NEUROSCIENCE

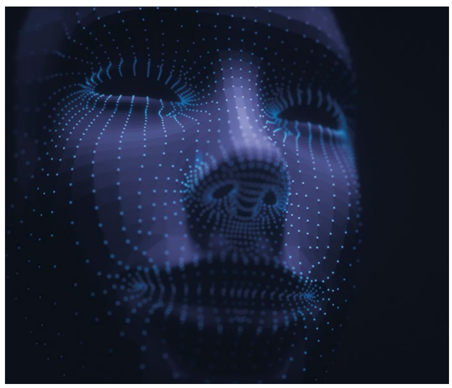
# Representing faces in 3D

How do we recognize the individual faces of our family members, friends and acquaintances across the variation that is common in daily life? Zhan and colleagues demonstrate the importance of three-dimensional structure in the representations of known individuals and argue that texture—the surface properties of faces—plays little role in representation.

Nicholas Blauch and Marlene Behrmann

Most humans effortlessly recognize the faces of hundreds or thousands of individuals, but we still do not know precisely how this is performed. Because faces share a common set of parts, one approach that has been applied in artificial vision systems involves aligning these parts into a common space and analyzing faces based on the resultant appearance. This is the idea behind two-dimensional (2D) active appearance models[1], which, beyond their influence in computer vision, have recently been hailed as an explicit theory of neural computations underlying face perception in the primate brain. Examining a series of face-selective brain regions in the macaque monkey, neuroscientists have demonstrated that earlier cortical regions respond along principal axes describing the location of a set of facial landmarks, while later cortical regions respond along principal axes describing the appearance of faces after a computational alignment of face landmarks to a common template[2]. Does this imply that the brain literally aligns retinal images of faces to a generic 2D template, followed by an analysis of the shape-free appearance? In *Nature Human Behaviour*, Zhan and colleagues[3] provide an intriguing study of human three-dimensional (3D) face perception that compels us to consider otherwise.

By using a database of 355 3D scans of human faces, varying in age, gender and ethnicity, with a general linear modelling (GLM) framework, the researchers developed a novel 3D face space capable of generating new faces along psychologically interpretable dimensions of age, sex, ethnicity, 3D rotation, and high-variance 'identity' dimensions. Their modelling framework was performed using a 3D version of active appearance models[4] that first registers faces into a common 3D structural mesh, retaining a description of how an individual face structure deviates from the common mesh, and then similarly registers their structure-free texture (the 3D equivalent of shape-free appearance). The researchers then scanned 4 additional human faces and recruited 14 participants familiar with these additional individuals as work colleagues. The researchers used their generative face model to generate random-identity faces matched to the test faces in terms of age, gender and race. Shown sets of 6 such faces at a time, participants were asked to rate the similarity of the face with the greatest similarity to a given test face. These ratings allowed Zhan and colleagues to use reverse correlation to recover maps of the structural and textural features used by the majority of their subjects in representing each face. With these maps, the researchers performed a component analysis to extract a small number of whole-face diagnostic and non-diagnostic components. Finally, they devised a task to validate the extent to which these components captured the information used by the participants to recognize their colleagues across various transformations, such as rotation or manipulation of gender or age, through the use of their generative face model. Given a certain transformation, the researchers added a specific amount of a diagnostic component, or its non-diagnostic counterpart, and measured its effect on the ability of their participants to recognize their colleagues across the transformation. They found that adding diagnostic information improved generalization across faces and tasks, suggesting that these components captured the mental representations used by their participants.



Credit: Science Photo Library / Alamy Stock Photo

Strikingly, the generalization task was based entirely on a structural description of the faces. When assessing the role of structure-free texture, the researchers found few to no consistent patterns across subjects, suggesting that their texture model did not faithfully represent identity. The dominant influence of shape highlights the often-overlooked point that 2D shape-free appearance representations are not truly devoid of shape. This poses a challenge for interpreting results concerning the representation of face surface properties separate from 2D face shape[5], and it poses a challenge for our interpretation of complementary shape and appearance representations in separate face-selective brain regions[2]. Zhan and colleagues suggest that the earlier 2D 'shape-based' representations may correspond to local structural properties, and that later 'appearance-based' representations may correspond to combinations of local structural properties, akin to the modelled multivariate dimensions.

It is interesting to note that feedforward convolutional neural networks, which have achieved great success both in computer vision applications and in modeling biological visual representations, do not exhibit a similar dependence on 3D structure. Multiple studies have demonstrated that these models display a greater dependence on texture[6,7], which can be partially corrected by training with stylized images that induce a reliance on shape[6]. Learning to extract 3D structural representations from 2D images may also benefit from signals in the temporal domain. Indeed some work in computational neuroscience has shown how smooth object rotations can benefit the development of robust structural representations[8], and computer vision researchers are actively investigating these ideas, for example in using temporal contiguity for action recognition[9] and other dynamic tasks. Finally, a recent 'analysis-by-synthesis' model called efficient inverse graphics[10] has accounted for some aspects of human 3D perception and primate neural activity by proposing that the brain stores a three-dimensional generative model akin to the one used by Zhan and colleagues, with recognition ensuing through model inversion.

Zhan and colleagues provide persuasive evidence for the importance of 3D structure in human face representation. However, their work does not provide a definitive case against the role of texture. While textural representations were found to be unreliable across subjects, some textural patterns were found within subjects (see Supplementary Information for texture analyses). More strikingly, constant-shape texture morphs appear to have a strong impact on perceived identity. Taking their results at face value, however, there remains an important question of how a structural template would be derived biologically, as the 3D registration approach used by the authors is more a mechanism for revealing a representation than it is a model of the mechanism that achieves the representation. Finally, more general questions of face perception may be extended into the 3D domain. How can 3D representations cope with dynamic aspects of faces, such as changes in expression? As we become familiar with a new face, how does learning its specific properties bootstrap and refine our 3D representations? From our perspective, it is clear that the approach and stimulus set developed by Zhan and colleagues will serve as an important human benchmark for evaluating models of the development and nature of representations underlying human face recognition and will be a powerful tool for studying faces in 3D. ❐

Nicholas Blauch[1] and Marlene Behrmann[1,2*]

[1]*Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, PA, USA.* [2]*Department of Psychology, Carnegie Mellon University, Pittsburgh, PA, USA.*

**e-mail: behrmann@cmu.edu*

Published online: 17 June 2019
https://doi.org/10.1038/s41562-019-0630-6

## References

1. Cootes, T.F., Edwards, G.J. & Taylor, C.J. Active appearance models. *European Conference on Computer Vision* (eds.) 484–498 (Springer, 1998).
2. Chang, L. & Tsao, D. Y. *Cell* **169**, 1013–1028.e14 (2017).
3. Zhan, J., Garrod, O.G.B., van Rijsbergen, N. & Schyns, P.G. *Nat. Hum. Behav*. https://doi.org/10.1038/s41562-019-0625-3 (2019).
4. Blanz, V. & Vetter, T. A morphable model for the synthesis of 3D faces. *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques* (eds.) 187–194 (1999).
5. Nemrodov, D., Behrmann, M., Niemeier, M., Drobotenko, N. & Nestor, A. *Neuroimage* **184**, 813–825 (2019).
6. Geirhos, R. et al. ImageNet-Trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. *International Conference on Learning Representations* (eds.) 1–22 (2019).
7. Xu, T. et al. Preprint at *arXiv* https://arxiv.org/abs/1811.07807 (2018).
8. Isik, L., Leibo, J.Z. & Poggio, T. *Front. Comput. Neurosci*. https://doi.org/10.3389/fncom.2012.00037 (2012).
9. Dave, A., Russakovsky, O. & Ramanan, D. Preprint at *arXiv* https://arxiv.org/abs/1704.03615 (2017).
10. Yildirim, I., Belledonne, M., Freiwald, W., Tenenbaum, J. Preprint at *bioRxiv* https://www.biorxiv.org/content/10.1101/282798v2 (2019).

## Competing interests

The authors declare no competing interests.