Cortical 3D Face and Object Recognition Using 2D Projections

João Rodrigues, University of the Algarve, Portugal
Roberto Lam, University of the Algarve, Portugal
Hans du Buf, University of the Algarve, Portugal

ABSTRACT

Empirical studies concerning face recognition suggest that faces may be stored in memory by a few canonical representations. In cortical area V1 exist double-opponent colour blobs, also simple, complex and end-stopped cells which provide input for a multiscale line/edge representation, keypoints for dynamic feature routing, and saliency maps for Focus-of-Attention. All these combined allow faces to be segregated. Events of different facial views are stored in memory and combined to identify the view and recognise a face, including its expression. In this paper, the authors show that with five 2D views and their cortical representations it is possible to determine the left-right and frontal-lateral-profile views, achieving a view-invariant recognition rate of 91%. The authors also show that the same principle with eight views can be applied to 3D object recognition when they are mainly rotated about the vertical axis.

Keywords: Face Recognition, Focus-of-Attention, Multi-Scale Keypoints, Multi-Scale Lines and Edges, Object Recognition, Visual Cortex

INTRODUCTION

One of the most important topics of image analysis is face detection and recognition. There are several reasons for this, such as the wide range of commercial vigilance and law-enforcement applications. Face recognition is also one of the most important capabilities of our visual system. Information about a person’s gender, ethnicity, age and emotions contribute to the recognition process. For instance, in court, a lot of credibility is placed on identifications made by eyewitnesses, although studies have shown that people are not always reliable when comparing faces with recollections (Smeets et al., 2010).

Recently, because of the limitations of 2D approaches and with the advent of 3D scanners, face-recognition research has expanded from 2D to 3D with a concurrent improvement in performance. There are many face-recognition methods in 2D and 3D; for detailed surveys see Abate et al. (2007) and Bowyer et al. (2006). Recently, Rashad et al. (2009) presented a
face-recognition system that overcomes the problem of changes in facial expressions in 3D range images by using a local variation detection and restoration method based on 2D principal component analysis. Ramirez-Valdez and Hasimoto-Beltran (2009) also related 3D facial expression to recognition. Berretti et al. (2010) took into account 3D geometrical information and encoded the relevant information into a compact graph representation. The nodes of the graph represent equal-width iso-geodesic facial stripes. The edges between pairs of nodes are labeled by descriptors, and referred to as 3D weighted walkthroughs that capture the mutual relative spatial displacement between all node pairs in the corresponding stripes.

For the recognition of general 3D objects, there are also several approaches which have been explored. Some of these are based on 3D shape retrieval methods (Tangelder & Velkamp, 2008), on 2D projected images (Su et al., 2006) and on multiresolution signatures (Lam & du Buf, 2011), including approaches that can be applied both to 3D objects and to faces (Pasqualis et al., 2007).

State-of-the-art recognition systems have reached a certain level of maturity, but their accuracy is still limited when imposed conditions are not perfect: view-invariant object recognition is still problematic (Pinto et al., 2010), and for faces all possible combinations of changes in illumination, pose and age, with artefacts like beards, moustaches and glasses, including different facial expressions and partial occlusions may cause problems. The robustness of commercial systems is still far away from that of the human visual system, especially when dealing with different views of the same person. For this reason, despite the fact that the human visual system may not be 100% accurate, the development of models of visual perception and their application to real-world problems is important and, eventually, may lead to a significant breakthrough.

In this paper we present a cortical model to recognise 3D faces from their 2D projections, i.e., exploiting their aspect ratio but without using stereo disparity. We consider all common degrees of rotation like pan (from frontal to lateral and profile views) and tilt (the face looking up or down). We study the number of 2D feature templates required to represent all views. In the recognition process we first detect the view and then match the input face with the view-based templates stored in memory. We also test the same model to elongated 3D objects, i.e., 4-legged animals, because their shape and view can also be inferred from the aspect ratio if they are rotated about the vertical axis (pan). The rest of this paper is organised as follows: we first discuss the cortical background of the model, before explaining the frameworks for 3D faces and then 3D objects. Recognition results are presented and discussed in the next section, and we end this paper with conclusions.

CORTICAL BACKGROUND

Face perception in humans is mediated by a distributed neural system which links multiple brain regions. The functional organisation of this system embodies a distinction between the representation of invariant aspects of faces, which is the basis for recognising individuals. A core system, consisting of occipitotemporal regions in extrastriate visual cortex, mediates visual analysis of faces (Haxby et al., 2002). Faces are represented as locations in a multidimensional space, where the distance separating representations is proportional to the degree of dissimilarity between faces, the structure of this face–space being tolerant to lighting and viewpoint transformations (Blank & Yove, 2011).

Invariant face and object detection, categorisation and recognition depends on a hierarchy of cortical stages that build invariance gradually (Grossberg et al., 2011), involving both bottom-up and top-down data streams in the so-called “what” and “where” subsystems (Deco & Rolls, 2004), including the integration of both subsystems (Farivar, 2009). In cortical area V1 there are simple and complex cells, which are tuned to different spatial frequencies (scales) and orientations, but also to disparity (depth) because of neighbouring
Related Content

Towards Multiple 3D Bone Surface Identification and Reconstruction Using Few 2D X-Ray Images for Intraoperative Applications

Implementation of an Immersive Videogame: Legends of Girona
[www.igi-global.com/article/implementation-of-an-immersive-videogame/130017?camid=4v1a](www.igi-global.com/article/implementation-of-an-immersive-videogame/130017?camid=4v1a)

Mixed Reality Boundaries in Museum Preservation Areas
[www.igi-global.com/article/mixed-reality-boundaries-in-museum-preservation-areas/90161?camid=4v1a](www.igi-global.com/article/mixed-reality-boundaries-in-museum-preservation-areas/90161?camid=4v1a)
Analysis of Musical Content in Digital Audio
www.igi-global.com/chapter/analysis-musical-content-digital-audio/6856?camid=4v1a