American Journal of Medical Genetics 126A:339–348 (2004)

3D Analysis of Facial Morphology

Peter Hammond,[1]* Tim J. Hutton,[1] Judith E. Allanson,[2] Linda E. Campbell,[3] Raoul C.M. Hennekam,[4] Sean Holden,[5] Michael A. Patton,[6] Adam Shaw,[6] I. Karen Temple,[7] Matthew Trotter,[8] Kieran C. Murphy,[9] and Robin M. Winter[10]

[1]*Eastman Dental Institute, UCL, London, United Kingdom*
[2]*Division of Genetics, Children's Hospital of Eastern Ontario & University of Ottawa, Ottawa, Ontario, Canada*
[3]*Institute of Psychiatry, King's College, London, United Kingdom*
[4]*Departments of Pediatrics and Clinical Genetics, Emma Children's Hospital, University of Amsterdam, Amsterdam, The Netherlands*
[5]*Computer Laboratory, Cambridge University, Cambridge, United Kingdom*
[6]*Department of Medical Genetics, St. George's Hospital Medical School, London, United Kingdom*
[7]*Wessex Clinical Genetics Service, Southampton University Hospitals Trust, Southampton, United Kingdom*
[8]*Wolfson Institute for Biomedical Research, UCL, London, United Kingdom*
[9]*Department of Psychiatry, Royal College of Surgeons in Ireland, Dublin, Ireland*
[10]*Institute of Child Health, UCL, London, United Kingdom*

Dense surface models can be used to analyze 3D facial morphology by establishing a correspondence of thousands of points across each 3D face image. The models provide dramatic visualizations of 3D face-shape variation with potential for training physicians to recognize the key components of particular syndromes. We demonstrate their use to visualize and recognize shape differences in a collection of 3D face images that includes 280 controls (2 weeks to 56 years of age), 90 individuals with Noonan syndrome (NS) (7 months to 56 years), and 60 individuals with velo-cardio-facial syndrome (VCFS; 3 to 17 years of age). Ten-fold cross-validation testing of discrimination between the three groups was carried out on unseen test examples using five pattern recognition algorithms (nearest mean, C5.0 decision trees, neural networks, logistic regression, and support vector machines). For discriminating between individuals with NS and controls, the best average sensitivity and specificity levels were 92 and 93% for children, 83 and 94% for adults, and 88 and 94% for the children and adults combined. For individuals with VCFS and controls, the best results were 83 and 92%. In a comparison of individuals with NS and individuals with VCFS, a correct identification rate of 95% was achieved for both syndromes. This article contains supplementary material, which may be viewed at the American Journal of Medical Genetics website at http://www.interscience.wiley.com/jpages/0148-7299/suppmat/index.html.

KEY WORDS: facial morphology; dense surface models; 3D analysis; dysmorphology; diagnosis; Noonan syndrome; velo-cardio-facial syndrome

INTRODUCTION

Many dysmorphic syndromes involve craniofacial abnormality [Winter, 1996]. Experienced geneticists often make an immediate diagnosis by recognizing characteristic facial features of a syndrome. Inexperienced clinicians may struggle to make such a Gestalt diagnosis, e.g., in very young children or when they have had limited exposure to a particular syndrome or to affected individuals of the same age or ethnicity. Thus, the objective analysis of dysmorphic facial growth is potentially useful in training clinical geneticists and in assisting clinical diagnosis.

Objective techniques for analyzing facial abnormality, e.g., anthropometry, cephalometry, and photogrammetry, were previously surveyed [Allanson, 1997]. Anthropometric studies of the face have documented characteristic features and their change over time for a number of dysmorphic syndromes, e.g., Down syndrome [Allanson et al., 1993], Rubinstein–Taybi syndrome [Allanson and Hennekam, 1997], and Sotos syndrome [Allanson and Cole, 1996]. An early study of the Noonan syndrome (NS) phenotype documented changes in facial form causing some characteristic features to become more subtle with age [Allanson et al., 1985]. This remodeling of the face was reconfirmed in a 2D photogrammetric study [Sharland et al., 1993] of 104 individuals with NS using an anthropometric approach [Stengel-Rutkowski et al., 1984]. Forty-four craniofacial and 26 other features were used in a study of patterns of dysmorphic morphology in schizophrenia [Scutt et al., 2001]. The study population included patients with velo-cardio-facial syndrome (VCFS), a subgroup that subsequently formed the bulk of one of the four major clusters identified. One hundred patients between 1 year and 17 years with VCFS were the focus of an anthropometric analysis in which a characteristic pattern of craniofacial dysmorphology was established [Minugh-Purvis et al., 2002]. A smaller study of 15 patients with 22q11 deletion made similar findings [Guyot et al., 2001]. A study of lateral cephalometric radiographs of eight children with Williams syndrome identified important skeletal features contributing to facial appearance but it was not possible to use them to characterize the facial morphology conclusively [Mass and Belostoky, 1993]. Twenty-nine children under 10 years of age took part in a photogrammetric study of Williams syndrome which established soft-tissue craniofacial indices outside normal ranges [Hovis and Butler, 1997].

Grant sponsor: Birth Defects Foundation (BDF); Grant number: 2000/27; Grant sponsor: Wellcome Trust; Grant number: TG02; Grant sponsor: PPP Healthcare Medical Trust; Grant number: 1206/188.

*Correspondence to: Prof. Peter Hammond, UCL (Eastman Dental Institute), 256 Gray's Inn Road, London WC1X 8LD, UK. E-mail: p.hammond@eastman.ucl.ac.uk

Received 23 April 2003; Accepted 22 September 2003

DOI 10.1002/ajmg.a.20665

Until relatively recently, most studies of facial morphology have concentrated on the delineation of characteristic features and not on the construction and testing of computational models of face-shape variation, to be used to visualize and discriminate facial differences between or within syndromes, or between groups with specific syndromes and the general population. The application of 2D face-shape analysis in fetal alcohol syndrome (FAS) has resulted in a diagnostic protocol that is used in a number of clinical centers [Sokol et al., 1991; Astley and Clarren, 1996; Sampson et al., 2000]. More recently, stereo-photogrammetry using multiple images to calculate 3D measurements has proved more consistent than direct measurement [Meintjes et al., 2002]. A technique for pixel level analysis of 2D facial images, Gabor wavelet transformation, showed potential in discriminating between individuals with mucopolysaccharidosis type III ($n = 6$), Cornelia de Lange ($n = 12$), fragile X ($n = 12$), Prader–Willi ($n = 12$) and Williams ($n = 13$) syndromes with a success rate of 76% [Horsthemke et al., 2002]. The small numbers and lack of confirmation of blind testing suggests that this approach may not yet be fully tested.

A 2D study of lateral head radiographs in the classification of vertical facial deformity in non-syndromic patients used a point distribution model (PDM), principal components analysis (PCA) and pattern-matching techniques (nearest mean, decision tree induction, and neural networks) to compare the diagnoses of different experts [Hammond et al., 2001a]. The PDM was generated using a template of landmarks identifying important craniofacial features. A similar PDM-based approach was initially applied in the delineation of NS using 2D face photographs but was transferred to 3D data once they became available [Hammond et al., 2001b]. Many other techniques have been applied in the computer-based diagnosis of dysmorphic syndromes, often using other phenotypic descriptors in addition to or instead of those of the face [Winter et al., 1988; Evans, 1995; Evans and Winter, 1995; Braaten, 1996; Pelz et al., 1998].

The technique of geometric morphometrics is a structured approach to the analysis of landmarks for shape variation [Kendall, 1984; Bookstein, 1997; Dryden and Mardia, 1998]. Applications include comparative studies of human skull-shape across regions and through evolutionary development [O'Higgins and Jones, 1998; Hanihara, 2000; O'Higgins, 2000; Hennessy and Stringer, 2002]. Typically, such studies use a limited set of reproducible landmarks that are biologically homologous. However, on soft-tissue surfaces such as the face there are few such landmarks. Across the cheek and forehead, for instance, there are no points that have an exact biological correspondence and yet aspects of their shape contain useful biological information. Our approach makes use of this extra data by interpolating a dense correspondence between a small set of reproducible landmarks [Hutton et al., 2001; Hutton et al., 2003].

The 3D surface scans of the human face are obtainable from a variety of sources. CT images are sometimes available for patients with syndromes, e.g., for those requiring surgical treatment of craniofacial dysostosis. More often, their use for the study of facial-shape would be considered unethical or too costly. Non-invasive acquisition methods include laser-scanning [Arridge et al., 1985] and stereo-photogrammetry [Ayoub et al., 1998]. Speed of capture is important when young and potentially uncooperative children form the major part of the study population. The photogrammetric devices we have used acquire 3D surface images of the face significantly quicker than laser-based systems and simultaneously capture the appearance of the face. The appearance of the face also assists in the location of anatomical landmarks on the face surface, an essential component of our data preparation and model building. The method of image capture, however, is irrelevant to the building of dense surface models.

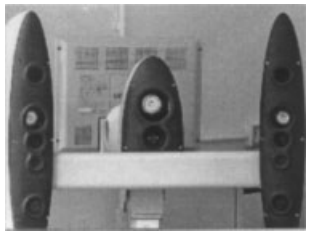

Fig. 1. Multiple cameras of DSP400 3D face scanner.

While the vast majority of VCFS individuals demonstrate an interstitial deletion of chromosome 22q11, the genetic defect underlying NS has yet to be identified conclusively. This results in difficult diagnostic dilemmas for clinical geneticists regarding some individuals who may present with unusual or atypical features. Using a new technique to analyze 3D facial morphology in patients with dysmorphic syndromes, we compared facial morphology in individuals with VCFS, NS, and normal controls. We hypothesized that 3D analysis of facial morphology: (1) predicts the clinical diagnosis of NS determined by clinical geneticists and (2) predicts the clinical diagnosis of VCFS in individuals demonstrating a chromosome 22q11 deletion.

TECHNIQUE

Image Acquisition

The 3D images used in this study were captured with the DSP400 and MU2 photogrammetric face scanners manufactured by 3dMD in the UK (http://www.3dMD.com). Both are non-contact scanners, and simultaneously capture photographic images of the face from four viewpoints using separate CCD cameras (Fig. 1).

The speckle pattern (Fig. 2a) in each of these four images is used to compute a 3D surface (Fig. 2b) that is overlaid with the subject's appearance using left and right three-quarter portraits to give the final result (Fig. 2c).

A photogrammetric scanner was preferred to a laser-based device because of its speed of capture (2 ms), essential for young children who cannot hold a pose for long. For children and adults with learning and/or physical disabilities resulting in limited motor control or involuntary body movements, it was necessary to wait until the subject became sufficiently calm or was momentarily in a suitable pose.

The DSP400 is too clumsy and heavy to be transported easily. The MU2 device and supporting computer hardware, although still heavy, is more modular, can be transported by car and is operable by one individual. The subject sits on a

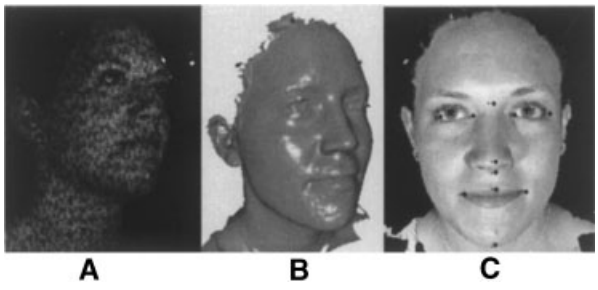

Fig. 2. 3D Face image capture. **A**: Random speckle pattern projected onto face showing one of four views/patches used to generate surface of the face. **B**: Surface of face computed from four patches. **C**: 11 landmarks typically located on each 3D image and used to build the dense surface models.

normal office-type chair, with height appropriately adjusted, in front of the scanner. The operator uses on-screen views to adjust the subject's position and need only press a mouse button to capture an image. In the vast majority of cases, it was possible to obtain a 3D face image with the subject maintaining a natural pose with neutral expression, as requested. Some subjects were either unwilling or unable to respond to this request. Where the pose distorted the face, or where the expression was not neutral, or at least not "natural" to the individual, the images were omitted. Some children, e.g. those with hypotonia, had a very relaxed lower jaw. As this is characteristic of NS, for example, such images were retained. Inevitably, facial expression may interfere with the accurate identification of facial morphology.

Image Processing

In order to build the dense surface models, 3D landmarks are located manually on each image. Typically 11 landmarks were used in the surface shape analyses: inner and outer canthi of both eyes, center of upper lip, outer corners of the mouth, nasion, pronasale, subnasale, and a chin point (Fig. 2c). Models used solely for visualizations may employ as many as 25 landmarks to give finer detail. Too few landmarks results in poor anatomical registration of the scans and too many introduces noise into the resulting model because of the inaccuracy in placing soft-tissue landmarks on a virtual image without the ability to palpate and locate bony landmarks as in conventional anthropometry. The landmarks were not employed to compute measurements, as in anthropometric studies, or inclinations of lines and planes as in cephalometric analyses. They could be used for this, but their primary use in our approach is to guide the formation of a dense correspondence between a common set of points across all face surfaces in the study group. Once the dense correspondence has been made, as many as 10,000 points are used as landmarks.

Full technical details of the generation of a dense surface model are provided elsewhere [Hutton et al., 2003]. Here we give an informal summary. Following the placing of landmarks on a given set of 3D face surfaces, the generalized Procrustes algorithm [Gower, 1975] is used to calculate the mean landmarks for the set. Each surface is then warped using the thin-plate spline (TPS) technique [Bookstein, 1997] to bring corresponding landmarks on each face into precise alignment with the mean landmarks. A dense correspondence (closest point) is then made with the vertices of a base mesh, selected arbitrarily from the dataset. The extent of the surface in the images is quite varied, and may include unwanted areas of the neck or clothing, a parent's face to one side or even a steadying hand on a shoulder. Such extraneous data is removed by including only those vertices whose distance from their base mesh location to every surface (after alignment) is at most 20 mm. Thus, the resulting collection of corresponded vertices constitutes an intersection of the meshes underlying the scans.

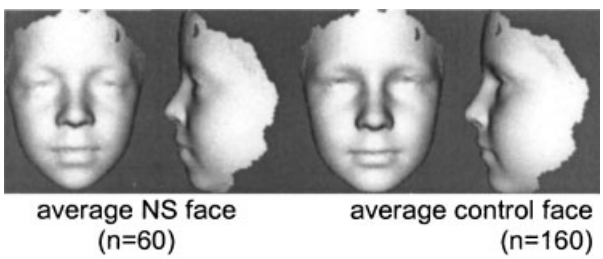

Fig. 3. Mean face surfaces for control and Noonan syndrome (NS) subgroups for individuals under 19.4 years VIDEO1 (see online addition containing VIDEOS 1–22 at http://www.interscience.wiley.com/jpages/0148-7299/suppmat/index.html).

This explains the loss of areas of the face not appearing in every 3D image (Fig. 3).

The mesh connectivity in the base mesh is then transferred to the densely corresponded meshes in the other surfaces. The original meshes/surfaces and the landmarks are then abandoned. Finally, the inverse of the TPS warp returns each surface to its original location. The points in the revised surface meshes can now be treated as landmarks to which we can apply Procrustes alignment to compute an average shape. The set of residuals of co-ordinate differences between points on individual faces and corresponding points on the average face are then subjected to a PCA to compute the major modes of shape variation. Henceforth, the phrase dense surface model refers to the set of PCA modes derived from a set of densely corresponded face surfaces.

Visualization and Pattern Recognition Testing

The dense surface models arising in this study were all computed using software produced in-house in the Biomedical Informatics Unit at UCL's Eastman Dental Institute. This software, ShapeFind, also provides a collection of tools with which to inspect the resulting set of PCA modes. The visualization in 3D of each separate mode can itself be illuminating, as is the ability to morph between average faces of subgroups of subjects within a single model. The former isolates major variations in face-shape. It should be emphasized, however, that these variations are specific to the dataset used and some overlapping face-shape variation may be seen in multiple modes. The morphing between averages of groups has great potential for highlighting important facial characteristics of a syndrome or between syndromes. Some still images are included in the "Results," but the visualizations are best appreciated dynamically by visiting the companion web page for this article.

For the comparison of controls and subjects with NS, the dataset was analyzed in three stages: under 19.4 years, over 19.4 years, and with no age restriction. This staggered approach was followed because of the previously reported observation that in NS abnormal facial characteristics become more subtle or even disappear completely in adulthood.

For each training-set, a dense surface model was computed and the top modes covering 98% of the shape variation (usually between 40 and 50) were exported for the pattern recognition component of the study. Because of the relatively small size of the subgroups under study, training-sets and test-sets of unseen examples were generated for a 10-fold cross-validation. For each pattern recognition experiment, the same proportion of subjects with syndromes and control subjects were employed for each training- and test-set pair.

The face surfaces used in the discrimination testing are synthesized from the dense surface model. The more faces included in the model, the better the synthesis. Therefore, we included as many control faces as were compatible with the age range of the group for the syndrome under scrutiny. Thus, in the models there are significant imbalances between the number of controls and the number of individuals with a syndrome.

Logistic regression, neural networks, and C5.0 decision trees were trained and tested within the Clementine data mining environment [CLEM, 2004]. Proximity to the nearest mean was evaluated within the ShapeFind system, and support vector machines were trained and tested using LIBSVM [Chang and Lin, 2001].

STUDY POPULATION

The collection of 3D face images available for this study included individuals with putative diagnoses for NS ($n = 146$) and VCFS ($n = 64$). A large collection of over 1,000 images of

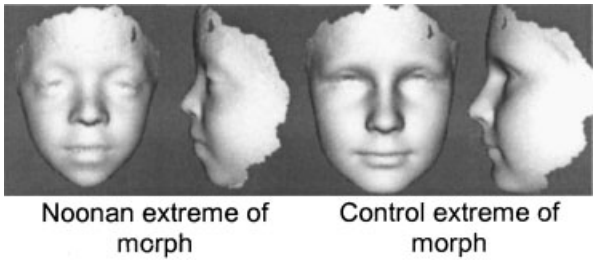

Fig. 4. Morph of overall mean along the mean hyperline VIDEO2.

subjects not known to have a syndrome was also available. We chose to demonstrate the potential use of dense surface models in facial morphology on these particular syndromes for two reasons. A diagnosis of NS still requires considerable clinical judgement because the recently established molecular test covers only about 35% of cases [Tartaglia et al., 2001]. In contrast, a diagnosis is more certain for VCFS patients. These two syndromes, therefore, provide a useful comparison when testing the discriminating performance of the pattern recognition algorithms. The images of individuals with NS have been inspected by four experienced clinical geneticists (J.E.A., R.C.M.H., I.K.T., R.M.W.) to identify individuals whose facial appearance casts doubt on their diagnosis or even suggests an alternative. A majority verdict of the experts suggested that 36 of the original 146 images be excluded from the study. Another 20 were excluded because of poor quality or non-European ethnicity.

The majority of the images of individuals with NS were captured at family meetings organised by specialist support groups such as TNSSG in the USA and Birth Defects Foundation in the UK. Some were recruited from a separate 10-year follow-on study of the NS phenotype [Shaw et al., 2002]. Most of the VCFS patients were already taking part in an existing study. Others were scanned at a conference organised in the UK by the VCFS Educational Foundation. Informed consent was obtained from all participants or their parents/guardians. All subjects selected for this study were of European ethnic background. A few individuals were scanned on more than one occasion, typically between 6 months or 1 year apart. Individuals with substantial facial hair were eliminated from the dataset.

The control population comprised individuals with no known syndrome, without obvious facial growth abnormality and with no previous maxillofacial surgery. They were either volunteers from staff and student bodies and their children or healthy siblings of children with a syndrome attending family support groups. A small group of healthy babies and young children was also recruited from a London postnatal clinic.

The study has involved building dense surface models of the face to make three comparisons:

- NS vs. controls;
- VCFS vs. controls; and
- NS vs. VCFS.

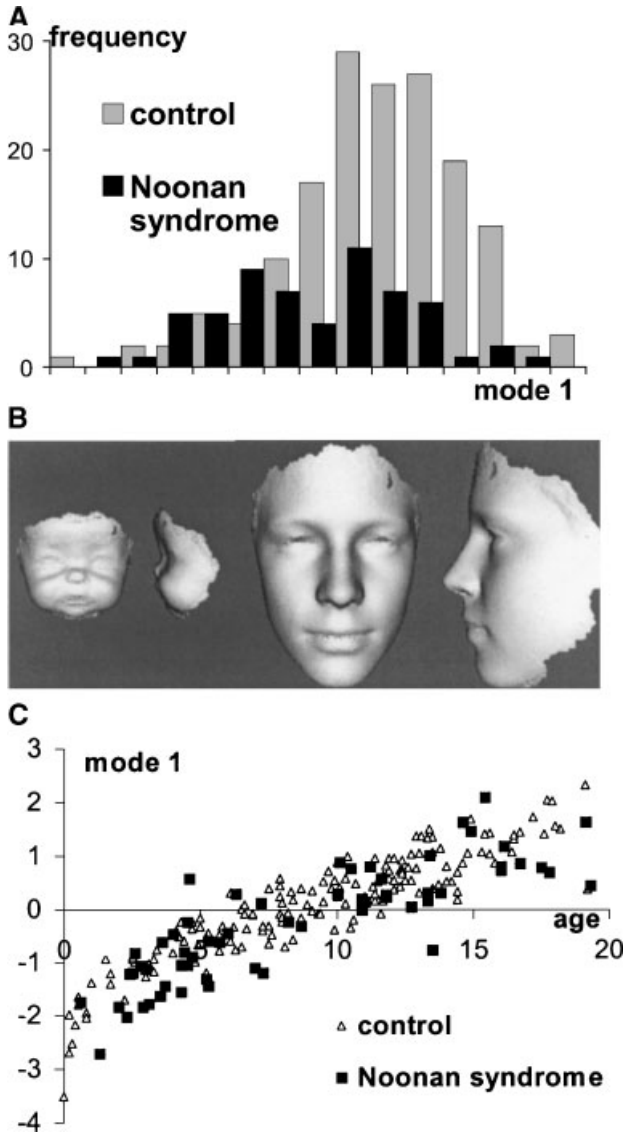

Fig. 5. Mode 1 of DSM for all NS and control individuals under 19.4 years. **A**: Distribution of mode 1 covering 73.1% of variation (mean, 0; min, −3.5 SD; max, +2.3 SD). **B**: Mode 1 variation at −3.5 and 2.3 SD; VIDEO3. **C**: Scatter plot of mode 1 against age.

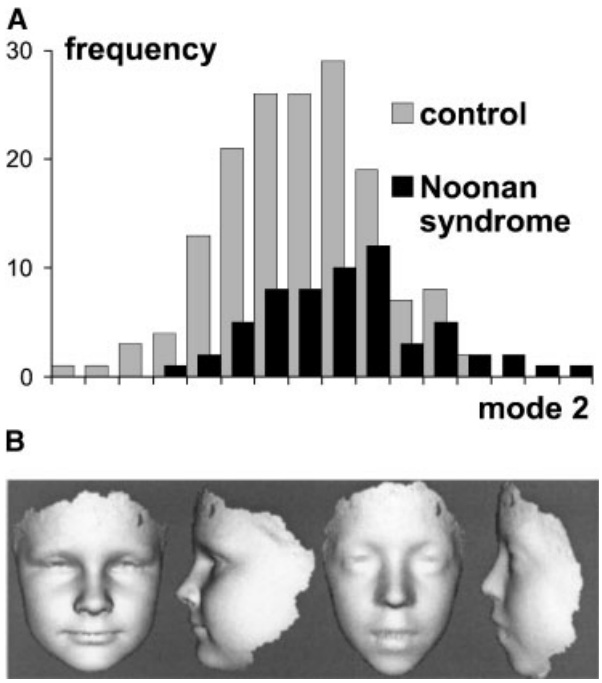

Fig. 6. Mode 2 of DSM for all NS and control individuals under 19.4 years. **A**: Distribution of mode 2 covering 6.5% of variation (mean, 0; min, −2.9 SD; mad, +3.3 SD). **B**: Mode 2 variation at −2.9 SD and +3.3 SD VIDEO4.

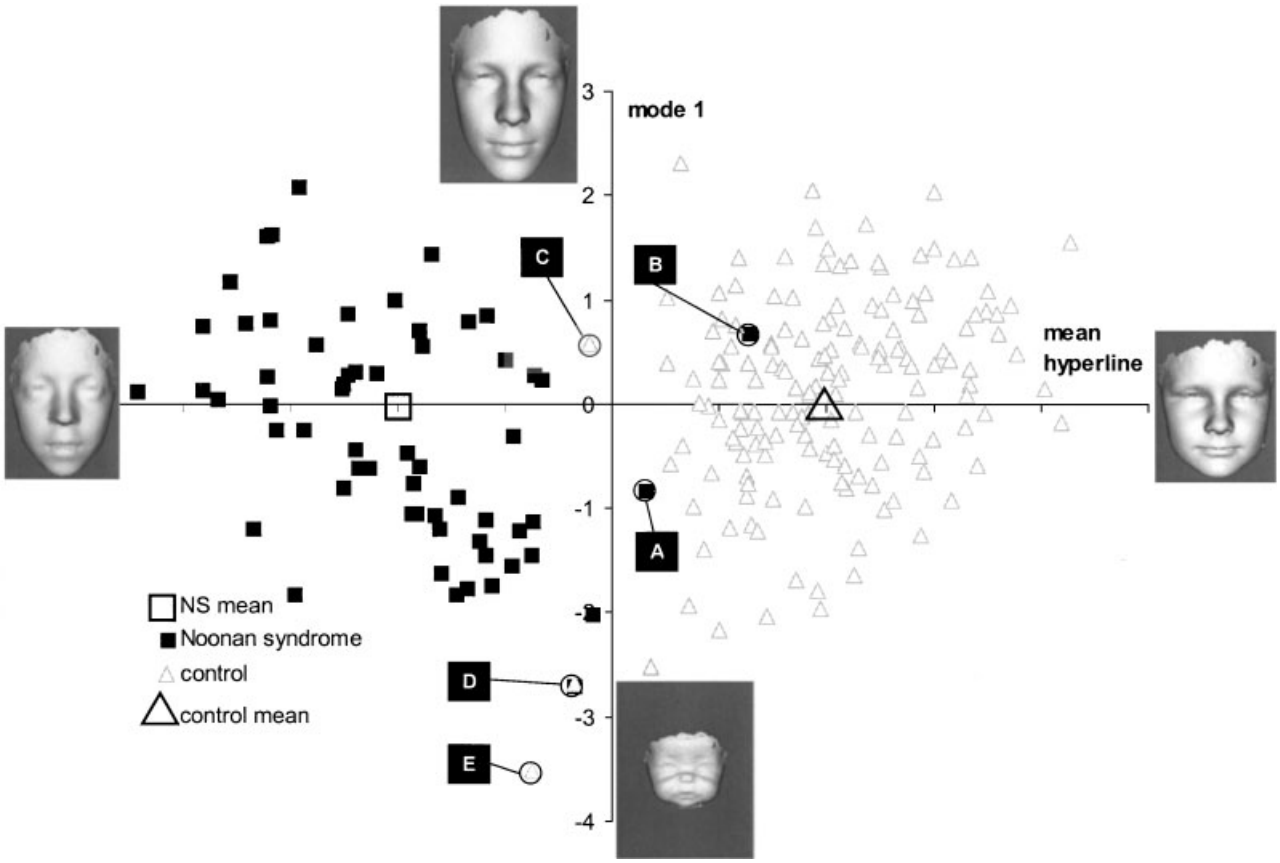

Fig. 7. Scatter plot of mean hyperline against mode 2 for non-adult controls (n = 160) and individuals with NS (n = 60). The labeled faces are misclassified according to the nearest mean discrimination test. Faces **D** and **E** are of very young babies and given the small number of such children in the database, their misclassification is not surprising. Faces **C** is just closer to the NS mean in this model. Faces **A** and **B** are misclassified as controls, but it happens that some of their features are quite uncharacteristic of NS. For ethical reasons, the faces of these individuals are not shown.

RESULTS

NS vs. Controls

A division at 19.4 years splits the NS and control dataset into 220 non-adults and 210 adults, both convenient sizes for the 10-fold cross-validation.

NS vs. controls under 19.4 years (n = 220)

Mean face comparison. A dense surface model was built using all 220 Noonan and control individuals under 19.4 years for visualization purposes only. The model used 25 landmarks and required 50 modes to cover 98% of the face-shape variation. The mean face surfaces for the Noonan and control subsets are shown in Figure 3 in full face and in profile.

The vector of PCA mode values corresponding to a face represents a point in a multi-dimensional space. Similarly, the means of the NS and control subgroups correspond to two points in the same "face-shape" space. Each point representing a face can be projected orthogonally onto a hyperline joining the two means, henceforth referred to as the "mean hyperline" when the two subgroups are obvious. The overall mean can then be morphed along this line to give a visualization of the face-shape variation across the two groups. We can exaggerate the overall mean as far as points on the mean hyperline that represent faces projected onto it. Figure 4 shows static images at each extreme. A dynamic morph can be viewed on the companion web page for this article.

PCA modes for visualizing the DSM. The first mode is similar to that found in all dense surface models of a mixed group of children and adults. It reflects overall size of the face and accounts for 73.1% of all shape variation in this particular model. Figure 5a shows the distribution of this mode (in terms of standard deviations relative to the mean), while Figure 5b shows faces computed by morphing the overall mean to the extremes of the range. The face on the left of Figure 5b corresponds to an age of about 2 weeks. Figure 5c is a scatter plot of mode1 against age, showing the strong positive correlation between the two (Pearson product moment = 0.882).

The second mode displays variation in the length and squareness of the face, and accounts for about 6.5% of all variation. Figure 6a shows the distribution and Figure 6b the extremes of the mode. Children with NS score relatively high

TABLE I. Average Sensitivity and Specificity of Cross-Validation*

Nearest mean (%)		Decision trees (%)		Neural networks (%)		Logistic regression (%)		Support vector machines (%)	
Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec
88	86	60	83	83	94	83	93	92	93

*Age <19.4 years; NS = 60; control = 160.

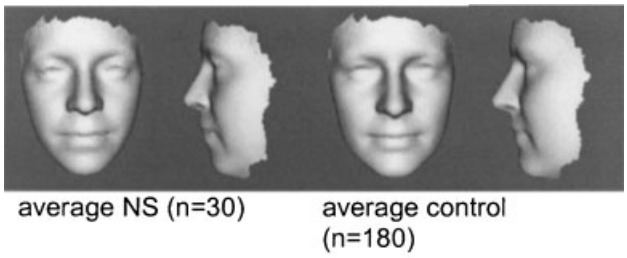

Fig. 8. Mean face surfaces for control and NS individuals over 19.4 years (VIDEO8).

on this mode. This is consistent with previous findings of reduced midfacial growth and triangular shape of the lower face. Of course, apparent face length will be accentuated for individuals with the characteristic slackness in the lower jaw.

The remaining modes are not discussed in detail here. Mode 3 (VIDEO5) contains elements of facial flatness as well as expression (smiling and grimacing). Amongst other features, modes 5 (VIDEO6) and 10 (VIDEO7) display strong elements of hypertelorism and ptosis, respectively—both characteristic features of NS. They also include elements of nose-shape change and aging.

Figure 7 is a 2D projection of the multi-dimensional face-shape space onto the hyperline between the control and NS means (as the x-axis) and mode 1 (as the y-axis), the latter being located mid-way between the two means. Some individual faces have been highlighted so as to emphasize their displacement from the means. Thus, the location of a face to the left or right of the y-axis, nearer to one or other of the mean faces, is a simple classification device. In general, this informal use of nearest mean as a discrimination test is unacceptable because it tests the dataset used to compute the two means in the first place. For a more valid evaluation of this, or any other, classification algorithm, disjoint training and test cases must be employed.

Ten-fold cross-validation testing. Table I summarizes the average sensitivity and specificity obtained in a 10-fold cross-validation of five different algorithms, including nearest mean. The 10 dense surface models generated from the 90% training-sets all used the same 11 landmarks that were described earlier. The best classification result, an average sensitivity of 92% and an average specificity of 93%, was achieved using support vector machines [Vapnik, 1995] with a radial basis function kernel classification and the Hinton heuristic with regularization parameter C = 100.

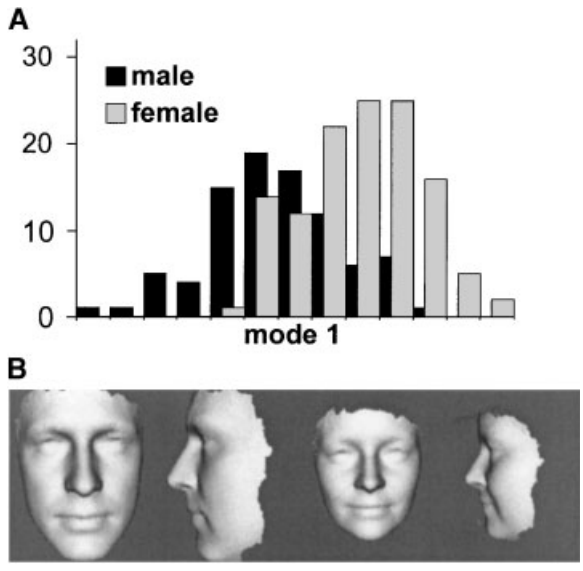

Fig. 9. Mode 1 of DSM for all 210 controls and individuals with NS over 19.4 years of age. **A**: Distribution (mean, 0; min, −2.9 SD; max, 2.1 SD). **B**: Mode 1 at −2.9 and 2.1 SD (VIDEO9).

Previously, a similar discrimination evaluation was carried out using 90 or so of the current NS study population—including some of those individuals rejected by the expert consensus. The sensitivity then, for a similar age range, was 85% [Hammond et al., 2002]. The improvement to 92% is likely to be due in part to the rejection of patients with a dubious diagnosis.

NS vs. controls over 19.4 years (n = 210)

Mean face comparison. The mean face surfaces for the Noonan (n = 30) and control subgroups (n = 180) over 19.4 years are shown in Figure 8 in full face and in profile.

The exaggerated control and NS means are not shown because they are only slightly different from the means themselves. The video is of the morph between the exaggerated means.

PCA modes for visualizing the DSM. As for the earlier group of non-adults, a single dense surface model was built using all 210 Noonan and control individuals over 19.4 years. Figure 9a shows the distribution of the first mode of this model while Figure 9b visualizes the faces at the extremes of the range. Thus, mode 1 for this group of adults still reflects variation in size of the face but here it also reflects gender—as can be seen from the distribution shown in Figure 9a and the faces in Figure 9b generated by morphing the mean to the extremes of the distribution of mode 1.

Mode 2 (VIDEO10) reflects the usual long/oval to short/square face-shape variation. Mode 3 (VIDEO11) involves a combination of feature changes including hypotelorism and nose width and length. Gender and facial expression variations are strong in mode 4 (VIDEO12). Hypotelorism features in mode 5 (VIDEO13) but alongside significant changes in nose length. Mode 6 (VIDEO14) is dominated by retrognathic–prognathic variation. Asymmetry of the face is a significant element of mode 9 (VIDEO15). None of these modes appears to be particularly noteworthy. Their dynamic morphs are available for inspection on the companion web site.

Ten-fold cross-validation testing. The average sensitivity and specificity of a 10-fold cross-validation for the same five pattern recognition algorithms tested on the non-adult population are shown in Table II.

By comparison with the non-adult group, discrimination between individuals with NS and controls in the adult group is considerably less successful. This is consistent with the previously cited diminution in adulthood of characteristic facial differences in NS.

NS vs. controls without age restriction (n = 430). The comparison of means and the visualization of modes is omitted here as they are similar to those of the previous two sections. Instead, we simply give the results of the 10-fold cross-validation for the same five pattern recognition algorithms in Table III below.

These results, generally intermediate between those for the separate child and adult groups, are consistent with diminishing differences in facial morphology during the maturing of individuals with NS.

VCFS vs. Controls

VCFS vs. controls between 3 years and 17 years (n = 190)

Mean faces. The mean face surfaces for the VCFS (n = 60) and control subgroups (n = 130) between 3 years and 17 years of age are shown in Figure 11 in full face and in profile.

TABLE II. Average Sensitivity and Specificity of Cross-Validation*

Nearest mean (%)		Decision trees (%)		Neural networks (%)		Logistic regression (%)		Support vector machines (%)	
Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec
80	94	20	92	60	96	60	93	83	94

*Age $\geq$19.4 years; NS = 30; control = 180.

TABLE III. Average Sensitivity and Specificity of Cross-Validation*

Nearest mean (%)		Decision trees (%)		Neural networks (%)		Logistic regression (%)		Support vector machines (%)	
Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec
81	82	44	85	68	96	84	93	88	94

*NS = 90; control = 340.

TABLE IV. Average Sensitivity and Specificity of Cross-Validation*

Nearest mean (%)		Decision trees (%)		Neural networks (%)		Logistic regression (%)		Support vector machines (%)	
Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec
82	85	58	81	57	97	75	90	83	92

*VCFS = 60; control = 130.

The exaggerated morphs of the overall mean along the mean hyperline are only marginally different from the means so their visualization is not included here but on the companion web site.

Modes for visualizing the DSM. A dense surface model was generated for the combined VCFS and control subgroups. The first mode of this model reflects the usual size of face variation (VIDEO17) and accounts for 60.9% of shape variation. The second mode, covering 12.9% of shape variation, reflects VCFS characteristics such as the longer nose, narrow nasal base, and somewhat tubular shape (Fig. 11), but also some facial expression.

Figure 12 illustrates the same 2D projection of the face-shape space using a mean hyperline vs. mode 1 scatter plot as shown previously for NS. Once again, such within training-set testing of the nearest mean algorithm gives excellent discrimination between the two subgroups, with just five faces misclassified.

Ten-fold cross-validation. The results of the unseen discrimination testing in Table IV are not as good as those for the NS–control comparison for a similar age range.

They emphasize the obvious fact that individuals with VCFS have more subtle facial differences from controls than is the case in NS. Moreover, the classification may also be impaired by the fact that all but a handful of the VCFS 3D scans were captured with the older scanner at a poorer resolution. Therefore, more subtle facial features may not be modeled well.

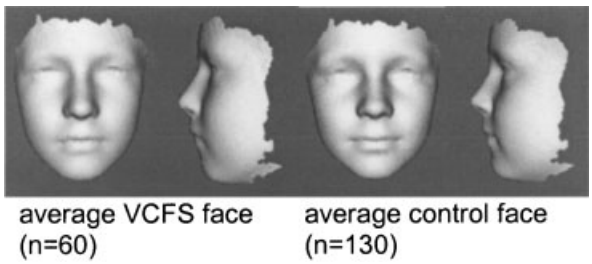

Fig. 10. Mean face surfaces for control and velo-cardio-facial syndrome (VCFS) subgroups for individuals between 3 years and 17 years (VIDEO16).

VCFS vs. NS

VCFS vs. NS between 2 years and 20 years (n = 120)

Exaggerated means. The means and exaggerated morph along the mean hyperline are not illustrated statically since they are very similar to images already shown in Figures 3, 4, and 10. The dynamic variation gives an excellent visualization

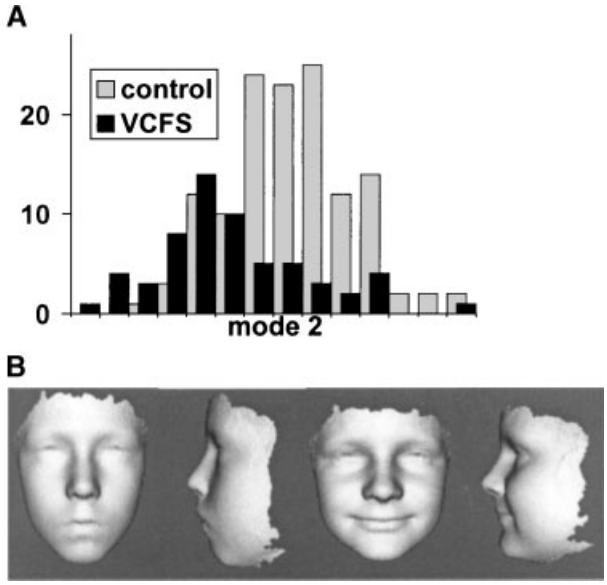

Fig. 11. Mode 2 of dense surface model for all 190 controls and VCFS individuals aged 3–17 years. **A**: Distribution of mode 2 covering 12.9% of variation (mean, 0; min, −2.4 SD; max, +2.6 SD). **B**: Mode 2 variation at −2.4 SD and +2.6 SD VIDEO18

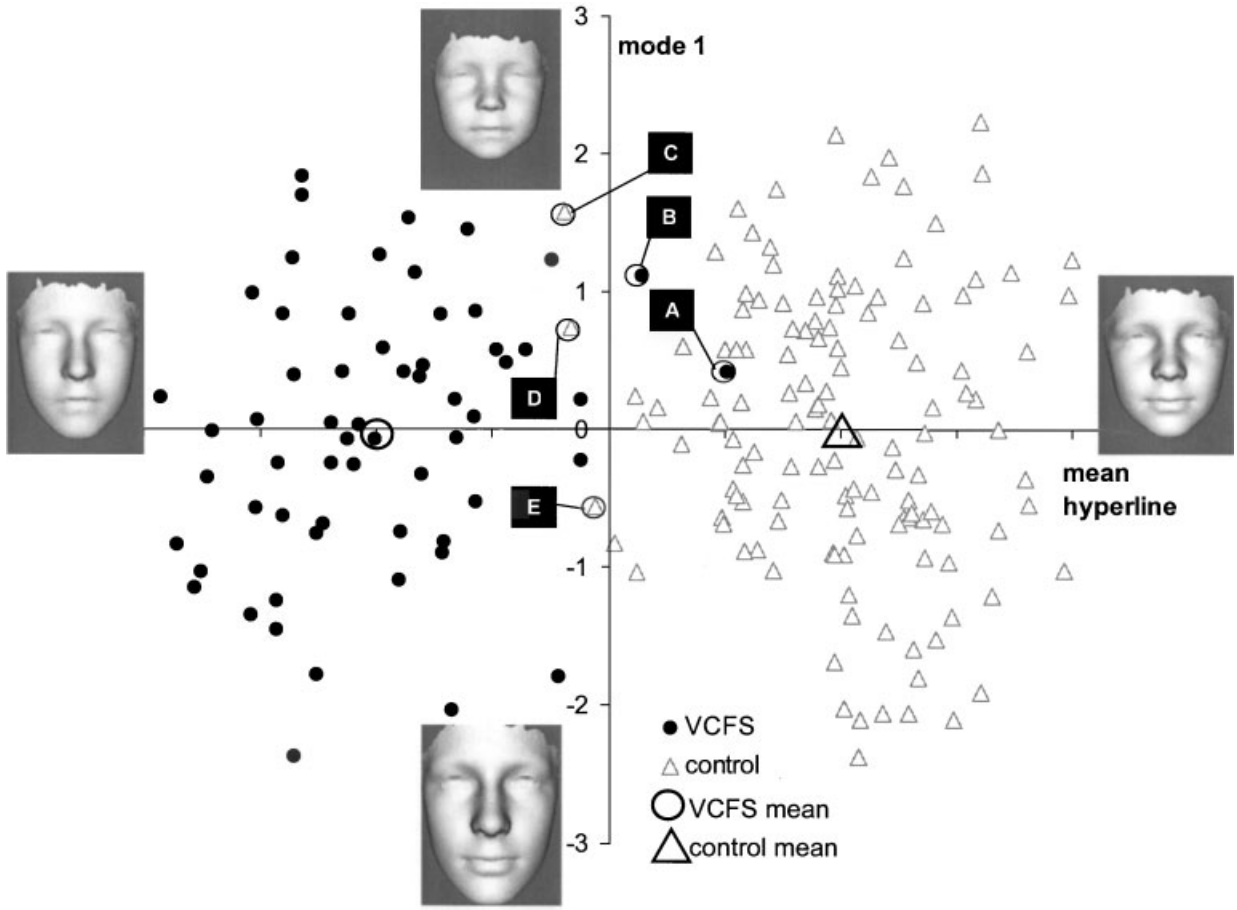

Fig. 12. Scatter plot of mean hyperline (x-axis) against mode 1 (y-axis) for controls (n = 130) and individuals with VCFS (n = 60). The five faces labeled **A–E** are misclassified according to the nearest mean discrimination test. As with Figure 7, individual faces are not shown. The images included illustrate the extremes for each of the axes.

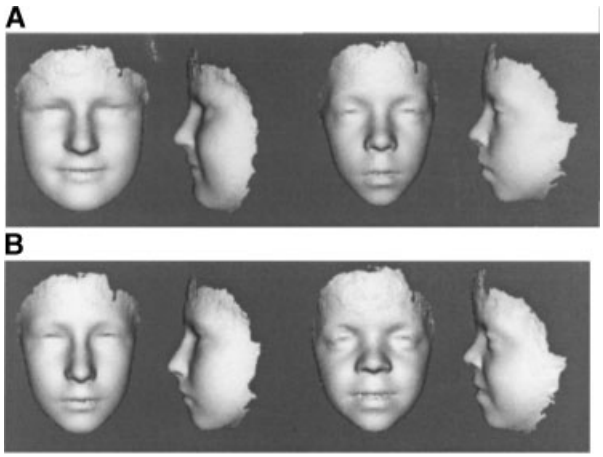

Fig. 13. Modes 3 and 4 of DSM for NS (n = 60) and VCFS (n = 60) individuals aged 2–20 years. **A**: Mode 3 at −3.3 and 2.2 SD—VIDEO20. **B**: Mode 4 at −2.1 and 3.1 SD—VIDEO21.

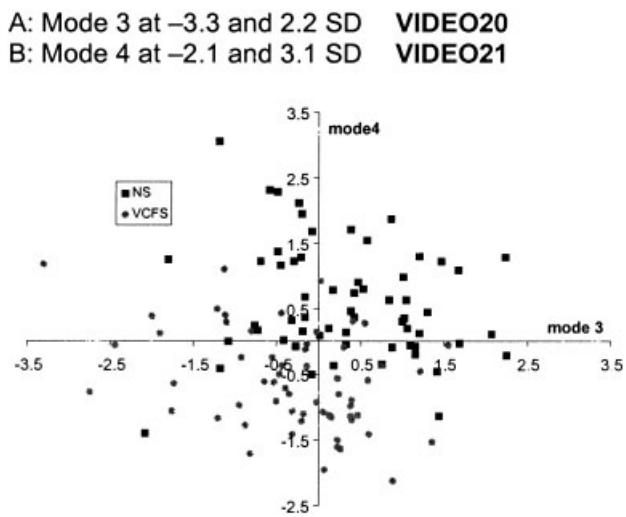

Fig. 14. Scatter plot of modes 3 and 4 of DSM for NS (n = 60) and VCFS (n = 60) individuals aged 2–20 years illustrating their partial discriminating ability.

of the overall difference in facial morphology between NS and VCFS (VIDEO19).

PCA modes for visualizing the DSM. A dense surface model was generated to visualize the PCA modes for the combined VCFS and NS subgroups. The first two modes of this model capture the usual face-size and -shape variations. Modes 3 and 4, shown in Figure 13 and scatter plotted in Figure 14, undergo the greatest change during the morph along the mean hyperline.

Mode 3 emphasizes differences in head width as well as mandible size and shape. Mode 4 strongly reflects eye separation as well as width of both nasal bridge and nasal base. The nasal shape differences are very noticeable, as would be expected with these two syndromes.

Ten-fold cross-validation. The results of the unseen discrimination testing are shown in Table V.

These results are generally better than for the syndrome groups vs. control discrimination tests. This is due in part to the more even balancing of the training- and test-sets and suggests that in future a more thorough investigation should be made of the interactions between training-test set balancing, accuracy of face synthesis, and discriminating ability.

DISCUSSION

In this study, we found that a novel method of 3D analysis demonstrates high levels of sensitivity (88%) and specificity (94%) in discriminating between controls and individuals previously diagnosed with NS and vetted visually by a further panel of four experienced clinical geneticists. In addition, we found high levels of sensitivity (83%) and specificity (92%) in discriminating between individuals with VCFS (diagnosed cytogenetically) and controls. We suggest that this novel technique may assist the clinical geneticist in making a diagnosis in individuals who present with complex or atypical facial features.

As the molecular defect in NS has yet to be characterized, the diagnosis is generally made by experienced clinical geneticists. However, it is probable that a small minority of patients, diagnosed with NS by clinical geneticists, do not in fact have NS and are phenocopies. Thus, the high sensitivity and specificity of 3D facial analysis reported in this study are potentially confounded by the possibility of diagnostic error. In view of this, we also performed 3D image analysis in people with a known chromosome 22q11 microdeletion. The high sensitivity and specificity we report in discriminating between people with VCFS and controls demonstrates that diagnostic error is an unlikely explanation for the high discriminating power of this technique in NS.

In comparison to the portable scanner (MU2), the older model (DSP400) does not capture both ear surfaces consistently and captures far fewer surface points. As was previously explained, the dense surface model involves computing an intersection of captured face surfaces, and so our derived models may not include any ear surface at all depending on the particular images included in the model. In NS, where low ear position is a characteristic trait, and in VCFS, where ear shape can be dysmorphic, this is potentially damaging to any analysis.

Fig. 15. Extreme faces on mean hyperline for NS vs. control using a subset of images with good ear coverage (n = 106) VIDEO22.

Figure 15 and the associated dynamic morph demonstrate the huge improvement in the visualization when ears are included in the captured surface. Moreover, dense surface models generated from such images are likely to discriminate with even greater success.

It is difficult to estimate the minimum number of 3D images needed before a dense surface model supports useful visualizations or discrimination. Further testing is required before firm conclusions can be drawn.

CONCLUSIONS

The still and dynamic images included here and on the accompanying website demonstrate that dense surface models can generate striking and informative visualizations of facial morphology in 3D. Any benefit they may provide in the training of clinical geneticists is yet to be evaluated.

Dense surface models when combined with state of the art pattern recognition algorithms achieve impressive results in discriminating between controls and individuals with a particular syndrome, as well as between individuals with different syndromes—at least for the two syndromes covered. It is also encouraging that the results presented here were obtained using 3D meshes of different resolutions and quite varied coverage, often missing anatomical features important in facial dysmorphology.

As was remarked earlier, the larger the number of examples included in a dense surface model the greater the accuracy of the synthesis of each face using the associated modes. The rarity of some syndromes will limit the available dataset and imbalance the mix of controls and individuals with a particular syndrome. Inevitably, this will reduce the discrimination accuracy. Therefore, it is important that image capture is undertaken on an international basis, with sharing of data once appropriate ethical approval and patient or parent consent is obtained.

Our collection of 3D face scans also includes groups of individuals with Angelman, Rett, Rubinstein–Taybi, Smith–Magenis, and Williams syndromes. Having demonstrated that dense surface models show promise in delineating and discriminating face shape associated with Noonan and VCFS, separate studies of the comparative 3D facial morphologies of individuals with these other syndromes have begun. As the data gathering for these becomes more internationally widespread, it will also become possible to address the issue of ethnic variation in facial morphology. Similar unavoidable confounding factors when considering facial morphology,

TABLE V. Average Sensitivity and Specificity of Cross-Validation*

Nearest mean (%)		Decision trees (%)		Neural networks (%)		Logistic regression (%)		Support vector machines (%)	
Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec
85	90	86	71	91	87	93	89	95	95

*Using 60 NS as −ve and 60 VCFS as +ve.

of course, are gender, age, and familial likeness. Before these can begin to be addressed, much more data needs to be gathered.

ACKNOWLEDGMENTS

The authors are thankful to the patients, families, and volunteers who took part in this study. Wellcome Trust provided a small travel grant. BDF also provided funding and opportunities for scanning attendees at their NS family meetings in UK, as did TNSSG in USA. We also thank PPP Healthcare Medical Trust whose funding supported the involvement of children with VCFS and the 22q11 (UK) Support Group and VCFSEF (USA) who kindly helped with recruitment. The St. Albans Clinic in Highgate, London clinic, and attending parents consented for their children to be included as controls in the study. Dr. Henry Potts and Dr. Caroline Deys are thanked for helping to arrange attendance at the clinic. Prof. Bernard Buxton (Dept. of Computer Science, UCL) was generous with his advice during discussions of the data analysis.

REFERENCES

Allanson JE. 1997. Objective techniques for craniofacial assessment: What are the choices? Am J Med Genet 70:1–5.

Allanson JE, Cole TRP. 1996. Sotos syndrome: Evolution of the facial phenotype subjective and objective assessment. Am J Med Genet 65: 13–20.

Allanson JE, Hennekam RCM. 1997. Rubinstein–Taybi Syndrome: Objective evaluation of craniofacial structure. Am J Med Genet 71:414–419.

Allanson JE, Hall JG, Hughes HE, Preus M, Witt RD. 1985. Noonan Syndrome: The changing phenotype. Am J Med Genet 21:507–514.

Allanson JE, O'Hara , Farkas, Nair RC. 1993. Anthropometric craniofacial pattern profiles in Down syndrome. Am J Med Genet 47:748–752.

Arridge S, Moss JP, Linney AD, James DR. 1985. Three dimensional digitisation of the face and skull. J Maxillofac Surg 13:136–143.

Astley SJ, Clarren SK. 1996. A case definition and photographic screening tool for facial phenotype of fetal alcohol syndrome. J Pediatr 129:33–41.

Ayoub A, Siebert P, Moos K, Wray D, Urquhart X, Niblett T. 1998. A vision-based thee-dimensional capture system for maxillofacial assessment and surgical planning. Brit J Oral Maxillofac Surg 36:353–357.

BDF. 2004. (Birth Defects Foundation). http://www.birthdefects.co.uk

Bookstein FL. 1997. Shape and the information in medical images: A decade of the morphometric synthesis. Comput Vis Image Process 33:33–80.

Braaten O. 1996. Artificial intelligence in pediatrics: Important clinical signs in newborn syndromes. Comput Biomed Res 29:153–161.

Chang C-C, Lin C-J. 2001. LIBSVM: A library for support vector machines. http://www.csie.ntu.edu.tw/~cjlin/libsvm

CLEM. 2004. http://www.spss.com/spssbi/clementine/

Dryden IL, Mardia KV. 1998. Statistical shape analysis. Chichester, UK: John Wiley and Sons.

Evans CD. 1995. A case-based assistant for diagnosis and analyis of dysmorphic syndromes. Med Inf 20:121–131.

Evans CD, Winter RM. 1995. MD Computing 12:127–136.

Gower JC. 1975. Generalized proscrustes analysis. Psychometrika 40: 33–51.

Guyot L, Dubuc M, Pujol J, Dutout O, Philip N. 2001. Craniofacial analysis in patients with 22q11 microdeletion. Am J Med Genet 100:1–8.

Hammond P, Hutton TJ, Nelson-Moon ZL, Hunt N, Madgwick AJA. 2001a. Classifying vertical facial deformity using supervised and unsupervised learning. Meth Inf Med 40:365–372.

Hammond P, Hutton TJ, Patton MA, Allanson JE. 2001b. Delineation and visualisation of congenital abnormality using 3D facial images. In: Bellazzi R, Zupan B, Liu X, editors. Proceedings of the Workshop Intelligent Data Analysis in Medicine and Pharmacology. London, UK: IDAMAP2001 at MedInfo2001.

Hammond P, Hutton T, Allanson JA, Shaw A, Patton MA. 2002. 3D Digital stereo photogrammetric analysis of face shape in Noonan Syndrome. J Med Genet 399(suppl 1):S35.

Hanihara T. 2000. Frontal and facial flatness of major human populations. Am J Phys Anthropol 111:105–134.

Hennessy RJ, Stringer CB. 2002. Geometric morphometric study of the regional variation of modern human craniofacial form. Am J Phys Anthropol 117:37–48.

Horsthemke B, Wieczorek D, Loos HS, von der Malsburg C. 2002. Computer-based recognition of syndromic faces. Baltimore, USA: Proc Am Soc Hum Gen 2002.

Hovis CL, Butler MG. 1997. Photoanthropometric study of craniofacial traits in individuals with Williams syndrome. Clin Genet 51:379–387.

Hutton TJ, Buxton BF, Hammond P. 2001. Dense surface point distribution models of the face. Kauai, Hawaii: Proc IEEE Workshop on Mathematical Methods in Biomedical Image Analysis. 153–160.

Hutton TJ, Buxton BF, Hammond P, Potts HWW. 2003. Estimating average growth trajectories in shape-space using kernel smoothing. IEEE Trans Med Imaging 22(6):747–753.

Kendall DG. 1984. Shape-manifolds, procrustean metrics, and complex projective spaces. Bull Lond Math Soc 16:81–121.

Mass E, Belostoky L. 1993. Craniofacial morphology of children with Williams syndrome. Cleft Palate Craniofac J 30:343–349.

Meintjes EM, Douglas TS, Martinez F, Vaughan CL, Adams LP, Stekhoven A, Viljoen D. 2002. A stereo-photogrammetric method to measure the facial dysmorphology of children in the diagnosis of fetal alcohol syndrome. Med Engl Phys 24:683–689.

Minugh-Purvis N, Kirschner RE, Slemp AE, McDonald-McGinn DM, Zackai EH, LaRossa D, Emanuel BS. 2002. Craniofacial dysmorphology in 22q11.2 deletion syndrome: An anthropometric analysis. Italy: Proc DELETION 22q11.2, Third International Meeting.

O'Higgins P. 2000. The study of morphological variation in the human fossil record: Biology, landmarks and geometry. J Anat 197:103–120.

O'Higgins P, Jones N. 1998. Facial growth in Cercocebus torquatus: An application of three-dimensional geometric morphometric techniques to the study of morphological variation. J Anat 193:251–272.

Pelz J, Arendt V, Kunze J. 1998. Computer assisted diagnosis of malformation syndromes: An evaluation of three datanases (LDDB, POSSUM, SYNDROC). Am J Med Genet 63:257–267.

Sampson PD, Streissguth A, Bookstein FL, Barr H. 2000. On categorizations in analyses of alcohol teratogenesis. Environm Health Perspect 108: 421–428.

Scutt LE, Chow EWC, Weksberg R, Honer WG, Bassett AS. 2001. Patterns of dysmorphic features in schizophrenia. Am J Med Genet (Neuropsychiatric Genetics) 105:713–723.

Sharland M, Morgan M, Patton MA. 1993. Photoanthropometric study of facial growth in Noonan syndrome. Am J Med Genet 45:430–436.

Shaw A, van der Burgt I, Brunner HG, Noordam K, Kalidas K, Crosby AH, Ion A, Jeffrey S, Patton MA, Tartaglia M, Gelb BD. 2002. Genotype and phenotype analysis of 127 patients with Noonan syndrome. Strasbourg: Proc ESHG.

Sokol RJ, Chik L, Martier SS, Salari V. 1991. Morphometry of the neonatal fetal alcohol syndrome face from 'snapshots.' Alcohol Suppl 1:531–534.

Stengel-Rutkowski S, Schimanek P, Wernheimer A. 1984. Anthropometric definitions of dysmorphic facial signs. Hum Genet 67:272–295.

Tartaglia M, Mehler EL, Goldberg R, Zampino G, Brunner HG, Kremer H, van der Burgt I, Crosby AH, Ion A, Jeffrey S, Kalidas K, Patton MA, Kucherlapati RS, Gelb BD. 2001. Mutations in PTPN11, encoding the protein tyrosine phosphatase SHP-2, cause Noonan syndrome. Nat Genet 29:465–467.

TNSSG. 2004. http://www.noonansyndrome.org/home.html

Vapnik V. 1995. The nature of statistical learning theory. New York: Springer.

VCFSEF. 2004. http://www.vcfsef.org/

Winter RM. 1996. What's in a face? Nat Genet 12:124–129.

Winter RM, Clark RD, Ashley K, Gibbs G. 1988. A combinatorial method for grouping cases with multiple malformations. J Med Genet 25: 118–121.