Pitfalls and Promise of 3-dimensional Image Comparison for Craniofacial Surgical Assessment

Harold S. Matthews, PhD*†‡§
Jonathan A. Burge, MBChB¶∥
Pieter-Jan R. Verhelst, MD§**
Constantinus Politis, MD, DDS, MHA, MM, PhD§**
Peter D. Claes, PhD*†‡††
Anthony J. Penington, MD*¶‡‡

Summary: Three-dimensional (3D) photography is becoming widely used in plastic surgery. It provides an accurate and reproducible record of the facial surface anatomy and could be a versatile tool for treatment planning and assessment. However, the existing software tools available for the assessment of 3D facial imaging often give highly misleading results. The goal of this special topic article is to give clinicians an insight into methods of 3D image assessment and explain the reasons why results may be misleading. We point toward the advantages of an alternative approach using “nonrigid surface registration” for the comparison of pre- and postsurgical images. This approach is compared with the regular rigid surface registration, and this is illustrated by the assessment of a child with Crouzon syndrome before and after LeFort III osteotomy and distraction. Findings of the standard method imply that changes have occurred that are anatomically not possible, whereas the alternative approach indicates realistic changes. Furthermore, we demonstrate an exciting capacity of 3D image analysis to construct reference populations of normal head size and shape. These can be used to assess the parts of the head that are normal and abnormal pre- and posttreatment of the same child. We conclude that, while 3D image analysis has great potential in surgical assessment, existing software does not always give an adequate assessment. Collaboration among surgeons and engineering and computer science specialists should be encouraged. This way, more comprehensive and accurate techniques in patient assessment and surgical planning can be developed and applied in clinical practice.

INTRODUCTION

Three-dimensional (3D) photography is widely used in plastic and craniomaxillofacial surgery as an accurate and reproducible record of surface anatomy. It has the potential to be a versatile tool for assessment of outcomes and is increasingly used in treatment planning. For example, it could be used to monitor facial growth in cleft patients or to estimate the volume of implant required to correct breast or other asymmetries. Not all surgeons, however, have found the technology immediately useful. Sometimes apparently simple image analysis gives anomalous answers for reasons that are not obvious. Previous reviews have considered potential applications of 3D image analysis in craniofacial surgical assessment1 and the use of color visualizations of the differences between images in particular2 but have focused more on the potential than on the problems. The aim of this special topic is to give surgeons an understanding of 3D image analysis and to explain the reasons why results may be misleading. We point toward the advantages of an alternative approach using “nonrigid surface registration” for the comparison of pre- and postsurgical images. This approach is compared with the regular rigid surface registration, and this is illustrated by the assessment of a child with Crouzon syndrome before and after LeFort III osteotomy and distraction. Findings of the standard method imply that changes have occurred that are anatomically not possible, whereas the alternative approach indicates realistic changes. Furthermore, we demonstrate an exciting capacity of 3D image analysis to construct reference populations of normal head size and shape. These can be used to assess the parts of the head that are normal and abnormal pre- and posttreatment of the same child. We conclude that, while 3D image analysis has great potential in surgical assessment, existing software does not always give an adequate assessment. Collaboration among surgeons and engineering and computer science specialists should be encouraged. This way, more comprehensive and accurate techniques in patient assessment and surgical planning can be developed and applied in clinical practice.

From the *Facial Sciences Research Group, Murdoch Children’s Research Institute, Melbourne, Australia; †Department of Human Genetics, KU Leuven, Leuven, Belgium; ‡Medical Imaging Research Group, Department of Imaging and Pathology, Faculty of Medicine, KU Leuven, Leuven, Belgium; §Department of Plastic and Maxillofacial Surgery, Royal Children’s Hospital, Melbourne, Australia; ¶Melbourne Medical School, University of Melbourne, Melbourne, Australia; **Department of Oral and Maxillofacial Surgery, University Hospitals Leuven, Leuven, Belgium; ††Department of Electrical Engineering, ESAT/PSI, KU Leuven, Leuven, Belgium; and ‡‡Department of Pediatrics, University of Melbourne, Melbourne, Australia.

Received for publication February 10, 2020; accepted March 23, 2020.

Supported by the Royal Children’s Hospital Foundation (2013–127), the Jigsaw Foundation, the Australian National Health and Medical Research Council, and the Victorian Government Operational Infrastructure Program.

Copyright © 2020 The Authors. Published by Wolters Kluwer Health, Inc. on behalf of The American Society of Plastic Surgeons. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal. DOI: 10.1097/GOX.0000000000002847

Disclosure: The authors have no financial interest to declare in relation to the content of this article.
not just in craniofacial surgery but across the whole spectrum of plastic surgery.

**WHAT IS A 3D IMAGE?**

Most 3D photographs are produced by a process called stereo-photogrammetry. Such 3D images are a composite of anywhere from two to ten 2-dimensional (2D) images taken by different cameras positioned around the subject. Proprietary software integrates the 2D images into a 3D model of the surface. The 3D “image” actually consists of many thousands of points in space, distributed across the surface of the subject, linked into a “mesh” that defines that surface (Fig. 1). Information about color and texture are then applied to produce the smooth 3D image seen on the screen. The location of each point in space is defined by 3 values: the distances, in x, y, and z coordinates, from an arbitrarily defined zero point somewhere in the image space.

**COMPARISON OF 2 IMAGES**

The most common application of 3D imaging technology is to compare 2 images of the same individual taken on separate occasions, such as pre- and postsurgery. In general, for 2 images to be comparable and for the images to accurately represent the anatomy, they should both be taken with the subject displaying a neutral facial expression. The regions of interest (eg, the face) should also be present in both images. An example

![Fig. 1.](image)
is given in Figure 1 of a child with Crouzon syndrome who has undergone a LeFort III osteotomy and distraction along with hydroxyapatite cranioplasty and lateral canthoplasties. Consecutive 3D photographs, taken just before surgery at the age of 6 years and a year later at 7 years, were taken to quantify how the soft tissues were affected by surgery.

Comparison of such images is a 2-stage process. First, the 2 images need to be aligned or overlaid in some way to expose any difference in surface contour. Once this has been achieved, some objective measurement or quantification of the difference can be performed. At either stage, subtle errors can occur, which can produce misleading or confusing results.

**IMAGE ALIGNMENT**

Most imaging systems are shipped with software that allows 2 images to be represented and aligned with each other in a single workspace. Alignment can be achieved manually by the user manipulating and moving 1 image onto the other until they believe the 2 images are appropriately overlaid. Most systems also have an algorithm for automated or semiautomated image alignment. Either way, image alignment involves moving and rotating one image in 3 dimensions until it accurately overlays the other.

Given an appropriate software interface, most users do not find it hard to overlay one image onto the other in a meaningful way. Even though changes in overall shape will exist between the 2 images, the operator will instinctively avoid using the areas altered by surgery to align the image, preferentially using anatomical features distant from the surgery site to make the images correspond. Difficulties are most likely to arise when general changes unrelated to surgery have occurred in the shape of the head in the interval between images, such as those due to growth or change in weight. Manual alignment introduces subjectivity into the process and the potential for human error, so preferred using anatomical features distant from the surgery site to make the images correspond. Difficulties are most likely to arise when general changes unrelated to surgery have occurred in the shape of the head in the interval between images, such as those due to growth or change in weight. Manual alignment introduces subjectivity into the process and the potential for human error, so an automated or semiautomated process would generally be a better option if it is available.

Unfortunately, the results of automated alignment are not necessarily more accurate than those of a well-performed manual alignment. Alignment software typically applies an algorithm that minimizes the overall differences between the surface contours of the 2 images. However, this does not distinguish, as an expert user would, between areas that have been altered by surgery and those that have not. Without a method to identify specific changes due to surgery, automatic alignment may introduce systematic errors into subsequent analysis.

**MEASURING THE DIFFERENCE BETWEEN IMAGES**

Once the images have been aligned, quantification and representation of the contour differences can be performed. In some circumstances, such as in breast surgery, all that is wanted is the overall change in volume, but, in most applications, the surgeon would like to know the magnitude and direction of change at each point across the surface. Such changes are usually presented in the form of a “color map,” in which the color of the image at each point indicates some measure of difference at that point, usually distance between the 2 surfaces. In Figure 2, the default method of comparison used in most software packages has been applied to the images in Figure 1. Here the color indicates the distance from each point in the preimage to the nearest point in the postimage. The accompanying scale indicates the distance and direction of difference (outward or inward) represented by each color: red where the postoperative contour is further out, green where it is further in. Because, using the default settings of the software, the color scale is not centered with blue at zero, a portion of the blue-green spectrum actually indicates a slight positive (outward) difference. The histogram indicates the proportion of points on the face that correspond with each color.

Surgeons often assume that the difference between surfaces represents the change that has occurred due to surgery at any particular point of the face. It is important to understand why this is not necessarily the case. The color indicates the distance from each point on the preimage to the closest point on the postimage, not the corresponding anatomical point on the postimage. These 2 distances are not the same. The inset in Figure 2 shows that for most points on the nose and lip, the distance to the closest point on the postimage, shown by the black arrows, is very different from the distance to the corresponding anatomical point, indicated by the red arrows. Only the red arrows suggest the actual change in position of the tissue achieved by surgery. The degree to which the discrepancy between these 2 values distorts the measurement of surgical outcomes depends on the contour of the surface being analyzed. In some circumstances, it can produce highly misleading results. An example in Figure 2 is that the tip of the nose is being matched to the bridge of the nose, and thereby the displacement of this point due to surgery is significantly underestimated. Also, the lower lip appears to be displaced as much as the upper lip, which is clearly incorrect. This problem can be addressed by applying a technique known in image analysis as image “nonrigid registration.”

**NONRIGID REGISTRATION**

The term “registration” usually refers to “rigid registration.” This is the manual, or automatic alignment of one image to another, discussed above. The purpose of nonrigid registration is to ensure that each point on 2 or more such images is linked to each other in a way that is anatomically meaningful. In other words, each point defines the same anatomical location in each image. In the last 2 decades, effective and reproducible methods of nonrigid registration of 3D biological images have been developed. Although currently used only in the research context, nonrigid registration of images can produce much more accurate, reproducible, and intuitively interpretable measurement of surgical outcomes. It also opens the door to more sophisticated techniques of image analysis.

To understand nonrigid registration, we need to return to the idea that the “image” consists of thousands
of points in 3D space distributed across the surface of the object. When each 3D photograph of the same individual is taken, not only do the images have different locations and orientations in the computerized 3D “world,” but the surface of the face or head is also represented by a different number of points in each image. These points are scattered essentially randomly and have no consistent anatomical meaning. For example, point 768 may be on the tip of the nose in one image and on the cheek in another. To compare the entire surface of one image to the entire surface of the other in a way that captures anatomical changes, each surface must be represented by the same number of points and each numbered point must be in the same anatomical location in each image. Achieving this is the key outcome of a nonrigid registration. It is, in effect, an automated method of identifying “landmarks” on the face as has traditionally been performed by an expert anatomist. However, the landmarks are not just at important anatomical sites, but evenly distributed across the entire surface to be studied.

Various nonrigid registration methods have been described. An example (MeshMonk https://github.com/TheWebMonks/meshmonk) is illustrated in Figure 3. First an identical copy of the preimage is created (Gray image in Fig. 3). An algorithm is then applied, which gradually changes its shape to match that of the postimage. The key to retaining the biological meaning of each point during this warping algorithm (eg, ensuring that point 786 remains on the tip of the nose) is in that the copy is deformed toward the postimage slowly over many iterations. At the end of the process, there is a new version of the postimage that has exactly the same number of points as the preimage, and each numbered point on that image is in a location that matches the location on the preimage. This image can now be aligned and compared with the preimage. Anatomical correspondence can be verified by looking at known anatomical landmarks (see the arrows of Fig. 4 and the study by White et al9), an important quality-control measure.

Automatic image alignment after nonrigid registration is much more accurate than a regular alignment because it minimizes the overall distances between corresponding, rather than the closest points. The alignment can be further improved by using an algorithm that iteratively identifies those parts of the image that have undergone the most change, due to surgery in this case, and weighting them least in the alignment process. This is essentially what a surgeon would do instinctively in a manual alignment, but in an automated, reproducible manner.9 More importantly, nonrigid registration allows the generation of a color map that corresponds more closely with the effects of surgery because it indicates the change

![Color Map](A) ![Alignment](B)

**Fig. 2.** Standard visualization of contour difference in 3dMD patient after automatic alignment. “Color map” shows the color map produced by the software and indicates the distance between the 2 surfaces at each point. “Alignment” shows the 2 aligned surfaces. The inset illustrates the discrepancy between the closest points and the anatomically corresponding points. The anatomically corresponding points are defined manually, by the authors, for illustrative purposes.
between anatomical points on the images, not the nearest points. The improved analysis is shown in Figure 4. The left-hand image shows again the comparison of the 2 images in Figure 1 using standard software without non-rigid registration. The standard approach appears to show that little or no change has occurred in the midface, the area where maximum advancement is expected, while the chin seems to have moved backwards. On the right side, analysis of exactly the same images using a nonrigid registration demonstrates a much more realistic picture of the changes due to surgery. Just as with standard software, the colors of the color map created using the nonrigid registration indicate the postoperative change in millimeters. The expected forward movement of the midface relative to the rest of the head is clearly shown, and the nose has been carried forward with it. The apparent backward movement of the chin shown using the standard approach is no longer seen.

COMBINING MULTIPLE IMAGES TO PRODUCE 3D POPULATION NORMS

A further advantage of nonrigid image registration is that it allows the simultaneous analysis of more than 2 images. This allows meaningful comparisons to be made between populations and between individuals and populations. When images of multiple faces are all registered with a single reference face, the average location of each point across the surface can be used to generate an “average” face shape of that population. Measures of variation, such as standard deviations (SDs), can also be calculated and displayed as a color map. Given a set of 3D images of a population of individuals, such as a normal population or a group of patients with a particular syndrome, a normal average or a syndrome-specific average face can be generated. Such “average” faces have many potential applications in both research and clinical practice. These can be used as a frame of references against which the outcomes of surgery can be assessed.

The most obvious application is to assess how close to the population norm a particular face is either before or following surgery. This can be used to measure how effective surgery has been in bringing the face shape closer to what is “normal” for the patient’s appropriate reference population. There are various ways in which an individual face can be compared with established population norms. The simplest way is to create a color map, which shows the absolute distance of each point in the face from the population mean (Fig. 5, third column). Alternatively, a color map can be created to show which areas of the face lie within or outside a certain threshold, such as 2 SDs from the population mean. A cutoff z score is therefore used to determine if a distance between anatomically corresponding points of the image surfaces lies within or outside the accepted range of values.

A significant issue in the analysis of the outcomes of surgery in children is the change in face and head shape that occurs with normal growth. Pre- and postsurgery images are typically taken several months or more apart, and during this interval, the shape of the face and head would be expected to change, even if there had been no surgical intervention. Using a database of 3D images of normal children, we previously developed a method of creating an appropriate age-specific average head shape along with SDs at each point. This can be used to
make an allowance for changes that would be expected to have occurred through normal growth in the interval between images. A similar technique could be used to model the effects of aging in adults.

Figure 5 shows an example of how this type of analysis could be used to audit the results of surgery. The images of the child from Figure 1 have been aligned to age-appropriate references and compared. The first column shows the average faces that have been calculated specifically for the 2 ages at which pre- and postoperative images were taken. The second column shows the patient aligned to the age-appropriate average face. The third column shows the distance between the patient and the population norm, and the fourth column indicates which regions are statistically abnormal. Those areas that are beyond the cutoff z scores are colored red (>2 SD) or green (<−2 SD) and those in the normal range are colored blue. We can now see how the LeFort III osteotomy and distraction has brought the midface from outside the normal range to somewhere close to the population mean (the “Distance” color map becomes less green and moves toward blue) and that this sits in the range of normal variation. The hydroxyapatite onlay graft to the forehead, as expected, has not been sufficient to bring the forehead shape into the normal range.

The change in the position of the chin, relative to population norms, may be due to a postural change in the mandible, possibly related to the altered occlusion. It may also be that the growth in the child’s lower face has failed to keep up with average population growth, which is significant at this age.15,16

CONCLUSIONS AND FUTURE WORK

The aim of many procedures in plastic and craniofacial surgery is to correct or enhance the human form, but assessment of the outcomes of such procedures is a significant challenge. CT scans are an excellent measure of outcome in craniofacial surgery, but radiation exposure and cost preclude their routine use and they are of little value in soft tissue procedures. Two-dimensional photographs are the most widely used
measure of outcome, but slight changes in photograph angle and lighting are known to significantly alter assessor interpretation. 3D photography has the potential to overcome most, if not all, of these problems. As this technology becomes cheaper and more widely available, it will inevitably replace other methods of assessment. It is vital that surgeons understand both its potential and its limitations.

A 3D image provides an accurate and objective measure of the surface contour of a body part, and so can act as a permanent record of pre- and postoperative body shape. Linear measurements and simple comparisons can be performed using software typically shipped with devices, but there are definite pitfalls that must be avoided. Some of these pitfalls have been described in this article, along with solutions that have recently been developed. Although these solutions are not able to be applied by a nonexpert user, the necessary software is open source and can be implemented by an individual with some programming expertise.

Among the most potentially powerful techniques described in this review are those that apply to populations of images. The ability to compare surgical outcomes to population norms has only previously been possible by comparing linear distances between landmarks against tabulated measurements. The comparison of the whole surface of the face and head, or indeed any body part, against population norms represents a significant advance in this field. This underlines the importance of the collection of databases of 3D images for different populations, across different ages, locations, and pathologies. Collecting such databases will require strong international collaborations.

Finally, 3D image analysis is playing an increasing role in surgical planning, making consequences of errors much more significant. This makes it all the more important for surgeons to understand the principles of 3D image analysis, rather than becoming passive users of proprietary software.

**CONCLUSIONS**

3D image analysis has huge potential in plastic and reconstructive surgery, but the techniques required to analyze 3D images are still in development. There are significant potential pitfalls in image comparison, which can be avoided by understanding the principles of image analysis and by implementation of new, automated methods of image analysis.

**REFERENCES**

1. Kau CH, Richmond S, Incrapera A, et al. Three-dimensional surface acquisition systems for the study of facial morphology and their application to maxillofacial surgery. *Int J Med Robot*. 2007;3:97–110.
2. Jayaratne YS, Zwahlen RA, Lo J, et al. Three-dimensional color maps: a novel tool for assessing craniofacial changes. *Surg Innov*. 2010;17:198–205.
3. Maal TJ, van Loon B, Plooij JM, et al. Registration of 3-dimensional facial photographs for clinical use. *J Oral Maxillofac Surg*. 2010;68:2391–2401.
4. Moghaddam MB, Brown TM, Clausen A, et al. Outcome analysis after helmet therapy using 3D photogrammetry in patients with deformational plagiocephaly: the role of root mean square. *J Plast Reconstr Aesthet Surg*. 2014;67:159–165.
5. van Heerbeek N, Ingels KJ, van Loon B, et al. Three dimensional measurement of rhinoplasty results. *Rhinology*. 2009;47:121–125.
6. Claes P, Walters M, Clement J. Improved facial outcome assessment using a 3D anthropometric mask. *Int J Oral Maxillofac Surg*. 2012;41:324–330.
7. Walters M, Claes P, Kakulas E, et al. Robust and regional 3D facial asymmetry assessment in hemimandibular hyperplasia and hemimandibular elongation anomalies. *Int J Oral Maxillofac Surg*. 2013;42:36–42.
8. White JD, Ortega-Castrillón A, Matthews H, et al. MeshMonk: open-source large-scale intensive 3D phenotyping. *Sci Rep*. 2019;9:6085.
9. Claes P, Daniels K, Walters M, et al. Dysmorphometrics: the modelling of morphological abnormality. *Theor Biol Med Model*. 2012;9:5.
10. Matthews HS. Changing the face of craniofacial growth curves: modelling growth and sexual dimorphism in children and adolescents using spatially dense 3D image analysis. [PhD thesis]. Melbourne: University of Melbourne;2018.
11. Shaweesh AI, Clement JG, Thomas CD, et al. Construction and use of facial archetypes in anthropology and syndrome diagnosis. *Forensic Sci Int*. 2006;159 (Suppl 1):S175–S185.
12. Kau CH, Zhurov A, Richmond S, et al. The 3-dimensional construction of the average 11-year-old child face: a clinical evaluation and application. *J Oral Maxillofac Surg*. 2006;64:1086–1092.
13. Hammond P. The use of 3D face shape modelling in dysmorphology. *Arch Dis Child*. 2007;92:1120–1126.
14. Hammond P, Suttie M. Large-scale objective phenotyping of 3D facial morphology. *Hum Mutat*. 2012;33:817–825.
15. Hammond P, Suttie M, Hennekam RC, et al. The face signature of fibrodyplasia ossificans progressiva. *Am J Med Genet A*. 2012;158A:1368–1380.
16. Hammond P, Hennes F, Suttie M, et al. Fine-grained facial phenotype-genotype analysis in Wolf-Hirschhorn syndrome. *Eur J Hum Genet*. 2012;20:33–40.
17. Matthews HS, Penington AJ, Hardiman R, et al. Modelling 3D craniofacial growth trajectories for population comparison and classification illustrated using sex-differences. *Sci Rep*. 2018;8:4771.
18. Fan Y, Matthews H, Kilpatrick N, et al. Facial morphology and growth following surgery for congenital midline cervical cleft patients. *Int J Oral Maxillofac Surg*. 2018;47:437–441.
19. Sommer DD, Mendelsohn M. Pitfalls of nonstandardized photography in facial plastic surgery patients. *Plast Reconstr Surg*. 2004;114:10–14.
