Face normalization using multi-scale cortical keypoints

João Cunha, João Rodrigues and J.M.H. du Buf
Vision Laboratory, Inst. for Systems and Robotics (ISR) – University of the Algarve, Faro, Portugal
{jrodrig, dubuf}@ualg.pt

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 extractions. Simple, complex and end-stopped cells tuned to different spatial frequencies (scales) and/or orientations provide input for line, edge and keypoint detection. This yields a rich, multi-scale object representation that can be stored in memory in order to identify objects. The multi-scale, keypoint-based saliency maps for Focus-of-Attention can be explored to obtain face detection and normalization, after which face recognition can be achieved using the line/edge representation. In this paper, we focus only on face normalization, showing that multi-scale keypoints can be used to construct canonical representations of faces in memory.

1. Introduction

Currently, one of the most investigated topics of image analysis is face detection and recognition [7, 8]. 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. The robustness of commercial systems is still far 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 important breakthroughs. In this paper we will only focus on a cortical model for face normalization, after which a cortical face recognition model can be applied [3].

2. Face normalization

Face normalization is done in two steps: **(i)** the detection of facial landmarks in a sequential order, eyes – nose – mouth, based on normal spatial relations (distances) [6], and on the principles proposed in [4] for quasi-normalized faces, i.e. detection of facial landmarks can be obtained by considering the most significant peaks of partial saliency maps (PSMs) which combine keypoints over scale intervals [4]. The top row of Fig. 1 shows from left to right a face plus keypoints detected at a fine and a coarse scale. The 2nd row shows PSMs at fine to coarse scales. In the following, all PSMs are tested independently except for the one at finest scales. The latter is used in combination with all other maps, but only for eye confirmation.

In the first step (for each PSM) the eyes are detected: **(a)** all existing PSM peaks are tested for representing one eye, but only if there also exist in the fine-scale PSM (Fig. 1, 2nd row, 1st image) two peaks in opposite directions in a small area related to the size of the dendritic fields (DFs) of the cells used to built the corresponding PSM. Figure 1, 3rd row, shows examples of peaks in the DF region for fine (left) and coarse (right) PSMs and possible eyes in between. **(b)** The next step is to search for a pair of eyes: two individual (possible) eyes have to obey criterion (a), the face distance criterion [6], and the line connecting the two eyes must form an angle $\leq 30^\circ$ to the horizontal axis (maximum allowed rotation). Yet another test verifies that the angles between this line and the ones that connect two lateral peaks at each eye candidate is $\leq 15^\circ$. In Fig. 1, 4th row, only the case labeled 8° obeys this criterion.

(c) For the nose, perpendicular to the eyes-line and obeying the normal face distance criterion [6] the PSM is checked for a peak inside a circle with a size related to the DF of the corresponding PSM (Fig. 1, 5th row, leftmost). Similarly, step **(d)** is applied for checking the presence of a mouth. This element has most variation as a function of facial expression. For this reason we apply a big rectangular area for checking two PSM peaks, one on each side of the mouth (Fig. 1, 5th row, right). This area has the following properties: parallel to the eyes-line; width equal to the distance between the eyes; height equal to half of the width; center of the rectangle at the normal distance between nose and mouth [6]. Two PSM peaks inside this area only correspond to a mouth if their distance is at least 1/5th of the

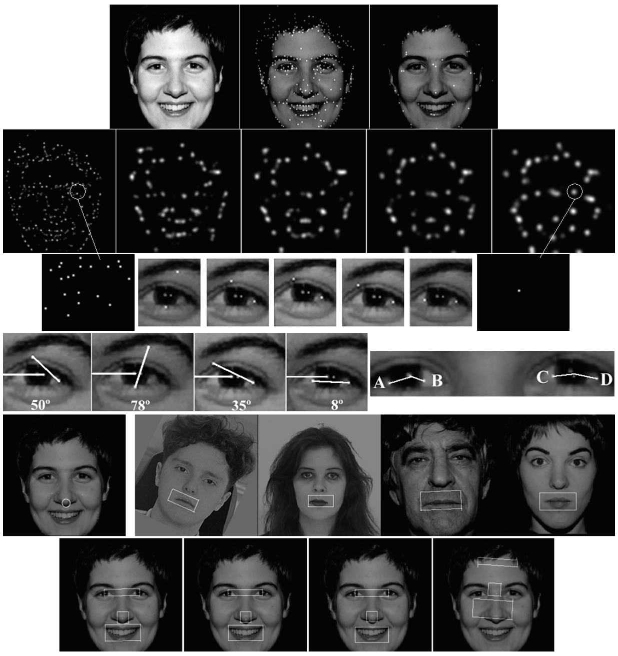

Figure 1. Eyes, nose and mouth detection.

rectangle's width and the angle between their connecting line and the orientation of the rectangle is at most $15°$. **(e)** The same processes are applied to all individual PSMs, detecting all possible face configurations. Figure 1, bottom row, shows examples of face configurations in four PSMs, at fine (left) to coarse (right) scales. As can be seen, there may be more face candidates that obey all criteria. Therefore, in the last step **(f)** only face candidates are selected with eyes, nose and mouth keypoints at the same positions in at least 3 PSMs, using small relaxation areas around the peaks with the size of the DF of the PSM at the finest scales.

(ii) After detection of all facial landmarks and entire face configurations, the positions of the eyes (corresponding PSM peaks) are mapped to predetermined (normalized) positions. The same mapping is applied to all cortical features (lines, edges, keypoints, etc.) for face identification (see [5] for a possible explanation of this cortical process).

3. Brief conclusions

Figure 2 shows input images on rows b and d and the corresponding normalized faces, obtained with bilinear interpolation, above on rows a and c. Of the 46 faces tested, 33 (71.7%) were correctly detected and normalized. The main problem proved to be correct detection of the eyes. This must be made much more robust, for instance by using a larger number of PSMs and/or combining keypoint information with multi-scale lines and edges. The latter is required in any case when dealing with 3/4 and lateral views of faces.

Bearing in mind that cortical invariant face and object

Figure 2. Results (see text).

detection and recognition are different, complex and at the same time complementary issues, which are too complicated to be explained in this very short paper, we refer to [1, 2, 3, 5] for relevant discussions.

Acknowledgements: This research is partly supported by the FCT (ISR/IST pluri-annual funding) through the POS-Conhecimento Program which includes FEDER funds.

References

[1] G. Deco and E. Rolls. Attention, short term memory, and action selection: a unifying theory. *Prog. in Neurobiol.*, 76:236–256, 2005.

[2] M. Riesenhuber. Object recognition in cortex: neural mechanisms and possible roles for attention. *Chapter in: Neurobiology of Attention, L. Itti, G. Rees, and J. Tsotsos (eds), Academic Press, Elsevier*, pages 279–287, 2005.

[3] J. Rodrigues and J. du Buf. Face recognition by cortical multiscale line and edge representations. *Proc. Int. Conf. Image Anal. Recogn., Póvoa do Varzim (Portugal)*, Springer LNCS Vol. 3211:329–340, 2006.

[4] J. Rodrigues and J. du Buf. Multi-scale keypoints in V1 and beyond: object segregation, scale selection, saliency maps and face detection. *BioSystems*, 86:75–90, 2006.

[5] J. Rodrigues and J. du Buf. Invariant multi-scale object categorisation and recognition. *Proc. 3rd Iberian Conf. on Patt. Recogn. and Image Anal., Girona (Spain)*, Springer LNCS 4477:459–466, 2007.

[6] E. Saber and A. Tekalp. Frontal-view face detection and facial feature extraction using color, shape and symmetry based cost functions. *Pattern Recogn. Letters*, 19(8):669–680, 1998.

[7] H. Yang, J. Kriegman, and N. Ahuja. Detecting faces in images: a survey. *IEEE Tr. PAMI*, 24(1):34–58, 2002.

[8] W. Zhao, et al. Face recognition: a literature survey. *ACM Computing Surveys*, 35(4):399–458, 2003.