Saccadic Search with Gabor features applied to Eye Detection

F. Smeraldi, A. Makarov and J. Bigun

Swiss Federal Institute of Technology (EPFL)
Microprocessor and Interface Laboratory
CH-1015 Lausanne

Abstract. We suggest an eye detection strategy based on an attention driven search mechanism inspired by human saccadic eye movements. Accordingly, the sampling strategy employed makes use of a retinotopical sampling grid obtained by log-polar mapping in the complex plane. At the grid points the local power spectrum is sampled instead of the conventional gray values. Experimental results are presented which show that robust left and right eye detection can be achieved after a comparatively low number of fixations.

1 Introduction

The human eye explores a visual scene by performing a sequence of large "jumps", known as saccades, between the different points of interest, on which fixation is maintained for a short while [3,6]. Our approach to feature detection consists in performing search based on a model of saccadic eye movement. Saccadic eye movements have been shown to play a central role not only in the exploration of a scene, but in the underlying cognitive processes proper, where there appears to be a selection mechanism for filtering task relevant information. Likewise, the algorithm we propose implements an attention-driven search strategy which exploits the retinal sampling of the image to minimise the number of sampled points required to detect the selected features. At each point of the grid, the Gabor decomposition of the image is sampled rather than the local intensity values. The particular choice of the eyes as a target for the search is motivated by their importance in the face recognition processes, which has repeatedly been assessed by both neurophysiological studies and computer vision results.

2 Gabor decomposition and retinal sampling

When using the Gabor decomposition, the frequency domain has to be partitioned in a way which would allow appropriate windowing. In this work, we suggest an equidistant partitioning applied to a conformal map of the frequency plane. Considering the frequency to be the complex quantity $\omega = \omega_x + j\omega_y =$

$\rho e^{j\varphi}$, we compute its complex logarithm and we partition the $\log \omega$ plane according to a uniform rectangular grid [2]. The windowing is performed by placing two dimensional Gaussians at the nodes of this grid. By inverse mapping, $\omega = \exp(\log \omega)$, octave frequency magnitude channels are obtained.

In the spatial domain, for each image pixel p, one obtains an array $\Gamma_p = (\gamma_{\rho\varphi})_p$ of values computed from the Gabor decomposition of the luminance. Each of these values is determined by the observed position p, frequency magnitude ρ and frequency orientation angle ϕ. We propose to sample such values according to a circular sampling grid composed of concentric rings, which can be created in the same way as the frequency and orientation partition in the frequency plane (Figure 1). The ratio of the radii of any two successive rings and the number of sampled points per ring are constant. This solution is inspired to the arrangement of receptors in the retina, which has already motivated some works in computer vision [7,5,1].

3 Eye localisation

In 1957, Yarbus analysed the stopping places of a subject's gaze exploring various images, among which the human faces [8,3]. The stopping places were most densely distributed in the high spatial frequency areas of explored images, such as eye regions of observed faces. This motivated us to develop an algorithm which would drive the retinal sampling grid to the eyes of a subject using saccadic eye motion. The eye detection procedure consists of three main steps. At first, local information driven saccadic eye movement is used to home on one of the eyes; following, the global information from all of the grid points is used to refine the search; finally, if detection is successful a saccade is performed to the assumed position of the other eye.

During each of the above step, several criteria are applied to check for the consistency of information. If a mismatch is detected, doubtful assignments are discarded.

3.1 Saccadic eye movements

At the beginning of the search, the retinal sampling grid is placed at a random position on the image and the corresponding set of Gabor features, $\mathcal{G}_0$, is extracted. Each vector in $\mathcal{G}_0$, after division by its Euclidean norm, is subsequently matched against a reference vector e_{av}. In order to construct the latter, the average Gabor responses from the centre of the right and left eye of six persons are computed. These two standard vector responses are then geometrically averaged component-wise so that e_{av} captures the features which are common to the right and the left eye. The point of the grid for which the Euclidean distance from e_{av} is minimal is selected as the target for the next saccade. The search is terminated when saccades become shorter than $1/6$ of the sampling grid's outer radius. If no saccade target whose distance from e_{av} is reasonably low can be found (which can be the case if the search starting point happens to fall in a blank region of the image), the search is restarted from a random position.

3.2 The eye model

The a priori knowledge about the appearance of the left and right eyes of the generic person is respectively encoded into a left eye model and a right eye model. The models are constructed from the sets $\mathcal{L} = \bigcup_p \Gamma_p$ and $\mathcal{R} = \bigcup_q \Gamma_q$ of Gabor features obtained by placing the retinal sampling grid on either of the eyes (Figure 1) and computing the Gabor responses Γ_p at each of its points.

Fig. 1. The retinal sampling grid placed on a person's right eye for model creation.

The features in $\mathcal{L}$ and $\mathcal{R}$ are then rearranged in a collection of matrices $\mathsf{M} = \{\mathsf{M}_{r\omega}\}_{r\omega}$ so that each one of the $\mathsf{M}_{r\omega}$ contains the responses for a fixed Gabor frequency radius ω and a given spatial circle with radius r of the sampling grid. The rows and columns of each $\mathsf{M}_{r\omega}$ therefore correspond to the variation of the angular coordinates in the spatial and frequency domains. Matrices $\mathsf{M}_{r\omega}$ are then normalised separately with respect to the norm defined by $|\mathsf{M}_{r\omega}| = \sqrt{\mathrm{Trace}(\mathsf{M}^t_{r\omega}\mathsf{M}_{r\omega})}$, which is equivalent to the Euclidean norm if $\mathsf{M}_{r\omega}$ is interpreted as a vector. All Gabor features from a single frequency channel ω belong to the same matrix $\mathsf{M}_{r\omega}$. Since each frequency channel is characterised by a specific bandwidth which is common to all the orientations, normalisation takes care of the variation of filter bandwidths across the frequency channels. Also, by grouping together all the points of the sampling grid circle of radius r in a single $\mathsf{M}_{r\omega}$ and then normalising, one makes sure that illumination changes are compensated for.

The eye model for the left eye, L, is computed by combining the collections of matrices M^i obtained by placing the grid manually on the left eye of six persons according to the relation

$$\mathsf{L} = \{\mathsf{L}_{r\omega}\}_{r\omega} = \left\{ \frac{\sum_i \mathsf{M}^i_{r\omega}}{|\sum_i \mathsf{M}^i_{r\omega}|} \right\}_{r\omega}$$

The same procedure is applied to obtain the eye model for the right eye.

Matching of the retinal grid samples I extracted from an image with the model is performed (e.g. in the case of a left eye) by minimising the value of the function $d(\mathsf{I}, \mathsf{L}) = \sum_{r\omega} |\mathsf{I}_{r\omega} - \mathsf{L}_{r\omega}|$.

3.3 Refining the search

After the saccadic phase of the search has converged to the target pattern, the Gabor responses in the points currently "viewed" by the grid are compared with both the left and the right eye models described in the preceding section. According to the model which obtains the best result, the candidate eye is assumed to be a left or a right eye. The appropriate model is then selected and the exact position of the local minimum is determined. If the resulting displacement is larger than a few pixels the saccadic search is restarted from a random position.

Experiments have shown that the saccadic search may detect some erroneous local minima (e.g. the corners of the mouth, ear-rings or details in the hair). In order to discriminate such fake targets, the difference is computed between the candidate's distance from the attributed eye model and its distance from the alternate model. The ratio of this difference to the minimum distance, which we call the *asymmetry*, measures the amount to which the chirality of the detected feature contributes to the match. In our experiments, the asymmetry always turned out to be grater than 0.1 for correct matches, while it generally dropped of one or two orders of magnitude in the case of spurious identifications. The errors thus detected are treated by restarting the search from a random position.

3.4 Looking for the other eye

After localisation of one eye, the system performs a saccade in the presumed direction of the other eye. Normal saccadic search is then performed until an eye is found. Due to scale differences between images, the initial saccade may not turn out to be long enough to prevent the system from finding again the same eye. In this case, further attempts are performed with an increasing starting distance from the known eye until the other eye is found. In case the search refinement detects a low asymmetry target, search is restarted with a random offset. If this condition persists for several attempts, it is assumed that the position of the first eye has been incorrectly assigned and eye detection is restarted from scratch. Although the assumption that faces are presented in an upright orientation is used to speed up the detection of the second eye, no strict constraint is imposed on its position relative to the first. Therefore, detection remains robust also in the case of subjects having their head tilted to one side.

3.5 Experimental results

The algorithm has been tested using a Gabor decomposition rosette consisting of six texture orientation sectors and five frequency magnitude octaves, ranging

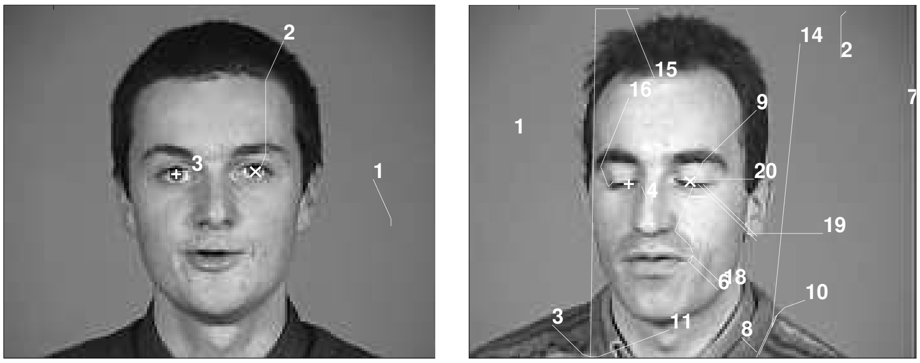

Fig. 2. The + and × signs denote the best match with the right and left eye models respectively. Numbers identify successive starting points for saccades. Eye detection for the left picture required 51 fixations. Note how saccadic search 1 was considered uninteresting and therefore discarded. A random restart (2) then lead to detection of the left eye, after which saccadic search resumed (3) near the location of the right eye. In the case of the right picture, information from the outline of the orbit allows eye detection even if the person's eyes are shut. During this trial the centre of the sampling grid explored 99 pixels and 14 targets were rejected after comparison with the eye models.

from $\frac{\pi}{16}$ to π. The retinal sampling grid employed had 5 rings and 16 rays, with the ring radii being distributed between $\rho_{\min} = 3$ and $\rho_{\max} = 30$ pixels.

Our image database consists of forty frontal shots of twenty different persons[1]. The image resolution is 143×175 pixels. Differences between the shots of the same persons consist in tan changes, haircut, makeup, eyelid position, head position (heads are often slightly rotated) and slight scale changes. Several persons in the database wear eyeglasses.

Single shots from six persons were used to extract the left and the right eye models. Repeated testing was then performed on the whole database without any mismatch being found. Information obtained from the outline of the orbit allows correct detection of the features even when the subject's eyes are closed (Figure 2). In our trials we found the median of the number of fixation points to be 49 for the detection of both eyes, that is to say that the centre of the retinal sampling grid explores 0.2% of the image pixels. The number of fixations is considerably increased (typically 100) for subjects wearing glasses with strong reflections or having their eyes shut. This is mainly due to the fact that since the algorithm knows nothing about facial features other than the eyes, no alternative cues can be used to infer their spatial position when their visibility is low. Nevertheless, detection is always correctly accomplished at the end.

[1] This image database is a part of an audiovisual database, developed in the framework of a European face recognition project M2VTS.

4 Conclusions

We have presented an attention driven search strategy mimicking the behaviour of the human saccadic system applied to an eye detection problem. The sampling strategy employed has accordingly been based on a rigid retinal sampling grid obtained by log-polar mapping. The sampled values were not gray level intensities but vectors representing the local power spectrum (Gabor decomposition), which are in turn obtained by another log-polar mapping.

The algorithm described has proved able to perform robust detection of eyes after as few as six images in the training set. The detected eyes have been used to recognise people [4]. Once the retinal sampling grid is implemented as a hardware device or a stand-alone unit, the main advantage of the presented eye detection technique is the reduced bandwidth requirements between the sensor and the processing unit. The worst case computational load could probably be reduced by implementing a description of other facial features, which would serve as auxiliary spatial cues in the case that the number of saccades be problematical.

5 Acknowledgement

This work has been supported by the VIRSBS project within the European IT-LTR programme.

References

1. J. Bigun. Gabor phase in boundary tracking and region segregation. In *proceedings of DSP & CAES conf. Nicosia, Cyprus*, pages pp. 229–237. Univ. of Nicosia, July 14-16 1993.
2. J. Bigun. Speed, frequency, and orientation tuned 3-d gabor filter banks and their design. In *Proceedings of International Conference on Pattern Recognition, ICPR, Jerusalem*, pages C–184–187. IEEE Computer Society, 1994.
3. D. Hubel. *Eye, brain and vision.* Scientific American Library, 1988.
4. A. Makarov, F. Smeraldi, and J. Bigun. Retinotopic Gabor features applied to face recognition. In *The Proceedings of ICCV*, volume submitted to ICCV98. IEEE Computer Society, 1998.
5. G. Sandini and V. Tagliasco. An antropomorphic retina-like structure for scene analysis. *Computer Graphics and Image Processing*, 14:365–372, 1980.
6. J. D. Schall, D. P. Hanes, K. G. Thompson, and D. J. King. Saccade target selection in frontal eye field of macaque. I. Visual and premovement activation. *The Journal of Neuroscience*, 15(10):6905–6918, 1995.
7. C. Weiman and G. Chaikin. Logarithmic spiral grids for image processing and display. *Computer Graphics and Image Processing*, 11:197–226, 1979.
8. A. L. Yarbus. *Eye movements.* Plenum, New York, 1967.