Vision Research 51 (2011) 969–977

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

The application of 3D representations in face recognition

Adrian Schwaninger [a,b], Jisien Yang [c,d,*]

[a] School of Applied Psychology, University of Applied Sciences Northwestern Switzerland
[b] Department of Informatics, University of Zurich, Switzerland
[c] Department of Psychology, University of Zurich, Switzerland
[d] Department of Psychology, National Chung Cheng University, Taiwan

ARTICLE INFO

Article history:
Received 14 May 2010
Received in revised form 5 February 2011
Available online 13 February 2011

Keywords:
Silhouette
Face recognition
3D face representation

ABSTRACT

Most current psychological theories of face recognition suggest that faces are stored as multiple 2D views. This research aims to explore the application of 3D face representations by means of a new paradigm. Participants were required to match frontal views of faces to silhouettes of the same faces. The formats of the face stimuli were modified in different experiments to make 3D representations accessible (Experiments 1 and 2) or inaccessible (Experiment 3). Multiple 2D view-based algorithms were not applicable due to the singularity of the frontal-view faces. The results disclosed the application and adaptability of 3D face representations. Participants can readily solve the tasks when the face images retain the information essential for the formation of a 3D face representations. However, the performance substantially declined when the 3D information in faces was eliminated (Experiment 3). Performance also varied between different face orientations and different participant groups.

1. Introduction

Face recognition plays an important role in social interaction. Despite all faces being composed of relatively few parts and sharing a similar configuration, humans are able to distinguish exceedingly subtle differences between them. Although much research has been conducted in recent decades to investigate the underlying processes and representations of faces (for a recent review see for example Schwaninger, Wallraven, Cunningham, & Chiller-Glaus, 2006), it remains debatable whether 3D representations and viewpoint transformations exist in human face recognition.

According to object-centered theories (Biederman, 1987, 2000; Marr & Nishihara, 1978), object recognition is based on structural descriptions which specify an object by its constituent parts, e.g. Geons and their spatial relations (Geon Structural Descriptions, GSD). Such descriptions are assumed to be object-centered, which provide the basis for view-invariant recognition. Biederman and Gerhardstein (1993) have enhanced the Recognition by Components (RBC) theory (Biederman, 1987) by specifying three prerequisites for viewpoint-independent recognition. First, the objects must be decomposable into their parts. Second, the GSD for different objects must be distinctive. Third, the same GSD of a specific object must be recoverable from different viewpoints.

In contrast, view-based theories propose that objects are not stored as object-centered structural descriptions but as a collection of 2D views (Biederman & Kalocsai, 1997; Bülthoff & Edelman, 1992; Bülthoff, Edelman, & Tarr, 1995; Tarr & Bülthoff, 1995; Tarr & Pinker, 1989). Object recognition relies on matching a novel view of an object to the stored views by using different mechanisms, such as linear interpolation between views (Poggio & Edelman, 1990), multiple views plus transformations (Tarr & Pinker, 1989), or linear combination of views (Ullman & Basri, 1991).

However, face recognition, it has been suggested, implicates mechanisms that are different from those applying to object recognition. The processes which distinguish object recognition from face recognition have been demonstrated in different research domains, such as behavioral studies (Yin, 1969), neuropsychological patients (Ellis & Florence, 1990; Farah, 1991; Farah, Levinson, & Klein, 1995; Hecaen & Angelergues, 1962; Yin, 1970), and cognitive neuroscience studies (Desimone, 1991; Kanwisher, Downing, Epstein, & Kourtzi, 2001; Kanwisher, McDermott, & Chun, 1997; Ojemann, Ojemann, & Lettich, 1992). However, in contrast to the debate on 2D vs. 3D representation in object recognition, it is generally assumed that faces are represented by a collection of 2D views (Biederman & Kalocsai, 1997; Bülthoff et al., 1995).

Although Biederman and Kalocsai (1997) propose that object recognition is viewpoint-independent, they further indicate that RBC applies only to basic-level object recognition but not to face recognition. Due to the fact that all faces share the same basic components (eyes, nose, mouth, chin, etc.) in the same basic arrangement (the eyes are above the nose which is above the mouth),

* Corresponding author at: Department of Psychology, University of Zurich, Thurgauerstrasse 39, CH-8050 Zurich, Switzerland. Fax: +41 43 336 01 00.
E-mail address: yangjisin@yahoo.com.tw (J. Yang).

0042-6989/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2011.02.006

faces cannot be distinguished based on structural descriptions. Violation of the prerequisites of viewpoint-independent recognition makes it difficult for face recognition to have 3D representations. Instead, Biederman and Kalocsai (1997) argue that a holistic, viewpoint-dependent system such as the models proposed by Christoph von der Malsburg and his colleagues (Lades et al., 1993; Wiskott, Fellous, Krüger, & von der Malsburg, 1997) explain human face recognition much more aptly. Moreover, several studies also suggest that face recognition is based on purely holistic and view-based processes using 2D representations rather than 3D representations (e.g., Lades et al., 1993; Tanaka & Farah, 1991, 1993; Wiskott et al., 1997). Using a computational model, Wallraven, Schwaninger, and Bülthoff (2005) implemented a view-based approach in which facial features and their spatial relations are stored in separate 2D views, which are temporally associated. As shown in a group of different studies, this model could explain various aspects of human face recognition such as processing component and configural information (Schwaninger, Wallraven, & Bülthoff, 2004; Schwaninger, Lobmaier, Wallraven, & Collishaw, 2009) as well as specific effects of viewpoint (Schwaninger, Schumacher, Wallraven, & Bülthoff, 2007; Wallraven, Schwaninger, Schuhmacher, & Bülthoff, 2002).

Accordingly, 3D face representations, which require the processes of utilizing the shading and shadow information to reconstruct the three-dimensional shape of faces, may be absent for humans in many face recognition models (Bruce, 1988; Bruce & Langton, 1994; Johnson, Hill, & Carman, 1992; Vetter, 1998). It is only in the domain of computer vision that 3D models and processes of faces have been implemented (Blanz & Vetter, 1999; Vetter, 1998; for a review of cognitive and computational models of face recognition see Schwaninger et al., 2006).

However, at least three different research lines provide converging results suggesting that human face recognition may involve 3D representation mechanisms rather than a mere match of multiple 2D face views. The first line of evidence comes from research regarding the recognition of one's own profile. Troje and Kersten (1999) have found that humans can recognize profile views of their own faces, even though such views are usually not encountered and thus are hardly available in visual memory. Tong and Nakayama (1999) report further interesting research. In a visual search task, their participants demonstrated an equivalent own-face advantage across frontal, three-quarter, and profile views. Their results are surprising when taking into account that the observers had equal amounts of visual experience of the stranger's frontal and profile view, but far greater experience of their own frontal face view than their profiles. They argue that people can develop robust representations for highly over-learned faces, such as one's own. This representation might involve viewpoint-independent 3D representations. However, participants in Tong and Nakayama's study might have relied on facial texture information to recognize the depth-rotated profiles of their own faces because it has been found that facial information, such as skin color, pigment, or texture features provide information which is important in reducing viewpoint dependence in face recognition (Hill, Schyns, & Akamatsu, 1997). Moreover, face features such as skin texture, blemishes, and dimples may be visible from largely different viewpoints (ÓToole, Bülthoff, Troje, & &Vetter, 1995).

The second line of evidence suggesting the application of 3D face representations in humans comes from haptic face recognition research. Kilgour and Lederman (2002) found that participants' performance (67.8%) was higher than chance (33.3%) when they were shown motionless live faces, and subsequently were required to recognize them by touching (haptic recognition). Their results imply that participants can construct a multimodal 3D representation of a face based on mere visual exposure. Casey and Newell's (2003) research regarding haptic own-face recognition also supports this assumption. They found that a greater amount of target faces were correctly identified when own-face masks were oriented towards participants, contrary to the orientation in which a haptic representation of one's own face is naturally generated. These findings suggest that humans might be able to construct 3D representations of their own faces via the large amount of visual experience. Viewers might then apply this 3D representation to the haptic recognition of representations of their own faces in which geometric properties are the only cues for correct identification. Although far from the same domain, Casey and Newell's results surprisingly correspond to the notion of Tong and Nakayama (1999) that a robust viewpoint-independent representation is formed for highly over-learned faces.

The third line of evidence comes from the studies associated with neuronal activation in the brain. Grill-Spector et al. (1999) found that both the caudal-dorsal (LO) region and the posterior fusiform (PF-LOa) region in the lateral occipital complex (LOC) are maximally activated by images of different individuals' faces. However, the activation is adapted by repeated presentation of identical individual faces, either in the original viewing condition or in a depth rotation condition. The adaptations for the original viewing condition and the depth rotation condition are equivalent in the PF-LOa region. Grill-Spector et al. argue that the PF-Loa is more invariant to changes in the object's position in the visual field compared to LO. Similarly, Chen, Kao, and Tyler (2006) also observed that brain activation is significantly different between frontal-view and inverted faces, but not between frontal-view and 3/4-view faces. These results suggest that there may be neural circuits responsible for the viewpoint invariance of face representation. In fact, a small portion of cells in the macaque superior temporal sulcus (STS) has been observed to respond equally to multiple views of a face (Perrett et al., 1991).

Although the results from the three lines of research imply the application of 3D face representations in humans, these studies did not directly examine the mechanisms of 3D face representations. In this research, we adopt the 'face silhouette vs. frontal-view faces matching' paradigm, in which a one-tone black silhouette is matched with a frontal-view face (Davidenko, 2007). This task can be solved by extracting 3D information, such as shading and shadow information contained in the face photographs (Bruce & Langton, 1994), reconstructing the 3D structure of faces (Bruce et al., 1991; Vetter, 1998), mentally rotating a 3D face model (Blanz & Vetter, 1999; Vetter, 1998), and matching it to the one-tone black silhouette. This task cannot be solved by matching 2D information contained in the faces, such as face configuration, texture, color, blemishes and dimples. Moreover, it cannot be solved by linear combination, because a set of 2D views is necessary for constructing 3D models in this way (Poggio, 1990; Ullman & Basri, 1991). Although, Poggio and his colleagues propose that only one non-accidental 2D view is sufficient for recognition in the case of bilaterally symmetrical objects such as faces, (Beymer & Poggio, 1995; Poggio, 1991; Poggio & Vetter, 1992), they constrain their conclusion by specifying that

> one should avoid to use in the data base a model view which is a fixed point of the symmetry transformations (since the transformation of it generates an identical new view). In the case of faces, this implies that the model view in the data base should not be an exactly front-view (Poggio & Vetter, 1992, p. 15).

Basically, the frontal-view face is singular, and a second view cannot be computed from it (Schyns & Bulthoff, 1993). As a result, a silhouette cannot be generated from just one frontal-view face image by algorithms of different 2D view-based models. According to view-based face recognition theories, matching a frontal-view face to its one-tone black silhouette would be improbable, because