Gabor Wavelet Transform and Application to Problems in Early vision

B. S. Manjunath
Center for Information Processing Research
Electrical and Computer Engineering Department
University of California, Santa Barbara, CA 93106

Abstract

An important goal of early vision is to transform the intensity data to a representation more suitable for further processing. Examples of such transformations include obtaining depth information from stereo images, shape from shading, image segmentation based on textural and intensity discontinuities, etc. It appears that one of the basic pre-processing steps in achieving these transformations is to detect oriented features at multiple scales. An approach that appears to be promising is the Gabor wavelet transformation of the intensity which results in such a multi-scale oriented feature representation. In this paper we review some recent work on the use of these wavelets in low level vision processing and illustrate some applications to problems such as boundary detection, feature detection and localization, and shape recognition.

1 Introduction

There has been a growing interest in the use of wavelets for multiscale representation of image data. The features of interest in an image are generally present at various scales. An efficient way to analyze such features is to have a multiscale decomposition of the image. Wavelets are families of basis functions generated by dilations and translations of a *basic wavelet*. The wavelet transform is thus a decomposition of the function (image intensity) in terms of these basis functions. One of the objectives of such a transformation is to provide a simultaneous description of the data in the frequency and spatial domains.

Let us first consider the one dimensional case. Let $g(x)$ be a wavelet, $x \in \mathbf{R}$ ($\mathbf{R}$ denotes the set of real numbers). Then the the family of basis functions corresponding to $g(x)$ can be generated by translations ($g(x-s)$) and dilations ($g(\alpha x)$), where s and α are the translation and scale parameters respectively. Let this family be denoted by $(g(\alpha(x-s)))$, $(\alpha, s) \in \mathbf{R}^2$. The wavelet transform of a function $f(x)$ (assuming that $f(x)$ is square integrable) is defined by

$$W_f(\alpha, s) = \int_{-\infty}^{\infty} f(x)\ g^*(\alpha(x-s))\ dx \qquad (1)$$

The (*) indicates complex conjugate. Wavelets can be discretized by a suitable sampling of the parameters α and s. For example we can write the scale parameter as α^j where $j \in \mathbf{Z}$, $\mathbf{Z}$ being the set of integers. This results in a class of discrete wavelets represented by $g(\alpha^j x - n), (j, n) \in \mathbf{Z^2}$. A function $f(x)$ can then be expanded in terms of the basis functions $g(\cdot)$ as

$$f(x) = \sum_{i,j} c_{ij} g(\alpha^j x - i)$$

The Laplacian pyramid [1] mentioned earlier is a wavelet decomposition based on Difference of Gaussian (DOG) wavelet and has found many applications in image processing [2]. Orthogonal wavelets are a special family of discrete wavelets corresponding to $\alpha = 2$, where the basis functions are mutually orthogonal, i.e., $\int g(x)\ g(2^j x - k)dx = 0$ for $((j,k) \in \mathbf{Z^2})$. An important feature of orthogonal wavelets is that the information at different resolutions is uncorrelated. Orthogonality, in general, is a strong condition, and is difficult to achieve if arbitrary orientation selectivity is desired.

This paper reviews some of the recent work on the use of Gabor wavelets, a non-orthogonal family of functions, in the analysis of image data. We begin with a brief introduction to Gabor wavelet transformation in the next section. Sections 3 and 4 deal with some application to problems in vision.

2 Gabor Functions and Wavelets

Gabor functions are Gaussians modulated by complex sinusoids. In its general form, the 2-D Gabor

1058-6393/92 $03.00 © 1992 IEEE

function and its Fourier transform can be written as [3],

$$g(x,y;u_0,v_0) = e^{(-[x^2/2\sigma_x^2+y^2/2\sigma_y^2]+2\pi i[u_0\ x+v_0\ y])} \quad (2)$$

$$G(u,v) = e^{(-2\pi^2(\sigma_x^2(u-u_0)^2+\sigma_y^2(v-v_0)^2))} \quad (3)$$

σ_x and σ_y define the widths of the Gaussian in the spatial domain and (u_0, v_0) is the frequency of the complex sinusoid. A well known property of these functions is that they achieve the minimum possible joint resolution in space and frequency domains [3]. A signal such as a delta function which is concentrated at a point in space has no frequency localization. Likewise, a function concentrated in frequency has no spatial localization. A good measure of localization in the two domains is the product of the bandwidths in space and frequency. The effective bandwidth of a signal is defined as the square root of the variance of the energy of the signal. Let δx and δy be the effective widths of the signal in the horizontal and vertical directions in space respectively and δu, δv denote the corresponding widths in frequency. Then the following inequalities (also called the uncertainty relations) hold: (a) $\delta x\ \delta u \geq 1/4\pi$ and (b) $\delta y\ \delta v \geq 1/4\pi$. Gabor family of functions are unique in attaining the minimum possible value of this joint uncertainty.

The Gabor functions form a complete but non-orthogonal basis set and any given function $f(x,y)$ can be expanded in terms of these basis functions. Such an expansion provides a localized frequency description and has been used in image compression [4] and texture analysis [5]. Local frequency analysis, however, is not suitable for feature representation as it requires a fixed window width in space and consequently the frequency bandwidth is constant on a linear scale. In order to optimally detect and localize features at various scales, filters with varying support rather than a fixed one are required. This would suggest a transformation similar to wavelet decomposition rather than a local Fourier transform. We now consider such a wavelet transform where the *basic wavelet* is a Gabor function of the form

$$g_\lambda(x,y,\theta) = e^{-(\lambda^2 x'^2+y'^2)+i\pi x'} \quad (4)$$

$$x' = x\cos\theta + y\sin\theta$$

$$y' = -x\sin\theta + y\cos\theta$$

where λ is the spatial aspect ratio, θ is the preferred orientation. To simplify the notation, we drop the subscript λ and unless otherwise stated assume that $\lambda = 1$. For practical applications, discretization of the parameters is necessary. The discretized parameters must cover the entire frequency spectrum of interest. Let the orientation range $[0, \pi]$ be discretized into N intervals and the scale parameter α be sampled exponentially as α^j. This results in the wavelet family

$$(g(\alpha^j(x-x_0, y-y_0), \theta_k)), \alpha \in \mathbf{R},\ j = \{0,-1,-2,..\} \quad (5)$$

where $\theta_k = k\pi/N$. The Gabor wavelet transform is then defined by

$$W_j(x,y,\theta) = \int f(x_1,y_1)\, g^*(\alpha^j(x_1-x, y_1-y), \theta)\, dx_1 dy_1 \quad (6)$$

3 Boundary Detection

In [6] a detailed discussion on the use of Gabor wavelets in localizing boundary information is provided. The even and odd symmetric filters respond maximally to line and step edges, respectively, in the image. These edges can be detected at the local maxima in the energy of the filtered outputs, where energy is the square root of the sum of the squares of the responses of the even and odd symmetric filters corresponding to the complex Gabor function. In the following discussion $I_{s,\theta}$ refers to the energy feature corresponding to the spatial location s and orientation θ. Local competitive interactions between these simple energy features help in localizing the texture boundary information. These interactions include competition between different orientations at each spatial location as well as competition between neighboring features within each orientation channel. Suppose the output of a cell at position s in a given frequency channel with a preferred orientation θ is denoted by $Y_{s,\theta}$, with $I_{s,\theta}$ being the excitatory input. $I_(s,\theta)$ could be the energy in the filter output corresponding to feature (s,θ) in a given frequency channel. Let N_s be the local spatial neighborhood of $s = (x,y)$. The competitive dynamics is then represented by

$$\dot{X}_{s,\theta} = -a_{s,\theta}X_{s,\theta} + I_{s,\theta} - \sum_{s'\in N_s} b_{s,s'}Y_{s',\theta} - \sum_{\theta'\neq\theta} c_{\theta,\theta'}Y_{s,\theta'} \quad (7)$$

$$Y_{s,\theta} = g(X_{s,\theta}) \quad (8)$$

and (a,b,c) are positive constants. In our experiments we have used a sigmoid non-linearity of the form $g(x) = 1/(1+exp(-\beta x))$. We assume that all these interactions are symmetric ($b_{s,s'} = b_{s',s}$ and $c_{\theta,\theta'} = c_{\theta',\theta}$). The competitive dynamics of the above system can be shown to be stable. The specific form of

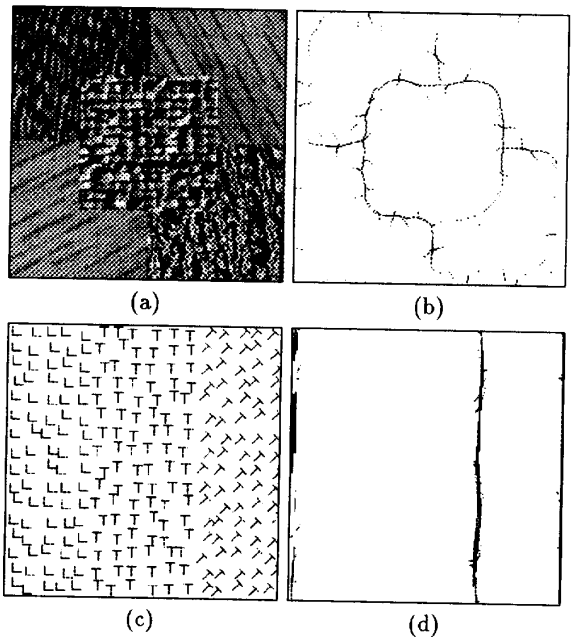

Figure 1: (a) Image consisting of four natural textures, water, wood, raffia and grass, (b) texture boundary detected, (c) texture consisting of three regions, L, T and tilted-Ts. The boundary between L and T s can not be easily detected. However the orientation difference between the two T regions is enough to discriminate between the two regions in almost all frequency channels, and (d) boundary detected is superimposed on the original. For both examples the results shown correspond to the combined output from channels $\alpha = \{1/2, 1/2\sqrt{2}, 1/4\}$ and $\sigma = 5$ pixels.

the dynamics such as the one in (7) is not very critical, as long as there is some form of local inhibition to suppress weak responses. Following these competitive interactions, a grouping process integrates information from similar features and completes the boundaries. We show some results of texture boundary detection in Figure 1.

4 Feature Detection and Localization

We now suggest a simple mechanism to model the behavior of end-inhibition, a property exhibited by certain cells in the visual cortex, using a local scale interaction model following the Gabor wavelet transformation. If $Q_{ij}(x, y, \theta)$ denotes the output of the end-inhibited cell at position (x, y) receiving inputs from two frequency channels i and j ($\alpha^i > \alpha^j$) with preferred orientation θ, then

$$Q_{ij}(x, y, \theta) = g(W_i(x, y, \theta) - \gamma W_j(x, y, \theta)) \qquad (9)$$

where $\gamma = \alpha^{-2(i-j)}$ is the normalizing factor, and $g(\cdot)$ a sigmoid non-linearity. The features represented by equation 9 typically correspond to sharp changes in the local curvature. Locations (x, y) in the image which are identified as feature locations satisfy the following:

$$Q_{ij}(x, y) = \max_{(x', y') \in N_{xy}} Q_{ij}(x', y') \qquad (10)$$

where

$$Q_{ij}(x', y') = \max_{\theta} Q_{ij}(x', y', \theta)$$

and $Q_{ij}(x', y', \theta)$ is given by (9). N_{xy} represents a local neighborhood of (x, y) within which the search is conducted. This feature detection scheme has been successfully applied to some practical problems, and in the following we briefly describe these applications.

4.1 Application to Motion Correspondence and Image Registration

The first example is illustrated in Figure 2. The goal is to detect and track features, and to compute motion and structure parameters. Feature points are first detected using the scheme outlined in the previous section. Some of these points are then tracked over successive frames. For tracking, the information contained in the Gabor wavelet transformation is used, and the correspondence problem is integrated with that of motion estimation. For details we refer to [7, 8]. Another application where this feature detection scheme has been quite effective is in obtaining correspondence in aerial images [9]. These aerial images do not contain any easily perceivable structures which can be detected and used in obtaining correspondence. Central to this is solving the correspondence problem. Figure 3 shows two images in the sequence and the features detected. The key to the matching problem is to identify a consistent set of feature locations in successive frames and the use of shading information to estimate the rotation between the images. Once rotation is estimated, an area based correlation around the feature points is used in obtaining the correspondence, details of which can be found in [9].

4.2 Application to Face Recognition

Here we discuss a potential application to the problem of face recognition (for details see [10]). After the

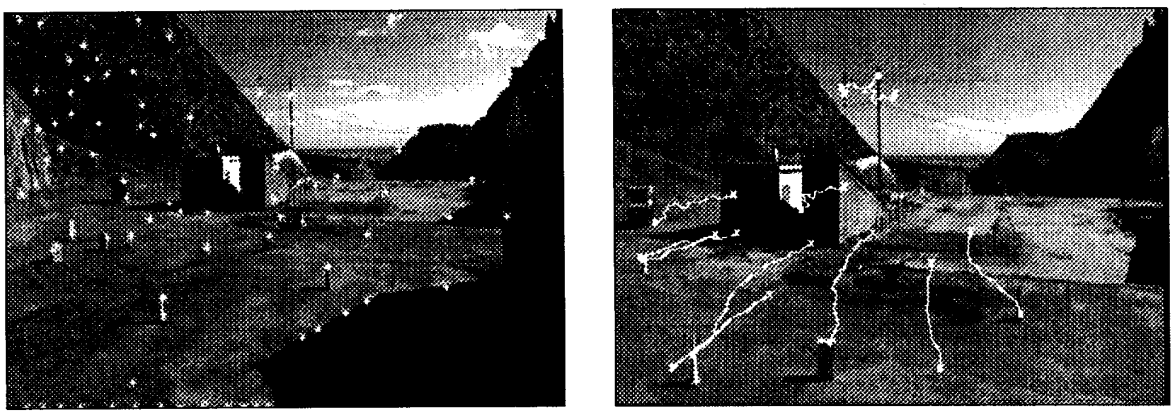

Figure 2: Tracking results for the Rocket sequence. (Courtesy: R. Dutta and R. Manmatha, University of Massachusetts at Amherst.)

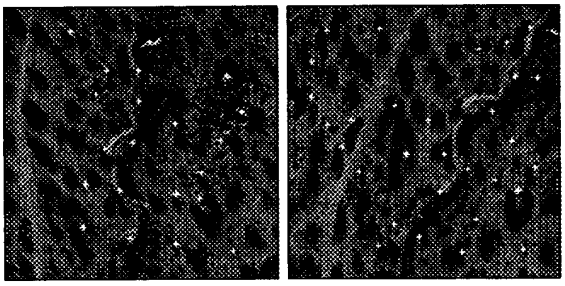

Figure 3: Two successive images from a motion sequence, and feature locations using our model (courtesy: Peter Kroger of JPL).

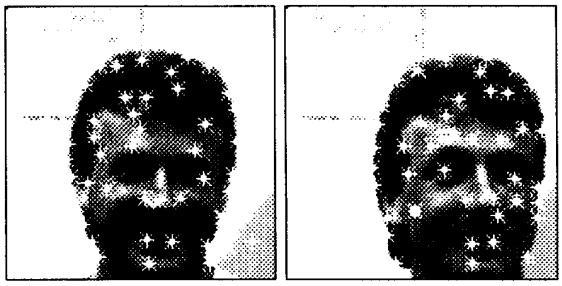

Figure 4: Feature locations marked for a pair of face images. The scales used in this case correspond to $i = -2, j = -5$ ($\alpha = \sqrt{2}$). Information at the feature locations is stored and used during the recognition process.

feature locations are identified in a given image (Figure 4), a representation of the shape information is built using a topological graph data structure. For convenience the features detected in a given image are numbered as {1,2,..} (in any arbitrary, but consistent way). The nodes V_i in the graph correspond to the feature points, and are characterized by $\{S, \mathbf{q}\}$, where S represents information about the spatial location, and

$$\mathbf{q}_i = [Q_i(x, y, \theta_1), \ldots, Q_i(x, y, \theta_N)] \qquad (11)$$

is the feature vector corresponding to the ith feature. Let N_i denote the set of neighbors of ith node. Directional edges connect the neighbors in the graph (i.e., the neighborhood is not symmetric). Neighborhood of a node is determined by taking into account both the maximum number of neighbors allowed as well as the distance between them. The Euclidian distance between two nodes V_i and V_j is denoted by d_{ij}. Such a topological graphs forms a representation of shape information in the database. When a given face image is to be identified, the steps involve: (a) Computing the Gabor wavelet transformation, and localizing the feature information using the scale interaction model, (b) Building a graph representation of the features so detected, encoding both the feature as well as topological information, and (c) Performing a sequential search over the database, matching each of the stored graphs with the input graph.

We have implemented a simple face recognition system based on the above principles. The input is

128×128 image, having very little background noise. In our current implementation, the feature responses are computed at only one scale, corresponding to the scale parameters $\alpha = \sqrt{2}$, $i = -2, j = -5$ in (9). Typical number of feature points detected in a face image using (10) vary from 35 to 50. Number of discrete orientations used was $N = 4$ corresponding to $\theta = \{0, 45, 90, 135\}$. One byte of information is stored for each of the components in the feature vector, or approximately about 200 bytes of information per face. This constitutes an order of magnitude savings in memory, from a 16K raw intensity data. The database we have used has face images of 86 persons, with two to four images per person, taken with different facial expressions, and/or orientations. Often there is a small amount of translation and scaling as well. The recognition accuracy (percentage of successful identifications) of the system is about 85%.

5 Discussion

In the past ten years considerable advances have been made in modeling the receptive field profiles of single cells in the visual cortex of mammals. One of the promising approaches is the use of Gabor functions to model these receptive fields, and the Gabor wavelet decomposition provides a good description of the first stage in the processing of intensity data. In this paper we have reviewed some of the recent work based on such a wavelet transformation of the intensity data. We discussed how textural and intensity discontinuities could be identified by considering local competitive interactions between features. A model for identifying salient image features is presented, and is based on local scale interactions. Several applications of this feature detections scheme are also provided.

Acknowledgements

This work is supported in part by a grant from the UCSB academic senate.

References

[1] P. J. Burt and E. H. Adelson, "The laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. COM-31, pp. 532–540, April 1983.

[2] P. J. Burt, "The pyramid as a structure for efficient computation," in *Multiresolution image processing and analysis* (A. Rosenfeld, ed.), pp. 6–35, Springer-Verlag, 1984.

[3] J. G. Daugman, "Uncertainty relations for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *J. Opt. Soc. Am. A*, vol. 2, no. 7, pp. 1160–1169, 1985.

[4] J. G. Daugman, "Relaxation neural network for non-orthogonal image transforms," in *Proc. Int. Conf. on Neural Networks*, vol. 1, (San Diego, CA), pp. 547–560, June 1988.

[5] A. C. Bovik, M. Clark, and W. S. Geisler, "Multichannel texture analysis using localized spatial filters," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-12, pp. 55–73, January 1990.

[6] B. S. Manjunath and R. Chellappa, "A computational model for boundary detection," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (Maui, Hawaii), pp. 358–363, June 1991.

[7] B. S. Manjunath, C. Shekar, R. Chellappa, and C. der Malsburg, "A robust method for detecting image features with application to face recognition and motion correspondence," in *Proc. International Conference on Pattern Recognition*, September 1992.

[8] S. Chandrashekhar and R. Chellappa, "Passive navigation in a partially known environment," in *IEEE Workshop on Visual Motion*, (Princeton, NJ), pp. 2–7, October 1991.

[9] Q. Zheng, R. Chellappa, and B. S. Manjunath, "Balloon motion estimation using two frames," in *Proc. 25th Asilomar Conf. on Systems and Signals*, November 1991. (invited paper).

[10] B. S. Manjunath and R. Chellappa, "A feature based approach to face recognition," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (Champaign, Illinois, June 1992), pp. 373–378, June 1992.