A Comparison of Gabor Filter Methods for Automatic Detection of Facial Landmarks

Ian R Fasel[1] Marian S Bartlett[2]
Javier R Movellan[2]
[1]Department of Cognitive Science, [2]Institute for Neural Computation
University of California, San Diego
La Jolla, CA, 92093-0515

E-mail: {ian,marni,javier}@inc.ucsd.edu

Abstract

This paper presents a systematic analysis of Gabor filter banks for detection of facial landmarks (pupils and philtrum). Sensitivity is assessed using the A' statistic, a non-parametric estimate of sensitivity independent of bias commonly used in the psychophysical literature. We find that current Gabor filter bank systems are overly complex. Performance can be greatly improved by reducing the number of frequency and orientation components in these systems. With a single frequency band, we obtained performances significantly better than those achievable with current systems that use multiple frequency bands. Best performance for pupil detection was obtained with filter banks peaking at 4 iris widths per cycle and 8 orientations. Best performance for philtrum location was achieved with filter banks with 5.5 iris widths per circle and 8 orientations.

Gabor filter banks [4] are reasonable models of visual processing in primary visual cortex [5, 3] and are one of the most successful approaches for processing images of the human face [6, 1, 2, 8]. The success of the approach parallels the success of bandpass filter banks, which approximate signal processing in the cochlea, in speech recognition problems. While the optimal filter bank characteristics have been extensively studied in the speech recognition literature, little work has been done to systematically explore which frequency and orientation bands are optimal for face processing applications. The goal of this paper is to start addressing this gap in the literature. To evaluate performance of the different filter bank approaches, we use a standard recognition engine (nearest neighbor) and measure sensitivity using the A' statistic. This is a non-parametric measure of sensitivity commonly used in the psychophisical literature.

1 Gabor Filters

In the space domain, the impulse response of Gabor filters is a Gaussian kernel modulated by a sinusoidal plane-wave

$$g(x,y) = s(x,y)w(x,y) \tag{1}$$

where $s(x,y)$ is a complex sinusoidal, known as *the carrier*,

$$s(x,y) = \exp(-j(2\pi(u_0 x + v_0 y))) \tag{2}$$

The parameters u_0 and v_0 define the spatial frequency of the sinusoidal, in Cartesian coordinates. This spatial frequency can also be expressed in polar coordinates as magnitude F_0 and direction ω_0. The function $w(x,y)$ is a 2-D Gaussian-shaped function, known as the *envelope*

$$w(x,y) = K \ \exp(-\pi(a^2(x-x_0)_r^2 + b^2(y-y_0)_r^2)) \tag{3}$$

where (x_0, y_0) is the peak of the function, a, b are scaling parameters, and the r subscript stands for a rotation operation

$$(x-x_0)_r = (x-x_0)\cos\ \theta + (y-y_0)\ \sin\theta \tag{4}$$

and

$$(y-y_0)_r = -(x-x_0)\sin\ \theta + (y-y_0)\ \cos\theta \tag{5}$$

The 2-D Fourier transform of the Gabor function is as follows (see Figure 4)

$$\begin{aligned}\hat{g}(u,v) = &\frac{K}{a\,b}\ \exp(-j(2\pi(x_{0r}\ u + y_{0r}\ v))) \\ &\cdot \exp(-\pi((\frac{(u-u_0)_r}{a})^2 + (\frac{(v-v_0)_r}{b})^2))\end{aligned} \tag{6}$$

Or in polar coordinates,

$$\begin{aligned}&\text{Magnitude}(\hat{g}(u,v)) = \\ &\frac{K}{a\,b}\ \exp(-\pi((\frac{(u-u_0)_r}{a})^2 + (\frac{(v-v_0)_r}{b})^2))\end{aligned} \tag{7}$$

$$\text{Phase}(\hat{g}(u,v)) = 2\pi(x_{0r}\ u + y_{0r}\ v)) \quad (8)$$

Thus, in the frequency domain, the impulse response is an oriented Gaussian centered at the frequency of the carrier. The iso-magnitude contours are ellipsoidal. It is convenient to characterize the impulse response via its half-magnitude contours. This requires 6 parameters (see Figure 1):

- F_0 and ω_0, the polar coordinates of the peak of the envelope.
- θ is the angle of rotation of the envelope.
- a and b are 1.06 times the lengths of the axis of the half-magnitude ellipse.
- θ is the angle of the axis controlled by the parameter a.

Gabor filters are good models of the receptive fields found in simple cells of cat and macaque striate cortex. The following constraints on the 6 parameters are known to hold, to a first approximation, in striate cortex [5, 3]: (1) $\theta = \omega_0$; (2) The half-magnitude spatial frequency bandwidth is about 1.4 octaves; (3) The half-magnitude orientation bandwidth is about 40 degrees. From these constraints it can be derived that a is about $0.95 \times F_0$ and b is about $a/1.24$. Using these constraints, we are left with two parameters to vary independently: the frequency F_0 of the carrier plane wave, and orientation of the carrier wave and envelope ω. Figures 4

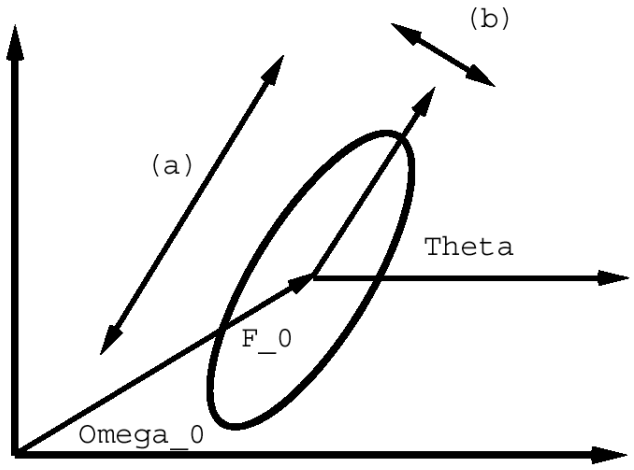

Figure 1. Parameters of the Gabor filter as reflected in the half-magnitude elliptic profile.

and 5 show examples of different Gabor kernels that adhere to the biological constraints. Since Gabors wavelets are localized in the frequency domain (they are bandpass filters) a search for optimal Gabor parameters can be interpreted as a search for spatial frequency regions that provide the greatest information for the task at hand.

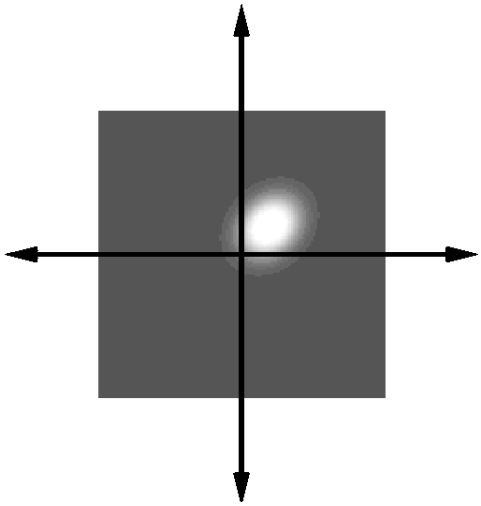

Figure 2. The Fourier transform of the Gabor filter. The peak response is at the spatial frequency of the complex sinusoidal: $u = v = 80\ pixels/cycle$.

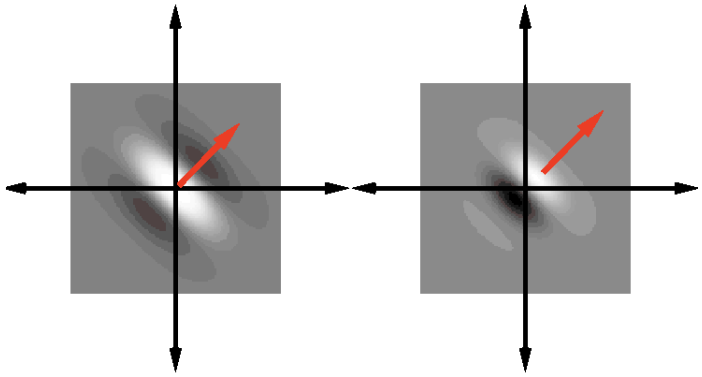

Figure 3. The real and imaginary parts of a complex Gabor. The images are 128x128 pixels. The parameters are as follows: $x_0 = y_0 = 0$, $a = 50$ **pixels,** $b = 40$ **pixels,** $\theta = -45$ **degrees,** $u_0 = v_0 = 80\ pixels/cycle$, $P = 0$ **degrees.**

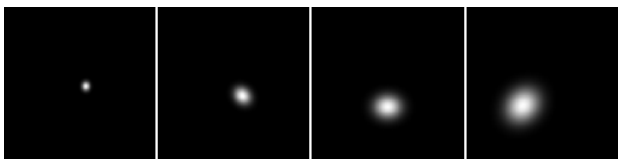

Figure 4. Gabor kernels at 4 orientations and 4 frequencies in frequency domain.

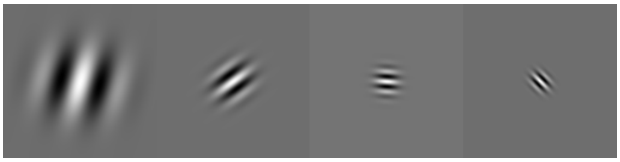

Figure 5. The corresponding Gabor kernels at 4 orientations and 4 frequencies in space domain.

1.1 Database

We used the FERET face database [7], a set of images collected by the US Army Research Laboratory to develop, test, and evaluate face recognition algorithms. A set of 446 frontal view images was randomly selected from this database and labeled by hand to locate key features. In our experiments we used pupils and philtrum as the features of interest.

2 Experimental Design

We randomly chose from each image three points from each of the features. These points could deviate from the center of the feature by a distance no less than the typical radius of an iris. For each image we also selected three non-feature points. The location of these points varied randomly from image to image. Thus for each test round there were 4284 test locations for features and non-features across all test images. (See Figure 6).

To evaluate performance of the different image processing approaches we used a nearest neighbor recognition engine: (1) We convolved the image with a bank of Gabor filters; (2) We stored a set of examples of filter bank outputs at the desired fiducial points; (3) When detecting whether a pixel in a new image contains a desired fiducial point, we computed the Euclidean distance from the output of the filter bank at that pixel and the output in the training database that it is closest to (See Figure 7). A point was recognized as containing a feature if its similarity value is below a certain threshold κ.

Each feature detector was tested on all test locations, both features and non-features. $P(\text{detect} \mid \text{feature absent})$ and $P(\text{detect} \mid \text{feature present})$ were then calculated based on the total number of test locations for feature versus non-feature.

We measured performance of each feature detector using standard ROC curves. For each value of the threshold parameter we plot $P(\text{detect feature} \mid \text{feature absent})$ vs. $P(\text{detect feature} \mid \text{feature present})$. The area under the ROC curve is called A' and is a non-parametric measure of sensitivity commonly used in the psychophysical literature. It can be shown that A' represents the optimal performance

Figure 6. For each test run, a new database of test locations is created by selecting three points from each feature, plus a set of random locations elsewhere in the image. These random locations vary from image to image. Thus for each test run the hit and false alarm rates of the system were estimated as a function of the threshold κ based on 4284 total tests.

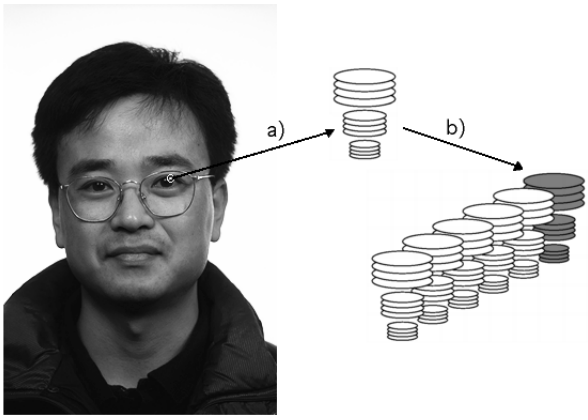

Figure 7. The process of matching features. First, a feature is convolved with a set of kernels to make a jet (a). Then that jet is compared with a collection of jets taken from previous images, and the similarity value for the closest one is taken (b).

achievable by a detector on a 2-alternative forced choice task. Thus an A' of 0.5 represents zero sensitivity, and an A' of 1.0 is perfect sensitivity (see Figure 8).

We systematically tested Gabor filter banks with different peak frequencies and number of orientation bands. We varied F_0 from 2 pixels/cycle to 50 pixels/cycle (about 6 to

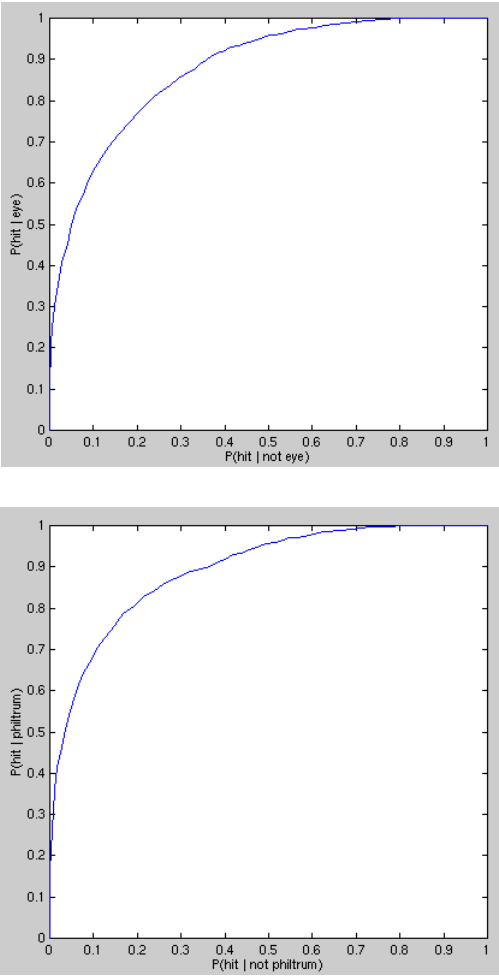

Figure 8. Top: ROC curve for best performing pupil detector (34 pixels/cycle, 8 orientations, $A' = .8781$). Bottom: ROC curve for best performing philtrum detector (44 pixels/cycle, 8 orientations, $A' = .8955$).

.12 cycles per iris width), and varied the number of orientations from 1 to 12, evenly distributed, starting at $\theta = 0$. This made a total of 576 different combinations of frequency vs. number of orientations. For each run, we tested generalization performance using a standard cross-validation approach, training on 357 images, testing on the remaining 89 images and repeating the procedure 5 different times, each time using different training and generalization sets.

3 Results

Peak performance was obtained using eight orientations and carriers of 32 pixels/cycle (about 4 iris widths per cycle) for pupils and 44 pixels/cycle (about 5.5 iris widths per cycle) for philtrum. The best A's were 0.8781 and 0.8664 respectively (see Figure 8). This performance was significantly better than that obtained using the set of Gabors used in [8, 1] which had in the order of 10 frequency bands and 8 orientations. When we used the Gabor filter bank described in [8], the A' was only 0.6141 for eye detection.

We found that as the peak frequency decreases, the optimal number of orientations also increases. For high frequency Gabors, the optimal number of orientations was two. As the carrier frequency was reduced, the optimal number of orientations increased to 8 (See Tables 1 and 2).

pix/cyc	6	10	18	22	28	38	44
A'	.7038	.7607	.8604	.8536	.8603	.8767	.8955
# orient	2	2	2	8	8	8	8

Table 1. Best A' for several example combinations of number of orientations and frequency for pupil detection

pix/cyc	6	14	22	28	34	38	44
A'	.5332	.7794	.8615	.8650	.8781	.8734	.8664
# orient	2	2	2	2	8	8	8

Table 2. Best A' for several example combinations of number of orientations and frequency for philtrum detection

4 Conclusions

While Gabor filter banks are an increasingly popular image processing technique for face processing tasks, little work has been done to identify which frequency and orientation bands are responsible for the success of the approach. Our results show that when the task at hand is locating facial landmarks, one of the first stages in most face recognition systems, current filter banks specifications may be sub-optimal and overly complex. Best performance may be achieved by concentrating a large number of orientations (8 orientation bands) on very low frequency carriers (spatial frequencies in the order of 5 to 8 iris widths per cycle). It should be noted that our results are specific to the task of locating features independent of the subject's identity. It is likely that if the task at hand is subject identification the optimal frequency bands may be different. However, most face identification systems tend to use the same filter bank architecture for feature location and face recognition. Our results suggest that better results may be obtained in the feature finding stage by using only a few filters in the lower end of the frequency spectrum.

References

[1] M. Bartlett. *Face Image Analysis by Unsupervised Learning and Redundancy Reduction*. PhD thesis, University of California, San Diego, 1998.

[2] V. Bruce. Human face perception and identification. In H. . Wechsler, P. Phillips, V. Bruce, F. Fogelman-Soulie, and T. Huang, editors, *Face Recognition: From Theory to Applications*, NATO ASI Series F. Springer-Verlag, 1998.

[3] R. DeValois and K. DeValois. *Spatial Vision*. Oxford Press, 1988.

[4] D. Gabor. Theory of communication. *Journal of the Institute of Electrical Engineers*, 93:429–549, 1946.

[5] J. Jones and L. Palmer. An evaluation of the two dimensional gabor filter model of simple receptive fields in cat striate cortex. *J. of Neurophysiology*, 58:1233–1258, 1987.

[6] C. Padgett and G. Cottrell. Representing face images for emotion classification. In M. Mozer, M. Jordan, and T. Petsche, editors, *Advances in Neural Information Processing Systems*, volume 9, Cambridge, MA, 1997. MIT Press.

[7] P. Phillips, H. Wechsler, J. Juang, and P. Rauss. The feret database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing Journal*, 16(5):295–306, 1998.

[8] L. Wiskott, J. M. Fellous, N. Krüger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):775–779, 1997.