Craniofacial Image Analysis

**Ezgi Mercan, Indriyati Atmosukarto, Jia Wu, Shu Liang,
and Linda G. Shapiro**

1 Introduction to Craniofacial Analysis

Craniofacial research focuses on the study and treatment of certain congenital malformations or injuries of the head and face. It has become a multi-disciplinary area of expertise in which the players consist of not only oral and maxillofacial or plastic surgeons, but also craniofacial researchers including a large array of professionals from various backgrounds: basic scientists, geneticists, epidemiologists, developmental biologists, and recently computer scientists.

It is important to represent the shape of the human face in a standard way that facilitates modeling the abnormal and the normal. Morphometrics, the study of shape, has been a crucial toolbox for craniofacial research. Classical morphometrics-based craniofacial analyses use *anthropometric landmarks* and require taking physical measurements directly on the human face. These measurements are then used in a numerical analysis that compares the patient's measurements with the normal population to detect and quantify the deformation. Another technique for measuring the severity of shape deformations involves having clinical experts qualitatively match the shape of the patient's head to a set of templates. Template-matching is a common method in clinical practice, but it heavily depends on human judgment.

E. Mercan (✉) • S. Liang • L.G. Shapiro
Computer Science and Engineering, University of Washington, Seattle, WA, USA
e-mail: ezgi@cs.washington.edu; liangshu@cs.washington.edu; shapiro@cs.washington.edu

I. Atmosukarto
Singapore Institute of Technology, Singapore, Singapore
e-mail: indria@cs.washington.edu

J. Wu
Electrical Engineering, University of Washington, Seattle, WA, USA
e-mail: jiawu@uw.edu

As the field of computer vision has progressed, its techniques have become increasingly useful for medical applications. Advancements in 3D imaging technologies led craniofacial researchers to use computational methods for the analysis of the human head and face. Computational techniques aim to automate and improve established craniofacial analysis methods that are time consuming and prone to human error and innovate new approaches using the information that has become available through digital data.

This paper describes a set of computational techniques for craniofacial analysis developed by the University of Washington Multimedia Group. Section 2 describes the craniofacial syndromes whose analyses we have performed. Section 3 summarizes our previous work in craniofacial research and describes our image analysis pipeline, including preprocessing, feature extraction, quantification, and classification. Section 4 introduces our new work on the comparison of different features in similarity-based retrieval, and Sect. 5 concludes the paper.

2 Craniofacial Syndromes

We will describe three relevant syndromes: deformational plagiocephaly, 22q11.2 deletion syndrome, and cleft lip and palate.

2.1 Deformational Plagiocephaly

Deformational plagiocephaly can be defined as abnormal head shape (parallelogram shaped skull, asymmetric flattening, misalignment of the ears) due to external pressure on the infant's skull [17]. Figure 1 shows photographs of several infants' heads from the top view, with and without plagiocephaly. Although considered a minor cosmetic condition by many clinicians, if left untreated, children with plagiocephaly may experience a number of medical issues, ranging from social problems due to abnormal appearance to delayed neurocognitive development.
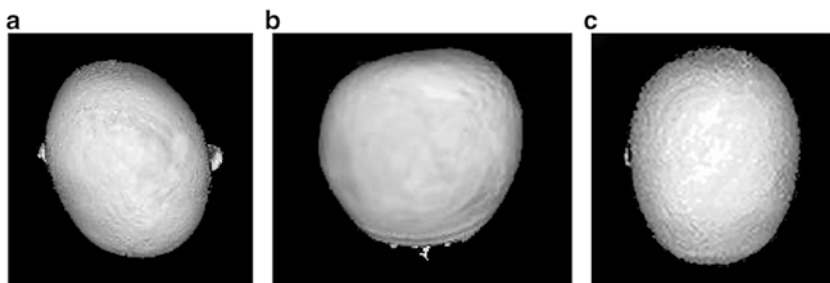
Fig. 1 *Deformational Plagiocephaly*—(**a,b**) *Top views* of heads of children with deformational plagiocephaly. (**c**) *Top view* of a child's head without deformational plagiocephaly

The severity of plagiocephaly ranges from mild flattening to severe asymmetry along a wide spectrum that is difficult to quantify. Clinical practices to diagnose and quantify plagiocephaly involve identifying anthropometric landmark points and taking measurements between the points. In one approach, the clinician determines the areas with the greatest prominence on the right and left sides of the head and measures diagonally the distances from these sites to the back of the head. The smaller length is subtracted from the larger resulting in an asymmetry number called the *transcranial diameter difference* [10]. Another technique compares the infant's skull shape to four templates: normal skull [score 0], mild shape deformation [score 1], moderate shape deformation [score 2], and severe shape deformation [score 3].

As an alternative to taking physical measurements directly on the infant's head, a technique called *HeadsUp* developed by Hutchinson et al. performs automated analysis of 2D digital photographs of infant heads fitted with an elastic head circumference band that has adjustable color markers to identify landmarks [13]. Although this semi-automatic approach is less intrusive and faster, it is still subjective, and the analysis is only 2D. There are some recently proposed techniques that use 3D surface data: Plank et al. [22] use a laser shape digitizer to obtain the 3D surface of the head, but still require manual identification of the landmarks. Lanche et al. [14] use a stereo-camera system to obtain a 3D model of the head and propose a method to compare the infant's head to an *ideal* head template.

2.2 22q11.2 Deletion Syndrome

22q11.2 deletion syndrome (22q11.2DS) is a disorder caused by a 1.5–3 MB deletion on chromosome 22 and occurs in 1 of every 4,000 individuals [21]. Over 180 phenotypic features are associated with this condition, including well-described craniofacial features such as asymmetric face shape, hooded eyes, bulbous nasal tip, tubular appearance to the nose, retrusive chin, prominent nasal root, small nasal alae, small mouth, open mouth and downturned mouth, among others. Some manifestations of facial features are very subtle, and even craniofacial experts find them difficult to identify without measurements and analysis. Figure 2 shows example manifestations of the syndrome on the face.

Fig. 2 *22q11.2DS craniofacial features*—Example 3D face mesh data of children with 22q11.2 deletion syndrome
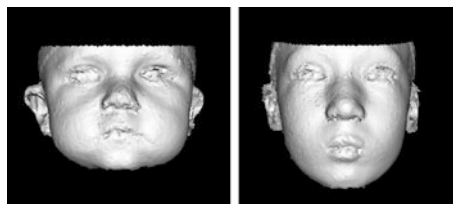
Early detection of 22q11.2DS is important, because the condition is known to be associated with cardiac anomalies, mild-to-moderate immune deficiencies and learning disabilities. Similar to the detection of deformational plagiocephaly, the assessment of 22q11.2DS has commonly been through physical examination and craniofacial anthropometric measurements. After identification of the symptoms, genetic tests can be conducted to confirm and complete the diagnosis.

There has been little effort to automate the diagnosis and analysis of 22q11.2DS. Boehringer et al. [7] used Gabor wavelets to transform 2D photographs of the individuals and PCA to classify the dataset. However, the method requires manual placement of anthropometric landmarks on the face. Hammond et al. [12] proposed a *dense surface model* method followed by the application of PCA on 3D surface mesh data, which also requires manually placed landmarks to align the meshes.

2.3 Cleft Lip and Palate

Cleft lip is a birth defect that occurs in approximately 1 in every 1,000 newborns and can be associated with cleft palate [2]. The deformity is thought to be a result of the failure of fusion *in utero* and may be associated with underdevelopment of tissues. Cleft lip and palate can range from multiple deep severe clefts in the palate to a single incomplete or hardly noticeable cleft in the lip. Figure 3 shows examples of deformations on infants' faces caused by cleft lip and/or palate. The condition can be treated with surgery and the treatment can produce a dramatic change in appearance of the lip depending on the severity of the cleft. Since the potential results and treatment options depend on the severity of the cleft, it is important to have an objective assessment of the deformity.

The assessment of cleft deformities relies on a clinical description that can be subjective and landmark-based measurements that can be time consuming and difficult to perform on young infants. Additionally, there is no "gold standard" for evaluation and the correlation between the scores given by different medical experts can be very low [26].

There has not been much computational work done towards the quantification of the cleft lip using the face shape. Nonetheless, some work has been done on face symmetry, which can be used in cleft assessment, in computer vision. Although not
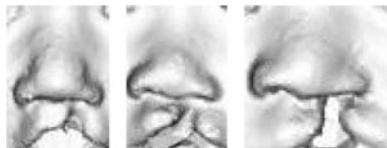
Fig. 3 *Cleft lip*—Example 3D face mesh data of children with cleft lip. Cleft of the lip can cause a wide range of deformations from mild to severe

applied to cleft assessment, Benz et al. introduced a method for 3D facial symmetry analysis using the iterative closest point algorithm to register the mirrored mesh to the original [5]. This method is reliable when the data is properly aligned and heavily depends on the choice of the initial plane about which the data is mirrored.

3 Craniofacial Image Analysis Pipeline

Our pipeline consists of data acquisition and preprocessing, feature extraction, and high-level operations of quantification, classification and content-based image retrieval. Figure 4 summarizes the steps of the pipeline and their relationships, as discussed in this section.

3.1 *Data Acquisition and Preprocessing*

With the developments in 3D imaging technologies, the use of 3D information has become widespread in research and applications. Craniofacial research greatly benefits from these developments due to the lower cost and higher accuracy of new imaging technologies like laser scanners and stereo-photography, in comparison to traditional methods such as direct measurements on the patient or 2D image-based techniques. Stereo imaging systems are popular among medical researchers, since they make it possible to collect large amounts of data in a non-invasive and convenient way.

The 3dMD® system is a commercial stereo-photography system commonly used for medical research. It uses texture information to produce a 3D mesh of the human face that consists of points and triangles. Figure 5 shows an example 3dMD setup where multiple pods of cameras are placed around a chair and simultaneously obtain photographs of the patient from different angles. Since the resulting mesh
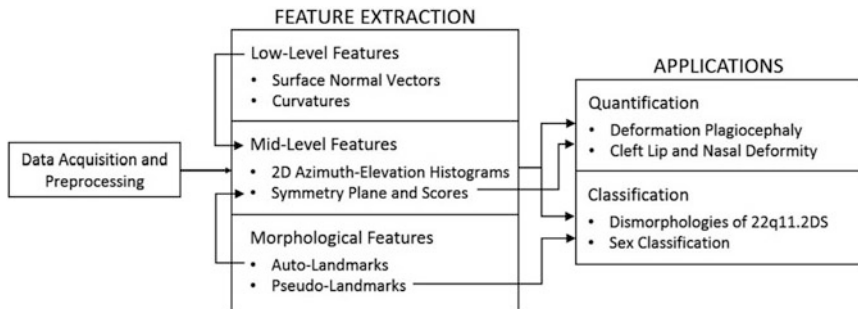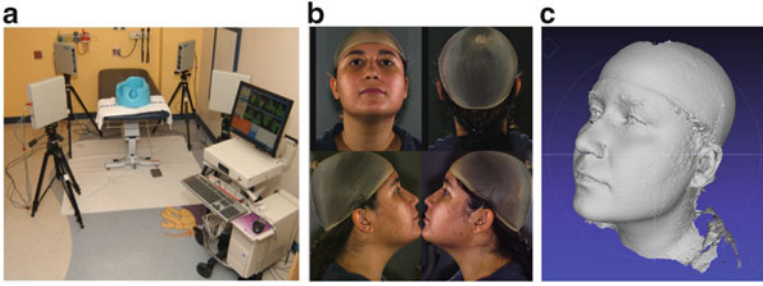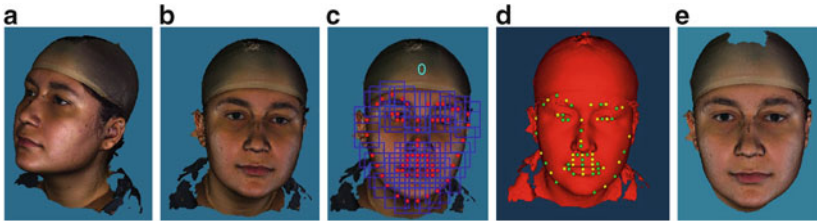

Fig. 4 The craniofacial image analysis pipeline overview

Fig. 5 *3D Face Mesh Acquisition*—(**a**) 3dMD® system with multiple cameras, (**b**) The texture images from four different cameras on front, back, left and right, (**c**) The 3D mesh produced with the stereophotography technique

Fig. 6 *Automated mesh cleaning*—(**a**) Original mesh acquired by 3DMD®, (**b**) Frontal snapshot of the rotated mesh, (**c**) Facial landmarks detected by 2D face detection algorithm, (**d**) Pose normalized mesh using landmarks and Procrustes superimposition, (**e**) Extracted face

is not aligned and contains pieces of clothing, it needs to be processed before further analysis. It is usually an expert who cleans the mesh to obtain the face and normalizes the pose, so that the head faces directly front. The accuracy and efficiency of this step is crucial to any analysis conducted on the data.

Wu et al. proposed a method for automated face extraction and pose-normalization from raw 3D data [25]. The method makes use of established face detection algorithms for 2D images using multiple photographs produced by the stereo system. Figure 6 shows the steps of the algorithm starting from original mesh (a). Using the local curvature around every point on the surface in a supervised learning algorithm, candidate points are obtained for the inner eye corners and the nose tip (Sect. 3.2.1 describes the calculation of the curvature values). The true eye corners and nose tip are selected to construct a triangle within some geometric limits. Using the eye-nose-eye triangle, the 3D mesh is rotated so that the eye regions are leveled and symmetric, and the nose appears right under the middle of the two eyes (Fig. 6b). The face detection algorithm proposed in [29] is used on a snapshot of the rotated data and a set of facial landmarks are obtained (not to be confused with anthropological landmarks used by medical experts) Fig. 6c. By projecting these 2D landmarks to the 3D mesh and using the Procrustes superimposition method [11], the mesh is rotated so that the distance between the landmarks of the head and the average landmarks of the aligned data is minimal (Fig. 6d). After alignment, the bounding box for the 3D surface is used to cut the clothing and obtain the

face region. Additionally, surface normal vectors are used to eliminate neck and shoulders from images where the bounding box is not small enough to capture only the face (Fig. 6e). The automation of mesh cleaning and pose normalization is an important step for processing large amounts of data with computational methods.

3.2 Feature Extraction

We describe both low-level and mid-level features used in craniofacial analysis.

3.2.1 Low-Level Features

A 3D mesh consists of a set of points, identified by their coordinates (x, y, z) in 3D space, and the connections between them. A cell of a mesh is a polygon defined by a set of points that form its boundary. The low-level operators capture local properties of the shape by computing a numeric value for every point or cell on the mesh surface. Low-level features can be averaged over local patches, aggregated into histograms as frequency representations and convoluted with a Gaussian filter to remove noise and smooth the values.

Surface Normal Vectors

Since the human head is roughly a sphere, the normal vectors can be quantified with a spherical coordinate system. Given the normal vector $n(n_x, n_y, n_z)$ at a 3D point, the azimuth angle θ is the angle between the positive x axis and the projection n' of n to the xz plane. The elevation angle ϕ is the angle between the x axis and the vector n.
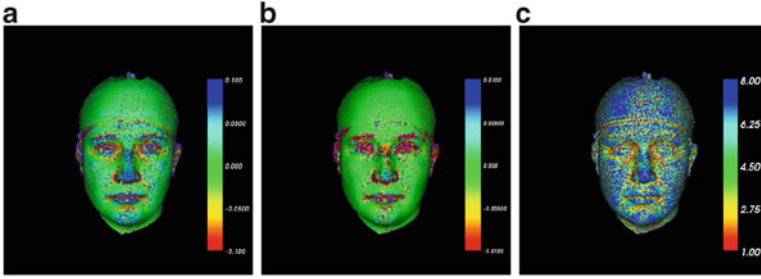
$$\theta = \arctan\left(\frac{n_z}{n_x}\right) \qquad\qquad \phi = \arctan\left(\frac{n_y}{\sqrt{(n_x^2 + n_z^2)}}\right) \qquad (1)$$

where $\theta \in [-\pi, \pi]$ and $\phi \in [-\frac{\pi}{2}, \frac{\pi}{2}]$.

Curvature

There are different measures of curvature of a surface: The *mean curvature H* at a point p is the weighted average over the edges between each pair of cells meeting at p.

$$H(p) = 1/|E(p)| \sum_{e \in E(p)} length(e) * angle(e) \qquad (2)$$

Fig. 7 *Curvature measures—*(**a**) Mean curvature, (**b**) Gaussian curvature and (**c**) Besl–Jain curvature visualized. Higher values are represented by cool (*blue*) colors while lower values are represented by warm (*red*) colors (Color figure online)

where $E(p)$ is the set of all the edges meeting at point p, and *angle(e)* is the angle of edge e at point p. The contribution of every edge is weighted by *length(e)*. The *Gaussian curvature K* at point p is the weighted sum of interior angles of the cells meeting at point p.

$$K(p) = 2\pi - \sum_{f \in F(p)} area(f)/3 * interior angle(f) \tag{3}$$

where $F(p)$ is the set of all the neighboring cells of point p, and *interiorangle(f)* is the angle of cell f at point p. The contribution of every cell is weighted by $area(f)/3$.

Besl and Jain [6] suggested a surface characterization of a point p using the sign of the mean curvature H and the Gaussian curvature K at point p. Their characterization includes eight categories: peak surface, ridge surface, saddle ridge surface, plane surface, minimal surface, saddle valley, valley surface and cupped surface. Figure 7 illustrates mean, Gaussian and Besl–Jain curvature on a head mesh.

3.2.2 Mid-Level Features

Mid-level features are built upon low-level features to interpret global or local shape properties that are difficult to capture with low-level features.

2D Azimuth-Elevation Histograms

Azimuth and elevation angles, together, can define any unit vector in 3D space. Using a 2D histogram, it is possible to represent the frequency of cells according to their orientation on the surface. On relatively flat surfaces of the head, all surface normal vectors point in the same direction. In this case, all vectors fall into the same bin creating a strong signal in some bins of the 2D histogram. Figure 8 shows the visualization of an example 8 × 8 histogram.
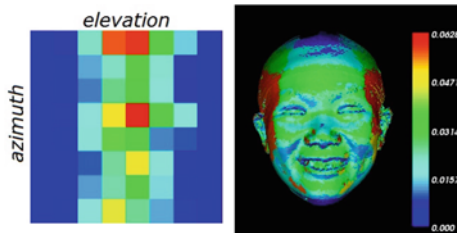
Fig. 8 Visualization of an 8×8 2D azimuth-elevation histogram. The histogram bins (*left*) with high values are shown with warmer colors (*red, yellow*). The image on *right* shows the localization of high-valued bins where the areas corresponding to bins are colored in a similar shade (Color figure online)

Fig. 9 (**a**) Symmetry plane, (**b**) *Front view* of grid showing θ and z for indexing the patches, (**c**) *Top view* of the grid showing the radius r and angle θ
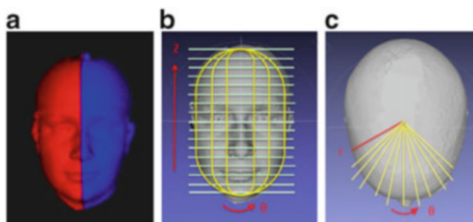
Symmetry Plane and Related Symmetry Scores

Researchers from computer vision and craniofacial study share an interest in the computation of human face symmetry. Symmetry analyses have been used for studying facial attractiveness, quantification of degree of asymmetry in individuals with craniofacial birth defects (before and after corrective surgery), and analysis of facial expression for human identification.

Wu et al. developed a two-step approach for quantifying the symmetry of the face [24]. The first step is to detect the plane of symmetry. Wu described several methods for symmetry plane detection and proposed two methods: learning the plane by using point-feature-based region detection and calculating the mid-sagittal plane using automatically-detected landmark points (Sect. 3.2.3). After detecting the plane, the second step is to calculate the shape difference between two parts of the face. Wu proposes four features based on a grid laid out on the face (Fig. 9):

1. Radius difference:

$$RD(\theta, z) = |r(\theta, z) - r(-\theta, z)| \qquad (4)$$

where $r(\theta, z)$ is the average radius value in the grid patch(θ, z), and $(-\theta, z)$ is the reflected grid patch of (θ, z) with respect to the symmetry plane.

2. Angle difference:

$$AD(\theta, z) = cos(\beta_{(\theta, z),(-\theta, z)}) \qquad (5)$$

where $\beta_{(\theta,z),(-\theta,z)}$ is the angle between the average surface normal vectors of each mesh grid patch (θ,z) and its reflected pair $(-\theta,z)$.

3. Gaussian curvature difference:

$$CD(\theta,z) = |K(\theta,z) - K(-\theta,z)| \tag{6}$$

where $K(\theta,z)$ is the average Gaussian curvature in the grid patch (θ,z), and $(-\theta,z)$ is the reflected grid patch of (θ,z) with respect to the symmetry plane.

4. Shape angle difference:

$$ED(\theta,z) = \left| \frac{\#points(\theta,z) > Th}{\#points(\theta,z)} - \frac{\#points(-\theta,z) > Th}{\#points(-\theta,z)} \right| \tag{7}$$

where Th is a threshold angle, and $\#points(\theta,z)$ is the total number of points with dihedral angle larger than Th in patch (θ,z).

These symmetry features produce a vector of length $M{\times}N$ where M is the number of horizontal grid cells and N is the number of vertical grid cells.

3.2.3 Morphometric Features

Most of the work on morphometrics in the craniofacial research community uses manually-marked landmarks to characterize the data. Usually, the data are aligned via these landmarks using the well-known Procrustes algorithm and can then be compared using the related Procrustes distance from the mean or between individuals [11]. Figure 10a shows a sample set of landmarks. Each landmark is placed by a medical expert using anatomical cues.
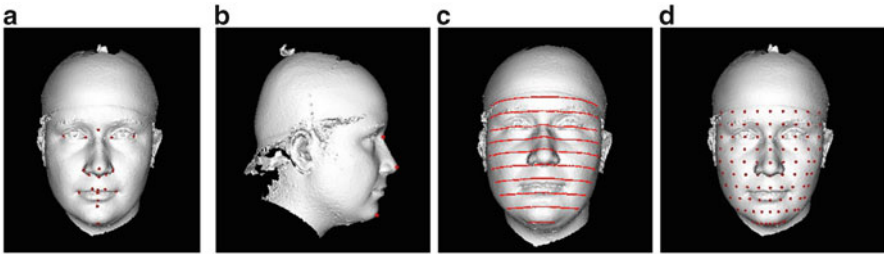
Fig. 10 (**a**) Twenty four anthropometric landmarks marked by human experts. (**b**) Sellion and tip of the chin are detected. (**c**) Two parallel planes go through chin tip and sellion, six parallel planes constructed between chin and sellion, and two above sellion. (**d**) On each plane, nine points are sampled with equal distances, placing the middle point on the bi-lateral symmetry plane. Ninety pseudo-landmark points calculated with ten planes and nine points

Auto-Landmarks

Traditional direct anthropometry using calipers is time consuming and invasive; it requires training of the expert and is prone to human error. The invasiveness of the method was overcome with the development of cost-effective 3D surface imaging technologies when experts started using digital human head data to obtain measurements . However, manual landmarking still presents a bottleneck during the analysis of large databases.

Liang et al. presented a method to automatically detect landmarks from 3D face surfaces [15]. The auto-landmarking method starts with computing an initial set of landmarks on each mesh using only the geometric information. Starting from a pose-normalized mesh, the geometric method finds 17 landmark points automatically, including 7 nose points, 4 eye points, 2 mouth points and 4 ear points. The geometric information used includes the local optima points like the tip of the nose (pronasale) or the sharp edges like the corners of the eyes. The sharp edges are calculated using the angle between the two surface normal vectors of two cells sharing an edge. The geometric method also uses information about the 17 landmarks and the human face such as the relative position of landmarks with respect to each other and the anatomical structures on the human face.

The initial landmark set is used for registering a template, which also has initial landmarks calculated, to each mesh using a deformable registration method [1]. The 17 landmark points (Fig. 11) provide a correspondence for the transformation of each face. When the template is deformed to the target mesh, the distance between the mesh and the deformed template is very small and every landmark point on the template can be transferred to the mesh. The average distance between the initial points generated by the geometric method and the expert points is 3.12 mm. This distance is reduced to 2.64 mm after deformable registration making the method very reliable. The method has no constraint on the number of landmarks that are marked on the template and transferred to each mesh. This provides a flexible work-flow for craniofacial experts who want to calculate a specific set of landmarks on large databases.
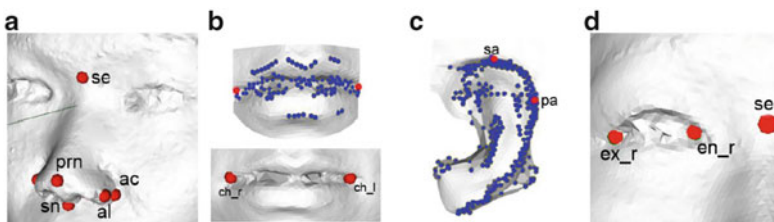
Fig. 11 Initial landmarks detected by the geometric method. (**a**) Nose points. (**b**) Mouth points. (**c**) Ear points. (**d**) Eye points

Pseudo-Landmarks

For large databases, hand landmarking is a very tedious and time consuming process
that the auto-landmarking method tries to automate. Moreover, anthropometric
landmarks cover only a small part of the face surface, and soft tissue like cheeks
or forehead do not have landmarks on them. This makes pseudo-landmarks an
attractive alternative. Hammond proposed a dense correspondence approach using
anthropometric landmarks [12]. The dense correspondence is obtained by warping
each surface mesh to average landmarks and finding the corresponding points using
an iterative closest point algorithm. At the end, each mesh has the same number of
points with the same connectivity and all points can be used as pseudo-landmarks.
Claes et al. proposed a method called the *anthropometric mask* [8], which is a
set of uniformly distributed points calculated on an average face from a healthy
population and deformed to fit onto a 3D face mesh. Both methods required manual
landmarking to initialize the process.

Motivated by the skull analysis work of Ruiz-Correa et al. [23], Lin et al.
[16] and Yang et al. [27], Mercan et al. proposed a very simple, but effective,
method that computes pseudo-landmarks by cutting through each 3D head mesh
with a set of horizontal planes and extracting a set of points from each plane [18].
Correspondences among heads are not required, and the user does no hand marking.
The method starts with 3D head meshes that have been pose-normalized to face
front. It computes two landmark points, the sellion and chin tip, and constructs
horizontal planes through these points. Using these two planes as base planes, it
constructs *m* parallel planes through the head and from each of them samples a set
of *n* points, where the parameters *n* and *m* are selected by the user. Figure 10 shows
90 pseudo-landmarks calculated with 10 planes and 9 points on a sample 3D mesh.
Mercan et al. show in [18] that pseudo-landmarks work as well as dense surface or
anthropometric mask methods, but they can be calculated without human input and
from any region of the face surface.

3.3   Quantification

Quantification refers to the assignment of a numeric score to the severity of a
disorder. We discuss two quantification experiments.

3.3.1   3D Head Shape Quantification for Deformational Plagiocephaly

Atmosukarto et al. used 2D histograms of azimuth-elevation angles to quantify
the severity of deformational plagiocephaly [4]. On relatively flat surfaces of the
head, normal vectors point in the same direction, and thus have similar azimuth and
elevation angles. By definition, infants with flat surfaces have larger flat areas on
their skulls causing peaks in 2D histograms of azimuth and elevation angles.
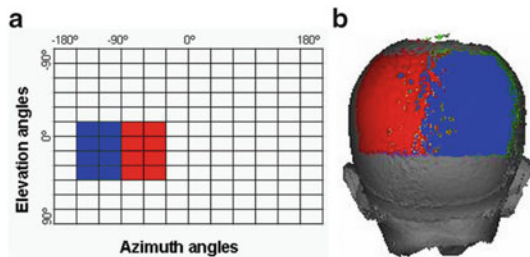
Fig. 12 On a 12 × 12 histogram (**a**), Left Posterior Flatness Score (*red*) and Right Posterior Flatness Scores (*blue*) are calculated by summing the relevant histogram bins. Selected bins correspond to points on the skull (**b**) that are relevant to the plagiocephaly (Color figure online)

Using a histogram with 12 × 12 bins, the method defines the sum of histogram bins corresponding to the combination of azimuth angles ranging from −90° to −30° and elevation angles ranging from −15° to 45° as the *Left Posterior Flatness Score* (LPFS). Similarly, the sum of histogram bins corresponding to the combination of azimuth angles ranging from −150° to −90° and elevation angles ranging from −15° to 45° gives the *Right Posterior Flatness Score* (RPFS). Figure 12 shows the selected bins and their projections on the back of the infant's head. The *asymmetry score* is defined as the difference between RPFS and LPFS. The asymmetry score measures the shape difference between two sides of the head, and the sign of the asymmetry score indicates which side is flatter.

The absolute value of the calculated asymmetry score was found to be correlated with experts' severity scores and the score calculated by Hutchinson's *HeadsUp* method [13] that uses anthropometric landmarks. Furthermore, the average flatness scores for left posterior flattening, right posterior flattening and control groups shows clear separation, providing a set of thresholds for distinguishing the cases.

3.3.2 Quantifying the Severity of Cleft Lip and Nasal Deformity

Quantifying the severity of a cleft is a hard problem even for medical experts. Wu et al. proposed a methodology based on symmetry features [26]. The method suggests that the asymmetry score is correlated with the severity of the cleft. It compares the scores with the severity of clefts assessed by surgeons before and after reconstruction surgery. Wu et al. proposed three measures based on asymmetry:

1. The *point-based distance* score is the average of the distances between points that are reflected around the symmetry plane:

$$PD_a = \frac{1}{n} \sum_p distance(p_s, q) \tag{8}$$

where n is the number of points and q is the reflection of point p.

2. The *grid-based radius distance* score is the average of the radius distance (RD) over the grid cells:

$$RD_a = \frac{1}{m \times m} \sum_{\theta, z} RD(\theta, z) \tag{9}$$

where m is the number of cells of a square grid, and RD is defined in (4).

3. The *grid-based angle distance* score is the average of the angle distance (AD) over the grid cells:

$$AD_a = \frac{1}{m \times m} \sum_{\theta, z} AD(\theta, z) \tag{10}$$

where m is the number of cells of a square grid, and AD is defined in (5).

Three distances are calculated for infants with clefts before and after surgery and compared with the rankings of the surgeons. The asymmetry scores indicate a significant improvement after the surgery and a strong correlation with surgeons' rankings. Figure 13 shows the visualization of RD_a scores for clefts with three different severity classes given by surgeons and the comparison of before and after surgery scores.
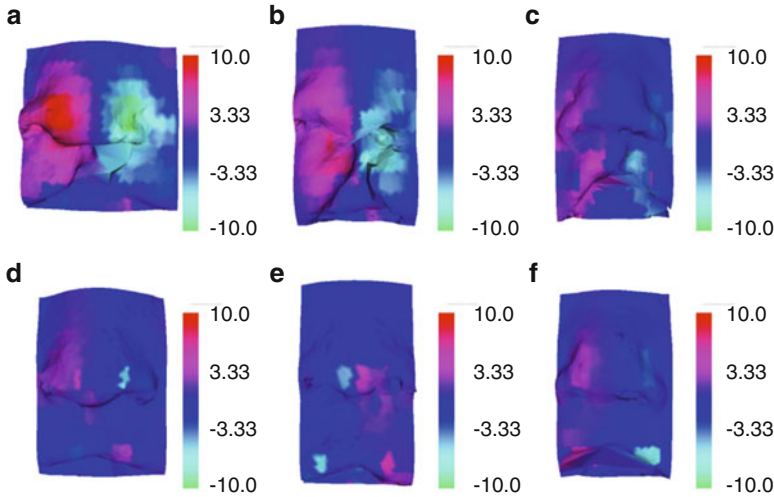
Fig. 13 RD_a reduction after the surgery for three cases. The *red* and *green* colors show the big difference between the *left* and *right* sides. *Red* means higher and *green* means lower. *Blue* means small difference between the two sides, (**a**) severe case pre-op $RD_a = 3.28$ mm, (**b**) moderate case pre-op $RD_a = 2.72$ mm, (**c**) mild case pre-op $RD_a = 1.64$ mm, (**d**) severe case post-op $RD_a = 1.03$ mm, (**e**) moderate case post-op $RD_a = 0.95$ mm, (**f**) mild case post-op $RD_a = 1.22$ mm (Color figure online)

3.4 Classification

We describe two classification experiments.

### 3.4.1 Classifying the Dismorphologies Associated with 22q11.2DS

The craniofacial features associated with 22q11.2 deletion syndrome are well-described and critical for detection in the clinical setting. Atmosukarto et al. proposed a method based on machine learning to classify and quantify some of these craniofacial features [3]. The method makes use of 2D histograms of azimuth and elevation angles of the surface normal vectors calculated from different regions of the face, but it uses machine learning instead of manually selecting histogram bins.

Using a visualization of the 2D azimuth-elevation angles histogram, Atmosukarto pointed out that certain bins in the histogram correspond to certain regions on the face, and the values in these bins are indicative of different face shapes. An example of different midface shapes is given in Fig. 14. Using this insight, a method based on sophisticated machine learning techniques was developed in order to learn the bins that are indicators of different craniofacial features.

In order to determine the histogram bins that are most discriminative in classification of craniofacial features, Adaboost learning was used to select the bins that give the highest classification performance of a certain craniofacial feature against others. The Adaboost algorithm is a strong classifier that combines a set of weak classifiers, in this case, decision stumps [9]. Different bins are selected for different craniofacial abnormalities. Note that the bins selected for each condition cover areas where the condition causes shape deformation.

After selecting discriminative histogram bins with Adaboost, a genetic programming approach [28] was used to combine the features. Genetic programming imitates human evolution by changing the mathematical expression over the selected histogram bins used for quantifying the facial abnormalities. The method aims to maximize a fitness function, which is selected as the F-measure in this work. The F-measure is commonly used in information retrieval and is defined as follows:

Fig. 14 Projections of 2D histograms of azimuth and elevation angles to the face. The projection shows discriminating patterns between individuals with and without midface hypoplasia
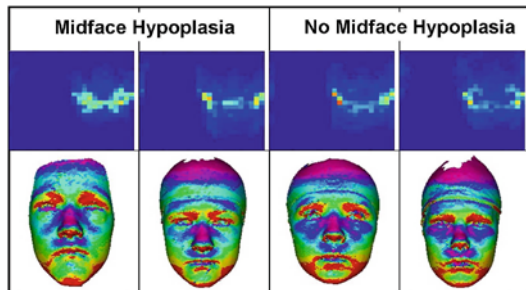
$$F(prec, rec) = 2 \times \frac{(prec \times rec)}{(prec + rec)} \tag{11}$$

where *prec* is the precision and *rec* is the recall metric. The mathematical expression with the highest F-measure is selected through cross-validation tests.

3.4.2 Sex Classification Using Pseudo-Landmarks

What makes a female face different from a male face has been an interest for computer vision and craniofacial research communities for quite some time. A great deal of previous work on sex classification in the computer vision literature uses 2D color or gray tone photographs rather than 3D meshes.

Mercan et al. used pseudo-landmarks in a classification setting to show their efficiency and the representation power over anthropometric landmarks [18]. L_1-regularized logistic regression was used in a binary classification setting where the features were simply the *x*, *y* and *z* coordinates of the landmark points. In a comparative study where several methods from the literature are compared in a sex classification experiment, it was shown that pseudo-landmarks (95.3 % accuracy) and dense surface models (95.6 % accuracy) perform better than anthropometric landmarks (92.5 % accuracy) but pseudo-landmarks are more efficient in calculation than dense surface models and do not require human input. L_1-regularization also provides feature selection and in the sex classification setting, the pseudo-landmarks around the eyebrows were selected as the most important features.

## 4 Content-Based Retrieval for 3D Human Face Meshes

The availability of large amounts of medical data made content based image retrieval systems useful for managing and accessing medical databases. Such retrieval systems help a clinician through the decision-making process by providing images of previous patients with similar conditions. In addition to clinical decision support, retrieval systems have been developed for teaching and research purposes [20].

Retrieval of 3D objects in a dataset is performed by calculating the distances between the feature vector of a query object and the feature vectors of all objects in the dataset. These distances give the dissimilarity between the query and every object in the dataset; thus, the objects are retrieved in the order of increasing distance. The retrieval performance depends on the features and the distance measure selected for the system. In order to evaluate the features introduced in Sect. 3.2, a synthetic database was created using the dense surface correspondence method [12]. 3D surface meshes of 907 healthy Caucasian individuals were used to create a synthetic database. The principle components of the data were calculated
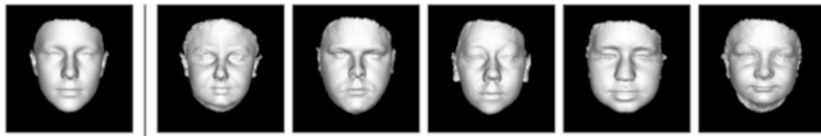
Fig. 15 Average face (*left*) and some examples (*right*) from the synthetic database
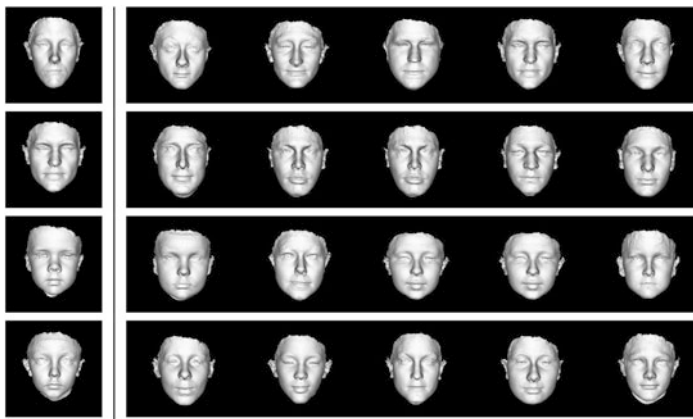

Fig. 16 Some queries made on the randomly produced synthetic dataset with the pseudo-landmark feature calculated from the whole face. The query is an adult female in the *first row*, an adult male in the *second row*, a young female in the *third row* and a young male in the *fourth row*

and 100 random synthetic faces were created by combining principle components with coefficients randomly chosen from a multivariate normal distribution modeling the population. Figure 15 shows the average face of the population and some example synthetic faces. Figure 16 shows four example queries made on the random dataset with adult female, adult male, young female and young male samples as queries. Although the retrieval results are similar to the query in terms of age and sex, it is not possible to evaluate the retrieval results quantitatively using randomly produced faces, since there is no "ground truth" for the similarity.

In a controlled experiment, the performance of the retrieval system can be measured by using a dataset that contains a subset of similar objects. Then, using the rank of the similar object in a query, a score based on the average normalized rank of relevant images [19] is calculated for each query:

$$score(q) = \frac{1}{N \times N_{rel}} \times \left(\sum_{i=1}^{N_{rel}} R_i - \frac{N_{rel} \times (N_{rel} + 1)}{2} \right) \qquad (12)$$

where N is the number of objects in the database, N_{rel} is the number of objects relevant to the query object q, and R_i is the rank assigned to i-th relevant object. The evaluation scores range from 0 to 1, where 0 is the best and indicates that all relevant
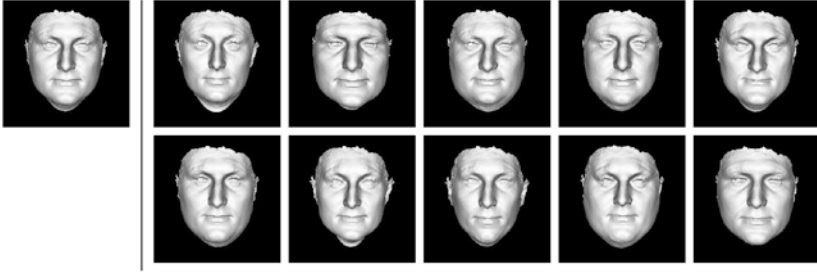
Fig. 17 A query face (*left*) and ten similar faces (*right*) produced by changing the coefficients of the principle components of the query face

Table 1 Evaluation results for different features and face regions in retrieval experiments on a synthetic database

	Face	Nose	Mouth	Eyes
Azimuth angles	0.077	0.150	0.226	0.168
Elevation angles	0.099	0.159	0.107	0.185
Gaussian curvature	0.223	0.386	0.333	0.385
2D azimuth-elevation histogram	0.046	0.109	0.087	0.142
Landmarks	0.164	NA	NA	NA
Pseudo-landmarks	0.102	0.028	0.041	0.094
Pseudo-landmarks (sized)	0.054	0.041	0.042	0.118

objects are retrieved before any other objects. To create similar faces in a controlled fashion, the coefficients of the principle components were selected carefully, and ten similar faces were produced for each query. For the synthesis of similar faces, we changed the coefficients of ten randomly chosen principle components of the base face by adding or subtracting 20 % of the original coefficient. These values are chosen experimentally by taking the limits of the population coefficients into consideration. Figure 17 shows a group of similar faces. Adding 10 new face sets with 1 query and 10 similar faces in each, a new dataset of 210 faces was obtained. The new larger dataset was used to evaluate the performance of shape features by running 10 queries for each feature-region pair. The features were calculated in four regions: the whole face, nose, mouth and eyes. Low-level features azimuth angles, elevation angles and curvature values were used to create histograms with 50 bins. 2D azimuth-elevation histograms were calculated at 8×8 resolution. Landmarks were calculated with our auto-landmarking technique. Pseudo-landmarks were calculated with 35 planes and 35 points. Both landmarks and pseudo-landmarks were aligned with Procrustes superimposition, and pseudo-landmarks were size normalized to remove the effect of shape size. Table 1 shows the average of the evaluation scores for each feature-region pair. Figure 18 shows a sample retrieval.
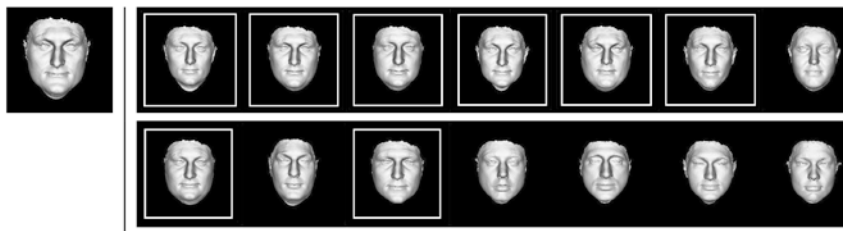
Fig. 18 Top 30 results of a query with the pseudo-landmark feature calculated from the whole face. The *top left* face is the query and the manually produced similar faces are marked with *white rectangles*. The faces without *white rectangle* are random faces in the database that happen to be similar to the query

The pseudo-landmarks obtained the best (lowest) retrieval scores for the nose, mouth and eyes, while the 2D azimuth-elevation angle histogram obtained the best score for the whole face. However, the sized pseudo-landmarks were a close second for whole faces.

5   Conclusions

This paper presents several techniques that automate the craniofacial image analysis pipeline and introduces methods to diagnose and quantify several different craniofacial syndromes. The pipeline starts with the preprocessing of raw 3D face meshes obtained by a stereo-photography system. Wu et al. [25] provided an automatic preprocessing method that normalizes the pose of the 3D mesh and extracts the face. After preprocessing, features can be calculated from the 3D face meshes, including azimuth and elevation angles, several curvature measures, symmetry scores [24], anthropometric landmarks [15] and pseudo-landmarks [18]. The extracted features have been used in the quantification of craniofacial syndromes and in classification tasks. Our new work, a content-based retrieval system built on multiple different features, was introduced, and the retrievals of similar faces from a synthetic database were evaluated.

Medical imaging has revolutionized medicine by enabling scientists to obtain lifesaving information about the human body—non-invasively. Digital images obtained through CT, MR, PET and other modalities have become standards for diagnosis and surgical planning. Computer vision and image analysis techniques are being used for enhancing images, detecting anomalies, visualizing data in different dimensions and guiding medical experts. New computational techniques for craniofacial analyses provide a fully automatic methodology that is powerful and efficient. The techniques covered in this paper do not require human supervision, provide objective and more accurate results, and make batch processing of large amounts of data possible.

References

1. Allen, B., Curless, B., & Popović, Z. (2003). The space of human body shapes: Reconstruction and parameterization from range scans. *ACM Transactions on Graphics, 22*, 587–594.
2. Ardinger, H. H., Buetow, K. H., Bell, G. I., Bardach, J., VanDemark, D., & Murray, J. (1989). Association of genetic variation of the transforming growth factor-alpha gene with cleft lip and palate. *American Journal of Human Genetics, 45*(3), 348.
3. Atmosukarto, I., Shapiro, L. G., & Heike, C. (2010). The use of genetic programming for learning 3d craniofacial shape quantifications. In *2010 20th International Conference on Pattern Recognition (ICPR)* (pp. 2444–2447). Los Alamitos, CA: IEEE Press.
4. Atmosukarto, I., Shapiro, L. G., Starr, J. R., Heike, C. L., Collett, B., Cunningham, M. L., et al. (2010). Three-dimensional head shape quantification for infants with and without deformational plagiocephaly. *The Cleft Palate-Craniofacial Journal, 47*(4), 368–377.
5. Benz, M., Laboureux, X., Maier, T., Nkenke, E., Seeger, S., Neukam, F. W., et al. (2002). The symmetry of faces. In *Vision Modeling and Visualization (VMV)* (pp. 43–50).
6. Besl, P. J., & Jain, R. C. (1985). Three-dimensional object recognition. *ACM Computing Surveys (CSUR), 17*(1), 75–145.
7. Boehringer, S., Vollmar, T., Tasse, C., Wurtz, R. P., Gillessen-Kaesbach, G., Horsthemke, B., et al. (2006). Syndrome identification based on 2d analysis software. *European Journal of Human Genetics, 14*(10), 1082–1089.
8. Claes, P., Walters, M., Vandermeulen, D., & Clement, J. G. (2011). Spatially-dense 3d facial asymmetry assessment in both typical and disordered growth. *Journal of Anatomy, 219*(4), 444–455.
9. Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences, 55*(1), 119–139.
10. Glasgow, T. S., Siddiqi, F., Hoff, C., & Young, P. C. (2007). Deformational plagiocephaly: Development of an objective measure and determination of its prevalence in primary care. *Journal of Craniofacial Surgery, 18*(1), 85–92.
11. Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika, 40*(1), 33–51.
12. Hammond, P., et al. (2007). The use of 3d face shape modelling in dysmorphology. *Archives of Disease in Childhood, 92*(12), 1120.
13. Hutchison, B. L., Hutchison, L. A., Thompson, J. M., & Mitchell, E. A. (2005). Quantification of plagiocephaly and brachycephaly in infants using a digital photographic technique. *The Cleft Palate-Craniofacial Journal, 42*(5), 539–547.
14. Lanche, S., Darvann, T. A., Ólafsdóttir, H., Hermann, N. V., Van Pelt, A. E., Govier, D., et al. (2007). A statistical model of head asymmetry in infants with deformational plagiocephaly. In *Image analysis* (pp. 898–907). Berlin: Springer.
15. Liang, S., Wu, J., Weinberg, S. M., & Shapiro, L. G. (2013). Improved detection of landmarks on 3d human face data. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 6482–6485). Los Alamitos, CA: IEEE Press.
16. Lin, H., Ruiz-Correa, S., Shapiro, L., Hing, A., Cunningham, M., Speltz, M., et al. (2006). Symbolic shape descriptors for classifying craniosynostosis deformations from skull imaging. In *27th Annual International Conference of the Engineering in Medicine and Biology Society, IEEE-EMBS 2005* (pp. 6325–6331). Los Alamitos, CA: IEEE Press.
17. McKinney, C. M., Cunningham, M. L., Holt, V. L., Leroux, B., & Starr, J. R. (2008). Characteristics of 2733 cases diagnosed with deformational plagiocephaly and changes in risk factors over time. *The Cleft Palate-Craniofacial Journal, 45*(2), 208–216.
18. Mercan, E., Shapiro, L. G., Weinberg, S. M., & Lee, S. I. (2013). The use of pseudo-landmarks for craniofacial analysis: A comparative study with l 1-regularized logistic regression. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 6083–6086). Los Alamitos, CA: IEEE Press.

19. Müller, H., Marchand-Maillet, S., & Pun, T. (2002). The truth about corel—evaluation in image retrieval. In *Proceedings of the Challenge of Image and Video Retrieval (CIVR2002)* (pp. 38–49).
20. Müller, H., Michoux, N., Bandon, D., & Geissbuhler, A. (2004). A review of content-based image retrieval systems in medical applications clinical benefits and future directions. *International Journal of Medical Informatics, 73*(1), 1–23.
21. Perez, E., & Sullivan, K. E. (2002). Chromosome 22q11. 2 deletion syndrome (digeorge and velocardiofacial syndromes). *Current Opinion in Pediatrics, 14*(6), 678–683.
22. Plank, L. H., Giavedoni, B., Lombardo, J. R., Geil, M. D., & Reisner, A. (2006). Comparison of infant head shape changes in deformational plagiocephaly following treatment with a cranial remolding orthosis using a noninvasive laser shape digitizer. *Journal of Craniofacial Surgery, 17*(6), 1084–1091.
23. Ruiz-Correa, S., Sze, R. W., Lin, H. J., Shapiro, L. G., Speltz, M. L., & Cunningham, M. L. (2005). Classifying craniosynostosis deformations by skull shape imaging. In *Proceedings of the 18th IEEE Symposium on Computer-Based Medical Systems, 2005* (pp. 335–340). Los Alamitos, CA: IEEE Press.
24. Wu, J., Tse, R., Heike, C. L., & Shapiro, L. G. (2011). Learning to compute the symmetry plane for human faces. In *Proceedings of the 2nd ACM Conference on Bioinformatics, Computational Biology and Biomedicine* (pp. 471–474). New York, NY: ACM Press.
25. Wu, J., Tse, R., & Shapiro, L. G. (2014). Automated face extraction and normalization of 3d mesh data. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Los Alamitos, CA: IEEE Press.
26. Wu, J., Tse, R., & Shapiro, L. G. (2014). Learning to rank the severity of unrepaired cleft lip nasal deformity on 3d mesh data. In *2014 24th International Conference on Pattern Recognition (ICPR)*. Los Alamitos, CA: IEEE Press.
27. Yang, S., Shapiro, L. G., Cunningham, M. L., Speltz, M., & Le, S. I. (2011). Classification and feature selection for craniosynostosis. In *Proceedings of the 2nd ACM Conference on Bioinformatics, Computational Biology and Biomedicine* (pp. 340–344). New York, NY: ACM Press.
28. Zhang, M., Bhowan, U., & Ny, B. (2007). Genetic programming for object detection: A two-phase approach with an improved fitness function. *Electronic Letters on Computer Vision and Image Analysis, 6*(1), 2007. URL http://elcvia.cvc.uab.es/article/view/135
29. Zhu, X., & Ramanan, D. (2012) Face detection, pose estimation, and landmark localization in the wild. In *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2879–2886). Los Alamitos, CA: IEEE Press.