Hierarchical Gabor Filters for Object Detection in Infrared Images

R. Neil Braithwaite and Bir Bhanu
College of Engineering
University of California
Riverside, CA 92521

Abstract

This paper presents a new representation called "hierarchical Gabor filters" and associated novel local measures which are used to detect potential objects of interest in images. The "first stage" of the approach uses a wavelet set of wide-bandwidth separable Gabor filters to extract local measures from an image. The "second stage" makes certain spatial groupings explicit by creating small-bandwidth, non-separable Gabor filters that are tuned to elongated contours or periodic patterns. The non-separable filter responses are obtained from a weighted combination of the separable basis filters, which preserves the computational efficiency of separable filters while providing the distinctiveness required to discriminate objects from clutter. This technique is demonstrated on images obtained from a forward looking infrared (FLIR) sensor.

1 Introduction

Automatic object recognition in the image understanding context attempts to find instances of modeled objects within an image. The steps leading up to recognition include the *detection* of interesting image features, the *indexing* of object models, and the *verification* of the hypothesized model-image correspondences. This paper focuses on the image processing used for object detection, while considering the requirements for model indexing. The proposed image processing approach uses hierarchical, multi-scale Gabor filters that are suitable for object detection in coarse-resolution infrared images.

Detection in images is complicated by the imaging process which involves a viewpoint-dependent 2D projection of an object. As a consequence, the appearance of an object in an image can vary greatly with its *aspect* and *scale*. The aspect is the rotational position of the object relative to the viewing direction; the scale is the object's image size which varies with the object's range from the viewer. Detection in infrared imagery is further complicated by coarse resolution and a multi-modal behavior that highlights only certain parts of an object [6]. The approach used in this paper is less sensitive to these effects because detection is based on distinctive local features.

This paper uses the Gabor representation for object detection. The Gabor representation of an image has many degrees of freedom that can be adjusted to highlight interesting local intensity patterns. A small number of wide-bandwidth separable filters, referred to as *basis filters*, are used for detection. For discriminating between object and clutter, small-bandwidth filters are created that are tuned to specific intensity patterns. The small-bandwidth filters are implemented as *hierarchical filters* where the bulk of the processing has already been performed by the basis filters. The tuning of the filters is driven by local measures. The approach adjusts with the specific image information instead of using a fixed trade-off between detection and discrimination.

The outline of this paper is as follows. Section 2 discusses the image processing performed by separable wavelet Gabor filters and the resulting local measures used to identify significant image features. Section 3 describes how non-separable filter responses are created from the output of the basis filters. Section 4 demonstrates this approach on infrared imagery.

2 Image Processing

The filter kernels used in this paper are 2D Gabor functions which are oriented sine-wave gratings that are spatially attenuated by an elliptical Gaussian window. The general form of the Gabor function $G_i(x,y)$ is given by [3] [5]

$$G_i(x,y) = \exp\left\{-\frac{1}{4\pi\sigma^2}\left[\acute{x}^2+\alpha^{-2}\acute{y}^2\right]\right\} \times \exp\left\{j\omega_i[x\cos\phi_i + y\sin\phi_i]\right\}, \quad (1)$$

$$\begin{bmatrix}\acute{x}\\ \acute{y}\end{bmatrix} = \begin{bmatrix}\cos\phi_G & \sin\phi_G\\ -\sin\phi_G & \cos\phi_G\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix}, \quad (2)$$

where x and y are the horizontal and vertical image coordinates, respectively; $\acute{x}$ and $\acute{y}$ denote the axes that are rotated by ϕ_G; ω_i and ϕ_i are the modulation frequency and orientation, respectively; and σ and α are the scale and the eccentricity of the elliptical Gaussian window, respectively. The parameters σ, α, ϕ_G, ω_i, and ϕ_i define the passband of the Gabor channel and the spectral characteristics of the Gabor kernel.

Scale and *aspect distortions* represent the image transformation undergone by the 2D projection of the object when it is viewed from different camera positions. Both scale and aspect distortions cause shifts in the spectral characteristics of an image pattern. For the case of scale distortion, the frequency shift is omnidirectional; for aspect distortion, the frequency shift is

along the foreshortened axis. Note that, for a Gabor function, the ratio of the channel bandwidth in the modulation direction and the modulation frequency is a scale/aspect invariant parameter because the "distortion transformation" affects the modulation bandwidth and frequency equally. This ratio is referred to as the "fractional bandwidth" and is given by

$$\lambda = \frac{\Delta\omega_i}{\omega_i}, \tag{3}$$

where $\Delta\omega_i$ is the channel bandwidth and ω_i is the channel modulation frequency. Note that $\Delta\omega_i = \sigma^{-1}$ when $\phi_i = \phi_G$.

When a viewed object comprises an image feature with a low fractional bandwidth or high eccentricity, the object can be detected reliably by a lone feature. During initial detection, when the appropriate values for α, ϕ_G, ω_i, and ϕ_i are unknown, a set of well-spaced Gabor channels with large fractional bandwidths are used. Large bandwidths allow for large scale or aspect mismatches, thereby simplifying object detection. After detection, a single small-bandwidth kernel, with spectral characteristics (α, ϕ_G, ω_i, and ϕ_i) tuned to the image intensity pattern, is then used to discriminate narrow-bandwidth object features from background clutter.

In this paper, the basis filter set used for initial object detection is comprised of separable wavelet Gabor kernels. Kernel separability reduces the computational requirements of the filtering. The separable wavelet restriction requires that λ is constant and $\alpha = 1$ for each kernel. The quadrature Gabor filter kernels [3] in a separable wavelet representation are given by

$$\begin{aligned} G_+(x,y,\omega_i,\phi_i) &= g(x,y)\cos[\omega_i(x\cos\phi_i + y\sin\phi_i)], \\ G_-(x,y,\omega_i,\phi_i) &= g(x,y)\sin[\omega_i(x\cos\phi_i + y\sin\phi_i)], \\ g(x,y) &= \exp\left(-\frac{\lambda^2\omega_i(x^2+y^2)}{4\pi}\right). \end{aligned} \tag{4}$$

The resulting Gabor channels are characterized by two responses: (a) a magnitude response that measures localized signal energy, and (b) a phase response that encodes the relational structure of an intensity pattern with respect to its spatial neighborhood. The magnitude response is useful for identifying significant features. The spatial phase gradient is used to estimate local spectral characteristics.

The spatially sampled output of the Gabor filter is referred to as a *Gabor coefficient*, and is given by

$$a_\pm(i) = \iint I(x,y)G_\pm(x_i - x, y_i - y, \omega_i, \phi_i)dx\,dy, \tag{5}$$

where $I(x,y)$ is the input image. The local magnitude m and the phase θ, obtained from the quadrature Gabor coefficients, are given by

$$m(x_i,y_i,\omega_i,\phi_i) = \sqrt{a_+^2(i) + a_-^2(i)}, \tag{6}$$

$$\theta(x_i,y_i,\omega_i,\phi_i) = \arctan\left[\frac{a_-(i)}{a_+(i)}\right]. \tag{7}$$

The spatial phase gradient, $[\omega_x,\omega_y] = [\partial\theta/\partial x, \partial\theta/\partial y]$, is used to estimate the *mean frequency* ω_n and *mean orientation* ϕ_n of the signal energy within a basis channel [4]:

$$\omega_n(x_i,y_i,\omega_i,\phi_i) = \sqrt{\omega_x^2(i) + \omega_y^2(i)}, \tag{8}$$

$$\phi_n(x_i,y_i,\omega_i,\phi_i) = \arctan\left[\frac{\omega_y(i)}{\omega_x(i)}\right]. \tag{9}$$

Multi-channel "local measures" can be obtained by combining the magnitude responses from a composite passband comprising basis channels with different orientations, but a common modulation frequency. These local measures are used for feature detection and channel selection. A *marginal magnitude* $\bar{m}$ is obtained by summing the magnitude responses within the composite passband: $\bar{m}(x_i,y_i,\omega_i) = \sum_{\phi_i} m(x_i,y_i,\omega_i,\phi_i)$. The *dominant spectral orientation* and variance are defined by the moment of inertia of the spectral energy within the composite passband. The dominant spectral orientation, ϕ_d, is the axis which produces the minimum moment of inertia:

$$\phi_d(x_i,y_i,\omega_i) = \frac{1}{2}\arctan\left\{\frac{\sum_{\phi_i} m(\phi_i)\sin[2\phi_n(\phi_i)]}{\sum_{\phi_i} m(\phi_i)\cos[2\phi_n(\phi_i)]}\right\}, \tag{10}$$

The variance in the orientation is defined as the *normalized minimum moment of inertia*, J:

$$J(x_i,y_i,\omega_i) = \frac{\sum_{\phi_i} m(\phi_i)\sin^2[\phi_n(\phi_i) - \phi_d(x_i,y_i,\omega_i)]}{\sum_{\phi_i} m(x_i,y_i,\omega_i,\phi_i)}. \tag{11}$$

The marginal magnitude measures local signal energy, highlighting image features such as contours, periodic patterns, and corners. The normalized minimum moment of inertia is used to discriminate between directional patterns ($J \approx 0$) and patterns comprising orthogonal, or no dominant, orientations ($J \approx 0.5$). When a directional pattern is detected, subsequent stages, such as the filter tuning discussed in Section 3, can be restricted to the basis channel containing the dominant orientation. It is important to note that (10) produces *two* orientations within the interval from $-\frac{\pi}{2}$ to $\frac{\pi}{2}$ that correspond to the minimum and maximum moments of inertia.

3 Creating New Tuned Filters

This section discusses a hierarchical filtering approach that can create a non-separable, small bandwidth filter response using a weighted combination of the outputs from the separable, wide-bandwidth basis filters. Since the bulk of the image processing is performed by the basis filters, the computational advantage of separable filtering is preserved. At the same time, the improved discrimination associated with smaller bandwidth kernels is obtained.

The initial stage of the tunable filter approach attempts to find regions in the basis output that exhibit

a Gaussian-shaped local magnitude response and a coherent phase response (constant phase gradient). Periodic patterns and elongated contours are identified by regions with elliptical magnitude responses and no rapid changes in the phase gradient (mean frequency and orientation). The orientation of the ellipse's major axis (ϕ_G) and the channel modulation (ϕ_l) are aligned for periodic patterns and orthogonal for elongated contours. Once these image regions are identified, the window shape and modulation parameters are extracted. If λ or α of the image feature is significant (i.e. $\lambda^{-2} \cdot \alpha$ is large), a sine-cosine pair of tuned Gabor kernels is formed.

The synthesis of new filters is performed using a variation of the Gabor expansion [1] [2]. The Gabor expansion represents an image pattern $P(x, y)$ by a weighted sum of basis functions G_{nmkl}:

$$P(x,y) = \sum_{nmkl} c_{nmkl} G_{nmkl}(x - x_n, y - y_m, \omega_k, \phi_l), \tag{12}$$

where c_{nmkl} is an expansion coefficient and x_n and y_m are spatial locations in the expansion lattice. In the hierarchical filter approach, the pattern of interest is the impulse response of the new filter. Since the new kernel $P(x, y)$ has a smaller fractional bandwidth than the basis kernel, it is possible to synthesize $P(x, y)$ by considering one basis channel only: the basis channel whose passband most overlaps with the desired kernel's passband. Instead of forming a new kernel $P(x, y)$ and then re-filtering the original image $I(x, y)$, it is possible to create the same filter response by a weighted summation of the basis filter outputs:

$$\iint I(x,y) P(x_i - x, y_i - y) dx\, dy = \sum_{y_m} \sum_{x_n} c_{nm} a_{nm} \tag{13}$$

where

$$a_{nm} = \iint I(x,y) G_{nmkl}(x_n - x, y_m - y) dx\, dy. \tag{14}$$

After selecting the appropriate basis channel (ω_k and ϕ_l), it is necessary to select the spatial lattice points (i.e., the set of x_n and y_m) to be used in the filter synthesis. The spatial lattice points are selected as a rectangular grid centered on the spatial position of the desired filter kernel (x_i, y_i). The spacing of the lattice points is much larger than the pixel spacing, but no larger than $(2\Delta\omega)^{-1}$, where $\Delta\omega$ is the bandwidth of the basis kernel. A reasonable choice for the lattice spacing, for basis kernels with $\lambda = 0.25\pi$, is $\delta x = \delta y = \pi(\omega)^{-1}$. Note that the number of lattice points must be increased in each of the x and y directions inversely with the bandwidth of the desired filter.

The inner product of the desired filter kernel with a selected basis function is given by

$$b_i = \iint P(x,y) G_i^*(x_i - x, y_i - y)\, dx\, dy. \tag{15}$$

The relationship between the expansion coefficients and b_i is given by $\bar{\mathbf{b}} = \mathbf{Q}\bar{\mathbf{c}}$ where $\bar{\mathbf{b}} = [b_0 \ldots b_{N-1}]^T$ and $\bar{\mathbf{c}} = [c_0 \ldots c_{N-1}]^T$. The elements of the matrix $\mathbf{Q}$ are the overlaps between pairs of basis functions. For two basis functions belonging to a common Gabor channel, but separated in x and y, the pair-wise overlap is given by

$$\begin{aligned} q_{i,j} &= \exp\left\{-\frac{\lambda^2\omega^2}{8\pi}\left[(x_i - x_j)^2 + (y_i - y_j)^2\right]\right\} \\ &\times \exp\left\{j\omega\left[(x_i - x_j)\cos\phi_i + (y_i - y_j)\sin\phi_i\right]\right\}. \end{aligned} \tag{16}$$

To ensure accuracy in the filter synthesis, the spatial lattice is slightly over-sampled with respect to the minimum sampling interval, $(2\Delta\omega)^{-1}$. As a result, $\mathbf{Q}$ does not have full rank. One solution to the Gabor expansion is given by $\bar{\mathbf{c}} = [\mathbf{Q} + r\mathbf{I}]^{-1}\bar{\mathbf{b}}$, where r is a regularization constant, and $\mathbf{I}$ is the identity matrix. Note that the solution to $[\mathbf{Q} + r\mathbf{I}]^{-1}$ can be calculated off-line, because it is only dependent on the number of lattice points, the lattice spacing, and the basis kernels, but not the new kernel.

To synthesize the new filter response, $\bar{\mathbf{b}}$ is required. This overlap between the basis functions and an elliptical, non-separable Gabor kernel is given by

$$\begin{aligned} b_i &= \exp\left\{-\frac{\lambda^2\omega^2}{4\pi}\left[\frac{\sigma_n^{-2}(\delta\acute{x})^2}{\lambda^2\omega^2 + \sigma_n^{-2}} + \frac{\alpha_n^{-2}\sigma_n^{-2}(\delta\acute{y})^2}{\lambda^2\omega^2 + \alpha_n^{-2}\sigma_n^{-2}}\right]\right\} \\ &\times \exp\left\{-\pi\frac{\left[\omega_{c(i)} - \omega_{c(n)}\right]^2}{\lambda^2\omega^2 + \sigma_n^{-2}} - \pi\frac{\left[\omega_{s(i)} - \omega_{s(n)}\right]^2}{\lambda^2\omega^2 + \alpha_n^{-2}\sigma_n^{-2}}\right\} \\ &\times \exp\left\{j(\delta\acute{x})\left[\frac{\sigma_n^{-2}\omega_{c(i)} + \lambda^2\omega^2\omega_{c(n)}}{\lambda^2\omega^2 + \sigma_n^{-2}}\right]\right\} \\ &\times \exp\left\{j(\delta\acute{y})\left[\frac{\alpha_n^{-2}\sigma_n^{-2}\omega_{s(i)} + \lambda^2\omega^2\omega_{s(n)}}{\lambda^2\omega^2 + \alpha_n^{-2}\sigma_n^{-2}}\right]\right\} \end{aligned} \tag{17}$$

where $\omega_{c(i)} = \omega\cos(\phi_i - \phi_G)$, $\omega_{s(i)} = \omega\sin(\phi_i - \phi_G)$, $\omega_{c(n)} = \omega_n\cos(\phi_n - \phi_G)$, $\omega_{s(n)} = \omega_n\sin(\phi_n - \phi_G)$,

$$\begin{bmatrix} \delta\acute{x} \\ \delta\acute{y} \end{bmatrix} = \begin{bmatrix} \cos\phi_G & \sin\phi_G \\ -\sin\phi_G & \cos\phi_G \end{bmatrix} \begin{bmatrix} x_i - x_n \\ y_i - y_n \end{bmatrix}, \tag{18}$$

and α_n, σ_n, ω_n, and ϕ_n are the eccentricity, scale (inverse of bandwidth), modulation frequency, and (modulation) orientation, respectively, of the new filter.

4 Results

This section demonstrates the use of local measures and the tuning of filters for detecting interesting features. Figure 1 shows a forward-looking infrared (FLIR) image of a tank seen at an oblique view. The filter will be tuned to the periodic pattern corresponding to the row of guide wheels. Figure 1 also shows the results of combining the four basis orientations into local measures: the marginal magnitude, the peak region of the marginal magnitude, and the dominant spectral orientation. Since a dominant orientation exists in the peak region, the subsequent analysis can be restricted to a single channel.

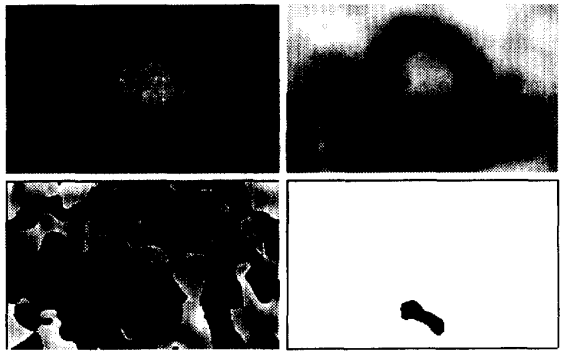

Figure 1: (*upper-left*) FLIR image of tank. (*upper-right*) Marginal magnitude. (*lower-left*) Dominant spectral orientation. (*lower-right*) Peak of marginal magnitude. Dark (light) values indicate high (low) values of local measures.

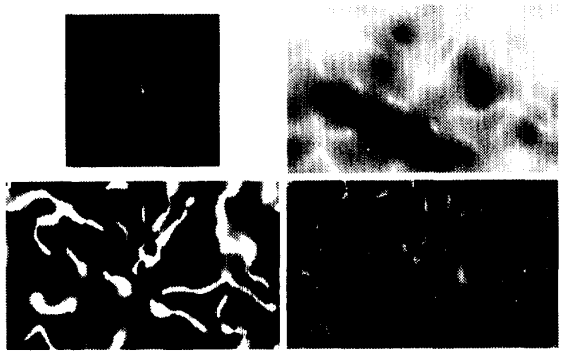

Figure 2: Tank responses for channel $\phi_1 = 0$: (*upper-left*) Basis kernel. (*upper-right*) Magnitude response. Dark (light) values have high (low) local magnitudes. (*lower-left*) Mean frequency. (*lower-right*) Mean orientation. Grey values in the lower images represent measurements within the channel bandwidth. Black (white) values denote measurements above (below) the channel passband.

Figure 2 shows the magnitude, mean frequency, and mean orientation of the tank for the selected basis channel ($\phi_1 = 0$). Note that the row of wheels appears as a region with a large magnitude response and a nearly constant frequency. The mean frequency in this region is 1.16 times the channel modulation frequency, and the mean orientation is -0.04 radians.

The impulse response of the tuned filter and the resulting magnitude response of the filtered tank image appear in Figure 3. The eccentricity of the elliptical window is $\alpha = 0.5$, the bandwidth is half the basis value, and the orientation of its major axis is $\phi_G = -0.44$ radians from horizontal. By comparing Figures 2 and 3, it is clear that the tuned magnitude response displays improved detection of the periodic pattern and suppression of secondary responses.

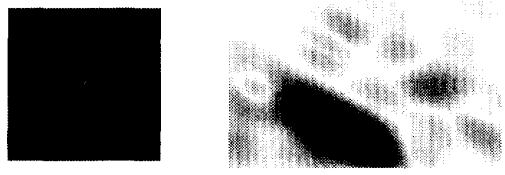

Figure 3: (*left*) Tuned filter kernel and (*right*) tuned magnitude response of the tank. Dark (light) values in the right figure denote high (low) magnitudes.

5 Conclusion

This paper has presented a hierarchical Gabor filtering approach for object detection in FLIR images. Hierarchical filtering allows one to build non-separable filter responses from a spatial combination of separable basis filter outputs, which results in significant computational savings. This paper has also presented local measures based on magnitude, phase, and frequency, which are used to extract important image features and provide information for tuning the non-separable filters to interesting image patterns. It allows the tasks of feature detection and feature-based model indexing to be decoupled, thereby avoiding the trade-off between the flexibility required for scale/aspect insensitive detection and the distinctiveness required for unique indexing and for discriminating objects from clutter.

Acknowledgements

The authors wish to thank Andy Forslund for programming support. This work is supported by ARPA grant MDA972-93-1-0010. The content of the information does not necessarily reflect the position or the policy of the Government.

References

[1] M. J. Bastiaans, "Gabor's expansion of a signal into Gaussian elementary signals," *Proc. IEEE*, vol. 68, pp. 538–539, 1980.

[2] R. N. Braithwaite and M. P. Beddoes, "Iterative methods for solving the Gabor expansion: considerations of convergence," *IEEE Trans. Image Processing*, vol. 1, no. 2, pp. 243–244, 1992.

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

[4] D. J. Fleet, A. Jepson, and M. Jenkin, "Phase-based disparity measurement," *CVGIP: Image Understanding*, vol. 53, no. 2, pp. 198–210, 1991.

[5] D. Gabor, "Theory of communication," *J. Inst. Elec. Eng.*, vol. 93, pp. 429–457, 1946.

[6] J. D. Wald, D. B. Krig, and T. DePersia, "ATR: problems and possibilities for the IU community," DARPA IU Workshop, pp. 255–264, 1992.