

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

DOI: 10.4018/jcicg.2012010104

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 (Passalis 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 com-

mon 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

left-right hypercolumns (Hubel, 1995). These cells provide input for grouping cells which code line and edge information (Rodrigues & du Buf, 2009a) and attribute depth information (Rodrigues & du Buf, 2004). In V1 there are also double-opponent colour blobs (Tailor et al., 2000) and end-stopped cells which, together with sophisticated inhibition processes, allow to extract keypoints, i.e., singularities, vertices and points of high curvature (Heitger et al., 1992; Rodrigues & du Buf, 2006) and optical flow (Farrajota et al., 2011).

On the basis of models of neural processing schemes, it is now possible to create a cortical architecture bootstrapped by global and local gist (Martins et al., 2009; Rodrigues & du Buf, 2011), with face and figure-ground segregation (Farrajota et al., 2011; Rodrigues & du Buf, 2006, 2009a), focus-of-attention (Martins et al., 2009; Rodrigues & du Buf, 2006), face/object categorisation and recognition (Rodrigues & du Buf, 2006, 2009a), including recognition of facial expressions (Sousa et al., 2010).

Recent work (Kourtzi & Connor, 2011) has provided a more precise picture of how 2D and 3D object structure is encoded in intermediate and higher-level visual cortices. Yet, other studies suggest that higher-level visual cortex represents categorical identity rather than structure (Kourtzi & Connor, 2011). Furthermore, object responses are surprisingly adaptive to changes in environmental factors, implying that learning through evolution, development, and also shorter-term experience in adulthood may optimise the object code. Future progress in reconciling these findings will depend on more effective sampling of the object domain and direct comparison of these competing hypotheses.

In conclusion, there are several open questions related to the perception and recognition of objects and faces in the brain. One of those is which and how many templates do we store of a person's face, and how those templates are related to create the notion of a 3D face in our brain. In the cortical hypercolumns in area V1, information from the left and right retinal projections is very close. As a consequence,

dendritic fields of cells can receive input from both projections, and simple and complex cells can be tuned to disparity such that depth can be attributed to detected lines and edges, mainly the vertical ones. This could yield a sort of 3D wireframe representation, like the one used in computer graphics for the modelling of 3D solid objects. This approach will not be applied here: instead of exploiting disparity information, we will restrict ourselves to 2D projections. The main three reasons are that (i) disparity, together with the vergence of the eyes, is mainly for estimating the distance to objects and for large objects,

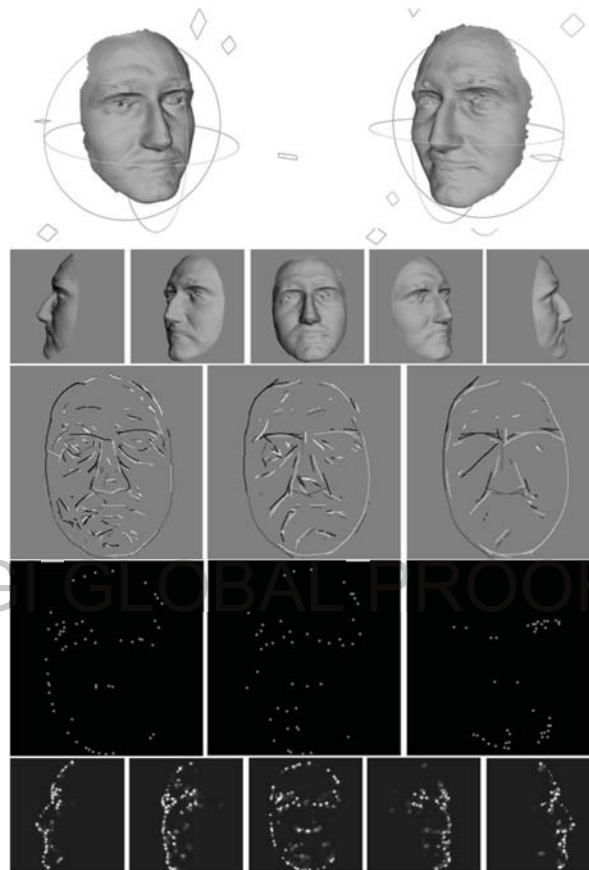
(ii) faces are normally rather small and their disparity information too, and (iii) persons with one eye have no problem telling people apart. However, such aspects are subject to ongoing research. In principle it is possible to combine disparity information with projected features.

In this paper we focus mainly on a cortical model for face recognition. The present model is based on a previous one (Rodrigues & du Buf, 2009b) which employs multiscale line/edge and keypoints representations based on cells in area V1 of the visual cortex. That model was shown to give good results for frontal and frontal-to-3/4 views, also with small occlusions. In the present paper we go much further. We test faces with more degrees of rotation¹ (y-rotated $\pm 90^\circ$, pan; x-rotated $\pm 10^\circ$, tilt), the number of 2D templates needed to represent a 3D face, the relation between them, and the detected view (left-right and profile-lateral-frontal). We also apply the same model to 3D objects (animals) with other degrees of rotation (y-rotated $\pm 180^\circ$, pan; x-rotated $\pm 10^\circ$, tilt).

CORTICAL LINE, EDGE AND KEYPOINT MODELS

There is extensive evidence that the visual input is processed at different spatial scales, from coarse to fine ones, and both psychophysical and computational studies have shown that different

Figure 1. Top row: two expressions of the same face with possible rotation intervals. Second row: profile, lateral and frontal 2D views of the neutral face (y-rotated). Third row: multiscale line and edge coding at three scales $\lambda = \{4, 12, 24\}$ of the frontal face view on the 2nd row. Fourth row: detected keypoints at the same scales. The bottom row shows saliency maps of the 2D views on the 2nd row.



scales offer different qualities of information (Bar, 2004; Oliva & Torralba, 2006).

Gabor quadrature filters provide a model of cortical simple cells (Rodrigues & du Buf, 2006). In the spatial domain (x, y) they consist of a real cosine and an imaginary sine, both with a Gaussian envelope. A receptive field (RF) is denoted by

$$G_{\lambda, \sigma, \theta, \varphi}(x, y) = \exp\left(-\frac{x^2 + \gamma y^2}{2\sigma^2}\right) \cdot \cos\left(\frac{2\pi x}{\lambda} + \varphi\right),$$

with

$$x = x \cos \theta + y \sin \theta$$

and

$$y = y \cos \theta - x \sin \theta,$$

the aspect ratio $\gamma = 0.5$ and σ the size of the RF. The spatial frequency is $1/\lambda$, with λ being the wavelength. For the bandwidth σ/λ we use 0.56, which yields a half-response width

of one octave. The angle θ determines the orientation (we use 8 orientations), and the phase φ the symmetry (0 or $-\pi/2$). Below, the scale of analysis will be given by λ expressed in pixels, where $\lambda = 1$ corresponds to 1 pixel. All tested images have 256×256 pixels.

Responses of even and odd simple cells, which correspond to real and imaginary parts of a Gabor kernel, are obtained by convolving the input image with the RFs, and are denoted by $R_{s,i}^E(x, y)$ and $R_{s,i}^O(x, y)$, s being the scale, i the orientation ($\theta_i = i\pi/N_\theta$) and N_θ the number of orientations (here 8) with $i = [0, N_\theta - 1]$. Responses of complex cells are then modelled by the modulus $C_{s,i}(x, y) = [\{R_{s,i}^E(x, y)\}^2 + \{R_{s,i}^O(x, y)\}^2]^{1/2}$.

A basic scheme for line and edge detection (LE) is based on responses of simple cells: a positive (negative) line is detected where R^E shows a local maximum (minimum) and R^O shows a zero crossing. In the case of edges, the even and odd responses are swapped. This gives four possibilities for positive and negative events (polarity). An improved scheme (Rodrigues & du Buf, 2009a) consists of combining responses of simple and complex cells, i.e., simple cells serve to detect positions and event types, whereas complex cells are used to increase the confidence. Lateral and cross-orientation inhibition are used to suppress spurious cell responses beyond line and edge terminations, and assemblies of grouping cells serve to improve event continuity in the case of curved events.

At each (x, y) in the multiscale line and edge event space, four gating LE cells code the 4 event types: positive line, negative line, positive edge and negative edge (Rodrigues & du Buf, 2009a). These are coded by different levels of gray, from white to black, in the 3rd row of Figure 1. It shows 3 scales of the face in the 2nd row, middle column. For the results presented in this paper we used $\lambda = [4, 24]$ and $\Delta\lambda = 1$, scale $s = 1$ corresponding to $\lambda = 4$. With this LE information plus the lowpass information available through special retinal

ganglion cells (Berson, 2003), we can reconstruct in our visual system the face; for details and illustrations see Rodrigues and du Buf (2009a).

Keypoints are based on end-stopped cells (Rodrigues & du Buf, 2006). They provide important information because they code local image complexity. There are two types of end-stopped cells, single (S) and double (D). If $[\cdot]^+$ denotes the suppression of negative values, then

$$\begin{aligned} S_{s,i}(x, y) &= [C_{s,i}(x + dS_{s,i}, y - dC_{s,i}) \\ &\quad - C_{s,i}(x - dS_{s,i}, y + dC_{s,i})]^+ \end{aligned}$$

and

$$\begin{aligned} D_{s,i}(x, y) &= [C_{s,i}(x, y) - \frac{1}{2}C_{s,i}(x + 2dS_{s,i}, y - 2dC_{s,i}) \\ &\quad - \frac{1}{2}C_{s,i}(x - 2dS_{s,i}, y + 2dC_{s,i})]^+ \end{aligned}$$

with $C_i = \cos \theta_i$ and $S_i = \sin \theta_i$. The distance d is scaled linearly with filter scale s : $d = 0.6s$. All end-stopped responses along straight lines and edges are suppressed, for which tangential (T) and radial (R) inhibition, $I_s = I_s^T + I_s^R$, are used (Rodrigues & du Buf, 2006). Keypoints are detected by the local maxima of $K_s(x, y)$ in x and y , where

$$\begin{aligned} K_s(x, y) &= \max \left\{ \sum_{i=0}^{N_\theta-1} S_{s,i}(x, y) - gI_s(x, y), \right. \\ &\quad \left. \sum_{i=0}^{N_\theta-1} D_{s,i}(x, y) - gI_s(x, y) \right\}, \end{aligned}$$

with $g \approx 1.0$. Keypoints are shown by the diamond symbols in the 4th row of Figure 1, at the same scales as the LE information in the 3rd row. For a detailed explanation with illustrations see Rodrigues and du Buf (2006).

The “what” and “where” subsystems are steered, top-down, on the basis of expected faces or objects and positions in the prefrontal (PF) cortex (Deco & Rolls, 2004). Our eyes are constantly moving in order to suppress static projections of blood vessels, etc., in our retinae. During a fixation, stable information propagates from the retinae via the LGN to V1, where first features are extracted, and then, also during the next saccade, to higher areas. Fixation points in regions where complex and therefore important information can be found are much more important than points in homogeneous regions. Focus-of-attention, for guiding the where system in parallel with the steering of our eyes, is thought to be driven by an attention component in the PF cortex because of overt attention: while strongly fixating our eyes to one point, we can direct mental attention to points in the neighbourhood (Parkhurst et al., 2002).

For modelling FoA we need a map, called saliency map S , which indicates the most important points to be analysed (fixated). We propose a simple scheme based on the multiscale keypoint representation, because keypoints code local image complexity. The activities of all keypoint cells at position (x, y) are summed over scales s by grouping cells, assuming that each keypoint has a certain Region-of-Interest (RoI). The size of this is coupled to the scale (size) of the underlying simple and complex cells. At positions where keypoints are stable over many scales, this summation map will show distinct peaks at the centres of faces, also at important facial and contour landmarks. This data stream is data-driven and bottom-up, and could be combined with top-down processing from the inferior-temporal cortex in order to actively probe the presence of faces (or facial landmarks) and objects in the visual field (Deco & Rolls, 2004). The bottom row of Figure 1 shows the saliency maps of the face views on the 2nd row. For more details and illustrations see Rodrigues and du Buf (2006).

FACE RECOGNITION FRAMEWORK

Humans can detect faces under many conditions, including poor lighting and at large distances. Colour is one of the primary attributes in detection, but needs to be integrated with other attributes like keypoints for the detection of facial landmarks and their spatial relationships (Rodrigues & du Buf, 2006). There are several methods based on skin colour for face detection (Kakumanu et al., 2007). Agbinya and Silva (2005), although without a biological background, presented interesting results. Their method is now being implemented by using a biologically-inspired representation based on double-opponent colour blobs in V1 (Tailor et al., 2000), also combining other attributes to achieve accurate face segregation.

These authors propose segmentation of skin colour by filtering the colours of all pixels in HSV colour space. First, they compute the probability of each pixel belonging to skin, using a number of samples of skin colours. This information is aggregated and thresholds are used to create binary images in which each zone is independently tagged. In the next step - they only considered frontal views - to each zone containing at least two holes (the eyes) a face template is applied and checked for subsequent validation. At the end of this process, face segregation is achieved using only the skin region and its location.

As face segregation is beyond the scope of this paper, we consider faces that are already segregated, as in the “GavabDB” database (Moreno et al., 2004). The scheme presented below is a simplification, because in real vision the system starts first with a categorisation, for example on the basis of the colour of the hair or gender. After having a first gist, the system will dynamically select (in PF cortex) a group of possible templates, optimising the recognition process by changing parameters. Here we skip categorisation and focus on recognition. For this reason we consider all faces in our database as possible templates.

The “GavabDB” 3D face database (Moreno et al., 2004) contains 549 three-dimensional images of facial surfaces. These meshes correspond to 61 different individuals (45 male and 16 female) with 9 meshes of each person. All individuals are Caucasian and their age is between 18 and 40 years. Each image is given by a mesh of connected 3D points of the facial surface without texture. The database provides systematic variations with respect to pose and facial expression. In particular, the 9 images corresponding to each individual are: 2 frontal views with neutral expression; 2 x-rotated views ($\pm 30^\circ$, looking up and looking down respectively) with neutral expression; 2 y-rotated views ($\pm 90^\circ$, left and right profiles respectively) with neutral expression; and 3 frontal non-neutral expressions (laugh, smile and a random one chosen by the individual).

Figure 1 shows on the top row two expressions of the same face with the possible rotation intervals. The second row shows, from left to right, five 2D views, i.e., left profile, left lateral, frontal, right lateral and right profile of the y-rotated neutral face. The following rows illustrate the multiscale feature extractions described in this section, but only at three of all scales, $\lambda = \{4, 12, 24\}$.

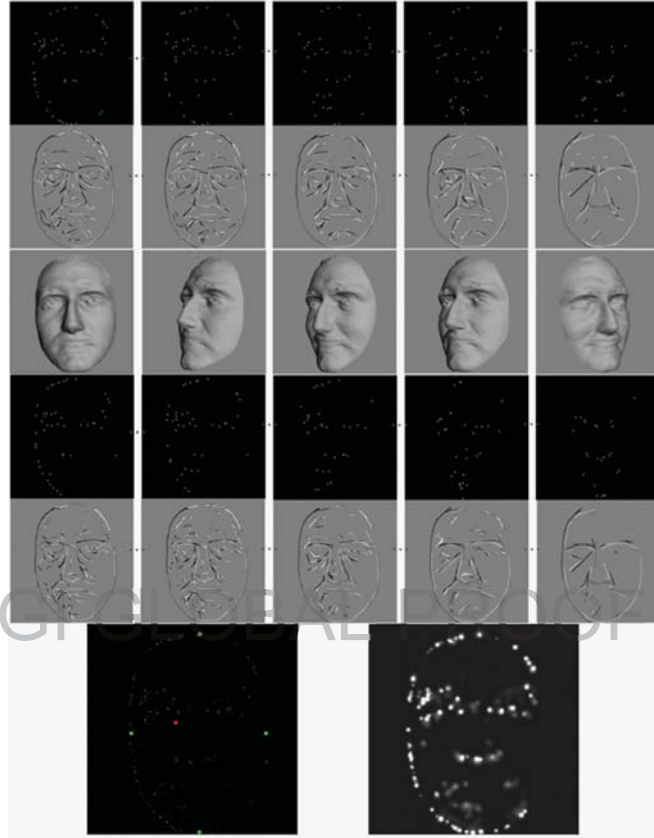
From the “GavabDB” database we randomly selected 10 individuals. Of each individual we took one 3D mesh with neutral expression to create five 2D views; see examples on the 2nd row in Figure 1 and the top three rows in Figure 4. These are used as templates stored in memory: frontal, lateral ($\pm 45^\circ$ y-rotated) and profile ($\pm 90^\circ$ y-rotated) (Valentin et al., 1997). For testing and for each face, 10 random images of each face were selected, considering: a) neutral or different expression, but discarding extreme expressions; b) any degree of y-rotation (pan); c) a maximum x-rotation of $\pm 10^\circ$ (tilt); and d) a maximum z-rotation of $\pm 2^\circ$. Images with the same rotation angles as the templates were excluded.

For each face, the templates stored in PF cortex are: the LE maps (in the present tests: 20 scales) with events characterised by type

and polarity (4: line/edge and positive/negative) for each view (5: frontal, lateral right/left and profile right/left), and the multiscale KPs maps (the same 20 scales used for LE). The last are used in conjunction with other processing schemes for dynamic routing to achieve normalisation of the pair to be matched (“face”/template) (Rodrigues & du Buf, 2009a, 2009b). Figure 2 shows on the first 2 rows part of the templates stored in memory, in the case of the frontal view shown in Figure 1: left to right, the multiscale KPs and LE maps at 5 of the 20 scales, equally spaced from fine to coarse scales on $\lambda \in [4, 24]$. The 3rd row shows examples of faces to be recognised. The fourth and fifth row show, for the leftmost image on the third row marked by a red square, the multiscale KPs and LE at the same scales as in the 1st row. The bottom row shows the summed KPs map with the accumulated keypoints marked in red. Also marked (in green) are the limits of the segregated face. On the right is the saliency map with the combined RoIs in white. The model consists of the following steps:

- (A) **Segregate the face from the scene:** This step consists of extracting the region where there is a face, for instance using colour information as briefly explained above. For small faces and/or z-rotated faces, size- and z-pose normalisation can be achieved by dynamic routing; see the small explanation in the section Object Recognition Framework and details in Rodrigues & du Buf (2009b). Here the faces are already segregated and normalised.
- (B) **Multiscale keypoint and line/edge detection:** For each input face we compute the keypoints, and lines and edges with their polarity. We use 20 scales $\lambda = [4, 24]$ with $\Delta\lambda = 1$. There is evidence that the horizontal tuning of face perception relies on the processing of intermediate and high spatial frequencies (Goffaux et al., 2011).
- (C) **Determine the view of the input face:** We compute the accumulated keypoints or AKPs. The AKPs are computed as follows: at each (x, y) in the multiscale keypoint

Figure 2. Top two rows, left to right: fine to coarse scales of the KPs and LE maps stored in memory in the case of the frontal view shown in Figure 1. Third row: examples of faces to be tested. Fourth and fifth rows: KPs and LE of the face marked by a red square in the 3rd row. Bottom row: the AKP marked in red on the left and saliency map on the right.



space, detected keypoints are first summed by grouping cells over all 20 scales, $mKP = \sum_s KP_s$. Then, by using two other grouping cells with large dendritic fields (DFs) the size of the segregated face, all existing mKP are summed over x and y,

$$AKP_x = \sum_x mKP * x$$

and

$$AKP_y = \sum_y mKP * y.$$

The two AKPs yield a single central position with coordinates x and y:

$$(x, y)_{AKP} = (AKP_x / mKP, AKP_y / mKP)$$

where $mKP = \sum_{DF} mKP$. The AKP position is marked in red in Figure 2 (bottom-left).

From the mKP map we compute the minimum and maximum coordinates in x and y, denoted by $CKP_{min/max, x/y}$. These are the first and last position in x and y where mKP has at least a value of 2; they are marked in green in Figure 2 (bottom-left). This means that at least two keypoint cells must have responded at the same position. With this information

we can compute the aspect ratio AR of the input face. Mathematically

$$AR = \frac{CKP_{\max,x} - CKP_{\min,x}}{CKP_{\max,y} - CKP_{\min,y}}$$

(C-i) Six gating cells are used to select the face view: frontal if $AR \in]0.61, 1]$, frontal-lateral if $AR \in]0.50, 0.61]$, lateral-frontal if $AR \in]0.40, 0.50]$, lateral-profile if $AR \in]0.33, 0.40]$, profile-lateral if $AR \in]0.31, 0.33]$ and profile if $AR \in]0, 0.31]$. These values were determined using the information from the templates, i.e., the ARs of the frontal, lateral and profile views were computed for each template and the average of all templates with the same view was calculated. The different levels of views were equally spaced between the anchor thresholds.

(C-ii) Two gating cells are used to select the lateral side: a face is seen from the right if coordinate AKP_x is closer to $CKP_{\max,x}$ or from the left if it is closer to $CKP_{\min,x}$.

The above processes may occur mainly in the dorsal where stream, i.e., the occipitoparietal area which exhibits object-selective responses and many 3D cues of shape, and can relay the information to cue- and view-invariant representations in the ventral what stream (Farivar, 2009).

(D) Construction of saliency and symbolic representation maps: For each input image we build the saliency map as explained in the previous section, and the symbolic representation maps where events (positions) are expanded by Gaussian cross-profiles (lines) and bipolar, Gaussian-truncated error-function profiles (edges). The sizes of these are coupled to the scale of the underlying simple and complex cells; see

Rodrigues and du Buf (2009a) for details and illustrations.

(E) Recognition process: We assume that templates (views) of faces are stored in memory and that these have been built through experience. Each face template in memory is represented by 5 views times 20 scales times 4 types of events (line/edge, positive/negative), which involves 400 maps.

The recognition scheme compares representations of input images (in the database) with those of templates (in memory). Depending on the input face view selected in step (C), the two most similar views of the templates are selected and weighted. If the input face is classified as frontal, the two selected templates are: the frontal and – depending on the detected face side, for instance the right one – the lateral-right template. The weight of each template is determined as a function of the aspect ratio AR, as shown in Table 1; the values of A and B, etc., are shown in Table 2.

At each scale (duplicated for the two selected views' templates), events in the 4 representation maps of the input image are compared with those in the corresponding maps of the templates, but only in the regions where the saliency map of the input image is active. These are the white regions in Figure 2 (bottom-right). Event co-occurrences are summed by grouping cells, which is a sort of event-type and scale-specific correlation. The outputs of the 4 event-type grouping cells are summed by another grouping cell (correlation over all event types). The global co-occurrence is determined by one more grouping cell which sums over all scales. A final grouping cell sums the results of the two views. The template (of the combined two views) with the maximum is selected by non-maximum suppression.

The multiscale line/edge representation is being exploited because this characterises facial features. Saliency maps, which have been used for Focus-of-Attention, are used to “gate” detected lines and edges in associated ROIs. This resembles the bottom-up data streams in the

Table 1. Weights applied to each template view as a function of the view assigned to the input image. Right is denoted by R and left by L. Values of weights are listed in Table 2

input face	Templ. view 1	weight	Templ. view 2	weight
Frontal R/L	Frontal	A	Lateral R/L	1-A
Frontal-Lateral R/L	Frontal	B	Lateral R/L	1-B
Lateral-Frontal R/L	Lateral R/L	C	Frontal	1-C
Lateral-Profile R/L	Lateral R/L	C	Profile R/L	1-C
Profile-Lateral R/L	Profile R/L	B	Lateral R/L	1-B
Profile R/L	Profile R/L	A	Lateral R/L	1-A

Table 2. Recognition rates (right column) using different weights for two views and using a single view (see text)

templates	weights	results
two views- f(input)	A=0.8;B=0.6;C=0.6	91%
two views- f(input)	A=0.5;B=0.5;C=0.5	90%
two views- f(input)	A=0.9; B=0.2; C=0.6	90%
two views- frontal & lateral	0.5	88%
two views- frontal & profile	0.5	73%
single view- frontal	-	56%
single view- lateral R/L	-	58%/23%
single view- profile R/L	-	37%/27%

where (FoA) and what (lines/edges) subsystems. However, it remains a simplification because processing is limited to cortical area V1, whereas in reality the two subsystems contain higher-level feature extractions in areas V2, V4, etc. (Hamker, 2005). The same way, top-down data streams are simplified by assuming that face templates held in memory are limited to lines, edges and keypoints, and 2D canonical views are limited to frontal plus left/right lateral and left/right profile.

OBJECT RECOGNITION FRAMEWORK

In Rodrigues and du Buf (2009b) we explain in detail how invariant object categorisation and recognition can be achieved by using a 2D

model with a single view of a template and an object. This process is called dynamic routing. It basically consists of matching central keypoints at very coarse scales – such central keypoints at coarse scales are always located at or close to an object's centroid – of input object and template. Hence, they are made to coincide by a virtual translation. After this, the same routing principle is applied to the two most significant peaks of the saliency maps, one of the input object and one of the template (virtual scaling and rotation).

The resulting routing is then applied to all significant peaks, because they belong to a single object/template pair, and all significant peaks of saliency maps of the input object and the template are tested in order to check whether sufficient coinciding pairs exist for a match. Up to here, only saliency maps are used

to find possibly matching templates, but mainly for dynamic routing which virtually “superimposes” the input object and templates. In the next step the dynamic routing of keypoints is also applied to the multi-scale line/edge representations in order to check whether an object and a template really correspond. This is done by many grouping cells with small DFs (local correlation of line/edge events) and one with a big DF (global

object/template correlation): two edges of object and template count for a match if they are at the same position but also if they are very close to each other. A similar process with line/edge position relaxation can be used for 3D face and object recognition. It should be stressed that the template information used in each step depends on the categorisation-recognition level. For example, in the case of a first, fast but coarse pre-categorisation, only line/edge events at coarse scales of the segregated (binary) object are used. In the subsequent finer categorisation we used line/edge events at all scales obtained from the entire object, instead of from the binary segregation. In final recognition, line and edge events are treated separately: object lines must match template lines and edges must match edges. Here we do not apply the principle of dynamic routing, because it is beyond the scope of this paper and our aim is to show that a predefined number of views (8 for animal objects) is enough to recognise such 3D objects if they are rotated 360° around the vertical axis.

From the AIM@SHAPE (Falcidieno, 2008) database we selected eight 3D mesh models of animals; see Figure 3 (top) and Figure 5. Of each model we created eight 2D views; see the two examples in Figure 3 (bottom). These are used as templates stored in memory: frontal, lateral ($\pm 45^\circ$ y-rotated), profile ($\pm 90^\circ$), rear (180°) and lateral rear ($\pm 135^\circ$). For testing and for each object, 12 random images of each object were selected, considering: a) any degree of y-rotation (pan); b) a maximum x-rotation of $\pm 10^\circ$ (tilt); and c) a maximum z-rotation of $\pm 2^\circ$; see examples in Figure 5. Images with the same rotation angles as the templates were excluded.

The framework used for object recognition is the same as that for faces, but with the following modifications: In step (C-i) six gating cells are used to select the object view: frontal if $AR=[0.0,0.6]$, frontal-lateral if $AR=]0.6,0.8]$, lateral-frontal if $AR=]0.8,0.9]$, lateral-profile if $AR=]0.9,1.0]$, profile-lateral if $AR=]1.0,1.1]$ and profile if $AR>1.1$. These values were determined using the information from the templates, i.e., the ARs of the frontal, lateral and profile views were computed for each template and the average of all templates with the same view was calculated (considering both frontal and rear views; see Figure 3, middle and bottom rows). The different AR levels of the views were equally spaced between the anchor thresholds.

In step (C-ii) two gating cells are used to select the lateral side: an object is seen from the right if coordinate AKP_x is closer to $CKP_{max,x}$ or from the left if it is closer to $CKP_{min,x}$. We must stress that this process only works for asymmetrical objects like animals, that all objects must be normalised in memory in the same way (in this case all animals with their heads to the left), and the reference side for the left/right view is taken to be the rear view, opposite to the head, because normally the keypoint density is much larger there.

In the recognition scheme, as in the case of faces, we compared representations of input images (in the database) with those of templates (in memory). Depending on the input view selected in step (C), the two most similar views of the templates are selected and weighted, but now in two pairs corresponding to the front and the rear of the object. If the input object is classified as “frontal” ($AR=[0.0,0.6]$), the two pairs of selected templates are the frontal and – depending on the detected object side, for instance the right one – the frontal-lateral-right template (first pair), and the rear and rear-lateral-right template (second pair). The weight of each template is determined as a function of the aspect ratio AR, as shown in Table 1, with the values $A = 0.8$ and $B = C = 0.6$. We used the same weights for both pairs.

Figure 3. Two examples of object templates (animals) from the AIM@SHAPE (Falcidieno, 2008) database

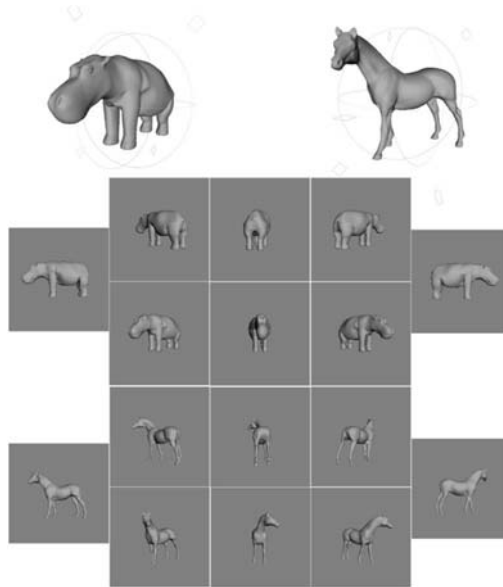


Figure 4. Examples of other templates (top 3 rows) and tested faces (bottom 3 rows)

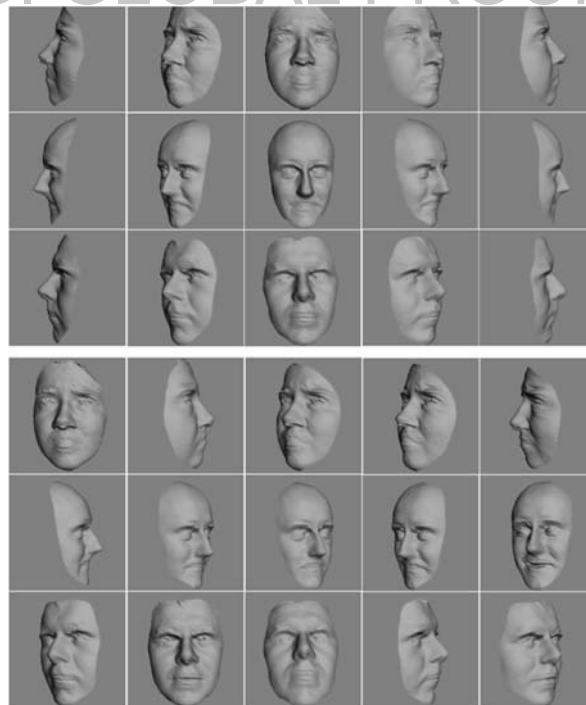
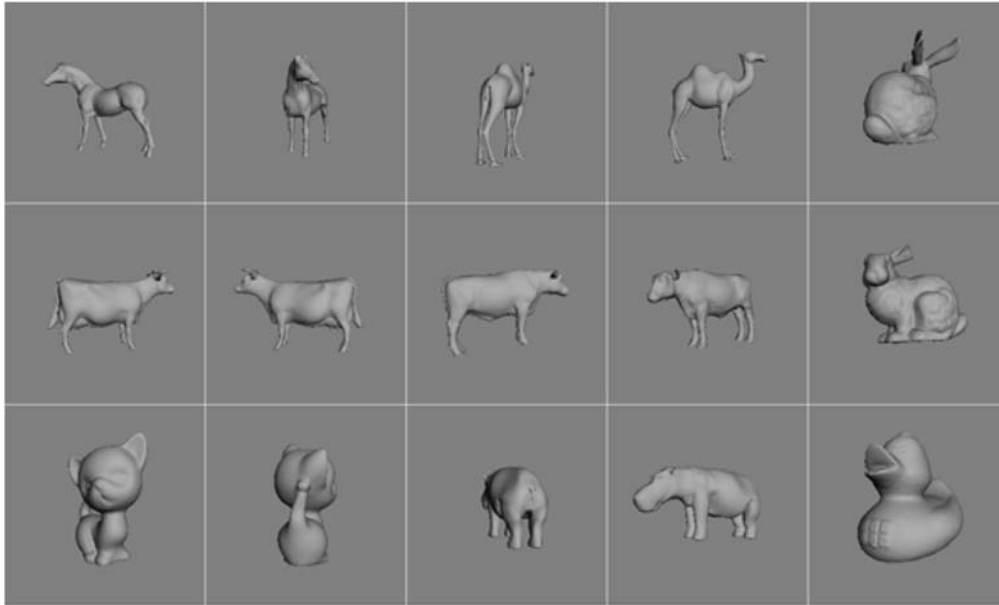


Figure 5. Examples of tested objects (animals)



At each scale, for each pair (front and rear), and in separate and duplicated templates for the two selected views, events in the 4 representation maps of the input image are compared with those in the corresponding maps of the templates, but only in the regions where the saliency map of the input image is active. Event co-occurrences are summed by grouping cells, which is a sort of event-type and scale-specific correlation. The outputs of the 4 event-type grouping cells are summed by another grouping cell. The global co-occurrence is determined by one more grouping cell which sums over all scales. A final grouping cell sums the results of the two views. The two templates for the front and one for the rear (of the combined two views) with the maximum are selected by non-maximum suppression. The selection of the final template (recognition) and view (front vs. rear) of the object is also selected by non-maximum suppression. In summary, in the first step we select the view of the object (left-right plus frontal-lateral-profile), and in the second step we check if the head of the animal (or a

reference view of an object) is facing to the front or to the back.

RESULTS AND DISCUSSION

Face Framework

For testing the face framework we used 100 images, 10 views per person, with different expressions (including the neutral) and also with different degrees of rotation: x axis $[-10^\circ, +10^\circ]$, y axis $[-90^\circ, +90^\circ]$ and z axis $[-2^\circ, +2^\circ]$. Figure 4 shows representative examples on the bottom three rows.

We tested the algorithm for: (a) The correct face side (left/right), which yielded a result of 99%. (b) The correct view must correspond to one of the three categories frontal, lateral or profile. For this we considered that an image returning frontal or frontal-lateral fits the frontal class, lateral-frontal and lateral-profile fit the lateral class, and profile and profile-lateral fit the profile class. The overall recognition rate was 93%, with the following misclassifications:

3% lateral assigned to profile and 4% profile assigned to lateral.

We also tested (c) different numbers of templates and different weights of the pairs of template views. The best result of 91% was achieved using the 5 views (templates) of each face, combined in pairs of two in function of the input view (Table 1), with $A = 0.8$ and $B = C = 0.6$. These three parameters fine-tune the model and should change from face to face in function of the initial gist, the gender and facial expression, etc. This dynamic weighting remains to be implemented. Table 2 summarises the most important tests, using either two combined views for each template face or a single view.

We must briefly explain how the tests reported in Table 2 were conducted. In the cases of "single view" step C (determine the view of the input face) was removed from the model. In the cases of "two views" frontal & lateral or frontal & profile, only the right/left detection of step C was applied, with both views equally weighted. In the case of "two views - f(input)" the entire model was applied. From the results we can see that using a single frontal view (56%) is not enough to recognise a face in different views. Nevertheless, using two views (frontal & lateral) plus the face side and 3 templates per face (frontal, lateral-right and -left), the results approach the best result achieved. As expected, this means that 5 templates give the best characterisation of the different face views, but if the view selection is undetermined or fuzzy, the most important templates to be used are the frontal and lateral ones.

It is possible to compare our results with those of other models which were tested on the GavabDB database. Moreno and Sanchez (2004), who created GavabDB, developed a feature-based model and reported a recognition rate of 78.0%. Celenk and Aljarrah (2006) projected the face scans to 2D range images and applied a PCA approach, achieving 92.0%. In their work only frontal projections of 60 persons were tested, the same projections as used by Li et al. (2009), but with a recognition rate of 94.7%. Li et al. (2009) also reported results

from 4 more authors who used the same database, the results ranging from 83.0% to 91.0%. Rashad et al. (2009) used 427 surface images of all 61 persons in the database, and achieved a recognition rate of 80.3%.

Although only based on 10 randomly selected persons from all 61 in the database, our best result of 91% is close to the best results achieved by the other groups, despite the fact that most only considered frontal views. In addition, our method is the only biologically-inspired one which can cope with different views of the same person.

Object Framework

For testing the model in case of objects, we used the eight 3D animal models camel, cow, cow2, horse, hippo, duck, bunny and kitten, with a total of 96 projections which included 12 views of each object with different degrees of rotation: x axis $[-10^\circ, +10^\circ]$, y axis $[-180^\circ, +180^\circ]$ and z axis $[-2^\circ, +2^\circ]$; see the examples in Figure 5. We tested the algorithms for: (a) The correct object side (left/right), which yielded a result of 84%. (b) The correct view. Like faces, these must correspond to one of the three categories frontal/rear, lateral and profile. This test resulted in 90% correct views. (c) Identifying the frontal or rear view, which yielded a result of 82%. The overall recognition rate of views and objects was 83%.

The misclassifications of test (a) were mainly due to the duck, because being a toy the duck's head is much bigger than its rear. This implies a larger keypoint density of the head view relative to the rear view; please recall that the reference to determine the animal's side is the rear view due to the normally larger keypoint density. In test (b) many misclassifications were lateral views assigned to profile views for several animals. In test (c) some animals were facing opposite sides (front to rear, rear to front), but in this case some misclassified animals in the overall recognition were still correctly classified because the front and the rear views, considering the overall shape, are quite similar, for example the camel. The object

recognition framework cannot yet be compared with other frameworks, because we tested only eight objects/animals and with rotations mainly about the y axis. Nevertheless, Lam and du Buf (2011) achieved 92.5% with 40 objects and in the SHREC contest of 2010, with 10,000 3D models, the recognition rates were between 45 and 70% (Veltkamp et al., 2010).

CONCLUSION

We presented a bio-inspired face-recognition model that can determine the side (left/right) and the view (frontal/lateral/profile) of a face to be recognised, as well as recognise faces with different views and facial expressions. Nevertheless, the presented model is a simplification, because in real vision the system starts with a first segregation and categorisation, for example on the basis of the colour of the hair and skin. After having a first gist (a group of possible face templates), the system will dynamically select the template views according to the view of the input face, and optimise the recognition process by changing parameters in relation to the gender and facial expression, etc. In view of the tremendous amount of data already involved in our simple experiments, the entire system has been developed in different modules which will be integrated in the future with GPU processing.

The system achieves good results mainly because the line/edge representation at coarser scales provides a stable abstraction of facial features. This explains, at least partly, the generalisation that allows us to classify faces with different expressions and views. The problem of normalisation, which is not addressed here, can be solved by using a segregated face based on colour with detected keypoints and dynamic routing (Rodrigues & du Buf, 2006). Keypoints can be used to determine facial landmarks (eye, nose, mouth), which was already implemented and tested for frontal views (Rodrigues & du Buf, 2006). In this paper we also showed that keypoints can be used to determine the view of the face. A complementary approach is to

compute the disparity, distances, angles and areas between points on the 3D facial surface. This procedure can also guarantee that templates in memory are really representative.

An interesting aspect for future research is the incorporation of age and biometric differences (e.g., gender, colour of the skin, age, birth marks, etc.), also expression classification already achieved by using multiscale lines and edges (Sousa et al., 2010). As for now, face recognition with extreme expressions or newly grown beards, etc., remains a big challenge. Furthermore, occlusions caused by objects like sunglasses must be addressed in a systematic way.

We also showed that the same framework, with small modifications, can be applied to 3D object recognition, although more tests need to be done and full 3D recognition can only be achieved by using more views (templates). The minimum number of views is still subject to research. Another open aspect is how objects are stored in memory: all animal views are defined by relating them to same side, like all heads to the left as in this paper. Instead, this could depend on the object type and be a function of one or several features, for instance keypoint density or focus-of-attention. It should be stressed that object recognition has been explored as a special case of face recognition because of the aspect ratios of 2D projections. Using the same framework does not mean that faces and general objects cannot be processed in different ways in the visual cortex. Indeed, there are several indications that there are differences (Biederman & Kalocsai, 1997). The facts that faces are very important in our social behaviour, and that there are cells which only respond when a face is present in their receptive field, point at the existence of a special face-processing subsystem, part of which may overlap the circuitry for dealing with general objects. The main difference is that faces always have the same geometry, more or less, which was exploited in face detection on the basis of the multi-scale keypoint representation by assuming standard relations between eyes, nose and mouth (Rodrigues & du Buf, 2006). A similar

geometry with standard relations between parts does not exist in case of general 3D objects and, unlike faces, most general 3D objects can be rotated arbitrarily. The latter requires storing many more feature templates in visual memory which cover all canonical views.

Despite the problems and possible solutions mentioned above, the results obtained are very encouraging. We expect significant improvements by implementing a dynamic system, in which successive tests are performed each time that more complete information is available, starting at coarse scales and adding then finer scales, such that all effort can be spent on scrutinising the images, which have not yet been identified with absolute certainty. This procedure simulates the processing in the bottom-up and top-down data streams in the what and where subsystems of our visual system.

ACKNOWLEDGMENT

This work was supported by the Portuguese Foundation for Science and Technology (FCT) through the PIDDAC Program funds (ISR/IST plurianual funding), EC project Neuro Dynamics (NeFP7-ICT-2009-6 PN: 270247) and FCT project Blavigator (RIPD/ADA/109690/2009).

REFERENCES

- Abate, A., Nappi, M., Riccio, D., & Sabatino, G. (2007). 2D and 3D face recognition: A survey. *Pattern Recognition Letters*, 28(14), 1885–1906. doi:10.1016/j.patrec.2006.12.018
- Agbinya, J. I., & Silva, S. D. (2005). Face recognition programming on mobile handsets. In *Proceedings of the 12th International Conference on Telecommunications*, Cape Town, South Africa (pp. 3-6).
- Bar, M. (2004). Visual objects in context. *Nature Reviews. Neuroscience*, 5, 619–629. doi:10.1038/nrn1476
- Berretti, S., Del Bimbo, A., & Pala, P. (2010). 3D face recognition using iso-geodesic stripes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(12), 2162–2177. doi:10.1109/TPAMI.2010.43
- Berson, D. (2003). Strange vision: ganglion cells as circadian photoreceptors. *Trends in Neurosciences*, 26(6), 314–320. doi:10.1016/S0166-2236(03)00130-9
- Biederman, I., & Kalocsai, P. (1997). Neurocomputational bases of object and face recognition. *Philosophical Transactions of the Royal Society B. Biological Science*, 352, 1203–1219. doi:10.1098/rstb.1997.0103
- Blank, I., & Yove, G. (2011). The structure of face-space is tolerant to lighting and viewpoint transformations. *Journal of Vision (Charlottesville, Va.)*, 11(8), 1–13. doi:10.1167/11.8.15
- Bowyer, K., Chang, K., & Flynn, P. (2006). A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition. *Computer Vision and Image Understanding*, 101(1), 1–15. doi:10.1016/j.cviu.2005.05.005
- Celenk, M., & Aljarrah, I. (2006). Interl shape-deformation invariant 3D surface matching using 2D principal component analysis. In *Proceedings of the SPIE-IS&T Conference on Electronic Imaging* (Vol. 6056, pp. 118-129).
- Deco, G., & Rolls, E. T. (2004). A neurodynamical cortical model of visual attention and invariant object recognition. *Vision Research*, 44(6), 621–642. doi:10.1016/j.visres.2003.09.037
- Falcidieno, B. (2008). *AIM@SHAPE*. Retrieved from <http://www.aimatshape.net>
- Farivar, R. (2009). Dorsal-ventral integration in object recognition. *Brain Research. Brain Research Reviews*, 61(2), 144–153. doi:10.1016/j.brainres-rev.2009.05.006
- Farrajota, M., Rodrigues, J. M. F., & du Buf, J. M. H. (2011, January 26-29). Optical flow by multi-scale annotated keypoints: A biological approach. In *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing*, Rome, Italy (pp. 307-315).
- Goffaux, V., van Zon, J., & Schiltz, C. (2011). The horizontal tuning of face perception relies on the processing of intermediate and high spatial frequencies. *Journal of Vision (Charlottesville, Va.)*, 11(10), 1–9. doi:10.1167/11.10.1
- Grossberg, S., Srinivasan, K., & Yazdanbakhsh, A. (2011). On the road to invariant object recognition: How cortical area V2 transforms absolute to relative disparity during 3D vision. *Neural Networks*, 24(7), 686–692. doi:10.1016/j.neunet.2011.03.021

- Hamker, F. (2005). The reentry hypothesis: the putative interaction of the frontal eye field, ventrolateral prefrontal cortex, and areas V4, IT for attention and eye movement. *Cerebral Cortex*, 15, 431–447. doi:10.1093/cercor/bhh146
- Haxby, J., Hoffman, E., & Gobbini, M. (2002). Human neural systems for face recognition and social communication. *Biological Psychiatry*, 51(1), 59–67. doi:10.1016/S0006-3223(01)01330-0
- Heitger, F., Rosenthaler, L., von der Heydt, R., Peterhans, E., & Kübler, O. (1992). Simulation of neural contour mechanisms: from simple to end-stopped cells. *Vision Research*, 32(5), 963–981. doi:10.1016/0042-6989(92)90039-L
- Hubel, D. H. (1995). *Eye, brain and vision*. New York, NY: Scientific American Library.
- Kakumanu, P., Makrogiannis, S., & Bourbakis, N. (2007). A survey of skin-color modeling and detection methods. *Pattern Recognition*, 40(3), 1106–1122. doi:10.1016/j.patcog.2006.06.010
- Kourtzi, Z., & Connor, C. (2011). Neural representations for object perception: structure, category, and adaptive coding. *Annual Review of Neuroscience*, 34, 45–67. doi:10.1146/annurev-neuro-060909-153218
- Lam, R., & du Buf, J. M. H. (2011). Retrieval of 3D polygonal objects based on multiresolution signatures. In G. Bebis, R. Boyle, B. Parvin, D. Koracin, S. Wang, K. Kyungnam, et al. (Eds.), *Proceedings of the International Symposium on Visual Computing, Part II*, Las Vegas, NV (LNCS 6939, pp. 136–147).
- Li, X., Jia, T., & Zhang, H. (2009). Expression-insensitive 3D face recognition using sparse representation. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition* (pp. 2575–2582).
- Martins, J. A., Rodrigues, J. M. F., & du Buf, J. M. H. (2009, February 5–8). Focus of attention and region segregation by low-level geometry. In *Proceedings of the International Conference on Computer Vision-Theory and Applications*, Lisbon, Portugal (Vol. 2, pp. 267–272).
- Moreno, A. B., & Sanchez, A. (2004). GavabDB: A 3D face database. In *Proceedings of the 2nd COST275 Workshop on Biometrics on the Internet*, Vigo, Spain.
- Oliva, A., & Torralba, A. (2006). Building the gist of a scene: the role of global image features in recognition. *Progress in Brain Research: Visual Perception*, 155, 23–26.
- Parkhurst, D., Law, K., & Niebur, E. (2002). Modelling the role of salience in the allocation of overt visual attention. *Vision Research*, 42(1), 107–123. doi:10.1016/S0042-6989(01)00250-4
- Passalis, G., Kakadiaris, I., & Theoharis, T. (2007). Intraclass retrieval of nonrigid 3D objects: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(2), 218–229. doi:10.1109/TPAMI.2007.37
- Pinto, N., Barhomi, Y., Cox, D., & DiCarlo, J. (2010). Comparing state-of-the-art visual features on invariant object recognition tasks. In *Proceedings of the IEEE Workshop on Applications of Computer Vision*, Kona, HI (pp. 463–470).
- Ramirez-Valdez, L., & Hasimoto-Beltran, R. (2009). 3D-facial expression synthesis and its application to face recognition systems. *Journal of Applied Research and Technology*, 7, 323–339.
- Rashad, A., Hamdy, A., Saleh, M., & Eladawy, M. (2009). 3D face recognition using 2DPCA. *International Journal of Computer Science and Network Security*, 9(12), 149–155.
- Rodrigues, J., & du Buf, J. M. H. (2004). Visual cortex frontend: integrating lines, edges, keypoints and disparity. In A. Campilho & M. Kamel (Eds.), *Proceedings of the International Conference on Image Analysis And Recognition* (LNCS 3211, pp. 664–671).
- Rodrigues, J., & du Buf, J. M. H. (2006). Multi-scale keypoints in V1 and beyond: object segregation, scale selection, saliency maps and face detection. *Bio Systems*, 2, 75–90. doi:10.1016/j.biosystems.2006.02.019
- Rodrigues, J., & du Buf, J. M. H. (2009a). Multi-scale lines and edges in V1 and beyond: brightness, object categorization and recognition, and consciousness. *Bio Systems*, 95, 206–226. doi:10.1016/j.biosystems.2008.10.006
- Rodrigues, J., & du Buf, J. M. H. (2009b). A cortical framework for invariant object categorization and recognition. *Cognitive Processing*, 10(3), 243–261. doi:10.1007/s10339-009-0262-2
- Rodrigues, J. M. F., & du Buf, J. M. H. (2011). A cortical framework for scene categorization. In *Proceedings of the International Conference on Computer Vision, Imaging and Computer Graphics-Theory and Applications*, Vilamoura, Portugal (pp. 364–371).

- Smeets, D., Claes, P., Vandermeulen, D., & Clement, J. (2010). Objective 3D face recognition: Evolution, approaches and challenges. *Forensic Science International*, (1-3): 125–132. doi:10.1016/j.forsciint.2010.03.023
- Sousa, R., Rodrigues, J. M. F., & du Buf, J. M. H. (2010, June 21-23). Recognition of facial expressions by cortical multi-scale line and edge coding. In *Proceedings of the International Conference on Image Analysis and Recognition*, Povoá do Varzim, Portugal (Vol. 1, pp. 415-424).
- Su, T., Lin, C., Lin, P., & Hu, J. (2006). Shape memorization and recognition of 3D objects using a similarity-based aspect-graph approach. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, Taipei, Taiwan (pp. 4920-4925).
- Taylor, D., Finkel, L., & Buchsbaum, G. (2000). Color-opponent receptive fields derived from independent component analysis of natural images. *Vision Research*, 40(19), 2671–2676. doi:10.1016/S0042-6989(00)00105-X
- Tangelder, J., & Veltkamp, R. (2008). A survey of content based 3D shape retrieval methods. *Multimedia Tools and Applications*, 39, 441–471. doi:10.1007/s11042-007-0181-0
- Valentin, D., Abdi, H., & Edelman, B. (1997). What represents a face? A computational approach for the integration of physiological and psychological data. *Perception*, 26(10), 1271–1288. doi:10.1068/p261271
- Veltkamp, R. C., Giezeman, G. J., Bast, H., Baumbach, T., Furuya, T., Giesen, J., et al. (2010). Shrec 2010 track: Large scale retrieval. In *Proceedings of the Eurographics/ACM SIGGRAPH Symposium on 3D Object Retrieval* (pp. 63-69).

ENDNOTE

- ¹ y-rotated stands for rotated about the vertical y-axis. The x-axis is horizontal “in the plane of the paper,” whereas the z-axis is “perpendicular to the paper.”

IGI GLOBAL PROOF

João Rodrigues, born in 1971, graduated in Electrical Engineering in 1993 at the University of Trás-os-Montes and Alto Douro (Portugal). He obtained his PhD degree in 2008 at the University of the Algarve in Faro, where he lectures computer science courses at the Instituto Superior de Engenharia. He is a senior researcher of the Vision Laboratory (UALg) since 1996, and is member of the Institute for Systems and Robotics (Lisbon) and the Portuguese chapter of the IAPR. His major research interests concern human vision: gist, attention and object categorisation and recognition.

Roberto Lam, born in 1960, graduated in Computer Science in 1995. In 2001 he obtained an MSc degree at the University of the Algarve in Faro, where he lectures computer science courses at the Instituto Superior de Engenharia. Presently he is pursuing a PhD degree in the Vision Laboratory (UALg). He is member of the Institute for Systems and Robotics (Lisbon) and the Portuguese Chapter of Eurographics. His major research interest is tridimensional modelling: 3D object representation, recognition and retrieval.

Hans du Buf was born in 1951. He received a PhD degree at the Technical University of Eindhoven (The Netherlands) in 1987. He then worked at the Swiss Federal Institute of Technology in Lausanne (Switzerland). In 1994 he was appointed associate professor at the Dept. of Electronics and Computer Science, University of the Algarve, in Faro. He conducts research in image processing, pattern recognition, computer graphics and visual perception. He is a member of the Portuguese chapter of the IAPR, and is associate editor of the International Journal of Pattern Recognition and Artificial Intelligence.

Haptic Interaction with Elastic Volumetric Structures

Francisco J. R. Prados, Universidad de Granada, Spain

Alejandro León Salas, Universidad de Granada, Spain

Juan Carlos Torres, Universidad de Granada, Spain

ABSTRACT

Considerable efforts have been done to produce realistic results when simulating interaction with elastic materials. Many applications such as surgery planning, medical training, or virtual sculpting would benefit from a plausible simulation scenario. However, even though many works have proposed very satisfactory results, realistic simulation of deformable bodies is still an open issue. One of the challenges when designing a realistic elastic body simulation is the huge amount of data that needs to be processed. For the inner properties of the material are crucial when it comes to reproduce the elastic problem, the simulation naturally calls for volumetric information. In this paper the authors propose a technique to interactively deform 3D images, such as those acquired by a CT scanner. While producing a physically plausible haptic feedback, deformation and visualization algorithms produce an efficient and natural feeling. Using a free form deformation structure as a wrapper, it is possible to deform complex structures at high frame rates, independently of the size of the volume.

Keywords: Computer Graphics, Computer Science, Deformable Models, Elastic Deformation, Haptic Interaction, Volume Graphics

INTRODUCTION

Simulation of rigid body deformation has been profusely studied by the computer graphics community. Realistic deformation models have immediate and obvious applications in many areas, such as virtual reality, surgery simulation or virtual sculpting. Since 1984, when Sederberg and Parry presented the Free Form Deformation method (Sederberg & Parry, 1986), a vast collection of works have aimed to create efficient

and realistic methods to create the illusion of elasticity or plasticity.

However, and similarly to any other physical simulation, realism and efficiency are confronted. On the one side, physically based simulations require many resources and computation time, which consequently limits the size of the model involved. On the other side, simplified or heuristic deformations are easier to compute, but obviously produce less robust results.

The proposed method tries to find an equilibrium between these two approaches. A natural tactile experience is generated using physical

DOI: 10.4018/jcicg.2012010105