Proceedings of the 1998 International Conference on Image Processing
Chicago (USA), vol. 3, pages 163 – 167, October 4–7, 1998

Facial feature detection by Saccadic Exploration of the Gabor Decomposition

F. Smeraldi, J. Bigün
Microprocessor and Interface Laboratory
Swiss Federal Institute of Technology (EPFL)
CH-1015 Lausanne

Abstract

The Gabor decomposition is a ubiquitous tool in computer vision. Nevertheless, it is generally considered computationally demanding for active vision applications. We suggest an attention–driven approach to feature detection inspired by the human saccadic system. A dramatic speedup is achieved by computing the Gabor decomposition only on the points of a sparse retinotopic grid. An application to eye detection is presented. Also, a real–time head detection and tracking system based on our approach is briefly discussed. The system features a novel eyeball-mounted camera designed to mimic the dynamic performance of the human eye and is, to the best of our knowledge, the first example of active vision system based on the Gabor decomposition.

1 Introduction

Gabor decomposition has long been known to be a powerful tool for pattern recognition tasks [2, 7, 3], and its use is motivated by strong biological analogies [8, 13]. Unfortunately, the calculation of Gabor filter responses is computationally demanding for active vision applications. In this paper, we propose a bio–inspired approach to circumvent this problem.

The human eye explores a visual scene by performing a sequence of large "jumps", known as saccades, between the different points of interest [5, 11]. Saccades play a central role also in the underlying cognitive processes, where there appears to be a selection mechanism that filters task relevant information [9, 10]. A dramatic reduction of the information flow is therefore achieved by the joint use of the saccadic system and of nonuniform image sampling, which is achieved at the retinal level. Our approach is an attention–driven search based on a model of saccadic eye movements. The algorithm is built around a sparse log–polar retinotopic grid. The Gabor decomposition is computed only on the points of the grid, so that the computational effort is greatly reduced.

An application of saccadic search to eye detection will be presented. We shall also briefly describe our real–time setup for head localisation and tracking that is, to the best of our knowledge, the first example of active vision system based on the Gabor decomposition.

2 The retinotopic sampling grid

Central to our attentional strategy is the use of a sparse retinotopic sampling grid which is rigidly displaced on the images. The grid has log-polar geometry, meaning that the density of sampling points decreases exponentially with the distance from the centre. In our approach, we limit the computation of the Gabor decomposition to the points of the retinal grid, and require the grid to be displaced in order for other image regions to be considered. This sampling topology automatically implements a "focus of attention" concept, concentrating the computational effort on the current fixation point. Furthermore, by keeping the global number of fixation low, it is possible to perform the feature extraction by direct filtering in the image domain, without the need of a Fourier transform. This brings about a dramatic increase in efficiency.

In analogy with the operation of the visual cortex of primates and humans [6], we found it beneficial, at least during the finest part of the search (section 6), to tune our frequency decomposition so that it matches the variable sampling rate of the retina. We therefore associate high frequency Gabor filters to the fovea of the retina, while low frequency responses are extracted at the periphery, where the sampling rate is coarser.

For our eye detection application, this allows a smooth integration of dense information from the centre of the eyes and global information from the outline of the orbit.

3 Log–polar frequency domain sampling

Complex valued Gabor functions in the frequency domain are scaled, translated and shifted versions of the following function:

$$\hat{\mathcal{G}}(\vec{\omega}|\sigma_x, \sigma_y, \omega_0) = \exp\left(-\frac{(\omega_x - \omega_0)^2}{2\sigma_x^2}\right) \cdot \exp\left(-\frac{\omega_y^2}{2\sigma_y^2}\right)$$

The parameters σ_x, σ_y, ω_0 and the rotation parameter are chosen to cover the frequency plane as completely at possible.

However, when only a small number of logarithmically spaced frequency channels is used,problems arise in obtaining a uniform coverage of the frequency plane. Since the spacing between the centres of the filters increases exponentially, the symmetric Gaussian shape doesn't appear to be optimal, since it extends the same distance towards the (well sampled) central region of the frequency space as well as towards the loosely sampled periphery. For these reasons, we choose to substitute for the standard Gabor function a modified filter

$$\hat{\mathcal{G}}'(\vec{\omega}|\sigma_\rho, \sigma_\phi, \rho_0) = \exp\left(-\frac{(\rho - \rho_0)^2}{2\sigma_\rho}\right) \cdot \exp\left(-\frac{\omega_\phi}{2\sigma_\phi^2}\right)$$

where $(\rho, \phi) = (\ln(|\vec{\omega}|), \tan^{-1}(\omega_y/\omega_x))$ is the conformal mapping of the frequency plane to log polar coordinates [1]. We therefore construct a uniform grid of Gaussian filters in the log–polar frequency domain, which in turn yields the desired uniform coverage of the Fourier plane (figure 1).

4 Eye localisation

When human subjects explore a natural scene, they do not use their eyes to scan it in a raster–like fashion. They rather perform rapid jumps between regions of interest, which they fixate for about 0.3 seconds. All the relevant information is acquired during such fixations, although much of the time is spent in deciding where the next saccade should be aimed. In 1957 Yarbus, who pioneered the study of the saccadic system, found that the stopping places of a subject's gaze exploring human faces were more densely distributed in the eye region [14, 5]. This motivated us to use saccadic search for eye detection, even because of the relevance of such region for face recognition.

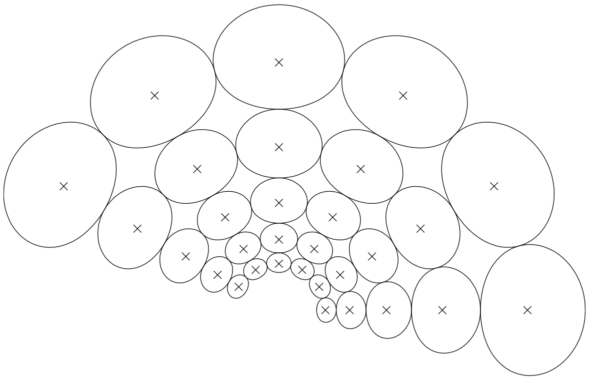

Figure 1: Iso–curves of the Gabor filters created by uniform sampling of the frequency plane in log–polar coordinates. The crosses represent maxima.

The procedure consists of three main steps. At first, local information driven saccadic eye movement is used to home the retinotopic grid on one of the eyes; following, the search is refined by pixel-wise displacement of the grid; finally, if detection is successful a saccade is performed to the assumed position of the other eye. During each of the above steps, several criteria are applied to check for the consistency of information. If a mismatch is detected, doubtful assignments are discarded.

5 Saccadic search

A local, appearance–based description of the search target (the eyes) is constructed by averaging the Gabor responses from the centre of the eyes of the persons in the training set. The resulting feature vector e_{av} consists of six orientation-selective responses for each one of the five frequency channels employed [12].

At the beginning of the search, the retinal sampling grid is placed at a random position on the image and the Gabor feature vectors are extracted for each of its points. Each of these vectors is subsequently matched against the reference e_{av}. The point of the grid for which the Euclidean distance from e_{av} is minimal is selected as the target for the next saccade. Saccadic search is assumed to have converged when saccades become shorter than a threshold. If no saccade target

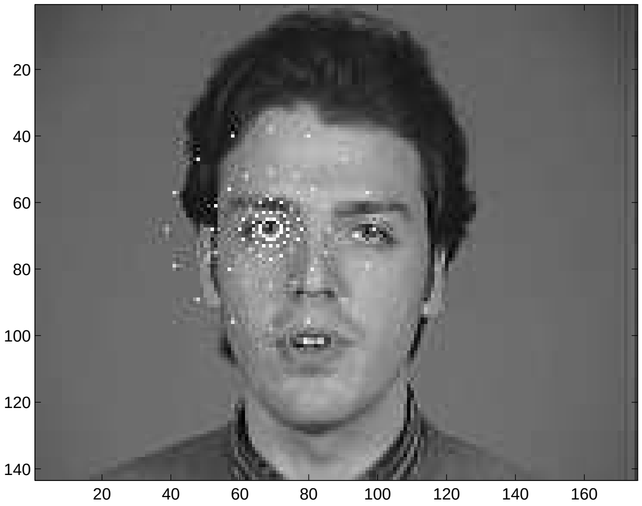

Figure 2: The retinal sampling grid placed on a person's right eye for model creation.

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.

6 Refining the search

A more complete description is obtained, for each eye, by placing the retinal sampling grid on the centre of each eye on the images of the training set (figure 2) and storing the Gabor responses from all of the retinal points. In order to reduce the sensitivity to positioning errors for small training sets, a relaxation procedure is used: user–supplied eye coordinates are employed to train a first version of the models, which is then used to perform a search on the training set itself. The eye coordinates thus detected are then used to retrain the system.

The two resulting "extended" eye models are used to distinguish left eyes from right eyes and to improve the precision of the localisation. A first comparison of the left and right eye models with the features currently "seen" by the retina is performed to state whether the spotted facial feature looks more like a left eye or a right eye. A gradient descent minimisation is successively performed by displacing the retina pixel-wise until the best match with the appropriate eye model is found.

The residual distance from the model is used to classify the detected feature as "eye" or "non–eye". The saccadic search is subsequently restarted in the expected direction of the other eye or, in the case that no eye has been found, 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.

7 Experimental results

7.1 Simulation

The algorithm has been tested using a retinal sampling grid with 5 rings and 16 rays. The relation between the dimension of the retina and the size of the facial images is evidenced in figure 2.

The image database employed consists of forty frontal shots of twenty different persons[1]. The image resolution employed 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 (figure 3). Information obtained from the outline of the orbit allows correct detection of the features even when the subject's eyes are closed (figure 4). 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

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

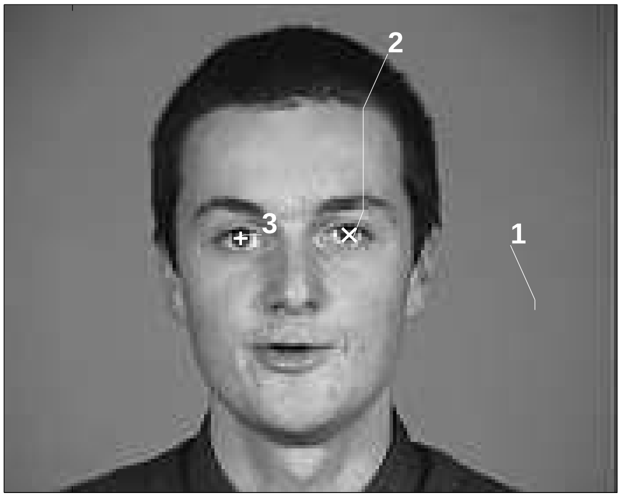

Figure 3: The + and × signs denote the best match with the right and left eye models respectively. Numbers identify successive starting points for saccades. Eye detection required 51 fixations. Note how saccadic search 1 was considered uninteresting and therefore discarded. A random restart (2) then lead to detection of the left eye, after which saccadic search resumed (3) near the location of the right eye.

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.

7.2 Real time head localisation and tracking

In order to demonstrate the flexibility and efficiency of the algorithm we implemented it into a real–time head localisation and tracking system. The retinotopic grid was “attached” to a b/w steerable camera developed at our laboratory 5. The camera had a spherical mount and was explicitly designed to mimic the performance of the human eye.

By substituting the eye model with an analogous head model, the saccadic search procedure described in section 5 could be used to perform head localisation and tracking. Real time performance was achieved on a 200 MHz Pentium processor PC [4], allowing the camera position to be adjusted every 0.5 seconds on the average.

The head localisation and tracking setup has been tested by using it to acquire 50 “passport size”images of each of 10 different subjects. The system was programmed to acquire a frame each time it believed the head of the person to be centred in the image. Acquisition of 50 frames took about one minute. Out of a total of 500 frames, 90% turned out to represent the head of the subject at “passport photo” quality.

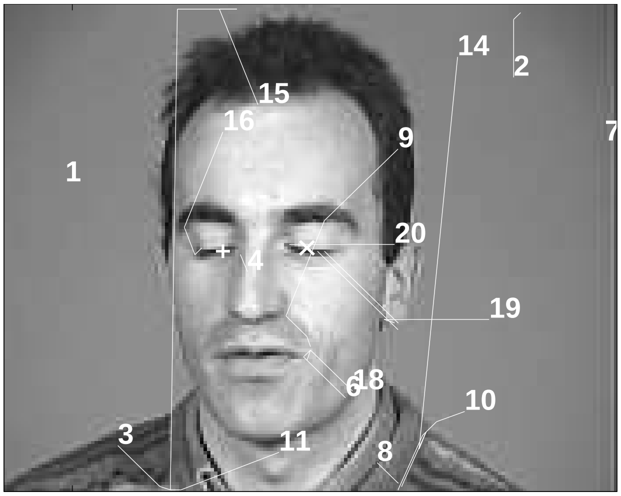

Figure 4: Information from the outline of the orbit allows eye detection even if the person’s eyes are shut. During this trial the centre of the sampling grid explored 99 pixels and 14 targets were rejected after comparison with the eye models.

8 Conclusions

We have presented an attention driven search strategy mimicking the behaviour of the human saccadic system. The main feature of this algorithm is the log–polar sampling of the Gabor decomposition. We have discussed two applications: eye detection on static images and real–time head detection and tracking. We believe that the main advantages of our approach are its generality and the dramatic reduction of the information processed in order to perform the task. This is crucial to allow the use of the Gabor decomposition in active vision applications. Our setup constitutes, as far as we know, a novelty in that sense. Validation of the feature detection algorithm and of the tracking system itself on a large database of subjects are in progress.

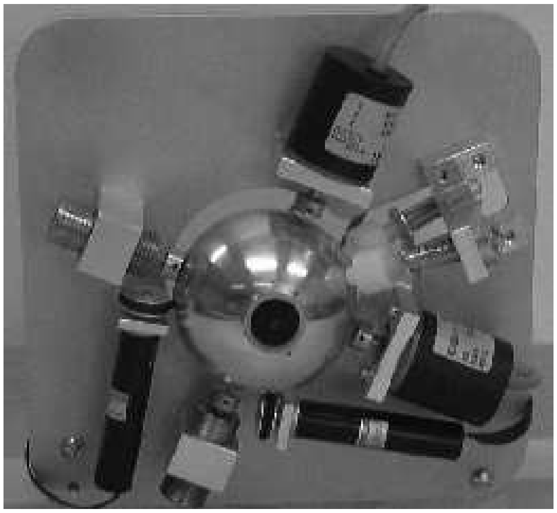

Figure 5: The eyeball mounted camera. The two narrow black cylinders are step motors; the larger ones are optical encoders. Position control is performed by a dedicated microcontroller card connected to the PC's serial port.

Acknowledgement

This work has been supported by the VIRSBS project within the European IT-LTR programme. We thank Mr. M. Bachmann and Mr. S. Menot for their help with programming and hardware.Special thanks to Mr. O. Carmona for making his eyeball mounted camera available to us.

References

[1] 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.

[2] J. Bigun and J. M. H. du Buf. N-folded symmetries by complex moments in gabor space. *IEEE-PAMI*, 16(1):80–87, 1994.

[3] B. Duc, S. Fischer, and J. Bigun. Face authentication with sparse grid gabor information. In *IEEE Proc. of ICASSP, Munich*, volume 4, pages 3053–3056, 1997.

[4] F.Smeraldi, O. Carmona, and J. Bigun. Saccadic search with Gabor features applied to eye detection and real-time head tracking. *Accepted by Image and Vision Computing*, 1998.

[5] D. Hubel. *Eye, brain and vision.* Scientific American Library, 1988.

[6] L. Maffei and A. Fiorentini. Spatial frequency rows in the striate visual cortex. *Vision Res.*, 1977.

[7] B. S. Manjunath, C. Shekhar, and R. Chellappa. A new approach to image feature detection with applications. *Pattern Recognition*, 31:627–640, 1996.

[8] G. A. Orban. *Neuronal operations in the visual cortex.* Studies of brain functions. Springer, 1984.

[9] J. B. Pelz. *Visual representations in a natural visuo-motor task.* PhD thesis, Carlson Center for Imaging Science, Rochester Institute of Technology, 1995.

[10] R. P. N. Rao, G. J. Zelinsky, M. M. Hayhoe, and D. H. Ballard. Eye movements in visual cognition: a computational study. Technical Report 97.1, National Resource Laboratory for the Study of Brain and Behavior, Department of Computer Science, University of Rochester, 1997.

[11] 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.

[12] F. Smeraldi, A. Makarov, and J. Bigün. Saccadic search with gabor features applied to eye detection. Technical Report 98/256, Swiss Federal Institute of Technology, Computer Science Department, CH-1015 Lausanne, January 1998. ftp://lamiftp.epfl.ch/pub/smeraldi/gaboreye.ps.gz.

[13] R. P. Würtz. Building visual correspondence maps — from neuronal dynamics to a face recognition system. In *Proceedings of the International Conference on Brain Processes, Theories and Models.* MIT Press, November 1995.

[14] A. L. Yarbus. *Eye movements.* Plenum, New York, 1967.