 3D snapshots, 2.5D depth images, and curved lines. Classification is performed on feature vectors comprised of the principal components analysis (PCA) coefficients of these representations. This paper describes our methodology and the classification experiments we have run.

2. RELATED WORK

Traditionally, the clinical approach to identify and study an individual with facial dysmorphism has been through physical examination combined with craniofacial anthropometric measurements [4]. These measurements are based on anatomic landmarks, which are located by visual inspection and/or palpation of the underlying skull.

Newer methods of craniofacial assessment involve using data from computerized tomography, magnetic resonance imaging, ultrasound studies, and stereoscopic imaging [5]. As the information in these data types is often hand measured, or at least hand-labeled, human effort in the use of these newer systems remains significant.

With respect to 22q11.2DS, craniofacial anthropometric measurements prevail as the standard manual assessment method. Automated methods for 22q11.2DS analysis are limited to just two. Boehringer et al. [6] applied a Gabor wavelet transformation to 2D photographs of individuals with ten different facial dysmorphic syndromes. The generated data sets were then transformed using principal component analysis (PCA) and classified using linear discriminant analysis (LDA), support vector machines (SVM), and k -nearest neighbors. The best prediction accuracy for 22q11.2DS was found to be 96% using LDA, dropping to 77% when using a completely automated system.

The second automated method is the Dense Surface Model (DSM) approach [7], which aligns training samples according to point correspondences and is thus able to produce an “average” face for each population being studied.

This work was supported by the National Science Foundation under Grant Number DBI-0543631, by the National Institute of Dental and Craniofacial Research under Grant Number 5K23DE17741-2, by the General Clinical Research Center under Grant Number # M01-RR 00037 and by the American Academy of Pediatrics Section on Genetics and Birth Defects.

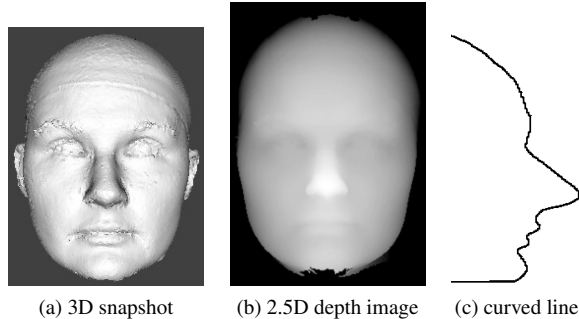


Fig. 1: Three data representations used in our experiments.

Once the average is computed, PCA is used to represent each face by a vector of coefficients. Multiple classifiers were tested, and the best sensitivity and specificity results for 22q11.2DS (0.83 and 0.92, respectively) were obtained using support vector machines [8]. Studying discrimination abilities of local features (face, eyes, nose, mouth) achieved a correct 22q11.2DS classification rate of 89% [9]. Neither Boehringer’s method or the Dense Surface Models methods are fully automatic; both require manual landmark placement.

3. METHOD

The 3D data used in our study was collected by the Craniofacial Center of Seattle Children’s Hospital (SCH) using the 3dMDcranial™ imaging system. All participants had a laboratory-confirmed 22q11.2 deletion. The SCH Institutional Review Board approved study procedures. The demographics of the study participants were age 0.8 - 39 (median 4.75) years, 51% female. Due to human subjects requirements, the data used in this research are 3D meshes without facial texture maps. An automated system to align the pose of each mesh was developed, using symmetry to align the yaw and roll angles and a height differential to align the pitch angle.

Symmetry between the left and right sides of face was used to determine the most central position of the face, with a misalignment error of 1%. Although faces are not truly symmetrical, the pose alignment procedure can be cast as finding the angular rotations of yaw and roll that minimized the difference between the left and right side of the face. The original 3D double precision mesh was interpolated to a 2.5D ordered grid. The resulting image was divided down the middle producing a left true image and a right mirrored image. These images were then overlaid and the difference was calculated by

$$Diff = \sum_{y=0}^{height} \sum_{x=0}^{width/2} |Image_{x,y} - Image_{width-x-1,y}| \quad (1)$$

Table 1: Line positions. Position (125,150) is the location of the nose tip in each individual.

Line type	# lines	Line coordinates
Vertical	1	125 (middle of width)
Vertical	3	75, 125, 175
Vertical	5	75, 100, 125, 150, 175
Vertical	7	50, 75, 100, 125, 150, 175, 200
Horizontal	1	150 (below middle of height)
Horizontal	3	100, 150, 200
Horizontal	5	100, 125, 150, 175, 200
Horizontal	7	75, 100, 125, 150, 175, 200, 225

The pitch of the head was aligned by minimizing the difference between the height of the chin and the height of the forehead. Note that if the rotation angle for pitch is set too wide, the top of the head can be selected as the optimal solution causing misalignment in 15% of the cases. One iteration of alignment for yaw/roll and pitch will not always yield a final alignment. This is solved by a second iteration of both yaw/roll and pitch alignment, with a much smaller search space; often 5° is sufficient.

3.1. Data Representation

Three representations (Fig. 1) were chosen: (1) frontal snapshots of the 3D meshes, (2) 2.5D depth images, and (3) 1D curved line segments. Snapshots of the 3D images were used as a starting point, while 2.5D depth images were used to retain the 3D aspect of the original mesh. Lastly, curved line segments obtained from the depth images were used to determine if there was any *affected* signal in the subsampled face data. For each individual, in all three data types, the information was normalized to the same height and width, and image size was set to 250x380 pixels.

The motivation for using 3D snapshots came from the eigenfaces [10] approach in computer vision. After neutral pose alignment, a set of frontal photographs of the 3D meshes was generated (Fig. 1a) using the visualization library VTK [11].

2.5D images are represented as pixels (Fig. 1b), where the original mesh data is rasterized to an integer-precision structured grid with the highest Z -value (the tip of the nose) placed at high illumination. The final width and height of each face is given by the X and Y axes, with final depth of the face given by Z . In order to properly scale in the Z direction, all of the data was hand clipped at the ears. For the X axis normalization, the face of each individual was scaled to be exactly 200 units wide; the Y axis information was unchanged.

Using the 2.5D images, curved lines can be extracted that may turn out to be descriptive of faces. For example, a vertical line down the middle of the face (Fig. 1c) becomes a wave-

Table 2: Choosing an appropriate data set. All data used is in 3D snapshot with ear cutoff threshold format classified using Naive Bayes.

Data Set	ALL	A106	AS106	W86	WR86
F-measure	0.53	0.65	0.66	0.68	0.60
Precision	0.56	0.74	0.78 ◦	0.82 ◦	0.71
Recall	0.52	0.60	0.61	0.62	0.56
Accuracy	0.75	0.69	0.71	0.74	0.66

◦, • statistically significant improvement or degradation

form (depth as a function of height) that can be analyzed. As seen in Table 1, four versions for both vertical and horizontal lines were selected for signal testing. Odd numbers of lines were used to maintain symmetry in the data. Finally, a combination of the horizontal and vertical lines was used to create grids of sizes 1x1, 3x3, 5x5, and 7x7.

4. EXPERIMENTS AND RESULTS

Each of the three data representations was transformed using Principal Components Analysis (PCA), which converted its original representation into a feature vector comprised of the coefficients of the principal component vectors. Since there were 189 individuals in our full data set, this allowed for a maximum 189 attribute representation for the entire data set. The attributes were assessed as to their ability to distinguish between *affected* and *control* individuals. The WEKA suite of classifiers [12] was used for all classification experiments. 10-fold cross validation was used for all classifiers and each training/testing set was executed ten times, for a result of 100 runs per data set per classifier.

4.1. Data Set Selection

Although testing on a balanced set is common practice in data mining, our data set only included a small number of *affected* individuals. The full data set included 189 individuals (53 affected, 136 control). Set *A106* matched each of the 53 affected individuals to a control individual of closest age. Set *AS106* matched each of the 53 affected individuals to a control individual of same sex and closest age. Set *W86* matched a subset of 43 affected self-identified white individuals to a control individual of same ethnicity, sex and closest age. Set *WR86* matched the same 43 affected white individuals to a control individual of same ethnicity, sex and age, allowing repeats of controls where same-aged subjects were lacking.

The *W86* dataset was chosen as the most appropriate to this work for the following reasons (Table 2). The racial and ethnic background influences the morphology of the face much more significantly than the effects of 22q11.2DS, causing a source of noise for the ethnically mixed sets (*A106* and

Table 3: Attribute selection of PCA vectors for data separation for sex, age and affected. Each attribute is named after it’s eigenvalue rank, i.e. 5 is the 5th eigenvector.

Data separation	# attributes	top 5 PCs	next 5 PCs
sex	64	1, 7, 8, 9, 10	11, 12, 14, 15, 16
age	47	2, 3, 5, 6, 9	13, 18, 20, 22, 23
affected	11	1, 5, 8, 15, 25	63, 66, 73, 75, 81

Table 4: Checking for data loss between data representations. All data shown here is from the *W86* dataset classified using Naive Bayes. Standard deviations shown.

Data Set	3Dsnap	3Dsnap cut	2.5D
F-measure	0.71±0.18	0.68±0.20	0.72±0.20
Precision	0.88±0.18	0.82±0.20	0.80±0.20
Recall	0.63±0.22	0.62±0.24	0.69±0.22
Accuracy	0.76±0.14	0.74±0.13	0.76±0.16

AS106). The performance of *WR86* as compared to *W86* is adversely influenced by the repetition of five *control* individuals (12% of the *control* dataset).

4.2. Attribute Selection

Since 22q11.2DS is associated with a subtle facial appearance and the data is varied in age and sex, the simple solution of examining only the top 10 principle components fails. This can be illustrated by using WEKA’s *Select Attributes* capability to find the attributes which best predict *age*, *sex* and *affected* in data set *W86*. Attributes used to best predict *affected* span the entire principle component list (Table 3).

4.3. Classifier Selection

Using the WEKA package, the performance of several classifiers was compared. The classifiers used were: Naive Bayes, decision trees, nearest neighbor classifier, neural networks, and support vector machines. The analysis of the results yielded Naive Bayes, one of the simplest classifiers, outperforming all other classifiers. Although surprising, such performance can be explained by the small size of the data set and the large number of descriptors for each individual [13].

4.4. Original 3D Snapshot versus 2.5D

For the human viewer, the 3D snapshot holds much more information than the 2.5D depth image. Additionally, since ears

Table 5: Classification of vertical curved lines using Naive Bayes on the W86 data set compared to 2.5D depth image.

	2.5D	1 line	3 lines	5 lines	7 lines
F-measure	0.72	0.71	0.76	0.78	0.67
Precision	0.80	0.81	0.88	0.88	0.79
Recall	0.69	0.68	0.70	0.73	0.62
Accuracy	0.76	0.75	0.80	0.82	0.72

are known as a signal carrier for 22q11.2DS and 2.5D does not include ears, it necessary to check if data loss is present.

The accuracy of the results ranged from 74% to 76%, and there were no statistically significant differences between the performance of the three data representations (Table 4). When the ear data is absent (*3Dsnap_cut*), there is a slight decline in both precision and recall which would follow the intuition that ear data is important in assessing 22q11.2DS. Examining the trend $3Dsnap \rightarrow 3Dsnap_cut \rightarrow 2.5D$, precision decreases while recall increases, implying that classification of *affected* individuals has improved.

4.5. Curved Lines

The purpose of this subanalysis was to discover if subsets of the data, such as curved lines, contain 22q11.2DS signal. The 3 and 5 vertical curved line representations performed the best, outperforming the more detailed 2.5D depth image (Table 5). Generally, horizontal lines performed poorly. Grid lines of 3x3 and 5x5 showed insignificant improvement over 2.5D, but with F-measure and recall both degrading as compared to vertical line performance, it is likely that the horizontal line aspect is introducing noise.

5. CONCLUSIONS

This paper has discussed a new methodology for quantifying 3D face data for research in 22q11.2DS. Two methods were designed to normalize the pose of each 3D head as part of image preprocessing. Three separate data types were chosen to represent information for each 3D head. The data set was examined for aspects of age, sex and race, data set consistency, and the quality of information gathered from each of the three data types selected. The 2.5D depth image was found to be a suitable representation, although 3 and 5 vertical curved lines outperformed it in classification. Based on known 22q11.2DS signals such as a hooded appearance of the eyes, prominent forehead profile, relatively flat midface or general hypotonic facial appearance, there is promise in using sparse vertical lines to describe one or more of these anthropometric features.

Although the focus of this work is on individuals with 22q11.2DS, the methods developed for the 22q11.2DS phe-

notype should be widely applicable to the shape-based quantification of other conditions associated with craniofacial dysmorphology.

6. REFERENCES

- [1] L Kobrynski and K Sullivan, “Velocardiofacial syndrome, digeorge syndrome: the chromosome 22q11. 2 deletion syndromes,” *Lancet*, 2007.
- [2] K Golding-Kushner and R Shprintzen, “Velo-cardio-facial syndrome volume 1,” *Plural Pub Inc*, 2007.
- [3] D Becker et al., “Accuracy in identification of patients with 22q11. 2 deletion by likely care providers using facial photographs,” *Plast Reconstr Surg*, 2004.
- [4] JT Richtsmeier, VB DeLeon, and SR Lele, “The promise of geometric morphometrics,” *Yearb Phys Anthropol*, vol. 45, pp. 63–91, 2002.
- [5] JE Allanson, “Objective techniques for craniofacial assessment: what are the choices?,” *Am J Med Genet*, vol. 70, pp. 1–5, 1997.
- [6] S Boehringer et al., “Syndrome identification based on 2d analysis software.,” *Eur J Hum Genet*, vol. 14, pp. 1082–1089, 2006.
- [7] P Hammond, “The use of 3d face shape modelling in dysmorphology,” *Arch Dis Child*, vol. 92, pp. 1120–6, 2007.
- [8] P Hammond et al., “3d analysis of facial morphology,” *Am J Med Genet*, 2004.
- [9] P Hammond et al., “Discriminating power of localized three-dimensional facial morphology,” *Am J Hum Genet*, vol. 77, pp. 999–1010, 2005.
- [10] M Turk and A Pentland, “Face recognition using eigenfaces,” *CVPR*, pp. 586–591, 1991.
- [11] WJ Schroeder, KM Martin, and WE Lorensen, “The design and implementation of an object-oriented toolkit for 3d graphics and visualization,” *IEEE Visualization*, vol. 96, pp. 93–100, 1996.
- [12] IH Witten and E Frank, “Data mining: Practical machine learning tools and techniques,” 2005.
- [13] P Domingos and M Pazzani, “On the optimality of the simple bayesian classifier under zero-one loss,” *Mach Learn*, 1997.