Localizing Eyes in Face Images through Saccadic Search

Plinio Moreno López and José Santos-Victor

Instituto Superior Técnico & Instituto de Sistemas e Robótica
1049-001 Lisboa - PORTUGAL
{plinio,jasv}@isr.ist.utl.pt

Abstract

We present an approach for detecting eyes in face images. This system is meant to provide input to a face detection system based on the combination of simpler facial features. The eyes are modeled as a feature vector collecting the response o Gabor filters at various orientations and scales. After learning the model from eye images, we search for new instances of the model, by evaluating the Gabor filter responses in image positions distributed according to a log-cartesian or log-polar sampling grids. The space variant resolution of these grids provides a way of locating the "focus of attention" in the center of each grid.

In addition to comparing different space variant sampling schemes, we investigate different metrics for establishing the distance between the eye model and image observations: Euclidean versus Mahalanobis distances..

Our results indicate that this saccadic search behaves well locally and that better results are obtained with log-Polar grid.

1 Introduction

Gabor filters have been extensively used as features in texture and object detection [3, 5, 7, 6]. Recently some research has attempted to explore the magnitude of filter's image convolution as raw features, arranged either as vectors [3] or matrices [7, 6] to achieve some rotational invariance.

In this work we evaluate the performance of Gabor filters as features in a saccadic search task. The overall goal consists in finding faces in images. We approach this problem by considering that a face is composed of a set of elementary facial features, such as the eyes, mouse, nose, eyebrows, etc. Hence, face detection would be achieved through the successful detection of these elementary features, provided that some geometric constraints, between these features, are met [8, 4, 2].

At this moment, our work focused on the use of Gabor features for the detection of individual features, namely, a person's eye. The idea is that a cognitive vision system would search a scene for faces or, in the current level of our implementation, for eyes. Thus, the system must search over a face image, performing a succession of saccades, in order to find the person's eye, as quickly as possible.

Our model for an eye consist of a vector, made up with the magnitude of Gabor filter outputs, evaluated at the eye location. These filters, tuned to different spatial frequencies and orientations are supposed to capture the particular texture distribution corresponding to a human eye (or any other particular facial feature). This model is learned through the observation of a set of facial features (eyes).

The search for an eye requires analyzing the filters output at various image locations. We have tested two different ways of sampling the image for the purpose of this visual search. In the first case, we use a log-cartesian sampling scheme to determine where the Gabor should be evaluated in the image. Instead, in the second approach, we use a log-polar sampling grid. The search is conducted by saccading the camera towards the direction in the field of view, yielding the Gabor responses closest to the eye model. Hence, the system will explore the image points according to the log-cartesian or log-polar grids, instead of searching the entire image.

The experiments described in this report aim to evaluate the following points

1. The usage of Gabor responses as models for visual features and

2. the comparison between log-cartesian and log-polar representations to guide the search in the image.

In Section 2 we make the method review, in Section 3 the discussion and some results and in Section 4 the conclusions.

2 Description of the work

2.1 Gabor Filters as features

Gabor filters have information about edges and ridges in all possible orientations and scales. The Gabor filter *kernel* is defined through the product of a gaussian term and a complex exponential. In the present case, we will consider that the Gaussian part is isotropic, thus being characterized by a single scale parameter, σ. The Gabor function can be expressed as:

$$g_{\theta,\lambda,\sigma}(x,y) = \exp\left\{-\frac{x^2+y^2}{2\sigma^2}\right\}\exp\left\{\frac{j\pi}{\lambda}(x\cos\theta + y\sin\theta)\right\} \tag{1}$$

Figure 1 shows an example of a Gabor filter with zero mean.

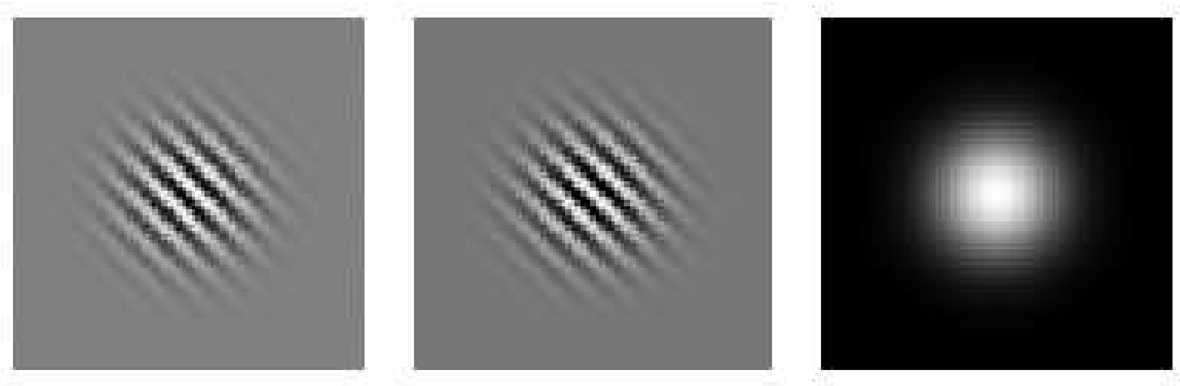

Figure 1: Left/center images real/imaginary part of the Gabor function(zero mean). Right: magnitude of Gabor function(zero mean). $\theta = \pi/4$, $\sigma = 9$ pixels, $\lambda = 5$ pixels

For our approach we use the absolute value of filter's output to form the image feature vectors.

2.2 Eye model: Feature Selection

As we said previously, the eye model consists of a vector of Gabor responses at the various scales and orientations. The values of σ, λ and θ in equation (1) must be chosen in such a way that they represent uniquely our detection target. Equally spaced angles between 0 and π have enough information about edges and ridges, and the eyes size must be related with λ's and σ's range.

In order to represent an eye, we use the Gabor filter response, $\mathbf{v}$, calculated at the eye's center

$$\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$$

$$v_i = |I(x_{ec}, y_{ec}) * g_{\theta_i,\lambda_i,\sigma_i}(x_{ec}, y_{ec})| \tag{2}$$

(x_{ec}, y_{ec}) is the eye's center position and n is the number of features. Finally, v is normalized ($|v| = 1$), in order to obtain image contrast invariance, gaining robustness to illumination changes.

The eye-model, $\bar{\mathbf{v}}$, is learned by observing a set of input eye patterns and the mean feature vector is kept as the eye model. In addition, the covariance of the observed feature vectors is also estimated. Since the number of training patterns is relatively modest (usually we use about 15 training patterns) we assume the covariance matrix is diagonal and only these elements are estimated.

2.3 Searching for eyes

Instead of evaluating the feature vector in every image pixel, we are going to use log-Polar and log-Cartesian grids. The idea is to reduce the number of points needed for reaching the eye and test the space variant sampling characteristics of the log-Polar and log-Cartesian grids, as shown in Figure 2. The space variant resolution provides a way of locating the "focus of attention" in the grid's center.

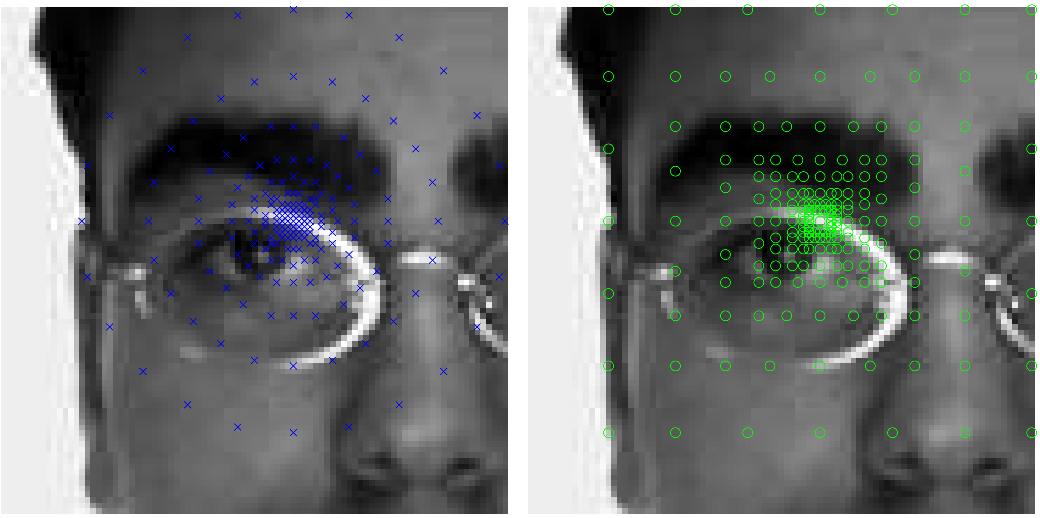

Figure 2: Log-Polar and Log-Cartesian grids samples

In the first "saccade" we compute the feature vector v for every grid point. The point that has minimum distance to eye's model is selected as the saccade objective. This procedure is done until a minimum is reached. The process depends on the metrics chosen for comparing the eye model and the observed image regions. We tested both the Euclidean and the Mahalanobis distances:

$$
\begin{aligned}
d_{Euclidean} &= \sqrt{\sum_i (\bar{v}_i - v_i)^2} \\
d_{Mahalanobis} &= \sqrt{(\mathbf{v} - \bar{\mathbf{v}})^T C^{-1} (\mathbf{v} - \bar{\mathbf{v}})}
\end{aligned}
$$

where the covariance matrix C required to determine the Mahalanobis distance, is assumed to be diagonal, $C_{ii} = \text{variance}(v_i)$.

3 Results and Discussion

We have applied the approach described before to the problem of finding (left) eyes in images. The λ, σ and eye's center position values were chosen by hand during the training. This process could be automatically, with certain criteria (i.e. to establish a cost function).

The experiments were performed with the AR face database[1] images. To obtain the Gabor feature vector, the values of σ and λ were chosen a shown in Table 1. For each pair in Table 1, $\theta = \{0, \pi/4, \pi/2, 3\pi/4\}$. The function expressed in Equation (1) allow us to make a dyadic recursive implementation of the image's convolution with the Gabor function in a fast way and remaining confidence.

σ(rows),λ(columns)	1.7	3.7	7.4	14.8	29.6
0.95	X	X			
2.12		X	X		
4.35			X	X	
8.75			X	X	
17.5			X	X	
35.04				X	X

Table 1: λ and σ pairs used in Gabor filters(selected cells, pixel values)

Figure 3 illustrates the "eyeness" (i.e. $d_{Euclidean}$ or $d_{Mahalanobis}$) evaluated at every pixel of a complete image.

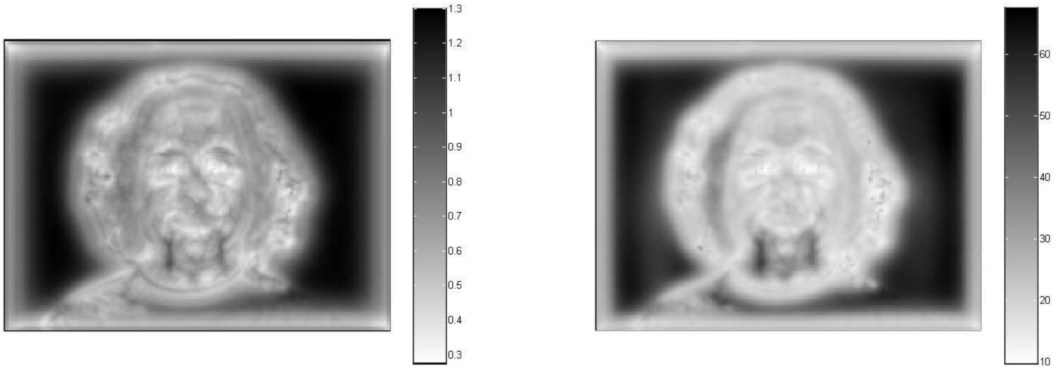

Figure 3: Eyeness of an image computing Euclidean (left) or the Mahalanobis (right) distances in feature space

The "eyeness" is given by $E(x, y) = dist(\bar{v}, v^{(x,y)})$, where $dist$ corresponds either to the Euclidean or Mahalanobis distances in the feature space, $\bar{v}$ is the feature vector's mean and $v^{(x,y)}$ is the feature vector at image position (x, y).

Trying to do a global search is not feasible. Clearly, if the initial point lies outside the face, the search is very likely to fail, as many image regions may have good eyeness scores (i.e. false detections).

We perform searches with the log-Polar grid, log-Cartesian grid, Euclidean distance and Mahalanobis distance. The results are plotted in Figure 4 and summarized in table 2. The Log-Polar grid performs better in terms of number of saccades required to conclude the search, and also in the number of false eyes. From table 2, the best combination is the log-Polar grid and Mahalanobis distance in feature space.

	Mean number of saccades	false positives
log-Polar-Euclidean distance	2.32	4.44%
log-Cartesian-Euclidean distance	2.42	4.44%
log-Polar-Mahalanobis distance	2.25	6.66%
log-Cartesian-Mahalanobis distance	2.26	13.33%

Table 2: Comparison between all experiments

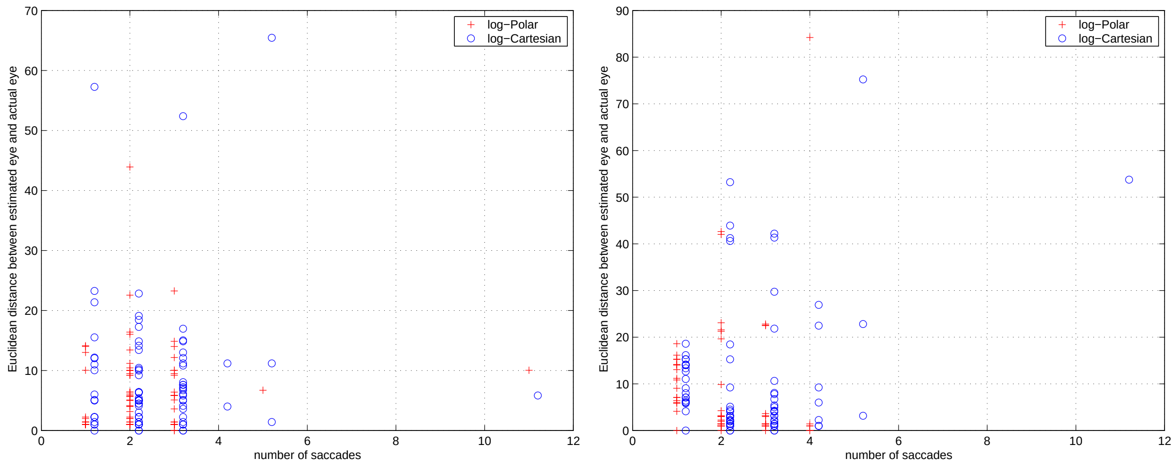

Figure 4: Experiments with Euclidean (left) and Mahalanobis (right) distances between vectors in feature space.

4 Conclusions

We have presented a method for searching specific visual features in the field of view through a sequence of saccades. The feature model consists of a collection of responses of Gabor filters. Our results indicate that such features are powerful enough to discriminate between the sought features and other textured regions.

The second main conclusion of our work is that there seems to be a (modest) gain in using log-polar sampling grids to evaluate where in the image to saccade in the following time step. Not only log-polar leads to smaller search steps, when compared to log-cartesian, but the rate of false positive is also smaller.

There are two differences between this work and [3]. Firstly, the feature vector normalization increase the image range and secondly, we use the Mahalanobis distance in the feature space, which further allows to decrease the number of saccades performed in the search.

Future work will further evaluate these results and also the use of area features, instead of point features, may increase the discriminative power of the approach. Finally, we believe that an adequate level of reliability in the detection process can only be attained once the macro-level (feature combination) is addressed.

References

[1] Martinez A.M. and Benavente R. The ar face database. Technical report, CVC, June 1998.

[2] Owen Carmichael and Martial Hebert. Object recognition by a cascade of edge probes. *Proceedings of the British Machine Vision Conference*, September 2002.

[3] Smeraldi F. and Bigun J. Retinal vision applied to facial features detection and face authentication. *Pattern Recognition Letters*, 23, 2002.

[4] R. Fegus and A. Zisserman P. Perona. Object class recognition by unsupervised scale-invariant learning. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2003.

[5] R. Feris, V. Krueger, and R. Cesar Jr. Efficient real-time face tracking in wavelet subspace. *ICCV2001 Workshop RATFG-RTS'01*, 2001.

[6] Kyrki V. Kamarainen J.-K. and Kälviäinen H. Invariant shape recognition using global gabor features. *Proceedings of the 12th Scandinavian Conference on Image Analysis*, 2001.

[7] Kyrki V. Kamarainen J.-K. and Kälviäinen H. Fundamental frecuency gabor filters for object recognition. *Proceedings of the 16th International Conference on Pattern Recognition*, 2002.

[8] M. Weber, M. Welling, and P. Perona. Towards automatic discovery of object categories. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2000.