

A Parallel Computer Preattentive-Vision System for Learning and Extracting special Features in 2-D Images

Anne M. Landraud-Lamole

Equipe RAPID: Recognition and Artificial Perception in Images from Data
Département Mathématiques-Informatique, Faculté des Sciences
97159 Pointe-à-Pitre, Guadeloupe, France (FWI)
alamole@netguacom.fr

Abstract

A robust parallel multichannel filtering method for modelling preattentive vision is presented in this paper. Our aim is to achieve an artificial perception such that its performance rather than its working be as close as possible to that of human vision. In establishing a rigorous theoretical basis for constructing original and efficient operators to be incorporated in a robust computer vision system, we drew our inspiration from existing computer methods, which inevitably imply approximate functioning, and also from the experimental results on the mechanisms of human vision, which a priori may be considered as optimum. Our model implements a "homothetic-filter bank" (HIFB) where each filter is selectively sensitive to a frequency-and-orientation pair just like a visual single cell. This model verifies both the Shannon theorem and what is known of the properties of human vision. The proposed system is independent of rotations and scale changes. Moreover, an appropriate choice of the frequency-filter function with its parameters allows us to provide for parallel processing using cascade computations.

Biological Basis

The perception faculty corresponds to the first stages of data processing by the visual system and corresponds to the so-called "preattentive vision" (Neisser 1967). In a performing artificial vision system, three important problems are to be solved: the rigorous sampling of the spatial frequencies, the scale related problem and parallel processing for real time applications. Experiences of biologists have shown that the perception of some phenomena, such as textures, seems to be made instantaneously - in parallel - by neurons localized in area-17 of the visual cortex (Julesz 1975). That preattentive vision is handled by single visual cells. It is generally admitted that these cells behave like filtering channels that are selectively sensitive to frequency-and-orientation pairs and that the frequency sensitivity is independent of the orientation sensitivity (Westheimer 1984). It has been shown that anisotropic single visual-cells sample orientations with a period of about ten degrees (Schiller, Finlay and Volman 1976). The angular bandwidth of each channel is between 30° and 50°, the average bandwidth

being equal to 40°. It has been also established that the spatial-frequency bandwidth is between 0.6 and 2 octaves, the frequency range being about 6 octaves (Movshen, Thompson and Tolhurst 1978). The average value of the spatial frequency bandwidth of single visual cells is slightly above one octave.

Whereas the angular sampling is relatively well-known, the frequency sampling is not yet cleared up and the changeover from 1-D to 2-D models raises some questions. In particular, we do not know any convincing experiment describing the interdependence between spatial frequencies and orientations. It has been observed that the sensitivity to frequency-and-orientation pairs does not imply any correlation between these two parameters. Since the construction of filters separable in frequencies and in orientations is not forbidden by any theoretical consideration, we have opted for that approach to build our model of vision.

Mathematical Model of Visual Perception

Our approach allows us to use, if necessary, different kinds of filters which could help lessen the computing complexity. Taking into account the research work done by psychophysicists, who have shown that the visual cells sensitivities to frequencies are independent of their angular sensitivities, we investigate separately the frequency behaviour of our proposed filters and their angular properties. We have first emphasized properties of the 1-D filters within the scope of frequency processing. It has been shown that the filter bandwidth to be considered will always be nearly proportional to its "preferential" (central) frequency and it was proposed to use real even functions corresponding to the real part of a Gabor function (Pollen and Ronner, 1983). We have shown that this kind of filter belongs to the more general following model. We call "homothetic filter bank" (HFB) any family of filters $\{f_i(x)\}$, where i belongs to an interval I of non-negative integers, having Fourier transforms $F_i(u)$, i belonging to I , that verify the following relations:

$$\exists F: \mathbb{R} \rightarrow \mathbb{C} \text{ and } \forall i \in I, \exists c_i \in \mathbb{R} \text{ and } p_i \in \mathbb{R}: F_i(u) = c_i F(p_i u). (1)$$

We call "profile of the filter bank" such a function F and we say that the HFB is generated by F . Reciprocally, let $F(u)$ be a continuous function defined in \mathbb{R} with values in \mathbb{C} , being the Fourier transform of a function $f(x)$, application of \mathbb{R} into \mathbb{C} . If $F(u)$ has a maximum value at $u = u_m$ and defines a bandpass as wide or lightly broader than u_m , then the HFB engendered by $F(u)$ can be considered for modelling all single visual cells. As a matter of fact, the filters $f_i(x)$ belonging to this HFB have bandwidths of about 1 octave. Subsequently, this set of filters is referred to as a "visual filter bank".

Most approaches using multichannel-filtering make use of energy measures after having filtered the image by Gabor functions for it has been shown that those functions allow an excellent localization in both image and frequency spaces (Bovik, Clark and Geisler 1987). Our model contains a product of 1-D functions which are frequency-and-orientation separable. Our system can be run with either Gabor functions or any other function provided they are well-fitted to frequency filters. We can prove that our model allows for optimal characterization. Our approach consists in characterizing images by energy measures taken as the output of a filter bank whose properties are the same as those of single visual cells.

We have shown that the energy of a signal $s(x)$ can be expressed as a convolution $T(\lambda) * G(\lambda)$ at $\lambda_i = -\log_e(p_i)$. In a "visual filter" bank, each filter $F_k(u)$ of the HFB has its own frequency selectivity (its bandwidth) and its own preferential frequency, noted u_k . Assuming that u_0 is the preferential frequency of the profile F , the following results can easily be inferred from the new expression of the frequency $\lambda = \log_e(u)$ and from the signal theory:

$$p_0 = 1/u_0 \quad (2)$$

$$u_k = u_0 h^k \quad (3)$$

$$\text{and} \quad F_k(u) = (u_k)^{-1/2} F(u/u_k) \quad (4)$$

where the sampling period of the function $e^{-\lambda} F(e^{-\lambda})$ is $\Delta\lambda = \log_e(h)$. This sampling period is calculated by applying the Shannon theorem. The whole filter bank, in the image space, should be $\{f_k(x)\}$, $k \in \mathbb{Z}$ where the functions $f_k(x)$ are the inverse Fourier transforms of the $F_k(u)$ functions. However, it is known that the maximum extent of the human visual system's sensitivity field to spatial frequencies is of six octaves. From these considerations, it can be deduced that in modelling the visual system, only the filters with a preferential frequency lying inside that six-octave range have to be kept so as to have a finite bank of filters.

The next important point is to treat the orientations while describing the changeover from the 1-D representation to the 2-D representation. Assuming that an appropriate function $G(\lambda, \theta)$ achieves a low-pass filtering in the plane (λ, θ) , we can again apply the Shannon theorem. The output energies of the filter bank are then interpreted as resulting from the 2-D sampling of the

function $T(\lambda, \theta) * G(\lambda, \theta)$. Taking into account notes by physiologists on the sensitivity both to angles and to frequencies, we assume that the filter function $G(\lambda, \theta)$ is a separable one:

$$G(\lambda, \theta) = G_{\lambda_i}(\lambda, \theta) = G_{\lambda_i}(\lambda) \cdot G_{\theta_j}(\theta) \quad (5)$$

where λ_i corresponds to the preferential frequency u_i , θ_j being the preferential orientation of the filter $G_{\theta_j}(\lambda, \theta)$. We choose the sampling ratio $\Delta\lambda$ of the λ values and, from $\Delta\lambda$, we infer the frequency-sampling ratio Δu exactly as done in the case of 1-D filters. We also apply the Shannon condition for selecting the θ -sampling ratio as a function of the width of the Fourier transform of $G_{\theta_j}(\theta)$. This is how we achieve a discrete representation in polar coordinates of the function T , i.e. of the energy spectrum in the Fourier space, where the frequency axis is graduated with logarithmic values. Owing to that representation, any scale change corresponds to a simple translation along the logarithmic frequency axis and any shift of orientation is expressed in a translation along the θ -orientation axis.

As we are aiming at practical efficiency, we make a few simplifications. The filters, constructed in the Fourier space, are made to be implemented in the image space for convolution filtering. For that purpose, we have investigated a good approximation making it possible to invert the given equations more easily. First, we consider the specific case where orientation $\theta_0 = 0$. The general shape of the filter being constructed in the Fourier plane is then:

$$F_A(u, v) = (u_0 h_i)^{-1/2} F[u/(u_0 h_i)] F_j[v/(u_0 h_i)]. \quad (6)$$

The expression of such a filter in cartesian coordinates is very simple. Moreover, this filter has the advantageous property to be separable in u and v coordinates, which makes the computation of its inverse Fourier transform easier, this transform being a simple product of two one-dimensional functions of x and y , respectively, in the image plane, instead of a convolution integral. On the other hand, the approximate filter $F_A(u, v)$ is no longer separable in polar coordinates and does not keep constant characteristics in frequencies or in orientations. Consequently, the frequency behaviour of the filter depends on the orientation and, conversely, its directional behaviour depends on the frequency. Because of this, we choose the shape of the filter so that its response is significant both in the neighbourhood of its preferential frequency and near its preferential orientation, while it quickly decreases outside of that neighbourhood. Under these conditions, we admit that the so-obtained approximation is valid. Filters with orientation θ_j will derive from horizontal ones through a simple rotation θ_j .

In a previous paper (Landraud-Lamole and Yum-Oh 1995), in order to make use of phase information, we have considered a classical 1-D complex Gabor function to represent the i th frequency filter. But here, we show that

in order to design a parallel vision system, it is more pertinent to use differences of gaussian functions.

A Parallel Vision System

A difference of gaussian ("dog") functions was suggested (Wilson and Bergen 1979) to simulate the response function of symmetrical single visual cells:

$$f_i(x) = c_1 \exp[-x^2/(2\sigma_1^2)] - c_2 \exp[-x^2/(2\sigma_2^2)] \quad (7)$$

The filter $f_i(x)$ is characterized by four parameters: c_1 , c_2 , σ_1 and σ_2 . We shall see that these filters are well adapted to parallel computing and make it possible to extract the most useful frequency properties of a given signal. The Fourier transform of equation (7) is:

$$F_i(u) = [c_1 \sigma_1 (2\pi)^{1/2}] \exp(-2\pi^2 \sigma_1^2 u^2) - [c_2 \sigma_2 (2\pi)^{1/2}] \exp(-2\pi^2 \sigma_2^2 u^2) \quad (8)$$

The response of the filter $f_i(x)$ to the zero frequency, i.e. to some constant image, is:

$$F_i(0) = [c_1 \sigma_1 - c_2 \sigma_2] (2\pi)^{1/2} \quad (9)$$

So as to make the filter responses independent of the average grey-level and dependent only on frequencies present in the image, $F_i(0) = 0$ is set. Constants c_1 and c_2 have then to verify:

$$c_1 \sigma_1 = c_2 \sigma_2 \quad (10)$$

On the other hand, it is assumed that the filter does not respond to zero frequencies. Indeed, in the case of a bandwidth of one octave centered at the frequency u , the interval $[u/2, 3u/2]$ never includes the zero frequency. In order to normalize gaussian functions that occur in each filter formula $f_i(x)$, the constants c_1 and c_2 are defined as:

$$c_1 = c_i / [\sigma_1 (2\pi)^{1/2}] \text{ and } c_2 = c_i / [\sigma_2 (2\pi)^{1/2}] \quad (11)$$

The determination of the standard deviations σ_1 and σ_2 for each of the two gaussian functions forming a filter makes it possible to infer the constant values c_1 and c_2 by using equation (11). The choice of σ_1 and σ_2 for each filter depends on its preferential frequency and on the size selected for its bandpass. A value of about one octave corresponds to the known properties of visual-cell bands. For the filter $f_i(x)$, we set:

$$\alpha = \sigma_2 / \sigma_1 = \text{constant} \quad (12)$$

The value α is the same for all the filters considered so that the bandwidth is constant. $F_i(u)$ is then written as:

$$F_i(u) = c_i [\exp(-2\pi^2 \sigma_1^2 u^2) - \exp(-2\pi^2 \alpha^2 \sigma_1^2 u^2)] \quad (13)$$

where σ_i stands for the parameter σ_i of Eq. (7). According to the theory presented above, the set $\{f_i(x)\}$ of real functions of a real variable x , having Fourier transforms $\{F_i(u)\}$ is a homothetic filter bank generated by the following profile F :

$$F(u) = \exp(-2\pi^2 u^2) - \exp(-2\pi^2 \alpha^2 u^2) \quad (14)$$

such that:

$$F_i(u) = c_i F(\sigma_i u) \quad (15)$$

Let us recall that these filters are to be used for characterizing filtered images by energy measures. The filters are normalized so that their responses to a white noise are always the same. In this way, none of the filters has a preponderant weight over the others. That normalization leads to the following result:

$$c_i = (\sigma_i)^{1/2} \quad (16)$$

We have found that the value of the bandwidth lies between 0.6 and 2 octaves for:

$$1.0625 \leq \alpha \leq 2 \quad (17)$$

and that the optimal value of α is about 1.5. We choose the value $\alpha = \sqrt{2}$ because it also allows us to implement a parallel vision system, as we shall see later.

Once the frequency filters are created, we build the 2-D separable filters by bringing in orientations which are to be sampled according to the conditions of the signal theory. Then at $\theta_0 = 0$, when the filter axis and the horizontal x -axis are identical, the general shape of the 2-D filter $f_i(x, y)$ in the spatial domain is defined as:

$$f_i(x, y) = (2\pi)^{-1/2} \{ (\sigma_{i1})^{-1} \exp[-x^2/(2\sigma_{i1}^2)] - (\sigma_{i2})^{-1} \exp[-x^2/(2\sigma_{i2}^2)] \} \cdot (2\pi)^{-1/2} \{ (\sigma_{i3})^{-1} \exp[-y^2/(2\sigma_{i3}^2)] \} \quad (18)$$

We have chosen the standard deviations so that, for all values i belonging to the set I : $\sigma_{i2} / \sigma_{i1} = \alpha = \text{constant}$ and $\sigma_{i3} / \sigma_{i1} = \beta = \text{constant}$. The constant β depends on the number of orientations being processed. By giving the value $\sqrt{2}$ to the constant α , a constant spatial-frequency bandwidth can be achieved, its width being slightly above one octave, according to visual system properties and to the signal theory. Moreover, this value $\sqrt{2}$ is the most suitable for computing convolutions, as recalled hereafter. The first gaussian function, with standard deviation σ_{i1} is:

$$g_{\sigma_{i1}, \theta}(x) = (2\pi)^{-1/2} \{ (\sigma_{i1})^{-1} \exp[-x^2/(2\sigma_{i1}^2)] \} \quad (19)$$

The filter can then be expressed as:

$$f_{\sigma_{i1}, \theta}(x) = [g_{\sigma_{i1}, \theta}(x) - g_{\alpha \sigma_{i1}, \theta}(x)] \cdot g_{\beta \sigma_{i1}, \theta}(y) \quad (20)$$

These filters, with an orientation $\theta_0 = 0$, are noted $f_{\theta_0}(x, y)$ or $f_{\alpha i}(x, y)$ indifferently changing expression (20) into:

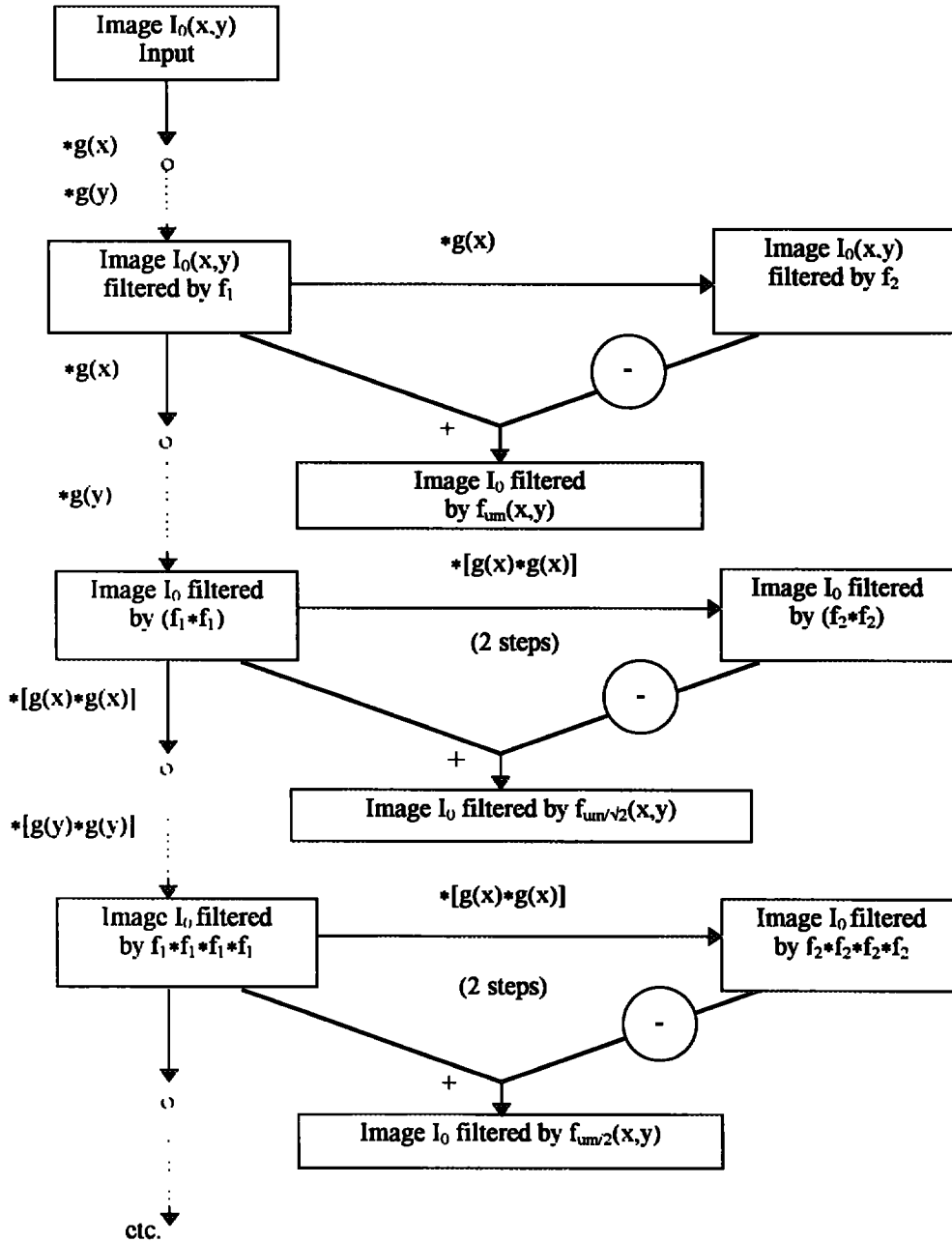
$$f_{\alpha i}(x, y) = [g_{\alpha i}(x) - g_{\alpha \alpha i}(x)] \cdot g_{\beta \alpha i}(y) \quad (21)$$

with:

$$g_{\alpha i}(x) = g_{\alpha i, \theta_0}(x), \quad g_{\alpha \alpha i}(x) = g_{\alpha \alpha i, \theta_0}(x) \quad \text{and} \quad g_{\beta \alpha i}(y) = g_{\beta \alpha i, \theta_0}(y)$$

The filters corresponding to any preferential orientation θ_j are deduced from filters with a zero preferential-orientation by executing a simple rotation of angle θ_j and are written $f_{\alpha i, \theta_j}(x, y)$. Subsequently, the filter $f_{\alpha i}(x, y)$ is written as a difference of two gaussian lowpass-filters:

$$f_{\alpha i}(x, y) = f_{1, \alpha i}(x, y) - f_{2, \alpha i}(x, y) \quad (22)$$



Sketch of cascade filtering

$$\text{with: } f_{1,\alpha}(x, y) = g_{\alpha}(x) \cdot g_{\beta\alpha}(y) \quad (23)$$

$$\text{and: } f_{2,\alpha}(x, y) = g_{\alpha\alpha}(x) \cdot g_{\beta\alpha}(y) \quad (24)$$

Three basic properties of the convolution integral make an interesting implementation of our filtering processes possible. First is the associativeness of the gaussian functions, whether f , g and h are 1-D or 2-D functions:

$$f*(g*h) = (f*g)*h \quad (25)$$

The two-dimensional convolution of separable functions is such that:

$$[f_1(x)g_1(y)]*[f_2(x)g_2(y)] = [f_1(x)*f_2(x)] \cdot [g_1(y)*g_2(y)] \quad (26)$$

The last is, the convolution of a gaussian function by itself is a similar gaussian function with a standard deviation multiplied by $\sqrt{2}$:

$$g_{\alpha}(x) * g_{\alpha}(x) = g_{\alpha\sqrt{2}}(x) \quad (27)$$

From these properties, the possibility of cascade filtering is inferred. Consider expressions (23) and (24) of the filters f_1 and f_2 and let us take into account the filter separableness. A convolution by f_1 (or by f_2) can be made in two steps: a convolution of the image lines by $g_{\alpha}(x)$ followed by a convolution of the image columns by $g_{\beta\alpha}(x)$. Selecting the value $\sqrt{2}$ for the parameter α is very appropriate for we can then write:

$$g_{\alpha\alpha}(x) = g_{\alpha\sqrt{2}}(x) = g_{\alpha}(x) * g_{\alpha}(x). \quad (28)$$

Cascade convolutions can then be executed according to the diagram of the figure. Cascade filtering of the image $I_0(x, y)$ uses relation (28). In this figure, the arrows with a continuous line represent a convolution of image lines, while the arrows with a broken line correspond to a convolution of columns. The notation $g(x)$ represents the function $g_{\alpha\alpha}(x)$ while $g(y) \equiv g_{\beta\alpha}(y)$. All filterings are made in a given direction $\theta_0 = 0$. The highest preferential frequency u_m is inversely proportional to the lowest standard deviation σ_m .

This work has a number of potential applications, especially in medical imagery and in the analysis of satellite images.

References

- Bovik, A. C., Clark, M., and Geisler, W. S., 1987. Computational Texture Analysis using Localized Spatial Filtering. In Proceedings of the IEEE Computer Society Workshop Computer Vision, Miami Beach, Florida.
- Julesz, B., 1975. Experiments in the Visual Perception of Texture. *Scientific American* 232: 34-43.
- Landraud-Lamole, A. M., and Yum-Oh, S., 1995. Texture Segmentation using Local Phase Differences in Gabor Filtered Images. *Lecture Notes in Computer Science, Image Analysis and Processing*. Springer-Verlag 974: 447-452.
- Movshen, J. A., Thompson, I. D., and Tolhurst, D. J., 1978. Spatial and Temporal Contrast Sensitivity of Neurons in Areas 17 and 18 of the Cat's Visual Cortex. *Journal of Physiology* 283: 101-120.
- Neisser, U., 1967. *Cognitive Psychology*. New York: Appleton-Century-Crofts.
- Pollen, D. A., and Ronner, S. F., 1983. Visual Cortical Neurons as Localized Spatial Frequency Filters. *IEEE Transactions on Systems, Man and Cybernetics* SMC-13, 5: 907-916.
- Schiller, P., Finlay, B. And Volman, S., 1976. Quantitative Studies of Single-Cell Properties in Monkey Striate Cortex - II. Orientation Specificity and Ocular Dominance. *Journal of Neurophysiology* 39: 1334-1351.
- Westheimer, G., 1984. Spatial Vision. *Annual Review of Psychology* 35: 201-226.
- Wilson, H. R., and Bergen, J. R., 1979. A Four Mechanism Model for Threshold Spatial Vision. *Vision Research* 19, 1: 19-32.