Cortical 3D Face Recognition Framework

J.M.F. Rodrigues and R. Lam and J.M.H. du Buf

Institute for Systems and Robotics (ISR), Vision Laboratory - University of the Algarve (ISE and FCT), 8005-139 Faro, Portugal
{jrodrig, rlam, dubuf}@ualg.pt

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 routing and saliency maps for Focus-of-Attention. All these combined allow us to segregate faces. Events of different facial views are stored in memory and combined in order to identify the view and recognise the face including facial expression. In this paper we show that with five 2D views and their cortical representations it is possible to determine the left-right and frontal-lateral-profile views and to achieve view-invariant recognition of 3D faces.

Categories and Subject Descriptors (according to ACM CCS): I.5.5 [Pattern Recognition]: Implementation—Special architectures

1. 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 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 eye-witnesses, although numerous studies have shown that people are not always reliable sources when comparing faces with recollections [SCVC10].

Recently, because of the limitations of 2D approaches and with the advent of 3D scanners, face recognition research has shifted from 2D to 3D with a concurrent improvement in performance. There are many face-recognition methods in 2D and 3D; for detailed surveys see [BCF06, ARS07]. Recently, Rashad et al. [RHSE09] 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 et al. [RVHB09] also related 3D facial expression to recognition. Berretti et al. [BDBP10] 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.

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: problematic are all possible combinations of changes in illumination, pose and age, with artefacts like beards, moustaches and glasses, including different facial expressions and partial occlusions. 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.

2. 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 [HHM02]. Face and object detection, categorisation and recognition can be obtained by means of bottom-up and top-down data streams in the so-called "what" and "where" subsystems [DR04], including the integration of both subsystems [Far09]. In cortical area V1 there are simple and complex cells, which are tuned to different spatial frequencies (scales) and orientations, but also disparity (depth) because of neighbouring left-right hypercolumns [Hub95]. These cells provide input for grouping cells which code line and edge information [RdB09b] and attribute depth information [RdB04]. In V1 there are also double-opponent colour blobs [TFB00]

and end-stopped cells which, together with complicated inhibition processes, allow to extract keypoints (singularities, vertices and points of high curvature) [F. 92, RdB06b] and optical flow [FRdB11].

On the basis of these models and neural processing schemes, it is now possible to create a cortical architecture bootstrapped by global and local gist [MRdB09, RdB11], with face and figure-ground segregation [RdB06b, RdB09b, FRdB11], Focus-of-Attention (FoA) [RdB06b, MRdB09], face/object categorisation and recognition [RdB06b, RdB09b], including recognition of facial expressions [SRdB10].

There are several open questions related to the perception and recognition of 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 this paper we focus on a cortical model for 3D face recognition. The present model is based on a previous one [RdB09b] which employs multiscale line/edge representations and keypoints 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 any degree of rotation (y-rotated $\pm 90^o$, pan; x-rotated $\pm 10^o$, 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).

The rest of this paper is organised as follows: In Section 3 the keypoint, line and edge extractions with the construction of the saliency maps for FoA are explained. Section 4 explains the face recognition model. Experimental results and discussion are reported in Section 5, and we conclude in Section 6.

3. Multiscale line, edge, keypoint and saliency maps

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 scales offer different qualities of information [Bar04, OT06].

Gabor quadrature filters provide a model of cortical simple cells [RdB06b]. 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

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

with $\tilde{x} = x\cos\theta + y\sin\theta$ and $\tilde{y} = y\cos\theta - x\sin\theta$, 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 $R^E_{s,i}(x,y)$ and $R^O_{s,i}(x,y)$, 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

$$C_{s,i}(x,y) = [\{R^E_{s,i}(x,y)\}^2 + \{R^O_{s,i}(x,y)\}^2]^{1/2}.$$

A basic scheme for line and edge detection (LE_s) 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 [RdB09b] 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 [RdB09b]. These are coded by different levels of gray, from white to black, in the 3rd row of Fig. 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 LEs information plus the low-pass information available through special retinal ganglion cells [Ber03], we can reconstruct in our visual system the face; for details and illustrations see [RdB09b].

Keypoints are based on end-stopped cells [RdB06b]. 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

$$S_{s,i}(x,y) = [C_{s,i}(x + d\mathcal{S}_{s,i}, y - d\mathcal{C}_{s,i}) - C_{s,i}(x - d\mathcal{S}_{s,i}, y + d\mathcal{C}_{s,i})]^+$$

and

$$D_{s,i}(x,y) = [C_{s,i}(x,y) - \frac{1}{2}C_{s,i}(x + 2d\mathcal{S}_{s,i}, y - 2d\mathcal{C}_{s,i}) - \frac{1}{2}C_{s,i}(x - 2d\mathcal{S}_{s,i}, y + 2d\mathcal{C}_{s,i})]^+$$

with $\mathcal{C}_i = \cos\theta_i$ and $\mathcal{S}_i = \sin\theta_i$.

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, $I_s = I^T_s + I^R_s$, are used [RdB06b]. Keypoints are detected by the local maxima of $K_s(x,y)$ in x and y, where

$$K_s(x,y) = \max\{\sum_{i=0}^{N_\theta-1} S_{s,i}(x,y) - gI_s(x,y), \sum_{i=0}^{N_\theta-1} D_{s,i}(x,y) - gI_s(x,y)\},$$

with $g \approx 1.0$. Keypoints are shown by the diamond shapes in the 4th row of Fig. 1, at the same scales as the LEs information in the 3rd row. For a detailed explanation with illustrations see [RdB06b].

The "what" and "where" subsystems are steered, top-down, on the basis of expected faces or objects and positions

in the prefrontal (PF) cortex [DR04]. 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 [PLN02].

For modeling 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 [DR04]. The bottom row of Fig. 1 shows the saliency maps of the face views on the 2nd row. For more details and illustrations see [RdB06b].

For illustration purposes and tests we used the "GavabDB" 3D face database [MS04]. It 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 ($\pm 30^o$, looking up and looking down respectively) with neutral expression; 2 y-rotated views ($\pm 90^o$, 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, $\lambda = \{4, 12, 24\}$.

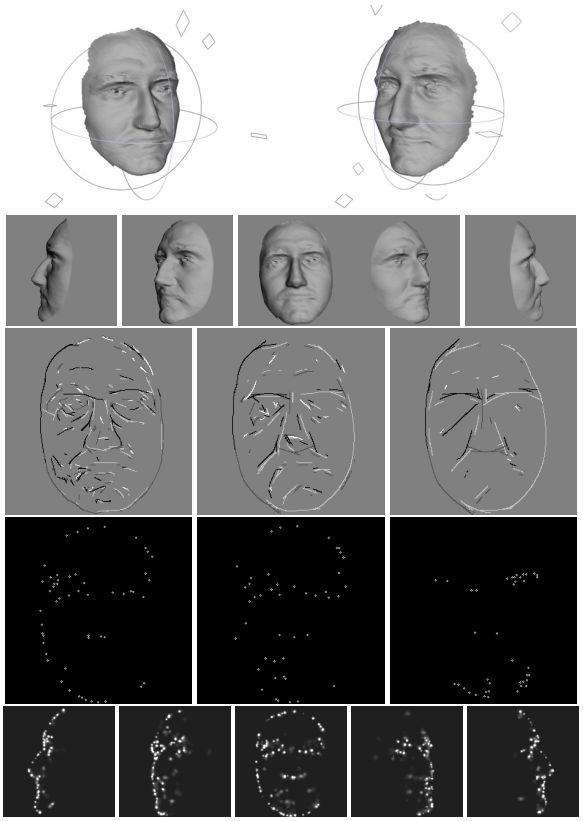

Figure 1: *Top row: two expressions of the same face with possible rotation intervals. Second row: profile, lateral and frontal 2D views of the neutral face (y-rotated). Third row: multiscale line and edge coding at three scales* $\lambda = \{4, 12, 24\}$ *of the frontal face view on the 2nd row. Fourth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

4. Face recognition framework

Humans detect a face in a wide range of conditions, such as poor lighting and/or 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 geometric relationships [RdB06b]. There are several methods based on skin colour for face detection [KMB07]. Agbinya and Silva [AS05], 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 [TFB00], 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. 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.

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 Fig. 1 and the top three rows in Fig. 3. These are used as templates stored in memory: frontal, lateral ($\pm 45^o$ y-rotated) and profile ($\pm 90^o$ y-rotated); see also [VAE97]. 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 $\pm 10^o$ (tilt); and d) a maximum z-rotation of $\pm 2^o$. Images with the same rotation angles as the templates were excluded.

For each face, the templates stored in PF cortex are: the LEs 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 LEs). The last are used in conjunction with other processing schemes for dynamic routing to achieve normalisation of the pair to be matched ("face"/template) [RdB09a]. Figure 2 shows on the first 2 rows part of the templates stored in memory, in the case of the frontal view shown in Fig. 1: left to right, the multiscale KPs and LEs 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 second row marked by a red quare, the multiscale KPs and LEs at the same scales as in the 1st row. The bottom row shows the summed KPs map with the accumulated keypoints (see below) 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 rotated (z-rotated) faces, size normalisation can be achieved by dynamic routing, see [RdB09b]. 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 $\Delta\lambda = 1$.

(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 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: $(x,y)_{AKP} = (AKP_y/\widetilde{mKP}, AKP_y/\widetilde{mKP})$ where $\widetilde{mKP} = \sum_{DF} mKP$. The AKP position is marked in red in Fig. 2 (bottom-left).

From the mKP map we compute the minimum and maximum coordinates in x and y, denoted by $CKP_{min/max,x/y}$.

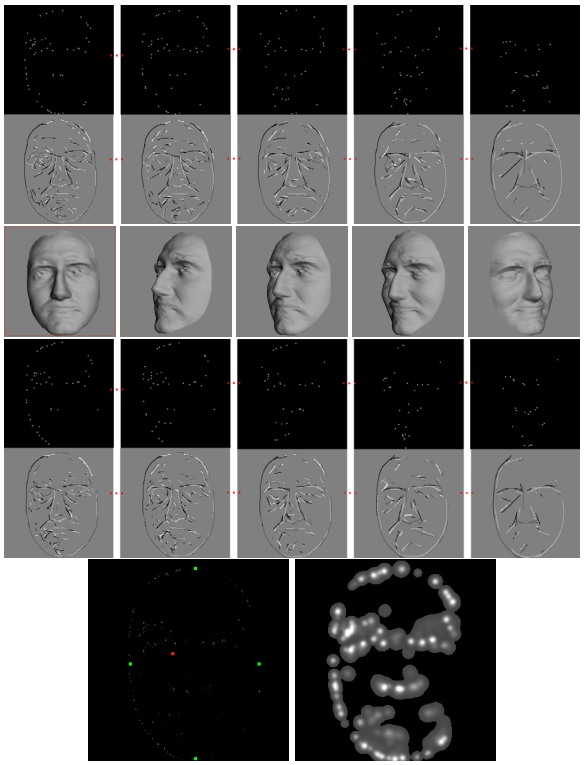

Figure 2: *Top two rows, left to right: fine to coarse scales of the KPs and LEs maps stored in memory in the case of the frontal view shown in Fig. 1. Third row: examples of faces to be tested. Fourth and fifth rows: KPs and LEs of the face marked by a red square in the 3rd row. Bottom row: the AKP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

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 Fig. 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 $AR =]0.61,1]$, frontal-lateral if $AR =]0.50,0.61]$, lateral-frontal if $AR =]0.40,0.50]$, lateral-profile if $AR =]0.33,0.40]$, profile-lateral if $AR =]0.31,0.33]$ and profile if $AR =]0,0.31]$. 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 occipito-parietal area which exhibits object-selective responses and many 3D cues of shape, and can relay the information to cue-invariant and view-invariant representations in the ventral what stream [Far09].

(D) Construction of saliency and symbolic representation maps: For each input image we build the saliency map as explained in Section 3, and the symbolic representation maps where events (positions) are expanded by Gaussian cross-profiles (lines) and bipolar, Gaussian-truncated errorfunction profiles (edges). The sizes of these are coupled to the scale of the underlying simple and complex cells; see [RdB06a] 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 more 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 Tab. 1; the values of *A* and *B* etc. are shown in Tab. 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 Fig. 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 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. [Ham05]. 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.

5. RESULTS AND DISCUSSION

For testing the model we used 100 images, 10 views per person, with different expressions (including the neutral) and also with different degrees of rotation: *x* axis $[-10^o, +10^o]$, *y* axis $[-90^o, +90^o]$ and *z* axis $[-2^o, +2^o]$. Figure 3 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.

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 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.*

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 the input view (see Tab. 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 Tab. 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 [MS04], who created GavabDB, developed a feature-based model and reported a recognition rate of 78.0%. Celenk and Aljarrah [CA06] 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 in [LJZ09], but now with a recognition rate of 94.7%. Li et al. [LJZ09] also reported results from 4 more authors who used the same database, the results ranging from 83.0% to 91.0%. Rashad et al. [RHSE09] 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

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%

Table 2: *Results using different weights for two views and using a single view; see text.*

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.

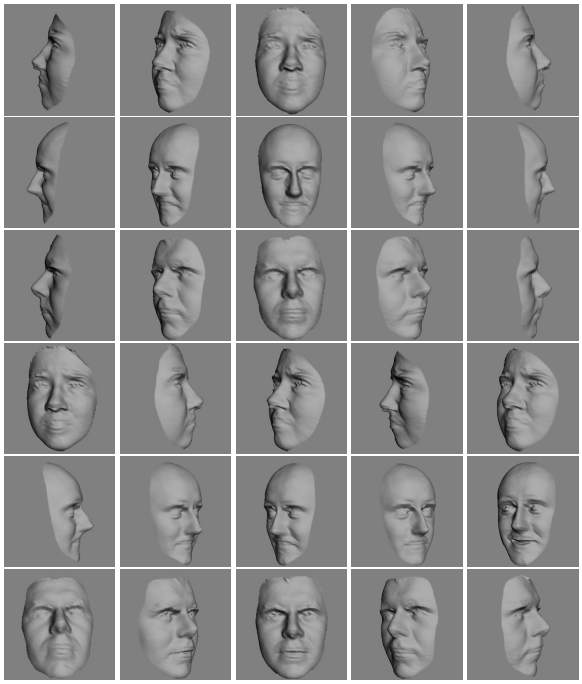

Figure 3: *Examples of other templates (top 3 rows) and tested faces (bottom 3 rows).*

6. Conclusions

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 model presented 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 [RdB06b]. Keypoints can be used to determine facial landmarks (eye, nose, mouth), which was already implemented and tested for frontal views [RdB06b]. In this paper we also showed that keypoints can also 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 [SRdB10]. 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.

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.

Acknowledgements:

Research supported by the Portuguese Foundation for Science and Technology (FCT), through the pluri-annual funding of the Inst. for Systems and Robotics (ISR/IST) through the POS_Conhecimento Program which (includes FEDER funds), and by the FCT project SmartVision: active vision for the blind (PTDC/EIA/73633/2006).

References

[ARS07] ABATE A. N. M., RICCIO D., SABATINO G.: 2D and 3D face recognition: A survey. *Pattern Recognition Letters 28*, 14 (2007), 1885–1906.

[AS05] AGBINYA J., SILVA S.: Face recognition programming on mobile handsets. *Proc. 12th Int. Conf. on Telecommunications, Cape Town, South Africa* (2005), 3–6.

[Bar04] BAR M.: Visual objects in context. *Nature Reviews: Neuroscience 5* (2004), 619–629.

[BCF06] BOWYER K., CHANG K., FLYNN P.: A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition. *Computer Vision and Image Understanding 101*, 1 (2006), 1–15.

[BDBP10] BERRETTI S., DEL BIMBO A., PALA P.: 3D face recognition using iso-geodesic stripes. *IEEE PAMI 32*, 12 (2010), 2162–2177.

[Ber03] BERSON D.: Strange vision: ganglion cells as circadian photoreceptors. *TRENDS in Neurosciences 26*, 6 (2003), 314–320.

[CA06] CELENK M., ALJARRAH I.: Interal shape-deformation invariant 3D surface matching using 2D principal component analysis. *Proc. of SPIE-IS&T Electronic Imaging 6056* (2006), 118–129.

[DR04] DECO G., ROLLS E.: A neurodynamical cortical model of visual attention and invariant object recognition. *Vision Res. 44*, 6 (2004), 621–642.

[F. 92] F. HEITGER, ET AL.: Simulation of neural contour mechanisms: from simple to end-stopped cells. *Vision Res. 32*, 5 (1992), 963–981.

[Far09] FARIVAR R.: Dorsal-ventral integration in object recognition. *Brain Research Reviews 61*, 2 (2009), 144–153.

[FRdB11] FARRAJOTA M., RODRIGUES J., DU BUF J.: Optical flow by multi-scale annotated keypoints: A biological approach. *Proc. Int. Conf. on Bio-inspired Systems and Signal Processing (BIOSIGNALS2011), 26-29 Jan., Rome, Italy* (2011), 307–315.

[Ham05] HAMKER F.: 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* (2005), 431–447.

[HHM02] HAXBY J., HOFFMAN E., M. G.: Human neural systems for face recognition and social communication. *Biol. Psychiatry 51*, 1 (2002), 59–67.

[Hub95] HUBEL D.: *Eye, brain and vision*. Scientific American Library, 1995.

[KMB07] KAKUMANU P., MAKROGIANNIS S., BOURBAKIS N.: A survey of skin-color modeling and detection methods. *Pattern Recognition 40*, 3 (2007), 1106–1122.

[LJZ09] LI X., JIA T., ZHANG H.: Expression-insensitive 3D face recognition using sparse representation. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* (2009), 2575–2582.

[MRdB09] MARTINS J., RODRIGUES J., DU BUF J.: Focus of attention and region segregation by low-level geometry. *Proc. Int. Conf. on Computer Vision Theory and Applications, Lisbon, Portugal, Feb. 5-8 2* (2009), 267–272.

[MS04] MORENO A., SANCHEZ A.: Gavabdb: A 3D face database. *Proc. 2nd COST275 Workshop on Biometrics on the Internet, Vigo, Spain* (2004).

[OT06] OLIVA A., TORRALBA A.: Building the gist of a scene: the role of global image features in recognition. *Progress in Brain Res.: Visual Perception 155* (2006), 23–26.

[PLN02] PARKHURST D., LAW K., NIEBUR E.: Modelling the role of salience in the allocation of overt visual attention. *Vision Res. 42*, 1 (2002), 107–123.

[RdB04] RODRIGUES J., DU BUF J.: Visual cortex frontend: integrating lines, edges, keypoints and disparity. *Proc. Int. Conf. Image Anal. Recogn., Springer LNCS Vol. 3211* (2004), 664–671.

[RdB06a] RODRIGUES J., DU BUF J.: Face recognition by cortical multi-scale line and edge representations. *Proc. Int. Conf. Image Anal. Recogn., Póvoa do Varzim (Portugal), Springer LNCS Vol. 3211* (2006), 329–340.

[RdB06b] RODRIGUES J., DU BUF J.: Multi-scale keypoints in V1 and beyond: object segregation, scale selection, saliency maps and face detection. *BioSystems 2* (2006), 75–90.

[RdB09a] RODRIGUES J., DU BUF J.: A cortical framework for invariant object categorization and recognition. *Cognitive Processing 10*, 3 (2009), 243–261.

[RdB09b] RODRIGUES J., DU BUF J.: Multi-scale lines and edges in V1 and beyond: brightness, object categorization and recognition, and consciousness. *BioSystems 95* (2009), 206–226.

[RdB11] RODRIGUES J., DU BUF J.: A cortical framework for scene categorization. *Proc. Int. Conf. on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP 2011), Vilamoura, Portugal* (2011), 364–371.

[RHSE09] RASHAD A., HAMDY A., SALEH M., ELADAWY M.: 3D face recognition using 2DPCA. *Int. J. of Computer Science and Network Security 9*, 12 (2009), 149–155.

[RVHB09] RAMIREZ-VALDEZ L., HASIMOTO-BELTRAN R.: 3D-Facial expression synthesis and its application to face recognition systems. *J. of Applied Research and Technology 7* (2009), 323–339.

[SCVC10] SMEETS D., CLAES P., VANDERMEULEN D., CLEMENT J.: Objective 3D face recognition: Evolution, approaches and challenges. *Forensic Science International 2010*, 1–3 (2010), 125–132.

[SRdB10] SOUSA R., RODRIGUES J., DU BUF J.: Recognition of facial expressions by cortical multi-scale line and edge coding. *Proc. Int. Conf. on Image Analysis and Recognition (ICIAR2010), Povoa do Varzim, Portugal, 21-23 June 1* (2010), 415–424.

[TFB00] TAILOR D., FINKEL L., BUCHSBAUM G.: Color-opponent receptive fields derived from independent component analysis of natural images. *Vision Research 40*, 19 (2000), 2671–2676.

[VAE97] VALENTIN D., ABDI H., EDELMAN B.: What represents a face? A computational approach for the integration of physiological and psychological data. *Perception 26*, 10 (1997), 1271–1288.