

Digital Exploration of Ethnic Facial Variation

Sawitree Wisetchat¹, Lisa DeBruine², and Daniel Livingstone³

¹ Glasgow School of Art, Glasgow, UK
sawitreeDesigns@gmail.com

² University of Glasgow, Glasgow, UK
lisa.debruine@glasgow.ac.uk

³ Glasgow School of Art, Glasgow, UK
d.livingstone@gsa.ac.uk

Abstract. The characteristic patterns of ethnic variation of the human face are seldom explicitly described. This paper addresses the question of how these characteristics may be explicitly described. A lexicon of 77 discrete 3D facial attributes has been developed for purposes of describing ethnic differences. While conventional descriptors are usually discrete-valued (categorical), these attributes are continuous-valued, each interpolated within a bounded range represented by a pair of basis shapes. They are also quasi-orthogonal, allowing for arbitrary combinations to reconstruct the shape of a mesh face to represent any of a large space of ethnicities. Unlike holistic representational schemes used for facial recognition, these attributes are local and ‘semantic’ (human-understandable), providing an intuitive basis for describing and visualizing the space of ethnic variation.

Keywords: ethnicity description • 3D modeling • face space.

1 Introduction

The human face varies with ethnicity around the world. Despite their individual differences, the individuals within each geographic locale share facial characteristics indicative of their ethnicity. The ethnic variation of the human face is seldom explicitly addressed in education. It would be of great value to foster the appreciation of the face as telling the story of the commonality of all of humankind and the diversity in our global distribution. Faces tell about origins and cultures. The language with which a face tells this story should be taught. What would constitute a descriptive language to express and communicate these differences? This paper summarizes a study by Wisetchat [1] as it applies to teaching an appreciation for how ethnicity is reflected in the human face. By what means are these characteristic ethnic differences to be described? It is a language not of words but of shapes, specifically three-dimensional shapes. Modern technology enables immersive visualization of three-dimensional shape in compelling ways that facilitate our learning a language with which to describe faces. An interactive animation framework is introduced that allows exploration of the space of ethnic variation via a set of intuitive, human understandable, facial shape properties.

This paper emphasizes a matter of some pedagogical importance: how to convey in concrete and understandable terms shape qualities that, while visually apparent and rec-

ognizable, are seldom described explicitly. The shape of the nose, for example, is usually described by reference to exemplars, typically extremes such as a ‘bulbous’ or a ‘beak-like’ nose. Similarly, terms such as ‘high cheekbones’ or a ‘square jaw’ may be helpful in describing the particularly distinctive features of an individual but of little help in developing an appreciation for how ethnicities differ visually, and in what regards individuals of a given ethnicity share some facial characteristics. Instead, there is heavy reliance on providing examples to supplement any text. Images and graphical depictions, while evocative, are not descriptions. A few more formal descriptive terms (e.g., *dolichofacial*, *mesognathous*) have been adopted to categorize facial dimensions and proportions based on anthropometric measurements between facial landmarks, however few such measurements reliably distinguish between even very different ethnicities [2-4]. Moreover, few facial features are diagnostic of one or another ethnicity by their presence or absence. The epicanthal fold, for example, is present in Asian eyes and generally absent in Europeans and Africans, however most other facial features differ only by degree across ethnicities, and often vary to a similar degree within an ethnicity. Virtually every aspect of a face is subject to variation, both among individuals within a population and across geographically distant populations: the dimensionality of the space of possible faces (however it is parameterized) is very great [5, 6].

In attempting to appreciate the differences in such high-dimensional data, one encounters many of the challenges common to data visualization in general. This is regarded as unavoidable since facial variation is both complex and subtle, and certainly ethnicity cannot be trivially reduced to a small number of factors. Fortunately, as discussed below, technology adopted from digital facial animation permits composing a specific face based on a *description* (a set of attribute-value pairs in a specialized lexicon) then *visualizing* that face in three dimensions. The appearance of a specific combination of facial attributes serves to capture characteristics of an ethnotype, and by ‘morphing’ from one to another, the differences between two ethnotypes can be very vividly conveyed.

1.1 An ‘Ethnicity Face Space’

Since faces generally vary simultaneously along many dimensions, a representation that spans the range of ethnicities would constitute a high-dimensional ‘face space’, where the specific choice of dimensions for this face space depends on the specific application. We distinguish two types of face space: an identity face space (IFS) for representing individual variations within a homogeneous population of faces, and an ethnicity face space (EFS) for representing variations of faces across ethnotypes. An IFS is intended for detecting the identity of an individual while an EFS is intended to represent facial characteristics. Both are similarity spaces [7, 8], but otherwise there is little reason to expect they share dimensions or indeed dimensionality. In our application of an EFS, we are not concerned with ethnicity *recognition* (the analog of IFS-based face recognition), but rather, ethnicity *description*. To visually explore facial variation across ethnicities, any proposed scheme needs to span the range of ethnic variation and to facilitate visualization of similarities and distinctions between ethnicities.

The concept of an *average* face is central to an IFS: it constitutes the origin of a space in which to map individual faces as *variations* on a mean [5], [7]. A core presumption of an IFS is that the sample set of faces is homogeneous and that individuals

are normally distributed about the sample mean along each IFS dimension, placing the average face at the origin of the IFS [7]. Since our EFS is not intended for ethnicity detection (the analog of face detection with an IFS), there is no need for a central-tendency presumption. Moreover, we avoid problems of measuring a global mean on which to map ethnicities as variations. A global mean across all facial ethnotypes would exhibit very large variance, since within-ethnicity variance in facial dimensions has been shown to obscure most across-ethnicity differences, even between very different selected ethnotypes [2]. Finally, in not attempting to measure a global mean, one avoids issues of sample bias, since some very populous ethnicities would greatly overshadow other important but smaller ethnicities. Fortunately, for our purposes, ethnotypes need not be described as variations on a global sample mean of ethnicities.

An EFS for face description has two core characteristics: 1) it represents faces with absolute rather than relative dimensions or axes, and 2) the dimensions are semantically meaningful, allowing the space to be ‘navigable’ by humans. In both regards the EFS is starkly different from an IFS, for which the dimensions correspond to relative differences, and being the eigenvectors determined by principal components analysis (PCA) on some training set, those dimensions are ‘holistic’ measurements derived across the entire face (e.g., ‘eigenfaces’) and not semantically meaningful [8], [10, 11].

Anthropometric face measurements (e.g., the width of the nose) are local, individually interpretable, and intuitive. They are also ad hoc, as they reflect what are regarded as useful as well as being efficiently and reliably measured [11], rather than any arguably complete, principled set of measurements. Likewise, fiducial points for image registration and delineation are based on an ad hoc set of convenient and conventional image landmarks [1]. In the same regard, a set of shape descriptors would be expected to be ad hoc, and while useful, not mathematically comprehensive. But unlike conventional anthropometry, the goal will be to capture subtleties of facial shape that are often salient indicators of ethnicity. The goal of creating such a set of descriptors is utility, not completeness. Advancements in scanning technology permit anthropometric data collection within increasingly fine resolution and spatial precision, but that data comprises measurements (e.g., two- or three-dimensional position information plus color and texture). While the shape of the face remains implicit in these dense sets of measurements, they are of great value in modeling (Section 4).

1.2 Summary of Development of the Descriptive Vocabulary

As surveyed below, a set of facial attributes with which to describe ethnic variation was identified (Section 2), then represented as three-dimensional basis shapes (Section 3), which were used as deformers of a polygonal mesh as controlled through an interactive user interface called the Ethnicity Modeler (Section 4). The development followed the spiral model methodology [12] since the process of converging on a representation scheme involved successive refinement and experimentation in order to defining a set of attributes that were additive, i.e., able to be assigned values in arbitrary combination and able to represent a broad range for each attribute, as discussed in Section 3. Once implemented, evaluation included a repeated-trial experiment to measure attribute precision, an accuracy study for some calibrated attributes, and user interface usability and applications (Section 5).

2 Identifying Facial Attributes

The human face can be regarded as having separable regions (eyes, nose, mouth, etc.) which can be measured and analyzed individually [13]. Likewise, the geometric complexity of the face is primarily contained within those regions with few facial characteristics defined along the borders between regions or spanning across regions. Therefore, an EFS can be regarded as composed of independent subspaces for the nose, eyes, and so forth. This simplified the search for attributes to one region at a time.

While individual faces vary considerably within an ethnicity, image averaging of multiple individuals of a given ethnotype reveals common characteristics associated with each ethnicity [14, 15]. To identify salient facial attributes, two-dimensional image averaging [14] was first performed on sets of photographs of individuals of differing ethnicities (primarily East Asian, African, and European). Movies were then created showing continuous blending transformation between the two images (e.g., an averaged East Asian ‘morphing’ into an averaged European), which were then analyzed region-by-region.

Consider for example the region of the eye in isolation. The image-averaged eyes of three ethnicities in Fig. 1 illustrate subtle, yet salient, ethnic *shape* differences. These attributes correspond to anatomical soft tissue shape features such as the epicanthal fold (ECF), the supratarsal fold (STF) and the superior palpebral sulcus (SPS), in addition to the conventional anthropometric properties of the palpebral fissure (width, length, inclination, canthal inclination, interpupillary distance, etc.). The eye region alone exhibits at least 20 such attributes that differ across ethnicities (as well as individuals), ten of which are identified in Fig. 1.)

In order that facial attributes may serve as EFS dimensions, they should be orthogonal (independent). Unlike dimensions based on individual fiducial (2D or 3D) sample points, facial attributes commonly share landmarks, and thus are not strictly independent. But as descriptive attributes, however, they are ‘quasi-orthogonal’, i.e., each can vary substantially independently of other attributes. For example, a supratarsal fold (STF) in the upper eyelid can co-occur with either a deep SPS or prominent SPC, despite their spatially overlapping (Fig. 1).

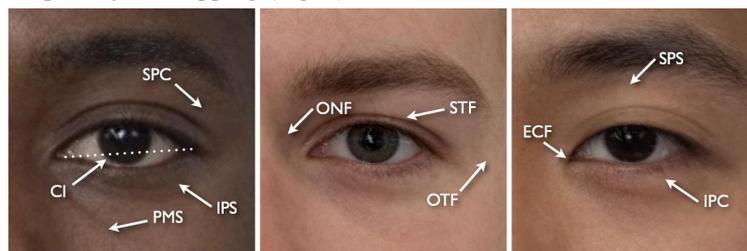


Fig. 1. Averaged images of three ethnicities (African, European, and East Asian), with ten salient shape attributes indicated: CI = canthal inclination, ECF = epicanthal fold, IPC = inferior palpebral convexity, IPS = inferior palpebral sulcus, ONF = orbitonasal fossa, OTF = orbitotemporal fossa, PMS = palpebromalar sulcus, SPC = superior palpebral convexity, SPS = superior palpebral sulcus, and STF = supratarsal fold.

Having identified a salient facial attribute for some facial region, we are concerned with capturing its absolute range of variation across a span of ethnicities, without an expectation that the distribution of this attribute exhibits a central tendency or mean. The origin of the EFS (corresponding to the default shape of the face) is simply a starting point for describing facial variations. Emphasis here is on creating a descriptive space for the nose, eyes, etc. that spans human ethnic (if not individual) variation.

3 Representing Facial Attributes by Basis Shapes

Linear-weighted summation underlies the synthesis of faces in 2D [16] or in 3D [17, 18] from a set of orthogonal basis functions derived from PCA. Linear-weighted summation also underlies the well-established technique of blendshape deformation in facial animation, wherein a ‘base’ 3D mesh (the default facial expression) is deformed by weighted combination of a number of basis shapes or ‘targets’ that share the same topology but with shape variations representing different facial expressions [19, 20]. Unlike the 3D ‘eigenmeshes’ derived from PCA, blendshape targets for digital animation are seldom orthogonal. Multiple basis shapes may influence the same vertices in the base and their combined effect on the base mesh frequently results in adverse interactions that require corrective measures [21]. The summation of multiple shapes, or shape-additivity, is a highly-intuitive means to create a shape as by graded combination of shape extremes. In facial animation, the set of targets constitute the dimensions of a multidimensional ‘expression space’, each dimension of which corresponds to local changes to the mesh topography, and are separately interpretable i.e., semantic.

In the application of basis shapes to represent the dimensions of ethnic variations in facial shape, the process of developing a set of meshes, each representing a facial attribute, shares many of the frustrations of creating blend shapes for digital animation, however the technique permits creating precise representations of shape features in various combination, matching samples of varying ethnicity, as will be discussed. The benefits of achieving continuous, graded ‘shape additivity’ outweigh the difficulties of overcoming the consequences of the basis shapes being not strictly independent in their combined effect.

To illustrate the representation of local shape attributes as basis shapes, eight of the ten attributes labeled in Fig. 1 are shown in the top and middle rows of Fig. 2. The extreme for each attribute is represented as a 3D mesh (such as the most pronounced ECF or STF). The bottom row shows four distinct linear combinations of the eight eye

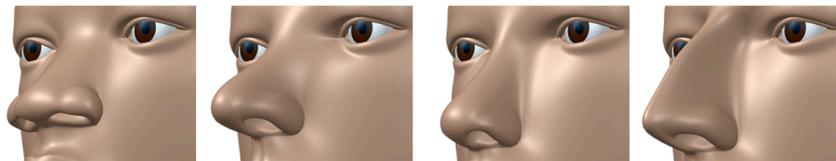


Fig. 3. Various combinations of nose attributes. About 20 nose attributes were identified, some corresponding to conventional scalar landmarks and others to local shape variations usually not measured anthropometrically. In combination, these attributes create a very large space of nose shapes and dimensions.

attributes. Note that these are unsigned attributes varying from 0.0 (absent) to 1.0 (maximum). Other attributes are signed, and vary between -1.0 (minimum) to 1.0 (maximum).

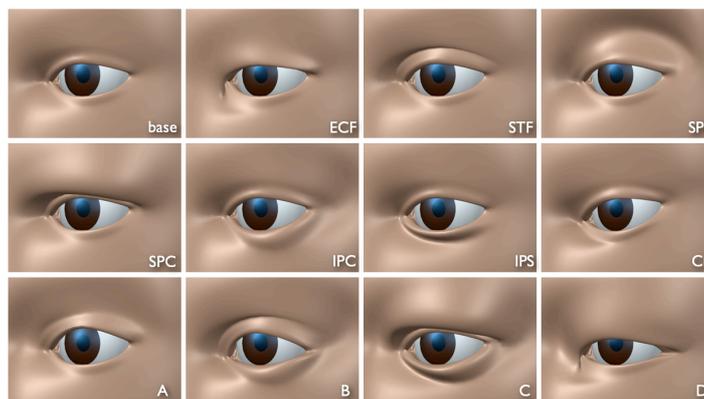


Fig. 2. The top and middle rows show representations of eight ethnically-significant eye attributes as basis shapes, each depicting an extreme value for that attribute (such as the pronounced depth of the superior palpebral sulcus, or SPS, above the upper eyelid). The bottom row shows the result of additivity of these shape attributes: A = STF + SPS; B = STF + OTF + IPC; C = SPC + IPC + IPS + ONF + OTF; D = ECF + SPC + ONF.

Ethnic variations in other facial regions are also amenable to modeling by linear summation of the quasi-orthogonal basis shapes. For example, Fig. 3 shows four distinct noses composed by combinations of different weights for twelve basis shapes. These models capture anthropometric dimensions between conventional landmarks, plus shape features such as alar contour depth, columellar show, alar prominence, etc. Quantitative calibration of the resultant shape permits matching the model to traditional anthropometric landmarks, whether derived from the literature (e.g., [1]), from averaged 2D images of ethnicity samples [14, 15], or by matching 3D data of ethnicity samples (Section 4).

Since facial regions are spatially separable to a significant degree, few basis shapes interfere across regions. Interference does arise within a facial region due to sharing surface patches, e.g., the complexity of folding in the tarsus of the upper eyelid (Fig. 2), but those issues were identified and eliminated, or at least minimized, through extensive testing of combinations of weights across the set of basis shapes. In total, about one hundred basis shapes were created across six facial regions that together influence the shape of a polygonal model for the entire face to visualize a large range of ethnicities. Additional attributes can be introduced as required in order to create finer shape distinctions.

4 Ethnicity Models

The *ethnicity modeler* EM provides an interface of about one hundred parametric controls, each directly corresponding to an EFS dimension, and implemented in Autodesk Maya using blendshapes (Fig. 4 shows the controls for the eye attributes, for example). The EM was then used to reconstruct the surface geometry of averaged face meshes representing different ethnicities. Each sample mesh consisted of the average of twenty individuals of common ethnicity and gender generated using the Di3D stereo-photogrammetry surface imaging system [22, 23] with delineation using MorphAnalyser 2.4.0 [24]. The scan data was imported into the EM and aligned such that the eyes in the scan were superimposed on those of the parametric model, then the shape parameters of the EM were iteratively adjusted manually, attribute-by-attribute, until the model's mesh approximated that of the sample mesh (Fig. 5). The required precision is practically determined by individual differences within a homogeneous sample population, which have standard deviations of typically 3-6 mm [2]. Here the model parameters were adjusted until the surface and the scan data deviated generally by less than $\pm 1-2$ mm.

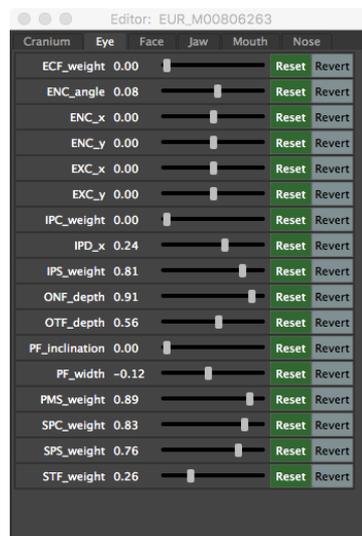


Fig. 4. The Ethnicity Modeler provides interactive controls for all 77 attributes, organized into tab panes. Here the eye attributes have been selected. These controls were then adjusted to create a close match with empirical data, such as provided by stereo-photogrammetry (Fig. 5).

Fig. 6 shows models of two African and two East Asian individuals. The modeling of individuals of various ethnicities was quite useful, as individuals often will reveal unexpected attribute extremes, that then required revising the underlying basis shapes in order to accommodate such extremes (and consequently revision of any models that had used the earlier version for those extremes, a consequence of the exploratory, spiral development process).

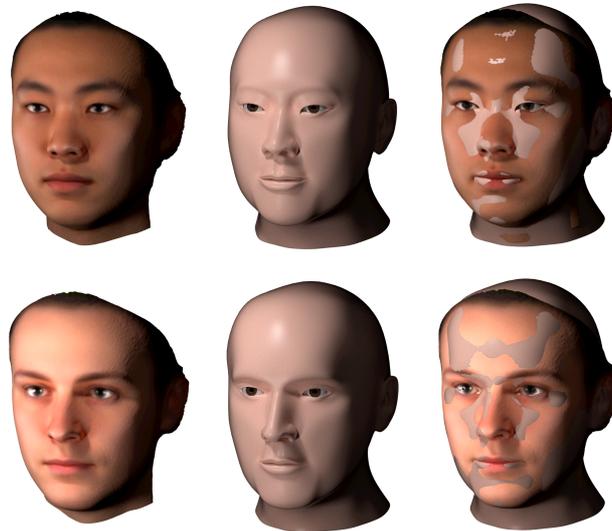


Fig. 5. In the upper row, the parametric model (center) has been adjusted to closely approximate the stereo-photogrammetric average of 20 male East Asian faces (left). The lower row shows the model matched against the averaged face of 20 European males. In each case, after a manual process of adjustment, the model mesh eventually matched the scan data to within $\pm 1\text{-}2$ mm generally, i.e., to within the uncertainty due to individual variation. The mottled patterns (right) shows this close correspondence (and reveals residual asymmetries in the scan data) when the two meshes are superimposed.

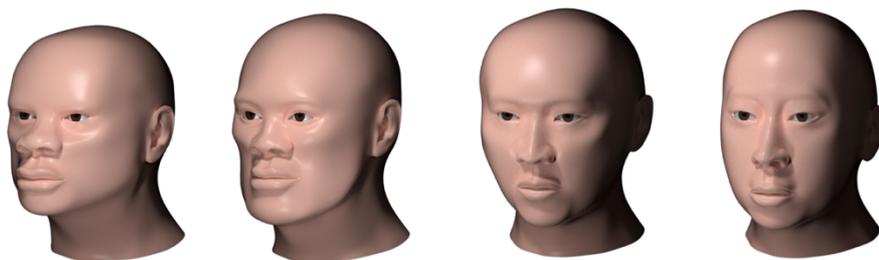


Fig. 6. Two African male individuals (left) and two East Asians individuals (right) were modeled, demonstrating the ability to parametrically capture a broad range of both ethnic and individual variation.

5 Evaluation

This modeling of a broad range of ethnic variation, as well as individual variation, would appear supported by a schema based on roughly 100-200 attributes (this study having implemented 77 attributes, but each facial region could have benefited from having additional, more subtle attributes). The schema was developed to support both representation and visualization. As a representational scheme, it was important to evaluate the precision and accuracy with which the model could match empirical data (such as stereo-photogrammetric scans). A study was pursued by the first author, wherein stereo-photogrammetric data for males of two ethnicities (East Asian and European) were matched repeatedly for 10 trials each, and the mean and standard deviation was then computed for each of the 77 attributes for each ethnicity. The standard deviations revealed that measurement uncertainty was generally less than 10% of the range of the corresponding attribute. Those signed attributes that correspond to anthropometric dimensions (such as mouth width or nose length) were repeatable to 5% or less, while some unsigned shape attributes (such as the depth of the shallow sulcus below the eyes or the roundness of the tip of the nose) showed larger trial-to-trial variances, found to be more difficult to judge, with a few such attributes having just noticeable differences of somewhat greater than 15% [1]. Regarding accuracy, five of the dimensional attributes were calibrated in millimeters. The mean attribute values were compared with the corresponding anthropometric measurements values (from [2]) for the two ethnotypes (East Asian and European), and the Ethnicity Modeler results generally matched to within one standard deviation, i.e., the variance due to individual variation [1].

The Ethnicity Modeler interface was developed only with the limited goals of supporting the development of the facial descriptive and visualization scheme. As shown in Fig. 4, each facial attribute was controlled directly through a slider widget. Adjusting the EM mesh in order to fit empirical data thus involved a substantial task of adjusting 15-20 attributes per facial region. An evaluation study involving five users (all familiar with user interfaces but only one was an expert in 3D sculpting software) tasked with adjusting the EM mesh to match 3D data, using the Thinking Aloud methodology [25]. The consensus was that the facial attributes were, as hoped, intuitive and understandable, and that they could be manipulated with some ease through the interface by slider widget [1]. The attributes were reported to be natural in for capturing, and communicating, facial shape. Negatively, however, it was immediately apparent that facial modeling by laborious manual adjustment of 77 attributes was an overwhelming task, especially for non-experts. The EM, however, was intended only as a proof of concept of the representation scheme, not as a turnkey modeling tool per se. This experience, nonetheless, raises the question of how to effectively explore such a high-dimensional EFS, a matter for future studies.

6 Discussion

Human observers naturally describe and appreciate faces in terms of local shape descriptors: visual attention is drawn to distinctive shape features of a face (either that of an individual or of an averaged representative of a given ethnotype), many of which are

associated with landmarks. There is also a universal tendency to base a description on exemplars, especially extreme examples of any given characteristic.

We have found that local shape descriptors can be represented by basis shapes with sufficient specificity to reconstruct (match) facial geometry across a range of ethnotypes, to within the uncertainty that is imposed by individual variations. The shape attributes are local, comprehensible, and consistent with the traditional anatomical nomenclature and literature. While it is challenging to model local facial attributes by basis shapes, and to achieve quasi-orthogonality among them, we can then match sample faces purely by parametric manipulation of mesh geometry in terms of those attributes. More generally, by regarding these shapes as defining an EFS, the resultant space permits exploration of ethnic differences directly in terms of shape differences, and that is its primary intention. Provide that the attributes are scaled to allow covariance to be meaningfully computed, a set of sample models across different ethnotypes would be amenable to PCA.

References

1. Wisetchat, S.: Description-based visualisation of ethnic facial types. The Glasgow School of Art School of Simulation and Visualisation. Ph.D. dissertation (2018)
2. Farkas, L.G.: Anthropometry of the head and face in clinical practice. In *Anthropometry of the Head and Face 2nd edition*, L.G. Farkas (Ed.). Raven Press, New York, 71-77 (1994)
3. Farkas, L.G., Katic, M.J., Forres, C.R., Alt, K.W., Bagic, I., Baltadjiev, G., et al.: International anthropometric study of facial morphology in various ethnic groups/races. *J Craniofac Surg.* 16(4), 615-46 (2005)
4. Franco, F.C., Arroyo, T.M., Vogel, C.J., Quintão, C.C.: Brachycephalic, dolichocephalic and mesocephalic: is it appropriate to describe the face using skull patterns? *Dental Press Journal of Orthodontics* 18(3), 159-163 (2013)
5. Valentine, T.: A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology*, 43(2), 161-204 (1991)
6. Valentine, T.: Face-space models of face recognition. Computational, geometric, and process perspectives on facial cognition: Contexts and challenges, 83-113 (2001)
7. Lewis, M.B.: Face-space-R: Towards a unified account of face recognition. *Visual Cognition* 11(1), 29-69 (2004)
8. Valentine, T., Lewis, M.B., Hills, P.J.: Face-space: A unifying concept in face recognition research. *The Quarterly Journal of Experimental Psychology*, 69(10), 1996-2019 (2016)
9. Lee, D., Seung, H.S.: Learning the Parts of Objects by Non-negative Matrix Factorization, *Nature* 401(6755), 788-791 (1999)
10. Turk, T., Pentland, A.: Eigenfaces for recognition. *Journal of Cognitive Neuroscience* 3(1), 71-86 (1991)
11. Swennen, G.R.J., Schutyser, F.A.C., Hausamen, J.-E.: *Three-Dimensional Cephalometry. A Color Atlas and Manual*. Berlin: Springer Verlag, 365pp (2005)
12. Boehm, B.W.: A spiral model of software development and enhancement. *Computer* 21(5), 61-72 (1988)

13. Zhang, Y., Badler, N.I.: Synthesis of 3D faces using region-based morphing under intuitive control. *Anim. Virtual Worlds*, 17, 421–432 (2006)
14. DeBruine, L.M., Tiddeman, B.P.: WebMorph. Retrieved from <http://webmorph.org> (2017)
15. Benson, P.J., Perrett, D.I.: Extracting prototypical facial images from exemplars. *Perception* 22, 257–262 (1993)
16. Lanitis, A., Taylor, C.J., Cootes, T.F.: Automatic interpretation and coding of face images using flexible models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(7), 743–756 (1997)
17. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3D faces. In *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques* 187-194, ACM Press/Addison-Wesley Publishing Co (1999)
18. Blanz, V., Vetter, T.: Face recognition based on fitting a 3d morphable model. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 25(9), 1063-1074 (2003)
19. Bergeron, P., Lachapelle, P.: Controlling Facial Expression and Body Movements in the Computer-Generated Short ‘Tony de Peltrie’, *Siggraph Tutorial Notes*, ACM Press (1985)
20. Lewis, J.P., Anjyo, K.I.: Direct manipulation blendshapes. *IEEE Computer Graphics and Applications* 30(4), 42-50 (2010)
21. Lewis, J.P., Mooser, J., Deng, Z., Neumann, U.: Reducing blendshape interference by selected motion attenuation. In *Proceedings of the 2005 Symposium on Interactive 3D Graphics and Games*, 25-29 (2005)
22. Holzleitner, I.J., Hunter, D.W., Tiddeman, B.P., Seck, A., Re, D.E., Perrett, D.I.: Men’s facial masculinity. When (body) size matters. *Perception* 43, 1191-1202 (2014)
23. Winder, R.J., Darvann, T.A., McKnight, W., Magee, J.D.M., Ramsay-Baggs, P.: Technical validation of the Di3D stereophotogrammetry surface imaging system. *British Journal of Oral and Maxillofacial Surgery*, 46(1), 33-37 (2008)
24. Tiddeman, B.P., Duffy, N., Rabey, G.: Construction and visualisation of three-dimensional facial statistics. *Computer Methods and Programs in Biomedicine* 63, 9–20 (2000)
25. Lewis, J.R.: Sample sizes for usability studies: Additional considerations. *Human Factors*, 36, 368-378 (1994)