Face recognition by cortical multi-scale line and edge representations

João Rodrigues[1] and J.M.Hans du Buf [2]

[1] Escola Superior de Tecnologia, Campus da Penha
[2] Vision Laboratory, Campus de Gambelas – FCT,
University of Algarve, 8000-117 Faro, Portugal

Abstract. Empirical studies concerning face recognition suggest that faces may be stored in memory by a few canonical representations. Models of visual perception are based on image representations in cortical area V1 and beyond, which contain many cell layers for feature extraction. Simple, complex and end-stopped cells provide input for line, edge and keypoint detection. Detected events provide a rich, multi-scale object representation, and this representation can be stored in memory in order to identify objects. In this paper, the above context is applied to face recognition. The multi-scale line/edge representation is explored in conjunction with keypoint-based saliency maps for Focus-of-Attention. Recognition rates of up to 96% were achieved by combining frontal and 3/4 views, and recognition was quite robust against partial occlusions.

1 Introduction

Currently, one of the most investigated topics of image analysis is face detection and recognition [23, 24]. There are several reasons for this trend, such as the wide range of commercial vigilance and law-enforcement applications. Although state-of-the-art recognition systems have reached a certain level of maturity, their accuracy is still limited when imposed conditions are not perfect: all possible combinations of illumination changes, pose (frontal vs. profile), beards, moustaches, glasses, caps, different facial expressions and partial occlusions are problematic. The robustness of commercial systems is still far away from that of the human visual system. For this reason, the development of models of visual perception and their application to real-world problems like face recognition is important and, eventually, could lead to a breakthrough.

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 the neighbouring left-right hypercolumns [12]. These cells provide input for grouping cells that code line and edge information and that attribute depth information. In V1 there also are end-stopped cells that, together with complicated inhibition processes, allow to extract keypoints (singularities, vertices and points of high curvature). Recently, models of simple, complex and end-stopped cells have been developed, e.g. [8], providing input for keypoint detection [8, 20] and line/edge detection [10, 22], including disparity extraction

[9, 20]. On the basis of these models and neural processing schemes, it is now possible to create a cortical architecture for figure-ground segregation [22, 25] and Focus-of-Attention (FoA) [4, 21], including object categorisation [18, 22] and face/object recognition [11, 13, 17, 19].

In this paper, we will focus on a cortical model for face recognition. This model employs the multi-scale line/edge representation based on simple and complex cells in area V1. We will also employ the multi-scale keypoint representation in V1, in order to study the possible importance of saliency maps for Focus-of-Attention (FoA). The rest of this paper is organised as follows: In Section 2 the line and edge representation is explained, and in Section 3 the extraction of keypoints and construction of saliency maps. Section 4 explains the face recognition model. Experimental results are reported in Section 5, and we conclude with a discussion in Section 6.

2 Line/edge representation and image reconstruction

In order to explain the face recognition model, it is necessary to illustrate how our visual system can reconstruct, more or less, the input image. Image reconstruction can be based on one lowpass filter plus a complete set of bandpass wavelet filters, such that the frequency domain is evenly covered. This concept is the basis of many image coding schemes; it could also be used in the visual cortex because simple cells in V1 are often modelled by complex Gabor wavelets. These are bandpass filters [8], and lowpass information is available through special retinal ganglion cells with photoreceptive dendrites [2]. Activities of all cells could be combined by summing them in one cell layer that would provide a reconstruction or brightness map. But then there is a big problem: it is necessary to create *yet another observer* of this map in our brain.

The solution is simple: instead of summing all cell activities, we can assume that the visual system extracts lines and edges from simple- and complex-cell responses, which is necessary for object recognition, and that responding "line cells" and "edge cells" are interpreted symbolically. For example, responding line cells along a bar signal that there is a line with a certain position, orientation, amplitude and scale, the latter being interpreted by a Gaussian cross-profile which is coupled to the scale of the underlying simple and complex cells. The same way a responding edge cell is interpreted, but with a bipolar, Gaussian-truncated, error-function profile.

Responses of even and odd simple cells, corresponding to the real and imaginary parts of a Gabor filter, are denoted by $R^E_{s,i}(x, y)$ and $R^O_{s,i}(x, y)$, i being the orientation (we use 8 orientations). The scale s will be given by λ, the wavelength of the Gabor filters, in pixels. Responses of complex cells are modelled by the modulus $C_{s,i}(x, y)$.

The basic scheme for line and edge detection 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 4 possibilities for positive and negative events.

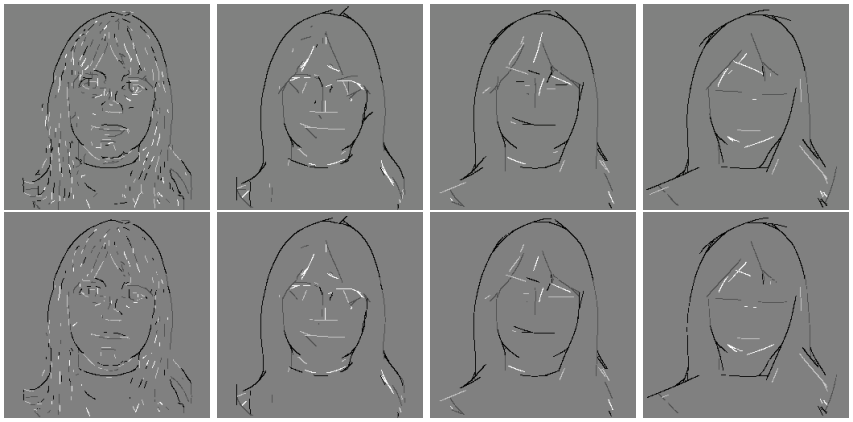

Fig. 1. Top: multi-scale line/edge coding of Kirsty, from fine (left) to coarse (right) scale. Bottom: results with multi-scale stabilisation.

The improved scheme [22] 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 is 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. In the case of multi-scale detection, all processes are linearly scaled with λ, the "size" of the simple and complex cells.

Figure 1 (top row) shows the multi-scale line and edge coding at four scales $\lambda = \{4, 12, 20, 28\}$; the same scales will be used below for the purpose of illustration. Different levels of grey, from white to black, are used to show detected events: positive/negative lines and positive/negative edges, respectively (see Fig. 2, top-right, for the input image). The bottom row in Fig. 1 shows detected events after applying a multi-scale stability criterion; see Section 4. Stabilisation leads to the elimination of events which are not stable over neighbouring scales, and therefore to less but more reliable events. As can be seen in Fig. 1, at fine scales many small events have been detected, whereas at coarser scales more global structures remain that convey a "sketchy" impression. Similar representations can be obtained by other multi-scale approaches [14].

One brightness model [5] is based on the symbolic line/edge interpretation referred to above. It is one of the very few models that can explain Mach bands [6, 16]. This model was tested in 1D and is now being extended to 2D. Here, we will not go into more detail; we will only illustrate the symbolic interpretation/reconstruction process in 2D that will be exploited in face recognition. The left part of Fig. 2 shows, top to bottom, symbolic edge and line representations at fine (left) and coarse (right) scales. The rightmost column illustrates visual reconstruction of the Kirsty image; from top to bottom: input image, lowpass-filtered image, the summation of symbolic line/edge representations (the sum of all images in the left part), and the final reconstruction. The use of more

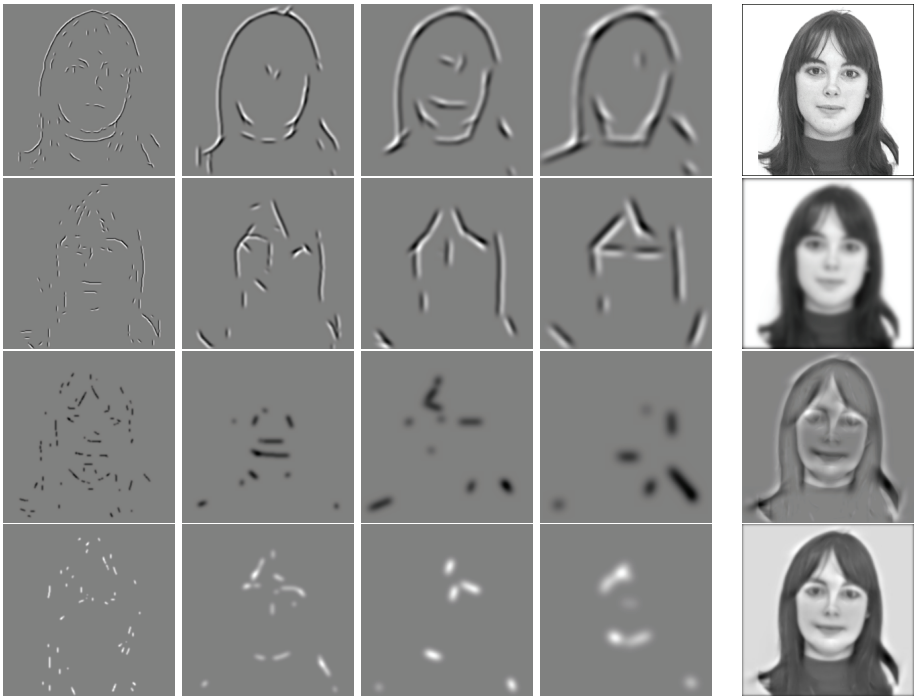

Fig. 2. Left part: multi-scale line/edge representation. Rightmost column: reconstruction of the Kirsty image (see text).

scales leads to better reconstructions, but the relative weighting of the lowpass and scale components is still under investigation. Summarising, the multi-scale line/edge representation allows to reconstruct the input image, and this representation will be used in face recognition.

3 Keypoints and saliency maps for FoA

Another important part of the model is based on responses of end-stopped cells in V1, which are very fuzzy and require optimised inhibition processes in order to detect keypoints at singularities. Recently, the original, single-scale model [8] has been further stabilised and extended to arbitrary scale, and the multi-scale keypoint representation has been used to detect facial landmarks and faces [21]. If we assume that detected keypoints are summed over all scales, which is a retinotopic (neighbourhood-preserving) projection by grouping cells, a saliency map can be created [21]: keypoints which are stable over many scales will result in large and distinct peaks. In other words, since keypoints are related to local image complexity, such a saliency map codes local complexity. In addition, different saliency maps can be created at different scale intervals, from fine to coarse scales, indicating interesting points at those scales with associated Regions-of-

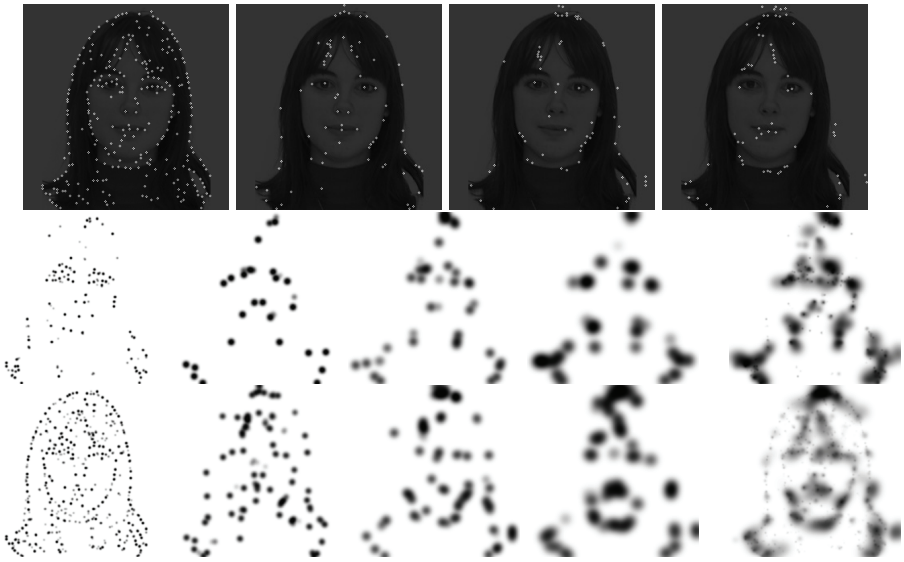

Fig. 3. Top row: keypoints detected at the four scales. Middle row: four partial saliency maps and the global map using $g = 1$. Bottom row: the same using $g = 0.25$. For explanation see text.

Interest (RoIs). Such information is very important in steering our eyes, because fixation points in complex regions (eyes, nose, mouth) are much more important than those in more homogeneous regions (forehead, cheeks).

There are two types of end-stopped cells: single and double. Responses of these are denoted by $S_{s,i}(x, y)$ and $D_{s,i}(x, y)$, which correspond to the first and second derivatives of the responses of complex cells $C_{s,i}(x, y)$. A final keypoint map K at scale s is obtained by $K_s(x, y) = \max\{\sum_{i=0}^{N_\theta - 1} S_{s,i}(x, y) - gI(x, y), \sum_{i=0}^{N_\theta - 1} D_{s,i}(x, y) - gI(x, y)\}$, in which I corresponds to the summation of tangential and radial inhibition and normally $g \approx 1.0$; see [21]. Figure 3 (top row) shows detected keypoints at fine (left) and coarse (right) scales.

Regions surrounding the peaks can be created by assuming that each keypoint has a certain RoI, the size of which is coupled to the scale (size) of the underlying simple and complex cells [21]. A global saliency map obtained by summing over all scales codes image complexity at all scales. Likewise, partial saliency maps can be constructed that code complexity at specific scale intervals. Figure 3 (middle row) shows, left to right, four partial saliency maps from fine to coarse scales, obtained by assuming 8 neighbouring scales at the centre scale (top row), plus the global saliency map, for $g = 1.0$. The bottom row shows the same in the case that $g = 0.25$. Summarising, less tangential and radial inhibition leads to more "complete" saliency maps, and in the global map we can "see" the structure of the input image (Fig. 3 bottom-right), in particular the regions around the eyes, nose, mouth etc. Actually, these regions correspond to the regions that contain many fixation points as measured by tracking the eyes of a person who is looking

at a face [15]. Below, in face recognition, these regions will be used to "gate" detected lines and edges, and we will experiment with different options.

4 Face recognition model

As was explained above, the multi-scale line/edge representation will be exploited, because this characterises facial features, and saliency maps will be used for Focus-of-Attention, i.e., to "gate" detected lines and edges in associated Regions-of-Interest. This resembles the bottom-up data streams in the where (FoA) and what (lines/edges) subsystems. However, it is 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. [11]. The same way, top-down data streams are simplified by assuming that stored face templates in memory are limited to lines and edges, and that a few canonical views[1] (frontal, 3/4) are normalised in terms of position, size and rotation (faces are expected to be vertical; for translation, size and rotation invariance see [4]).

In our experiments we use 8 primary scales $\lambda_1 = \{4, 8, 12, 16, 20, 24, 28, 32\}$ with $\Delta\lambda_1 = 4$. Each primary scale is supplemented by 8 secondary scales with $\Delta\lambda_2 = 0.125$, such that, for example, $\lambda_{2,\lambda_1=4} = \{4.125, 4.250, ..., 5.000\}$. These secondary scales are used for stabilisation and the construction of partial saliency maps. The model consists of the following steps:

(A) Multi-scale line/edge detection and stabilisation. To select the most relevant facial features, detected events must be stable over at least 5 scales in a group of 9 (1 primary plus the 8 secondary); see Fig. 1 bottom row.

(B) Construction of four symbolic representation maps. At each primary scale, stable events (positions) are expanded by Gaussian cross-profiles (lines) and bipolar, Gaussian-truncated error-function profiles (edges), the sizes of which being coupled to the scale of the underlying simple and complex cells; see Fig. 2. Responses of complex cells are used to determine the amplitudes of the profiles. As a result, each face image is represented by 4 maps at each of the 8 primary scales.

(C) Construction of saliency maps. Two types of maps are created: (1) one global saliency map (GSM), combining all keypoints at all 72 scales, and (2) eight partial saliency maps (PSM) at the primary and its secondary scales; see Fig. 3. The GSM can be used to gate all representation maps (point B above) at all scales, whereas PSMs are used to gate the maps at the same primary scales.

(D) Recognition process. We assume that templates (views) of faces are stored in memory and that these have been built through experience. Template images of all persons are randomly selected from all available images: either one *frontal* view or two views, i.e. one frontal plus one 3/4 view; see also [3]. Each template in memory is represented by 32 line/edge maps (point B above). Two recognition schemes will be tested, in which representations of input images (database) are compared with those in memory (templates):

[1] We will not go into the view-based vs. mental-rotation discussion.

Scheme 1: At each scale, events in the 4 representation maps of the input image are compared with those in the corresponding maps of a template. Co-occurrences are summed by grouping cells, which is a sort of event-type- and scale-specific correlation. Then, the outputs of the 4 event-type grouping cells are summed by another grouping cell (correlation over all event types). This results in 8 correlation factors. These factors are compared, scale by scale, over all templates in memory, and the template with the maximum number of co-occurrences over the 8 scales will be selected (in the case of equal co-occurrences we simply select the second template).

Scheme 2: Instead of comparing representations scale by scale, the global co-occurrence is determined by more levels of grouping cells, i.e., first over maps of specific event types, then over event types and finally over scales. The template with the maximum is selected by non-maximum suppression.

The above schemes are simplifications, because in real vision the system starts with a first categorisation, for example on the basis of the colour of the hair. After having a first gist (a group of possible templates), the system will dynamically optimise the recognition process by changing parameters. In view of the tremendous amount of data already involved in our simple experiments, this cannot (yet) be simulated.[2] We only simulate event selections by employing global and partial saliency maps obtained with different inhibition constants g in keypoint detection: the higher g, the more precise detection will be and, consequently, less line/edge events will be available for categorisation and/or recognition. The reason for this choice is the fact that coarse-scale information from area V1 propagates to IT (inferior-temporal) cortex first, for a very coarse first categorisation, after which information at increasingly finer scales arrive at IT [1]. This means that the system starts with coarse line/edge representations and partial saliency maps at those scales, to be simulated with $g = 1.0$, then refines the search with partial maps with $g = 0.25$, and can finish recognition with a fine-scale and/or global saliency map, also with $g = 0.25$. Figure 4 illustrates information available for recognition.

5 Experimental results

From the Psychological Image Collection at Stirling University (UK), we selected 100 face images of 26 persons in frontal or frontal-to-3/4 view, with different facial expressions. From those, 13 persons are seen against a dark background, with a total of 53 images, of which 40 images are in frontal view, 11 images are in (very near) 3/4 view (4 persons), and 2 frontal images with added Gaussian and speckle noise (one person). The other 13 persons (47 images) are seen against a light background, in frontal or near-frontal view. For typical examples see Fig. 6. All persons are represented with at least 3 different facial expressions.

All recognition tests involved the entire set of 100 images, although results will also be specified in terms of the subsets of 53 and 47 images in order to

[2] Hundred images of 256×256 pixels, with 72 scales and at each scale 4 event and 9 saliency maps, require more than 24 GBytes of storage.

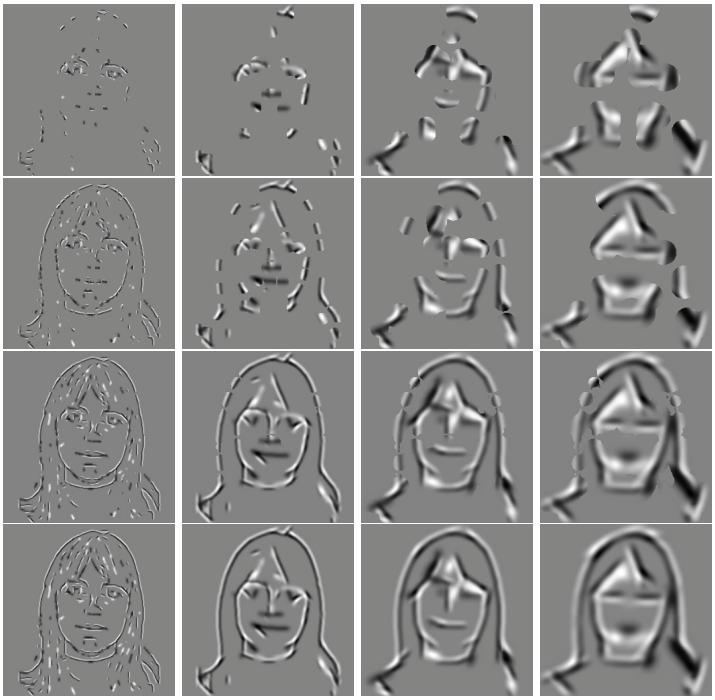

Fig. 4. Combined event representations gated by different saliency maps, from top to bottom: PSM with $g = 1.0$, PSM with $g = 0.25$, GSM with $g = 0.25$ and all detected events.

analyse the influence of the different backgrounds. For each person we used two different types of templates: (1) only one frontal view, and (2) two views, frontal and 3/4, but only in the case of 4 persons with images in frontal and 3/4 views. In all cases, template images were selected randomly. In order to study robustness with respect to occlusions, a second set of tests was conducted in which partially occluded representations of input images were matched against complete representations of templates.

Table 1 presents the results obtained by using partial saliency maps ("PSM") with $g = 1.0$ and 0.25, global saliency maps ("GSM") with $g = 0.25$, and all detected events, i.e. without applying saliency maps. We present the results by testing all images ("all"), and specify (split) these in the case of a dark ("black)" or light ("white") background. The penultimate column ("scales") lists the percentage of correct scales that lead to correct recognition in the case of "all" and scheme 1, where 100% corresponds to 800 because of 8 scales and 100 images. The last column ("base line") lists the number of all 100 images that have been recognised with absolute certainty, i.e. when scheme 1 and 2 and all scales point at the same person.

Comparing columns "all," "black" and "white," there are significant differences because dark and blond hair against dark and light backgrounds cause

templates	only frontal view							
recogn. scheme	2	2	2	1	1	1	1	base
images	all	black	white	all	black	white	scales	line
PSM $g = 1.0$	86.0	83.0	89.4	87.0	83.0	91.5	82.8	61
PSM $g = 0.25$	89.0	86.8	91.5	88.0	86.8	89.4	83.0	63
GSM $g = 0.25$	90.0	90.6	89.4	89.0	90.6	89.4	85.9	70
all events	91.0	90.6	91.5	89.0	86.8	91.5	85.5	71
templates	frontal plus 3/4 view							
PSM $g = 1.0$	94.0	98.1	89.4	94.0	96.2	91.5	88.6	67
PSM $g = 0.25$	95.0	98.1	91.5	93.0	96.2	89.4	89.1	69
GSM $g = 0.25$	95.0	100.0	89.4	95.0	100.0	89.4	92.5	79
all events	96.0	100.0	91.5	96.0	100.0	91.5	91.8	81

Table 1. Results obtained without occlusions.

different events, or even no events, at the hairline. Although the "all" results are reasonably close to the best results, separation of different backgrounds can lead to better but also worse results. This aspect certainly requires more research. Using no saliency maps, i.e. all detected events, yields best results, which was expected. Also expected was the increasing rates in the four lines, because the use of different saliency maps implies more or less information available for recognition, see Fig. 4.

Best results were obtained when using two templates with frontal and 3/4 views. Using all events, both recognition schemes resulted in 96%, whereas 81 was the base line with absolute certainty. The difference of 15% is due to relative ranking with some uncertainty. In future research it will make sense to increase the base line, especially when larger databases with more variations are considered. The increasing base line (67, 69, 79, 81) implies that a system simulating dynamic processing may have an easier task: after the first step, 67 of 100 images have been identified and the remaining 33 must be scrutinised in step 2.

Our best result of 96.0% is a little bit better than the 94.0% obtained by Petkov et al. [17] and 93.8% by Hotta et al. [13], and very close to the 96.3% reported by Ekenel and Sankur [7], despite the fact that in all studies the number of tested faces and the databases were different.

Partial occlusions: We tested the 5 occlusions as shown in Fig. 5, using all 100 images and applying recognition scheme 2 with templates that combine frontal and 3/4 views. Because of the tremendous amount of storage space and CPU time, representations were not re-computed (500 images!) but the occlusions were applied to the already computed representations, thereby suppressing event information. This is an approximation of real occlusions, but it indicates the relative importance of different facial regions in the recognition scheme. Table 2 presents results in terms of "rate (base line)," which must be compared with the bottom part of Table 1, i.e. the first and last columns.

In the case of PSM with $g = 0.25$, the base line of 69 (Tab. 1) drops to 54 with the most severe occlusion 4 (eyes, nose and mouth). In the "all events" case, instead of 81 only 64 was obtained. But this is the base line: still 54 or 64

Fig. 5. Occlusions 1 to 5 from left to right.

	frontal plus 3/4 views; recogn. scheme 2				
occlusion type	1	2	3	4	5
PSM $g = 0.25$	95.0 (68)	96.0 (66)	92.0 (62)	83.0 (54)	93.0 (66)
all events	96.0 (80)	95.0 (74)	96.0 (67)	93.0 (64)	97.0 (75)

Table 2. Results obtained with partial occlusions.

of all 100 images are classified with absolute certainty. The maximum rate for this occlusion (all events, 93%) is very close to the maximum without occlusion (Tab. 1, 96%), and slightly worse if compared to the other occlusions. This shows that the multi-scale representation, in particular the shape of the head and hair at the coarser scales, is very robust and contributes most in the recognition. The reason for this can be seen in Fig. 1: the stable and "sketchy" information without too much detail at coarse scales.

6 Discussion

The line/edge representation at coarser scales provides a stable abstraction of facial features (Figs 1,2). This explains, at least partly, the generalisation that allows to classify faces with noise, glasses, relatively normal expressions and views (Fig. 6). The main problems were: (1) a change of hairstyle and extreme expression (Fig. 2 top-right with long hair was recognised, but not Fig. 6 bottom-right with short hair and big smile); and (2) insufficient image normalisation (Fig. 6: the bottom images in the 2nd, 3rd and 4th columns were problematic because of different pose; however, the last image on the 3rd row was recognised!). These four images (2nd to 5th at the bottom of Fig. 6) were the only ones which were not recognised; hence, the overall recognition rate of 96 in 100.

The problem of insufficient normalisation can be solved because faces can be detected by grouping keypoints at eyes, nose and mouth [21]. By using detected keypoints at the two eyes and mouth corners, images can be morphed such that the central part of a face is normalised in terms of size and position. This procedure can also guarantee that templates in memory are really representative. However, similar solutions for hairlines and non-frontal views must be developed; see also [3]. As for now, correct recognition in the case of a drastic change of hairstyle and expression remains a Holy Grail.

Despite the problems and possible solutions mentioned above, our results obtained with a completely new approach are very encouraging. We expect sig-

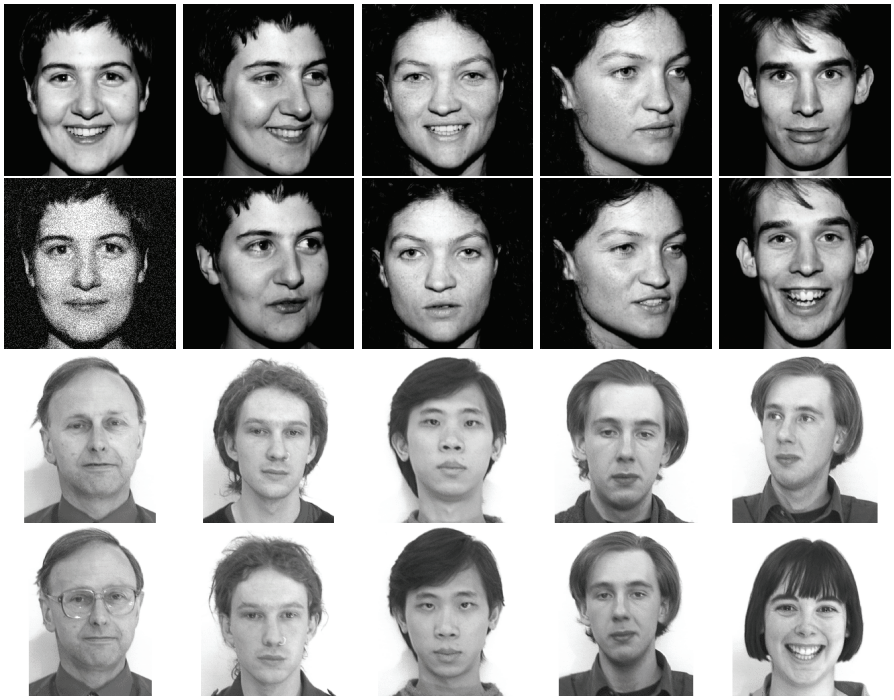

Fig. 6. Examples of images of eight persons.

nificant improvements by implementing a dynamic system, in which successive tests are performed with each time more complete information available, as illustrated in Fig. 4, such that all effort can be spent on scrutinising the images which have not yet been identified with absolute certainty. This procedure may simulate the processing in the bottom-up and top-down data streams in the what and where subsystems of our visual system. However, before departing on this track, logistic problems related to storage space and CPU time on a new, multi-processor computer system must be solved first.

Acknowledgements: This investigation is partly financed by PRODEP III Medida 5, Action 5.3, and by the FCT program POSI, framework QCA III. The Stirling images are available at http://pics.psych.stir.ac.uk/

References

1. M. Bar. A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, (15):600–609, 2003.
2. D. Berson. Strange vision: ganglion cells as circadian photoreceptors. *TRENDS in Neurosciences*, 26(6):314–320, 2003.
3. D. Valentin, et al. What represents a face? A computational approach for the integration of physiological and psychological data. *Perception*, 26(10):1271–1288, 1997.

4. G. Deco and E.T. Rolls. A neurodynamical cortical model of visual attention and invariant object recognition. *Vision Res.*, (44):621–642, 2004.
5. J.M.H du Buf. Ramp edges, Mach bands, and the functional significance of simple cell assembly. *Biol. Cybern.*, 70:449–461, 1994.
6. J.M.H du Buf and S. Fischer. Modeling brightness perception and syntactical image coding. *Optical Eng.*, 34(7):1900–1911, 1995.
7. H. Ekenel and B. Sankur. Multiresolution face recognition. *Image Vision Comp.*, 23(5):469–477, 2005.
8. F. Heitger, et al. Simulation of neural contour mechanisms: from simple to end-stopped cells. *Vision Res.*, 32(5):963–981, 1992.
9. D.J. Fleet, A.D. Jepson, and M.R.M. Jenkin. Phase-based disparity measurement. *CVGIP: Image Understanding*, 53(2):198–210, 1991.
10. C. Grigorescu, N. Petkov, and M.A. Westenberg. Contour detection based on nonclassical receptive field inhibition. *IEEE Tr. IP*, 12(7):729–739, 2003.
11. F. Hamker. The reentry hypothesis: The putative interaction of the frontal eye field, ventrolateral prefrontal cortex, and areas V4, IT for attention and eye movement. *Cerebrel Cortex*, 15:431–447, 2005.
12. D.H. Hubel. *Eye, brain and vision.* Scientific American Library, 1995.
13. K. Hotta, et al. Face matching through information theoretical attention points and its applications to face detection and classification. *Proc. 4th IEEE Int. Conf. Automatic Face and Gesture Recogn.*, 34-39, 2000.
14. T. Lindeberg. *Scale-Space Theory in Computer Vision.* Kluwer Academic Publishers, Dordrecht, Netherlands, 1994.
15. M. Pomplun, et al. Augenbewegungen als kognitionswissenschaftlicher forschungsgegenstand. In *Kognitionswissenschaft: Strukturen und Prozesse intelligenter Systeme.* Deutscher Universitätsverlag, 65-106, 1997.
16. L. Pessoa. Mach bands: how many models are possible? Recent experimental findings and modeling attemps. *Vision Res.*, 36:3205–3227, 1996.
17. N. Petkov, P. Kruizinga, and T. Lourens. Biologically motivated approach to face recognition. *Proc. Int. Worksh. Artif. Neural Networks*, 68-77, 1993.
18. M. Riesenhuber and T. Poggio. CBF: A new framework for object categorization in cortex. *Proc. 1st IEEE Int. Worksh. Biol. Motivated Comp. Vision*, 1-9, 2000.
19. M. Riesenhuber and T. Poggio. Models of object recognition. *Nature Neuroscience*, 3:1199–1204, 2000.
20. J. Rodrigues and J.M.H. du Buf. Visual cortex frontend: integrating lines, edges, keypoints and disparity. *Proc. Int. Conf. Image Anal. Recogn.*, Springer LNCS Vol. 3211:664–671, 2004.
21. J. Rodrigues and J.M.H. du Buf. Multi-scale keypoints in V1 and face detection. *Proc. 1st Int. Symp. Brain, Vision and Artif. Intell., Naples (Italy),*, Springer LNCS Vol. 3704:205–214, 2005.
22. J. Rodrigues and J.M.H. du Buf. Cortical object segregation and categorization by multi-scale line and edge coding. *Accepted for: Inter. Conf. Computer Vision Theory and Applications, 25-28 February, Setúbal, Portugal*, 2006.
23. M.H. Yang, D.J. Kriegman, and N. Ahuja. Detecting faces in images: A survey. *IEEE Tr. PAMI*, 24(1):34–58, 2002.
24. W. Zhao, R. Chellappa, A. Rosenfeld, and P. Phillips. Face recognition: A literature survey. *UMD CfAR Technical Report CAR-TR-948*, 2000.
25. L. Zhaoping. V1 mechanisms and some figure-ground and border effects. *J. Physiology*, Paris 97:503–515, 2003.