Design of Steerable Filters for Feature Detection Using Canny-Like Criteria

Mathews Jacob, *Member*, *IEEE*, and Michael Unser, *Fellow*, *IEEE*

Abstract—We propose a general approach for the design of 2D feature detectors from a class of steerable functions based on the optimization of a Canny-like criterion. In contrast with previous computational designs, our approach is truly 2D and provides filters that have closed-form expressions. It also yields operators that have a better orientation selectivity than the classical gradient or Hessian-based detectors. We illustrate the method with the design of operators for edge and ridge detection. We present some experimental results that demonstrate the performance improvement of these new feature detectors. We propose computationally efficient local optimization algorithms for the estimation of feature orientation. We also introduce the notion of shape-adaptable feature detection and use it for the detection of image corners.

Index Terms—Steerable, feature, edge, detection, ridge, contours, boundary, lines.

1 Introduction

In his seminal paper on computational edge detection, Canny identified the desirable qualities of a feature detector and proposed an appropriate optimality criterion. Based on this criterion, he developed a general approach to derive the optimal detector for specific image features such as edges [1]. This work had a great impact on the field and stimulated further developments in this area, particularly on alternate optimality criteria and design strategies [2], [3].

All of the above authors considered the derivation of optimal 1D operators. For 2D images, they applied the optimal 1D operator orthogonal to the feature boundary while smoothing in the perpendicular direction (along the boundary). This extension is equivalent to computing inner-products between the image and a series of rotated versions of a 2D reference template (tensor product of the optimal 1D profile and the smoothing kernel). With this detector, the rotation angle of the template that yields the maximum inner product, gives the feature orientation. Since the optimal 1D template did not have explicit formulae, they were typically approximated by simple first or second order differentials of a Gaussian. In practice, they were extended using Gaussian kernels of the same variance since the resulting 2D template could be applied in a directional manner inexpensively via the computation of smoothed image gradients or Hessians.

An alternative to these differential approaches to rotation independent feature detection is provided by the elegant work of Freeman and Adelson on steerable filters [4]. The underlying principle is to generate the rotated version of a filter from a suitable linear combination of basis filters; this sets some angular bandlimiting constraints on the class of admissible filters. Perona et. al., Manduchi et al., Simoncelli and Farid, and Teo and Hel-Or used this framework to approximate and design orientation-selective feature detectors [5], [6], [7], [8]. The concept of steerablity was also applied successfully in other areas of image processing such as texture analysis [9], [10] and image denoising [11].

In this paper, we propose to reconcile the two methodologies—computational approach and steerable filterbanks—by presenting a general strategy for the design of 2D steerable feature detectors. We derive the filter directly in 2D as opposed to the 1D schemes (1D optimization followed by an extension to 2D) of Canny and others. Moreover, in contrast with the work of Perona [5], we do not approximate a given template within a steerable solution space, but search for the filter that gives the best response according to an optimality criterion. Our filter is specified so as to provide the best compromise in terms of signal-to-noise ratio, false detections, and localization. We illustrate the method with the design of optimal edge and ridge templates. The detectors that we obtain analytically have better performance and improved orientation selectivity, yet they are still computationally quite attractive.

The paper is organized as follows: In Section 2, we introduce the concept of steerable matched filtering and reinterpret some of the classical detectors within this framework. In Section 3, we propose an optimality criterion and show how to determine the best filter from a class of steerable functions. In Section 4, we concentrate on specific 2D feature detectors and demonstrate their use in different applications. Though our algorithm is general, in this paper, we focus only on the detection of edge and ridge features. In Section 5, we introduce the concept of shape adaptive feature extraction and illustrate it with an example.

• *The authors are with the Biomedical Imaging Group, Swiss Federal Institute of Technology Lausanne, CH-1015, Switzerland. E-mail: michael.unser@epfl.ch, Mathews.Jacob@ieee.org.*

Manuscript received 7 Feb. 2003; revised 20 July 2003; accepted 13 Aug. 2003.
Recommended for acceptance by W. Freeman.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 118259.

2 Orientation Independent Matched Filtering

2.1 Detection by Rotating Matched Filtering

Suppose our task is to detect some feature in an image $f(x, y)$ at some unknown position and orientation. The detection procedure can be formulated as a rotated matched filtering. It involves the computation of inner-products with the shifted

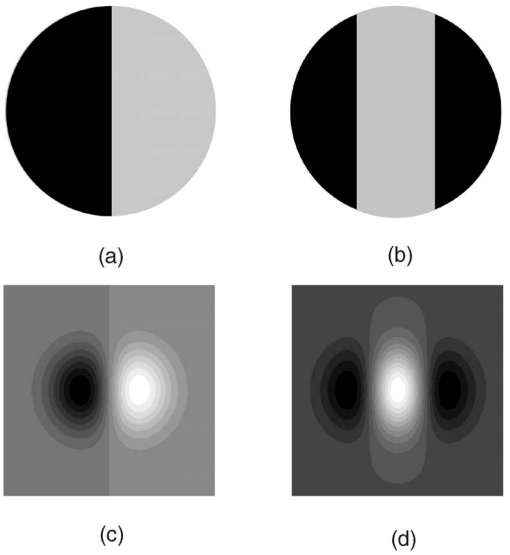


Fig. 1. Examples of feature templates. Feature detection is performed by convolution of the rotated versions of the template with the image. (a) Idealized edge template. (b) Idealized ridge template. (c) Popular edge template. (d) Popular ridge template.

and rotated versions of a 2D feature template $f_0(x,y) = h(-x,-y)$ at every point in the image. A high magnitude of the inner-product indicates the presence of the feature and the angle of the corresponding template gives the orientation. Some simple examples of templates are shown in Fig. 1. Mathematically, the estimation algorithm is

$$\theta^*(\mathbf{x}) = \arg\max_{\theta}(f(\mathbf{x}) * h(\mathbf{R}_\theta\,\mathbf{x})) \tag{1}$$

$$r^*(\mathbf{x}) = f(\mathbf{x}) * h(\mathbf{R}_{\theta^*}\,\mathbf{x}), \tag{2}$$

where r^* is the magnitude of the feature and θ^* its orientation at the position $\mathbf{x} = (x,y)$; $\mathbf{R}_\theta$ is the rotation matrix

$$\mathbf{R}_\theta = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \tag{3}$$

and $u * v$ stands for the convolution between u and v. Equations (1) and (2) correspond to the matched filter detection. They give the maximum-likelihood estimation of the angle θ and weight r for the signal model

$$f(\mathbf{x}) = r \cdot f_0(\mathbf{R}_\theta(\mathbf{x} - \mathbf{x}_0) + \mathbf{x}_0) + n(\mathbf{x}_0),$$

where $n(\mathbf{x})$ denotes Gaussian white noise. However, this scheme of detection is not very practical, for it requires the implementation of a large number of filters (as many as the quantization levels of the angle).

2.2 Steerable Filters

To cut down on the computational load, we select our detector within the class of steerable filters introduced by Freeman and Adelson [4]. These filters can be rotated very efficiently by taking a suitable linear combination of a small number of filters. Specifically, we consider templates of the form

$$h(x,y) = \sum_{k=1}^{M}\sum_{i=0}^{k} \alpha_{k,i}\,\frac{\partial^{k-i}}{\partial x^{k-i}}\frac{\partial^{i}}{\partial y^{i}}\; g(x,y), \tag{4}$$

where $g(x,y)$ is an arbitrary isotropic window function. We call such a $h(x,y)$ an Mth order detector.

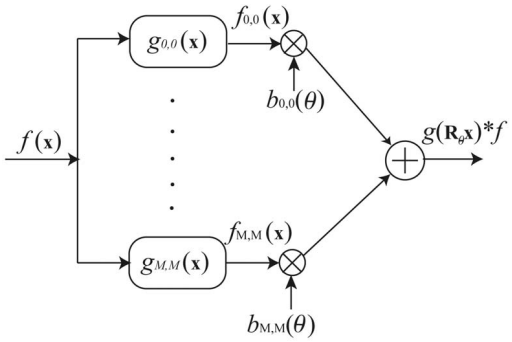


Fig. 2. Implementation of steerable filtering (cf. (5)).

Proposition 1. *The filter $h(x,y)$ is steerable. In other words, the convolution of a signal $f(x,y)$ with any rotated version of $h(x,y)$ can be expressed as*

$$f(\mathbf{x}) * h(\mathbf{R}_\theta\mathbf{x}) = \sum_{k=1}^{M}\sum_{i=0}^{k} b_{k,i}(\theta) f_{k,i}(\mathbf{x}), \tag{5}$$

where the functions $f_{k,i}(x,y)$ are filtered versions of the signal $f(x,y)$

$$f_{k,i}(x,y) = f(x,y) * \underbrace{\left(\frac{\partial^{k-i}}{\partial x^{k-i}}\frac{\partial^{i}}{\partial y^{i}}\; g(x,y)\right)}_{g_{k,i}(x,y)}. \tag{6}$$

The orientation-dependent weights $b_{k,i}(\theta)$ are given by

$$b_{k,i}(\theta) = \left(\sum_{j=0}^{k}\alpha_{k,j}\sum_{l,m\in\mathcal{S}(k,j,i)}\binom{k-j}{l}\binom{j}{m} (-1)^m \cos(\theta)^{j+(l-m)}\sin(\theta)^{(k-j)-(l-m)}\right), \tag{7}$$

where $\mathcal{S}(k,i,j)$ is the set $\mathcal{S}(k,i,j) = \{l,m \mid 0 <= l <= k-i;\ 0 <= m <= i; k-(l+m) = j\}$.

The proof is given in the Appendix A. A graphical representation of the implementation is given in Fig. 2. Once the $f_{k,i}(x,y)$ are available, $f(\mathbf{x}) * h(\mathbf{R}_\theta\mathbf{x})$ can be evaluated very efficiently via a weighted sum with its coefficients that are trigonometric polynomials of θ. Since the number of partial differentials in (5) for a general Mth order template is $M(M+3)/2$, $h(\mathbf{x})$ is steerable in terms of as many individual separable functions. Using some simplification, we can show that such a general $h(\mathbf{x})$ can also be rotated using $2M+1$ nonseparable filters[1] (an example of such a simplification is given by (39)-(42)).

A case of special interest corresponds to $g(\mathbf{x})$ being the Gaussian; indeed, the Gaussian is optimally localized in the sense of the uncertainty principle and the corresponding filters in (6) are all separable. Interestingly, the Gaussian family is equivalent to the class of moment filters (polynomials multiplied by Gaussian window) discussed in [4], but the filters are not identical. We will now show that the family described by (4) includes some popular feature detectors as particular cases.

1. This is the minimum number of filters required to steer a general Mth order tempate.

2.3 Conventional Detectors Revisited

2.3.1 Canny's Edge Detector

As already observed by Freeman and Adelson, the widely-used Canny edge detection algorithm can be reinterpreted in terms of steerable filters [4]. This algorithm involves the computation of the gradient-magnitude of the Gaussian-smoothed image. The direction of the gradient gives the orientation of the edge. Mathematically,

$$\theta^* = \arctan\left(\frac{(f*g)_y}{(f*g)_x}\right) \tag{8}$$

$$r^* = \sqrt{\left((f*g)_x\right)^2 + \left((f*g)_y\right)^2}, \tag{9}$$

where $g_x = \partial g/\partial x$ and $g_y = \partial g/\partial y$; g is a 2D Gaussian of a specified variance. The above set of equations can be shown to be the solution of (1) and (2), with $h = g_x$. Substituting $M = 1;\ \alpha_{1,0} = 1, \alpha_{1,1} = 0$ in (7), we get $b_{1,0}(\theta) = \cos(\theta)$, $b_{1,1}(\theta) = \sin(\theta)$. Thus,

$$\theta^*(\mathbf{x}) = \arg\max_{\theta}(f(\mathbf{x}) * g_x(\mathbf{R}_\theta \mathbf{x})) \tag{10}$$

$$= \arg\max_{\theta}(f * (g_x\cos(\theta) + g_y\sin(\theta))). \tag{11}$$

Here, we used the steerability of g_x from (5). To compute the maximum of the above expression, we set the differential of (11) with respect to θ to zero:

$$(f * g_x)\sin(\theta) - (f * g_y)\cos(\theta) = 0, \tag{12}$$

which results in (8) and (9). The corresponding feature template is shown in Fig. 1c.

2.3.2 Ridge Detector

Less well-known is the fact that a popular ridge estimator based on the eigen-decomposition of the Hessian matrix [12], [13], [14] can also be interpreted in terms of steerable filters. Assuming the template to be g_{xx} (the second derivative of a Gaussian), ridge detection can be formulated exactly as (1) and (2). The corresponding detector is shown in Fig. 1d. In this case, the steerability relation (5) can be expressed in a matrix form as

$$g_{xx}(\mathbf{R}_\theta \mathbf{x}) = \mathbf{u}_\theta^{\mathrm{T}} \underbrace{\begin{bmatrix} g_{xx}(\mathbf{x}) & g_{xy}(\mathbf{x}) \\ g_{xy}(\mathbf{x}) & g_{yy}(\mathbf{x}) \end{bmatrix}}_{\mathbf{H}_g} \mathbf{u}_\theta, \tag{13}$$

where $\mathbf{H}_g$ is the Hessian matrix and $\mathbf{u}_\theta = (\cos(\theta), \sin(\theta))$. Using the linearity of convolution, $f(\mathbf{x}) * g_{xx}(\mathbf{R}_\theta \mathbf{x}) = \mathbf{u}_\theta^{\mathrm{T}}\, \mathbf{H}_{f*g}\, \mathbf{u}_\theta$. We would like to obtain the maximum of $\mathbf{u}_\theta^{\mathrm{T}}\, \mathbf{H}_{f*g}\, \mathbf{u}_\theta$, subject to the constraint $\mathbf{u}_\theta^{\mathrm{T}}\mathbf{u}_\theta = 1$. We solve this constrained optimization problem using Lagrange's multiplier method by setting the gradient of $\mathbf{u}_\theta^{\mathrm{T}}\mathbf{H}_{f*g}\mathbf{u}_\theta + \lambda\, \mathbf{u}_\theta^{\mathrm{T}}\mathbf{u}_\theta$ to zero:

$$\mathbf{H}_{f*g}\mathbf{u}_\theta = -\lambda\, \mathbf{u}_\theta. \tag{14}$$

This implies that $-\lambda$ is an eigenvalue of $\mathbf{H}_{f*g}$; the corresponding normalized eigenvectors are the possible solutions to the problem. Since we are looking for the maximum of $\mathbf{u}_\theta^{\mathrm{T}}\, \mathbf{H}_{f*g}\, \mathbf{u}_\theta$, the optimal response and the angle are given by

$$r^* = \lambda_{\max} \tag{15}$$

$$\mathbf{u}_{\theta^*} = \mathbf{v}_{\max}. \tag{16}$$

Here, $\lambda_{\max}$ and $\mathbf{v}_{\max}$ are the maximum eigenvalue and the corresponding eigenvector, respectively.

It can be seen from Figs. 1c and 1d that these classical detectors do not have a good orientation selectivity. In the next section, we propose a new approach for the design of detectors that attempts to correct for this deficiency.

3 Design of Steerable Filters for Feature Detection

The widely-used contour extraction algorithm [1] has three steps: 1) feature detection, 2) nonmaximum suppression, and 3) thresholding. In this section, we present a general strategy for the design of steerable filters for feature detection, while keeping in mind the subsequent steps. We propose a criterion similar to that of Canny and we analytically derive the optimal filter—or, equivalently, the optimal weights—within our particular class of steerable functions specified by (4).

3.1 Optimality Criterion

We now review Canny's criterion and modify it slightly to enable analytical optimization. To derive the optimal 2D operator, we assume that the feature (edge/ridge) is oriented in some direction[2] (say, along the x axis) and derive an optimal operator for its detection. As the operator is rotation-steerable by construction, its optimality properties will be independent of the feature orientation.

The three different terms in Canny's criterion are as follows.

3.1.1 Signal-to-Noise Ratio

The key term in the criterion is the signal-to-noise ratio. The response of a filter $h(\mathbf{x})$ to a particular signal $f_0(\mathbf{x})$ (e.g., an idealized edge) centered at the origin is given by

$$S = \int_{\mathcal{R}^2} f_0(x,y)h(-x,-y)dx\,dy \tag{17}$$

S is given by the height of the response at its maximum. If the input is corrupted by additive white noise of unit variance, then the variance of the noise at the output is given by the energy of the filter:

$$\text{Noise} = \int_{\mathcal{R}^2} |h(x,y)|^2 dx\,dy. \tag{18}$$

We desire to have a high value of S for a given value of Noise; $\frac{S^2}{\text{Noise}}$ is the amplification of the desired feature provided by the detector.

3.1.2 Localization

The detection stage is preceded by nonmaximum suppression. The estimated feature position corresponds to the location of the local maximum of the response in the direction orthogonal to the feature boundary (y axis in our case). The presence of noise can cause an undesirable shift in the estimated feature location. The direct extension of Canny's

2. In 2D, the features of interest have boundaries of dimension 1.

expression for the shift-variance (due to white noise of unit variance) to 2D gives

$$E\Big[(\Delta y)^2\Big] = \frac{\int_{\mathcal{R}^2} |h_y(x,y)|^2 dx\, dy}{|\int_{\mathcal{R}^2} f_0(x,y) h_{yy}(-x,-y) dx\, dy|^2}. \quad (19)$$

Canny has proposed to maximize the reciprocal of this term. The numerator of (19) is a normalization term which will be small automatically if the impulse response of the filter is smooth along the y axis (low norm for the derivative). Since we are imposing this type of smoothness constraint elsewhere via an additional regularization term (see next section), it is not necessary to optimize this term here, which also keeps the effects well separated. Therefore, we propose to maximize the second derivative of the response, orthogonal to the boundary, at the origin

$$\begin{aligned} \text{Loc} &= -\frac{d^2}{dy^2}(f_0 * h) \\ &= -\int_{\mathcal{R}^2} f_0(x,y) h_{yy}(-x,-y) dx\, dy \end{aligned} \quad (20)$$

which is the square-root of the denominator in (20). The above expression is ensured to be positive because the second derivative of the response is negative at the maximum (assuming $S > 0$). Note that the new localization term is a measure of the width of the peak. The drift in position of the maximum due to noise will decrease as the response becomes sharper. In this work, we are neglecting the effect of neighboring features in deriving the localization term.

3.1.3 Elimination of False Oscillations

Canny observed that when the criterion is optimized only with the SNR and the localization constraint, the optimal operator has a high bandwidth; the response will be oscillatory and, hence, have many false maximas. In 2D, we desire that the response be relatively free of oscillations orthogonal to the feature boundary. This can be achieved by penalizing the term:

$$R_o = \int_{\mathcal{R}^2} |h_{yy}(x,y)|^2 dx\, dy. \quad (21)$$

Note that this term is the numerator of the expression for the mean distance between zero crossings proposed by Canny. It is a thin-plate spline-like regularization which is a standard technique to constrain a solution to be smooth (low bandwidth).

The thresholding step is easier if the response is flat along the boundary. The oscillation of the response along the boundary (x axis) can be minimized by penalizing

$$R_p = \int_{\mathcal{R}^2} |h_{xx}(x,y)|^2 dx\, dy \quad (22)$$

These terms will force the filter to be smooth making the response is less oscillatory, thus resulting in fewer false detections.

3.2 Derivation of the Optimal Detector

We combine the individual terms to obtain a single criterion

$$C = S \cdot \text{Loc} - \mu \underbrace{(R_o + R_p)}_{R}. \quad (23)$$

The filter in the family described by (4) that maximizes this criterion, subject to the constraint[3] $\text{Noise} = 1$, is our optimal detector. The free parameter $\mu > 0$ controls the smoothness of the filter; a high value makes the response less prone to false maxima and reduces oscillation along the ridge. However, these properties impose a trade off on the localization of the response.

In this work, we are also interested in performing a scale-independent design. In other words, if we dilate the window by a factor σ, using $g_\sigma(\mathbf{x}) = \sigma^{-\frac{1}{2}} g(\frac{\mathbf{x}}{\sigma})$, we want our solution to retain the shape independently of σ. This requires that we weight each of the terms in (23) using an appropriate power of the dilation factor. This issue is discussed later for each feature model separately.

For the ease of notation, we collect the component functions of (4) into a function vector $\mathbf{g}$ of length $(\frac{M(M+3)}{2})$, whose components are

$$[\mathbf{g}]_i(x,y) = \frac{\partial^{k-n}}{\partial x^{k-n}} \frac{\partial^n}{\partial y^n}\, g(x,y) \text{ with } i = \frac{(k-1)(k+2)}{2} + n$$
$$k = 0 \ldots M,\ n = 0..k.$$

Hence, an arbitrary function in the family is represented in a compact form as

$$h(\mathbf{x}) = \mathbf{a}^{\mathrm{T}} \mathbf{g}(\mathbf{x}), \quad (24)$$

where $\mathbf{a}$ is the vector containing the $\alpha_{i,k}$s in (4); it has the same length as the function vector. Now, we express the terms of the criterion in a matrix form as $S = \mathbf{a}^{\mathrm{T}}\mathbf{s}$, $\text{Loc} = \mathbf{a}^{\mathrm{T}}\mathbf{q}$, $\text{Noise} = \mathbf{a}^{\mathrm{T}}\mathbf{P}\,\mathbf{a}$, and $R = \mathbf{a}^{\mathrm{T}}\mathbf{R}\,\mathbf{a}$, where

$$[\mathbf{s}]_i = \langle f_0(\mathbf{x}), [\mathbf{g}(-\mathbf{x})]_i \rangle \quad (25)$$

$$[\mathbf{q}]_i = \Big\langle f_0(\mathbf{x}), ([\mathbf{g}(-\mathbf{x})]_i)_{yy} \Big\rangle \quad (26)$$

$$[\mathbf{P}]_{i,j} = \langle [\mathbf{g}]_i, [\mathbf{g}]_j \rangle \quad (27)$$

$$[\mathbf{R}]_{i,j} = \Big\langle ([\mathbf{g}]_i)_{yy}, ([\mathbf{g}]_j)_{yy} \Big\rangle + \Big\langle ([\mathbf{g}]_i)_{xx}, ([\mathbf{g}]_j)_{xx} \Big\rangle. \quad (28)$$

$g_{yy}(x,y)$ and $g_{xx}(x,y)$ denote $\partial^2 g(x,y)/\partial y^2$ and $\partial^2 g(x,y)/\partial x^2$, respectively. $\mathbf{P}$ and $\mathbf{R}$ are matrices of size $\frac{M(M+3)}{2} \times \frac{M(M+3)}{2}$, while the vectors $\mathbf{q}$ and $\mathbf{s}$ are of length $\frac{M(M+3)}{2}$. Here, $\mathbf{P}$ is ensured to be nonsingular. In the above expressions, the inner product of two functions is defined as

$$\langle f_1, f_2 \rangle = \int_{\mathcal{R}^2} f_1(x,y) f_2(x,y) dx\, dy.$$

Thus, the criterion (23) can be expressed in the matrix form as

$$C = \mathbf{a}^{\mathrm{T}}[\mathbf{Q} - \mu\,\mathbf{R}]\mathbf{a}, \quad (29)$$

where

$$\mathbf{Q} = \mathbf{s}\,\mathbf{q}^{\mathrm{T}}. \quad (30)$$

Since all the terms in the criterion are quadratic, the solution for the optimal parameters can be found analytically by using Lagrange's multiplier method. To maximize the

3. This constraint is just a normalization factor. Setting Noise to another constant will give detectors of the same shape, but with a different energy.

criterion subject to the constraint, we set the gradient of $C + \lambda\,\mathrm{Noise}$ to zero:

$$2[\mathbf{Q} - \mu\mathbf{R} + \lambda\,\mathbf{P}]\mathbf{a} = 0. \tag{31}$$

Rearranging the terms, we get

$$\mathbf{P}^{-1}[\mathbf{Q} - \mu\mathbf{R}]\;\mathbf{a} = -\lambda\;\mathbf{a} \tag{32}$$

which implies that λ is an eigenvalue of the matrix $(-\mathbf{P}^{-1}[\mathbf{Q} - \mu\mathbf{R}])$. The total number of eigenvalues is given by the dimension of $\mathbf{a}$. The corresponding eigenvectors $\mathbf{a}_{\lambda_i}$ need to be scaled so that the constraint $\mathbf{a}_{\lambda_i}^{\mathrm{T}}\mathbf{P}\mathbf{a}_{\lambda_i} = 1$ is satisfied. The optimal solution is therefore given by

$$\overline{\mathbf{a}} = \max\Big\{\mathbf{a}_{\lambda_i}^{\mathrm{T}}[\mathbf{Q} - \mu\mathbf{R}]\mathbf{a}_{\lambda_i}; i = 0 \ldots M(M+3)/2\Big\}. \tag{33}$$

Thus, the design of the optimal feature detector boils down to an eigen-decomposition followed by an appropriate weighting of the eigenvectors so as to satisfy the constraint.

3.3 Feature Detection by Local Optimization

Due to (5), the optimal angle θ^* in (1) is obtained as the solution of

$$\frac{\partial}{\partial\theta}(f(\mathbf{x}) * h(\mathbf{R}_{\theta^*}\mathbf{x})) = \sum_{k=1}^{M}\sum_{i=0}^{k} f_{k,i}(x,y)\underbrace{\frac{\partial}{\partial\theta}(b_{k,i}(\theta))\;|_{\theta=\theta^*}}_{c_{k,i}(\theta^*)} \tag{34}$$
$$= 0.$$

It is easy to see from (7) that each of the terms in $b_{k,i}(\theta)$ are of degree k in $\cos(\theta)$ and $\sin(\theta)$; $c_{k,i}(\theta)$ is of degree k as well. Hence, (34) is a polynomial of order M (in $\cos(\theta)$ and $\sin(\theta)$) and, thus, the estimation of the optimal angle involves the solution of an Mth order polynomial in two variables.

If $h(x,y)$ has only odd/even order partial derivatives (this is the case for many detectors), then $b_{k,i}(\theta)$ will be a polynomial with only odd/even degree terms (of $\cos(\theta)$ and $\sin(\theta)$) present. Consequently, (34) can be reduced[4] to a form where only terms of degree M are present. In this case, (34) can be further simplified (by dividing both the sides by $(\cos(\theta))^M$) to a polynomial in only one variable—$\tan(\theta)$. We then have an analytic solution if $M <= 3$ [15]. This case is illustrated in Section 4.1.3. When $M = 2$, the solution can also be computed as an eigen-decomposition of the Hessian matrix, which is better known (but, also, boils down to the above mentioned solution). This case is described in Section 4.2.2. When the solution of (34) is not trackable analytically, it can be solved numerically using an an iterative root finder such as the Newton-Raphson method.

4 Two-Dimensional Feature Detectors

We now design operators optimized for the detection of different 2D features. We chose the window function to be a Gaussian[5] $g(\mathbf{x};\sigma)$, where σ is the standard deviation. When it is clear from the context, we will suppress the dependence on σ to simplify the notation.

4.1 Edge Detection

As model for the edge, we choose the ideal step function

$$f_0(x,y) = \begin{cases} 1 & \text{if } y \geq 0 \\ 0 & \text{else.} \end{cases} \tag{35}$$

Since it is an odd function of y, the even order derivatives do not contribute to the signal energy; we therefore ignore[6] them in (4).

4.1.1 *Case 1:* $M = 1$

To illustrate the derivation of the optimal filter, we explain all the steps in detail in this simple case. Substituting the function vector $\mathbf{g} = [g_x, g_y]$ in the corresponding expressions, we get

$$\mathbf{s} = -\sigma\sqrt{\pi}\;[0,1]$$
$$\mathbf{q} = -\frac{2\sqrt{\pi}}{\sigma}\;[0,1]$$
$$\mathbf{P} = \frac{\pi}{2}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$$
$$\mathbf{R} = \frac{9\pi}{\sigma^4}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}.$$

Thus,

$$\mathbf{Q} = \mathbf{q}^{\mathrm{T}}\mathbf{s} = 2\pi\begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}.$$

The matrices $\mathbf{Q}$ and $\mathbf{P}$ are independent of σ, while $\mathbf{R}$ is inversely proportional to σ^4. So, we weigh $\mathbf{R}$ by σ^4 to have a scale-invariant solution. Hence,

$$\mathbf{P}^{-1}\left[\mathbf{Q} - \mu\sigma^4\mathbf{R}\right] = \begin{bmatrix}-18\mu & 0\\ 0 & 4 - 18\mu\end{bmatrix}. \tag{36}$$

The eigenvalues of $\mathbf{P}^{-1}\left[\mathbf{Q} - \mu\sigma^4\mathbf{R}\right]$ are $\lambda_1 = -18\mu$ and $\lambda_2 = 4 - 18\mu$, respectively. The corresponding scaled eigenvectors (so, as to satisfy the constraint) are

$$\left[0, -\sqrt{\frac{2}{\pi}}\right]$$

and

$$\left[-\sqrt{\frac{2}{\pi}}, 0\right],$$

respectively. When substituted in the criterion, they yield $4 - 18\mu$ and -18μ, respectively. Thus, the optimal solution is

$$\overline{\mathbf{a}} = \left[0, -\sqrt{\frac{2}{\pi}}\right]$$

(as $\mu > 0$), which corresponds to Canny's edge detector (cf. Fig. 1c).

4. If there is a term of degree $M - 2n$, we can multiply it by $(\cos(\theta)^2 + \sin(\theta)^2)^n$ to make it of degree M.

5. It is the only function that is isotropic and separable.

6. If we were to include them in the solution, their optimal coefficients would turn out to be zero anyway.

TABLE 1
Edge Detectors for Different Parameters

	S^2/Noise	Loc	# Basis filters	Expression	Implementation
$M=1$	$2\,\sigma^2$	1.63	2 separable	$-\sqrt{\frac{2}{\pi}}g_y$	Analytic
$M=3$			4 nonseparable		
$\mu=0.09$	$2.93\,\sigma^2$	1.98	6 separable	$-0.966\,g_y-0.256\,\sigma^2\,g_{xxy}$	Analytic
$M=3$			4 nonseparable	$-1.0655\,g_y-0.20\,\sigma^2\,g_{xxy}$	
$\mu=0.2$	$3.01\,\sigma^2$	1.83	6 separable	$-0.042\,\sigma^2\,g_{yyy}$	Analytic
				$-1.1215\,g_y-0.5576\,\sigma^2\,g_{xxy}$	
$M=5$	$3.69\,\sigma^2$	2.15	6 nonseparable	$-0.018\,\sigma^2\,g_{yyy}-0.0415\,\sigma^4\,g_{xxxxy}$	Sampling/
$\mu=0.15$			12 separable	$-0.0038\,\sigma^4\,g_{xxyyy}$	Iterative

4.1.2 Higher Order Cases

For higher M, we obtain a family of solutions that are increasingly smooth when μ goes up. A few examples of higher order templates are given in Table 1 with the filter impulse responses shown in Fig. 3. By comparing Figs. 3b and 3c, we observe that, as μ increases, the filter becomes smoother at the cost of directionality. The higher order templates are more elongated thus having higher SNR and localization (cf. Table 1); they should therefore result in better detections, at-least for idealized edges. The dependence of SNR on σ^2 implies that this figure can also be improved by increasing the variance of the Gaussian. However, the ability to resolve two adjacent parallel edges decreases as σ increases.

4.1.3 Implementation

Here, we develop the implementation procedure mentioned in Section 3.3 for the special case of third order edge detection. A general third order edge template (for different values of μ) is given by

$$h(\mathbf{x}) = \alpha_{1,0}\; g_x + \alpha_{3,0}\; g_{xxx} + \alpha_{3,2}\; g_{xyy}. \tag{37}$$

The rotated version[7] of this template h_θ is given by

$$\begin{aligned} h_\theta = {} & \alpha_{1,0}\;(g_x\cos(\theta)+g_y\sin(\theta))+ \\ & \alpha_{3,0}\left(g_{xxx}\cos^3(\theta)+3\,g_{xxy}\cos^2(\theta)\sin(\theta)+\right. \\ & \left. 3\,g_{xyy}\cos(\theta)\sin^2(\theta)+g_{yyy}\sin^3(\theta)\right)+ \\ & \alpha_{3,2}\left(g_{xyy}\cos^3(\theta)+(-2g_{xxy}+g_{yyy})\cos^2(\theta)\sin(\theta)+\right. \\ & \left.(-2g_{xyy}+g_{xxx})\cos(\theta)\sin^2(\theta)+g_{xxy}\sin^3(\theta)\right). \end{aligned}$$

Convolving the rotated template by f and simplifying, we get

$$\begin{aligned} (f*h_\theta)(\mathbf{r}) = {} & q_1(\mathbf{r})\;\cos(\theta)^3+q_2(\mathbf{r})\;\cos(\theta)^2\sin(\theta)+ \\ & q_3(\mathbf{r})\;\cos(\theta)\sin(\theta)^2+q_4(\mathbf{r})\;\sin(\theta)^3, \end{aligned} \tag{38}$$

where

$$q_1(\mathbf{r}) = \alpha_{3,0}\,f_{3,0}(\mathbf{r})+\alpha_{3,2}\,f_{3,2}(\mathbf{r})+\alpha_{1,0}\,f_{1,0}(\mathbf{r}) \tag{39}$$

$$q_2(\mathbf{r}) = (3\alpha_{3,0}-2\alpha_{3,2})f_{3,1}(\mathbf{r})+\alpha_{3,2}\,f_{3,3}(\mathbf{r})+\alpha_{1,0}\,f_{1,1}(\mathbf{r}) \tag{40}$$

$$q_3(\mathbf{r}) = (3\alpha_{3,0}-2\alpha_{3,2})f_{3,2}(\mathbf{r})+\alpha_{3,2}\,f_{3,0}(\mathbf{r})+\alpha_{1,0}\,f_{1,0}(\mathbf{r}) \tag{41}$$

$$q_4(\mathbf{r}) = \alpha_{3,0}\,f_{3,3}(\mathbf{r})+\alpha_{3,2}\,f_{3,1}(\mathbf{r})+\alpha_{1,0}\,f_{1,1}(\mathbf{r}). \tag{42}$$

We multiplied the single degree terms in $\cos(\theta)$ and $\sin(\theta)$ with $(\cos^2(\theta)+\sin^2(\theta))$ so that we get a polynomial with

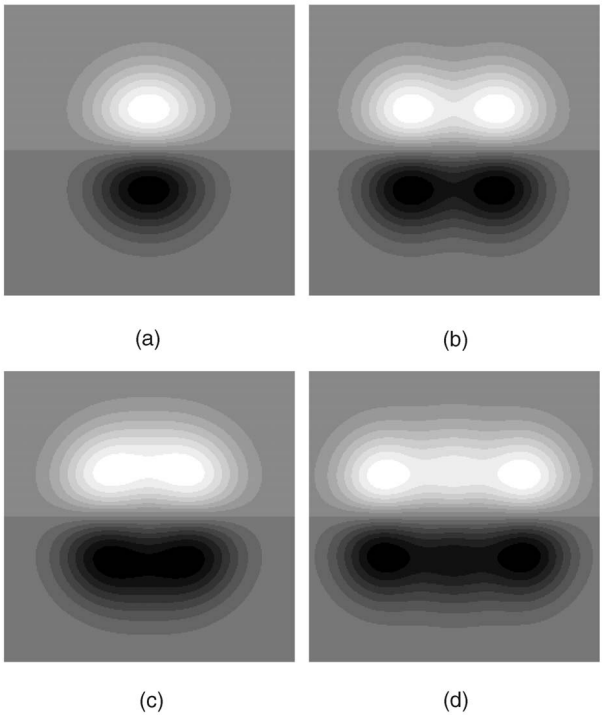


Fig. 3. Edge Detectors for different parameters. The detectors become more orientation selective as M increases. (a) Canny's edge detector. (b) $M=3$; $\mu=0.09$. (c) $M=3$; $\mu=0.2$. (d) M=5; $\mu=0.15$.

7. The expression for a general rotated template is given by (5) and (7). However, for simple templates, it may be easier to derive it directly in the Fourier domain as in (60).

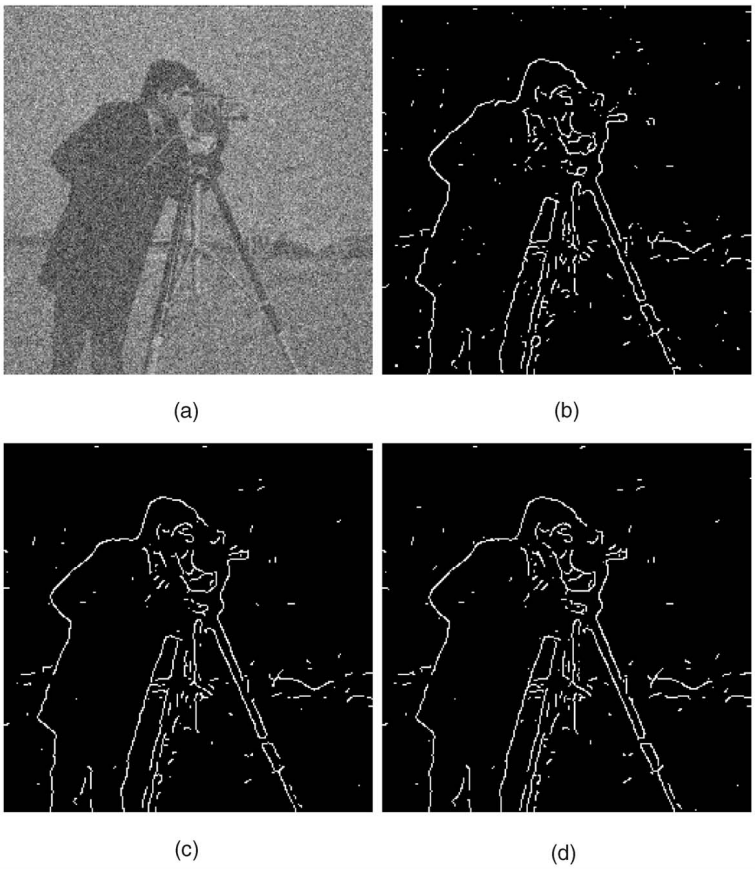


Fig. 4. Edge detection on a 256×256 noisy image (Gaussian white noise of variance 85). The thresholding is performed such that there are 2,000 detected pixels in each image. The variance of the Gaussian window is chosen as 1.7. Note that the higher order detectors give less wiggly contours with fewer breaks. The algorithm was implemented in Java as a plugin for ImageJ. The experiments were performed on an Intel Pentium processor at 2.66 GHz. (a) Noisy image. (b) Canny (time taken 141 ms). (c) $M = 3;\ \mu = 0.09$ (time taken 414 ms). (d) $M = 5;\ \mu = 0.15$ (time taken 1,995 ms).

only third degree terms. Note that the six functions $f_{k,i};\ k = \{1, 3\}, i = 0 \ldots k$, obtained by separable filtering, are combined to derive $q_i;\ i = 1 \ldots 4$. They can also be obtained by nonseparable filtering:

$$q_1(\mathbf{r}) = f * \big(\alpha_{3,0}\, g_{3,0} + \alpha_{3,2}\, g_{3,2} + \alpha_{1,0}\, g_{1,0}\big)(\mathbf{r}) \tag{43}$$

$$q_2(\mathbf{r}) = f * \big((3\alpha_{3,0} - 2\alpha_{3,2}) g_{3,1} + \alpha_{3,2}\, g_{3,3} + \alpha_{1,0}\, g_{1,1}\big)(\mathbf{r}) \tag{44}$$

$$q_3(\mathbf{r}) = f * \big((3\alpha_{3,0} - 2\alpha_{3,2}) g_{3,2} + \alpha_{3,2}\, g_{3,0} + \alpha_{1,0}\, g_{1,0}\big)(\mathbf{r}) \tag{45}$$

$$q_4(\mathbf{r}) = f * \big(\alpha_{3,0}\, g_{3,3} + \alpha_{3,2}\, g_{3,1} + \alpha_{1,0}\, g_{1,1}\big)(\mathbf{r}). \tag{46}$$

We use the separable approach due to its computational efficiency. The nonseparable approach may be profitable for large values of M.

For a particular value of $\mathbf{r}$, $(f * h_\theta(\mathbf{r}))$ is a function of only one variable—θ. At the local maxima and the minima of $(f * h_\theta(\mathbf{r}))$, we have $\frac{\partial}{\partial\theta}(f * h_\theta(\mathbf{r})) = 0$. Substituting for $f * h_\theta$ from (38), we get

$$q_2\, \cos(\theta)^3 + (2q_3 - 3q_1) \cos^2(\theta) \sin(\theta) + (3q_4 - 2q_2) \cos(\theta) \sin^2(\theta) - q_3\, \sin^3(\theta) = 0.$$

We divide both sides of this equation by $\cos(\theta)^3$ to get a cubic polynomial in one variable—$\tan\theta$:

$$q_2 + (2q_3 - 3q_1) \tan(\theta) + (3q_4 - 2q_2) \tan^2(\theta) - q_3\, \tan(\theta)^3 = 0. \tag{47}$$

The roots of this equation can be obtained analytically [15]. Since $\tan(\theta) = \tan((\theta + \pi) \bmod 2\pi)$, there are six possible values of θ in the range $[0, 2\pi]$ that satisfy (47). One of these values of θ correspond to the global maximum; it can be found out by substituting all them into (38) and picking the one which gives the maximum value. We breifly describe the steps of the local optimization algorithm in Appendix B.

For $M > 3$, the θ^* estimated for $M = 3$ can act as an approximate solution. This initial guess is further refined by performing a golden search [15] around the approximate solution.

4.1.4 Results

Because the scheme is optimized for noisy data, we perform edge detection on the cameraman image corrupted with additive white noise (cf. Fig. 4a). The size of the Gaussian window is the same in all the experiments. The detected edges after nonmaximum suppression and thresholding are presented in Fig. 4. It is seen that Canny's edge detector has a lot of false detections. Moreover, the detected edges are wiggly due to poor localization. The new detectors have significantly lower false detections and better localization, thus confirming the theoretical improvement.

Note the time taken for the various edge detection schemes from Figs. 4b, 4c, and 4d. The third order scheme only takes around 2.5 times the time as the Canny's

TABLE 2
Ridge Detectors for Different Parameters

	S^2/Noise	Loc	# Basis filters	Expression	Implementation
$M=2$ $\mu=2$	2.67	$4.38/\sigma$	3 separable	$-\sqrt{\frac{2}{3\pi}}\sigma\ g_{yy}$	Analytic
$M=2$ $\mu=0$	3	$4.64/\sigma$	3 separable	$-\sqrt{\frac{3}{4\pi}}\sigma\ (g_{yy}-g_{xx}/3)$	Analytic
$M=4$ $\mu=0.1$	4.302	$6.41/\sigma$	5 nonseparable 8 separable	$-0.204\sigma\ g_{yy}+0.059\sigma\ g_{xx}$ $+0.063\sigma^3\ g_{yyyy}-0.194\sigma^3\ g_{xxyy}$ $+0.024\sigma^3\ g_{xxxx}$	Analytic
$M=4$ $\mu=1/4$	4.47	$6.14/\sigma$	5 nonseparable 8 separable	$-0.392\sigma\ g_{yy}+0.113\sigma\ g_{xx}$ $+0.034\sigma^3\ g_{yyyy}-0.184\sigma^3\ g_{xxyy}$ $+0.025\sigma^3\ g_{xxxx}$	Analytic

detector. We believe that, for the performance improvement achieved, it is a quite reasonable price to pay. Since we resorted to a naive optimization algorithm using dichotomy, the fifth order method took more time. We believe that a better optimization scheme could drastically improve the computational efficiency.

4.2 Ridge Detection

For simplicity, we choose the idealized line model as:

$$f_0(x,y)=\delta(y), \tag{48}$$

where δ denotes the Dirac delta function. A more realistic model can be assumed without any change in the computational strategy. Here, $\mathbf{Q}$, $\mathbf{P}$, and $\mathbf{R}$ are inversely proportional to σ^4, σ^2, and σ^6, respectively. Hence, we scale $\mathbf{Q}$ by σ^2 and $\mathbf{R}$ by σ^4.

4.2.1 Optimized Detectors

Some examples of optimal templates are shown in Table 2 and Fig. 5. Interestingly, we see from the table that the optimal detector for $M=2$ and $\mu=0$ is better than the classical detector, both in terms of SNR and localization, at no additional cost. Also, note that the template in Fig. 5b is more directional than the classical one in Fig. 5a. The high value of $\mu=2$ (adjusted to get the equivalence) overconstrains the optimization, resulting in a lower performance.

Two cases for $M=4$ are also shown. It is seen that for small μ, the template oscillates along y producing undesirable sidelobes. However, it has a better localization at the expense of a lower SNR and R.

In general, we found that it is better to have a low value of μ for lower order templates; the model has few degrees of freedom and, hence, a high value of μ will overconstrain the system. On the other hand, for higher order templates, we need a higher value of μ to make them less oscillatory.

4.2.2 Implementation

Any second order detector can be implemented as an eigendecomposition, similar to the classical Hessian (described in Section 2.3). For example, the detector with $\mu=0$ can be implemented as

$$\begin{aligned}\theta^* &= \arg\max_{\theta} f * \left(g_{\mathbf{u}_\theta,\mathbf{u}_\theta}\frac{g_{\mathbf{u}_{\theta+\frac{\pi}{2}},\mathbf{u}_{\theta+\frac{\pi}{2}}}}{3}\right)\\ &= \arg\max_{\theta}(\mathbf{u}_\theta^{\mathrm{T}}\ \mathbf{H}_{\mathrm{mod}}\ \mathbf{u}_\theta),\end{aligned}$$

where $\mathbf{H}_{\mathrm{mod}}=\mathbf{H}_{f*g}\frac{1}{3}(\mathbf{P}^{\mathrm{T}}\ \mathbf{H}_{f*g}\ \mathbf{P})$; here, $\mathbf{P}$ is the rotation matrix

$$\mathbf{P}=\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}, \tag{49}$$

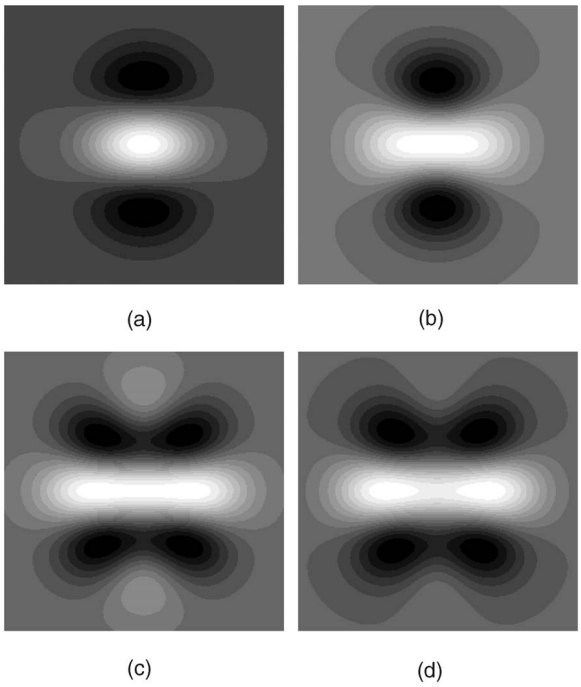


Fig. 5. Ridge Detectors corresponding to different orders and parameters. (a) $M=2$; $\mu=2$ (classical detector). (b) $M=2$; $\mu=0$. (c) $M=4$; $\mu=0.1$. (d) $M=4$; $\mu=0.25$.

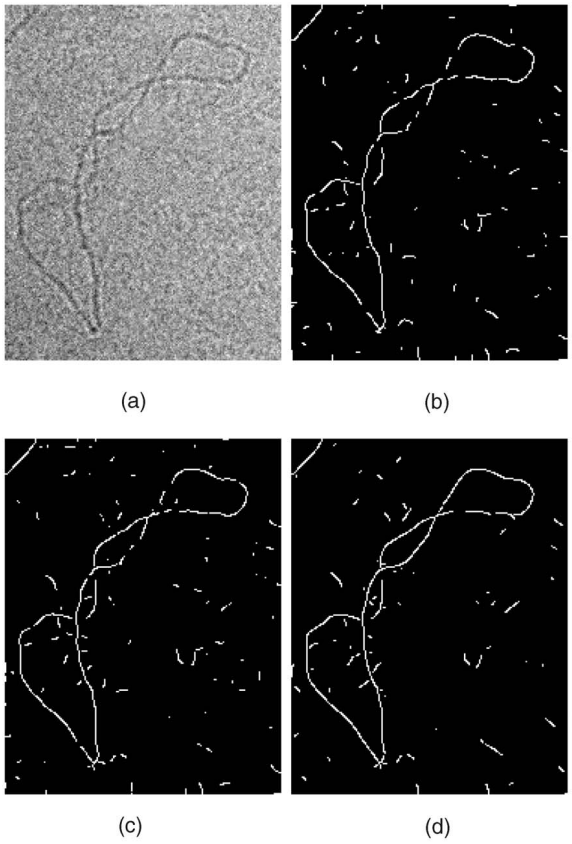


Fig. 6. Detection of DNA filament from its noisy cryo-electron micrograph. The features were ridges that were roughly 2-3 pixels wide. We chose the standard deviation of the Gaussian window to be 3. The images were thresholded such that there are 1,000 detected pixels. (a) DNA micrograph. (b) Classical detector (time: 260 ms). (c) $M = 2; \mu = 0$ (time: 260 ms). (d) $M = 4; \mu = 0.25$ (time: 590 ms).

such that $\mathbf{P}\mathbf{u}_\theta = \mathbf{u}_{(\theta+\frac{\pi}{2})}$. Thus, the optimal direction and ridge magnitude can be computed with the eigen-decomposition of $\mathbf{H}_{\text{mod}}$; the computational complexity is the same as with the classical scheme.

For the fourth order detector, we proceed exactly as in the case of the third order edge template. The computation of the optimal angle involves the solution of a quartic polynomial, which is also performed analytically [15].

4.2.3 Results

An interesting application, which motivated this whole development, is the detection of DNA filaments (cf. Fig. 6a) from their stereo cryo-electron micrographs [13]. The difficulty with these data is that the micrographs are extremely noisy because they are exposed to a low electron dose to avoid the degradation of the specimen. The results (Figs. 6b, 6c, and 6d) correspond to the output of ridge detection algorithm followed by nonmaximum suppression and thresholding. Overall, the $M = 4$ detector gives the best qualitative results: there are few breaks in the filament and the detection is less wiggly. Note that the performance improvement cost only two times the time taken for the classical approach. The optimal second order detector gave better results for the same computational complexity as the classical approach.

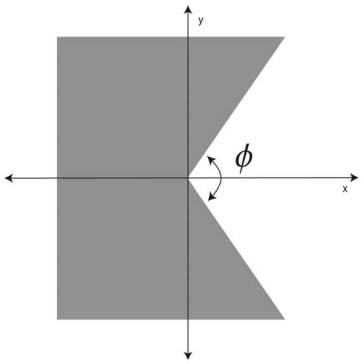


Fig. 7. Model of an ideal wedge.

5 Shape Adaptable Feature Detection

Steerability in rotation involves the representation of a template as a weighted linear combination of a few filters; the weights are nonlinear functions of a single parameter—the angle. This leaves us with extra degrees of freedom which can be utilized effectively. Perona used it to make the template steerable in scale [5]. We propose to utilize this freedom for the design of a shape-adaptable filter, thus making the system respond to different shapes depending on the parameters.

In Section 4.14, we designed templates for the detection of ideal step edges. However, as mentioned in [16], the edges are sometimes wedge shaped (close to image corners). Since this contradicts our assumption, we have low SNR at the corners. A bias in the position of the corner is also reported in the context of conventional corner detectors [17].

Corners are image regions with high surface curvature. They convey a lot of information about the image shape [18], [19], [20], [21]. Hence, we propose a new shape-adaptable, steerable corner detector that addresses these issues.

5.1 Derivation of the Wedge Detector

We model a corner as a wedge shown in Fig. 7, where the wedge angle ϕ is a variable. Analytically, we have

$$f_0(x,y) = \begin{cases} 1 & \text{if } -x\sin\left(\frac{\phi}{2}\right) \le y\cos\left(\frac{\phi}{2}\right) \le x\sin\left(\frac{\phi}{2}\right) \\ 0 & \text{otherwise.} \end{cases} \tag{50}$$

We focus on the derivation of a third order corner detector. Since the third order detectors cannot oscillate much, we set $\mu = 0$. We also get rid of the localization term—to obtain a simple expression, we optimize the detector only with respect to the SNR.

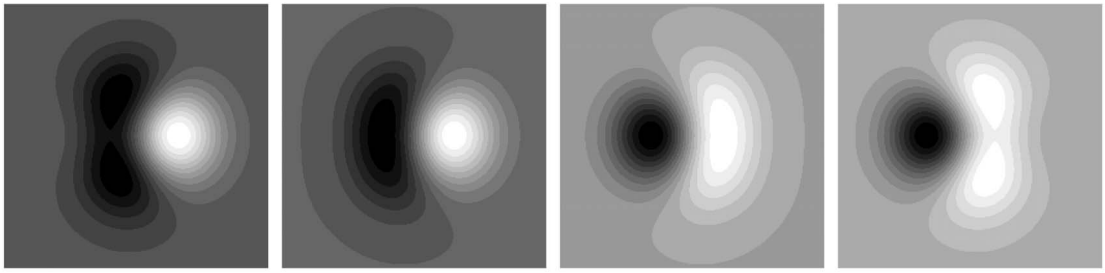
Setting the gradient of $S + \lambda\,\text{Noise}$ to zero (to maximize S subject to $\text{Noise} = 1$), we get

$$2\lambda\,\mathbf{P}\,\mathbf{a} = -\mathbf{s}, \tag{51}$$

from which we obtain the optimal solution as

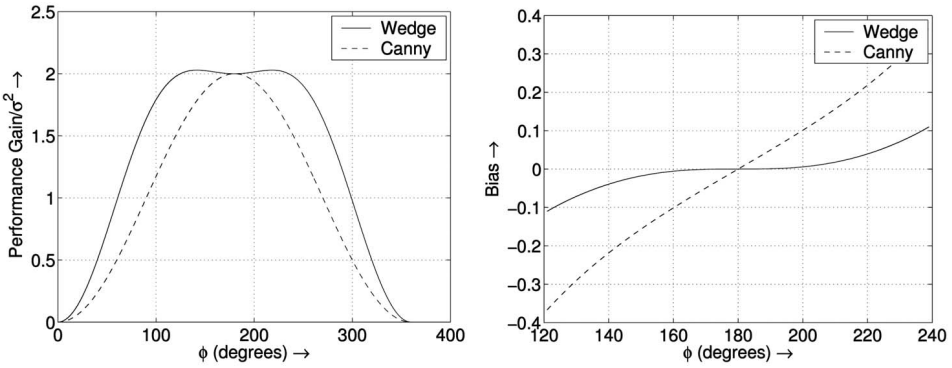
$$\overline{\mathbf{a}} = -\frac{\mathbf{P}^{-1}\mathbf{s}}{\sqrt{\mathbf{s}^{\mathrm{T}}\mathbf{P}^{-1}\mathbf{s}}}. \tag{52}$$

For a third order detector ($\mathbf{g} = \left[g_x, g_y, g_{xx}, g_{yy}, g_{xy}\right]$) and the idealized wedge model, $\mathbf{P}$ and $\mathbf{s}$ defined by (25) and (27) are given by



(a) (b) (c) (d)

Fig. 8. Wedge detectors for different wedge angles. (a) $\phi = 0.6\pi\ (108^\circ)$. (b) $\phi = 0.8\pi\ (144^\circ)$. (c) $\phi = 1.2\pi\ (216^\circ)$. (d) $\phi = 1.4\pi\ (252^\circ)$.



(a) (b)

Fig. 9. (a) S^2/Noise versus wedge angle. (b) Measure of $\frac{\overline{y}}{\sigma}$ (ratio of the bias and the standard deviation of the Gaussian window).

$$\mathbf{P} = \int_{\mathcal{R}^2} \mathbf{g}^T(x,y)\,\mathbf{g}\,(x,y)dxdy \tag{53}$$

$$= \begin{bmatrix} \frac{\pi}{2} & 0 & 0 & 0 & 0 \\ 0 & \frac{\pi}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{3\pi}{2\sigma^2} & \frac{\pi}{2\sigma^2} & 0 \\ 0 & 0 & \frac{\pi}{2\sigma^2} & \frac{3\pi}{2\sigma^2} & 0 \\ 0 & 0 & 0 & 0 & \frac{\pi}{2\sigma^2} \end{bmatrix} \tag{54}$$

$$\mathbf{s} = \int_{\mathcal{R}^2} f_0(x,y)\,\mathbf{g}\,(x,y)dxdy \tag{55}$$

$$= \left[-\sigma\sqrt{\pi}\sin(\tfrac{\phi}{2}) \quad 0 \quad \sin(\phi) \quad -\sin(\phi) \quad 0 \right]^{\mathrm{T}}. \tag{56}$$

Substituting the above in (52), we obtain the SNR-optimized third order template as

$$h(\mathbf{x}) = -\sqrt{\frac{2}{2+\pi+2\cos\phi}}\left(g_x + \frac{\sigma\cos\frac{\phi}{2}}{\sqrt{\pi}}(g_{xx} - g_{yy}) \right). \tag{57}$$

It is interesting to note that the optimal corner detector is Canny's edge detector when $\phi = \pi$. Some examples of detectors for different values of ϕ are shown in Fig. 8.

5.2 Implementation

We have a template $h_{\theta,\phi}$ which is now parametrized by two variables: θ—the orientation—and ϕ—the wedge angle. Hence, the detection procedure involves a two variable optimization. For our experiments, we resort to a slightly suboptimal solution where θ is estimated from the $\phi = \pi$ solution and the optimal ϕ is estimated by sampling. This approach is justifiable as the optimal angle does not change much with respect to ϕ.

5.3 Results

We now study the theoretical performance improvement of the wedge detector over Canny's edge detection scheme. We consider the responses of Canny's edge detector and the optimal wedge detector (designed for a specific ϕ) to the wedge. In Fig. 9, we show the variation of the SNR with respect to the wedge angle. Note that for Canny's edge detector, the SNR falls off much more rapidly as compared to the wedge detector. The SNR of the wedge detector has a flat zone around $\phi = \pi$ for roughly a span of 140 degrees.

To analyze the bias in the position, we consider the response $r(x,y)$ of the wedge $f_0(x,y)$ (shown in Fig. 7) to a template $h(x,y)$. The position of the maximum will be displaced from the origin, along the y axis. A first order approximation of the displacement can be obtained by using the Taylor series expansion of the response $r(x,y) = f_0(x,y) * h(x,y)$ along the y axis.

$$r(0,y) = r(0,0) + r_y(0,0)y + \frac{r_{yy}(0,0)}{2}y^2 + \mathcal{O}(y)^3. \tag{58}$$

We look for the point $\overline{y}$ such that $r_y(0,\overline{y}) = 0$. From the above expression, we obtain the first order expression of $\overline{y}$ as $r_y(x_0,y_0)/r_{yy}(x_0,y_0)$. Substituting $r = f * h$ and by using the commutativity of convolution and differentiation, we get

$$\overline{y} = -\frac{f * h_y\,|_{0,0}}{f * h_{yy}\,|_{0,0}}.$$

The plot of the bias ($\overline{y}$) for different wedge angles is shown in Fig. 9b. It is seen that, for Canny's edge detector, the wedge is displaced from the actual location much more than for the wedge detector tuned to the corresponding angle.

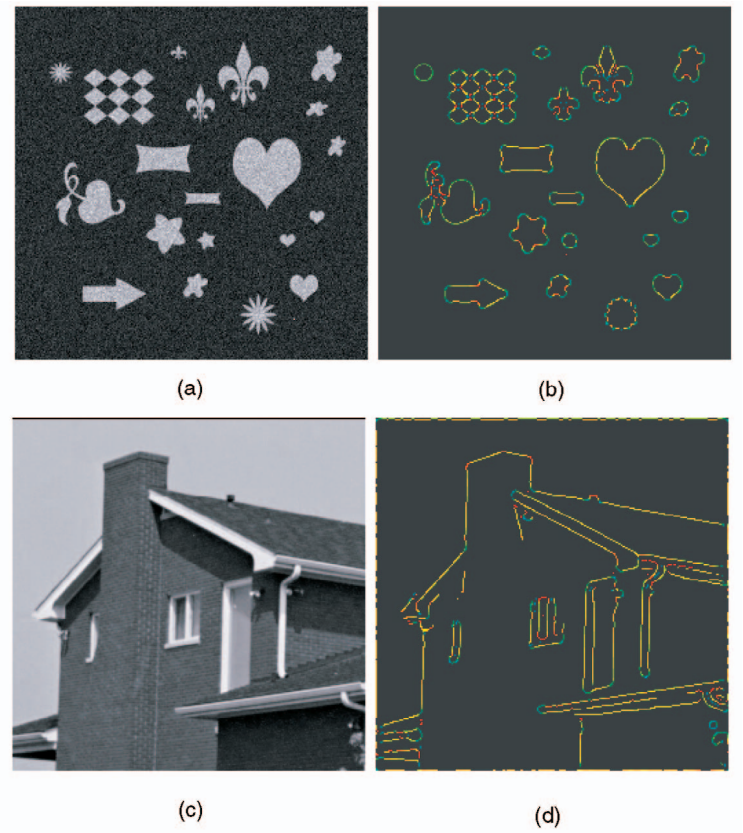


Fig. 10. Detected wedge angle. Here, red stands for $\phi = \frac{3\pi}{2}$, yellow for $\phi = \pi$, and cyan for $\phi = \frac{\pi}{2}$. Here, the corners are the points which are either in red or in cyan. Note that at the straight edges, the optimal wedge angle is π; the optimal detector is equivalent to the Canny's edge detector. In this experiment, we have chosen $\sigma = 3$. (a) Noisy image. (b) Angle output of wedge detector. (c) House image. (d) Angle output of wedge detector.

In short, the wedge detector performs better than the edge detector for nonideal step edges (wedges) for a range of angles; this range can be increased by considering higher order detectors.

To demonstrate the practical utility of the algorithm, we consider the synthetic pattern shown in Fig. 10a and the real image shown in Fig. 10c. We estimate the optimal parameters (θ and ϕ) and the response. We perform nonmaximum suppression of the response and keep only the values above a certain threshold. The estimated value of ϕ where the response is greater than the threshold are shown in Figs. 10b and 10d. Note that the detector can distinguish between convex and concave wedges based on the difference in the estimated angles. The estimated position of the wedge is also a reasonable fit to their true positions. Since Canny's detector is also in the family of wedge detectors, this scheme works well for straight edges as well.

6 Conclusions

We have proposed a general approach to derive optimal 2D operators for the detection of image features. We chose the optimal template from a family of steerable functions using an analytical optimization scheme based on a slight modification of Canny's criterion. In contrast to classical approaches, where the optimization is performed in 1D, we specified the filter directly in 2D. We derived optimal operators for a variety of image features and demonstrated their utility in various applications. We also introduced the notion of shape-adaptable feature detection and used it for the detection of image corners.

We now discuss a few issues that were not dealt with in this paper and are still open for further investigation.

1. Class of steerable functions: Although we have concentrated on the space of Gaussian derivatives as the steerable family, the design methodology is applicable to other classes as well. Interesting variations may be obtained by changing the window function or using by other known families of steerable functions [7], [22].
2. Discretization: We have derived the optimal operators in continuous space, neglecting discretization issues. It could be interesting to address the discretization effects as in [23] to be closer to practical situations.

Even though further research is required to address these issues, the results presented here are promising enough to justify the use of the proposed detectors in a variety of practical applications. The methodology is also general enough to allow for the design of application-specific templates.

The implementation of the algorithm is available as a Java plugin for ImageJ [24] at http://bigwww.epfl.ch/demo/steerable/.

Appendix A

Proof. Using the linearity of the Fourier transform and the property that differentiation corresponds to a

multiplication with $\mathrm{j}\omega$ in the Fourier domain, it is easy to derive the transfer function of the filter h:

$$\hat{h}(\omega_x, \omega_y) = \sum_{k=1}^{M} \sum_{i=0}^{k} \alpha_{k,i} (\mathrm{j}\omega_x)^{k-i} (\mathrm{j}\omega_y)^i \, \hat{g}(\omega_x, \omega_y), \tag{59}$$

where $\mathrm{j} = \sqrt{-1}$. Since the rotation of a filter in space corresponds to a rotation of its Fourier transform, we get

$$\begin{aligned} \Im(h(\mathbf{R}_\theta \, \mathbf{x})) &= \sum_{k=1}^{M} \sum_{i=0}^{k} \alpha_{k,i} (\mathrm{j}\omega_x \cos(\theta) + \mathrm{j}\omega_y \sin(\theta))^{k-i} \\ &\quad \Big(-\mathrm{j}\omega_x \sin(\theta) + \mathrm{j}\omega_y \cos(\theta)\Big)^i \, \hat{g}(\omega_x, \omega_y) \\ &= \sum_{k=1}^{M} \sum_{i=0}^{k} \alpha_{k,i} \sum_{l=0}^{k-i} \sum_{m=0}^{i} \binom{k-i}{l} \binom{i}{m} (-1)^m \\ &\quad \cos(\theta)^{i+(l-m)} \sin(\theta)^{(k-l)-(i-m)} (\mathrm{j}\omega_x)^{l+m} \\ &\quad (\mathrm{j}\omega_y)^{k-(l+m)} \hat{g}(\omega_x, \omega_y). \end{aligned} \tag{60}$$

Note that the window function is left unchanged because we are assuming that it is isotropic. Now, multiplying both sides by $\hat{f}$ and computing the inverse Fourier transform, we get

$$\begin{aligned} f(\mathbf{x}) * h(\mathbf{R}_\theta \mathbf{x}) &= \sum_{k=1}^{M} \sum_{i=0}^{k} \alpha_{k,i} \sum_{l=0}^{k-i} \sum_{m=0}^{i} \binom{k-i}{l} \binom{i}{m} (-1)^m \\ &\quad \cos(\theta)^{i+(l-m)} \sin(\theta)^{(k-i)-(l-m)} f_{k,k-(l+m)}(\mathbf{x}), \end{aligned} \tag{61}$$

where

$$f_{k,i}(\mathbf{x}) = f(\mathbf{x}) * \left(\frac{\partial^{k-i}}{\partial x^{k-i}} \frac{\partial^i}{\partial y^i} g(\mathbf{x}) \right).$$

Note that the component indices of f are dependent only on k and $l+m$. We collect the terms with the same values of $k-(l+m)$ and we define $\mathcal{S}(k,i,j)$ as

$$\begin{aligned} S(k,i,j) = \{ & l, m \mid 0 <= l <= k-i; \\ & 0 <= m <= i; k-(l+m) = j \}. \end{aligned} \tag{62}$$

Using this definition, we rewrite the right-hand side of (61) as

$$\sum_{k=1}^{M} \sum_{i=0}^{k} \alpha_{k,i} \Big(\sum_{j=0}^{k} \sum_{l,m \in \mathcal{S}(k,i,j)} \tbinom{k-i}{l} \tbinom{i}{m} (-1)^m \cos(\theta)^{i+(l-m)} \sin(\theta)^{(k-i)-(l-m)} f_{k,j}(\mathbf{x}) \Big)$$

$$\sum_{k=1}^{M} \sum_{j=0}^{k} f_{k,j}(\mathbf{x}) \underbrace{\Big(\sum_{i=0}^{k} \alpha_{k,i} \sum_{l,m \in \mathcal{S}(k,i,j)} \tbinom{k-i}{l} \tbinom{i}{m} (-1)^m \cos(\theta)^{i+(l-m)} \sin(\theta)^{(k-i)-(l-m)} \Big)}_{b_{k,j}(\theta)}.$$

□

APPENDIX B

In this section, we briefly outline the steps involved in the thrid order edge detection algorithm. We denote the 1D Gaussian of a specified variance, its first, second, and third derivatives sampled on a certain grid by g, g', g'', and g''', respectively.

Algorithm

f_{10} = `filterSeparable(image,` g', g);
f_{11} = `filterSeparable(image,` g, g');
f_{30} = `filterSeparable(image,` g''', g);
f_{31} = `filterSeparable(image,` g'', g');
f_{32} = `filterSeparable(image,` g', g'');
f_{33} = `filterSeparable(image,` g, g''');
for `i=0 to Nrows-1` **do**
  for `j=0 to Ncols-1` **do**
    $q_1 = \alpha_{3,0} f_{3,0}(\mathtt{i},\mathtt{j}) + \alpha_{3,2} f_{3,2}(\mathtt{i},\mathtt{j}) + \alpha_{1,0} f_{1,0}(\mathtt{i},\mathtt{j})$;
    $q_2 = (3\alpha_{3,0} - 2\alpha_{3,2}) f_{3,1}(\mathtt{i},\mathtt{j}) + \alpha_{3,2} f_{3,3}(\mathtt{i},\mathtt{j}) + \alpha_{1,0} f_{1,1}(\mathtt{i},\mathtt{j})$;
    $q_3 = (3\alpha_{3,0} - 2\alpha_{3,2}) f_{3,2}(\mathtt{i},\mathtt{j}) + \alpha_{3,2} f_{3,0}(\mathtt{i},\mathtt{j}) + \alpha_{1,0} f_{1,0}(\mathtt{i},\mathtt{j})$;
    $q_4 = \alpha_{3,2} f_{3,1}(\mathtt{i},\mathtt{j}) + \alpha_{3,0} f_{3,3}(\mathtt{i},\mathtt{j}) + \alpha_{1,0} f_{1,1}(\mathtt{i},\mathtt{j})$;
    `solset = solveCubic` $(q_2, 2q_3 - 3q_1, 3q_4 - 2q_2, -q_3)$;
    `thetaset = {atan(solset),atan(solset)+`π`}`;
    [`optmag(i,j), optangle(i,j)`] = `giveMaximumRoot(thetaset,` q_1, q_2, q_3, q_4);
  end for
end for

The routine `giveMaximumRoot` substitutes the θ values into (38); it returns the maximum value and the corresponding angle.

ACKNOWLEDGMENTS

This work was supported by the Swiss National Science Foundation under grant 2100-053540. The authors would like thank the reviewers for their constructive comments that made the paper more readable and easy to follow.

REFERENCES

[1] J. Canny, "A Computational Approach to Edge Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 8, no. 6, pp. 679-698, 1986.
[2] S. Sarkar and K.L. Boyer, "On Optimal Infinite Impulse Response Filters," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 13, no. 11, pp. 1154-1170, Nov. 1991.
[3] R. Deriche, "Fast Algorithms for Low-Level Vision," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 12, no. 1, pp. 78-87, Jan. 1990.
[4] W.T. Freeman and E.H. Adelson, "The Design and Use of Steerable Filters," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 13, no. 9, pp. 891-906, Sept. 1991.
[5] P. Perona, "Deformable Kernels for Early Vision," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 17, no. 5, pp. 488-499, May 1991.
[6] R. Manduchi, P. Perona, and D. Shy, "Efficient Deformable Filter Banks," *IEEE Trans. Signal Processing,* vol. 46, no. 4, pp. 1168-1173, 1998.
[7] E.P. Simoncelli and H. Farid, "Steerable Wedge Filters for Local Orientation Analysis," *IEEE Trans. Image Processing,* vol. 5, no. 9, pp. 1377-1382, 1996.
[8] P.C. Teo and Y. Hel-Or, "Design of Multi-Parameter Steerable Functions Using Cascade-Basis Reduction," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 21, no. 6, pp. 552-556, June 1999.
[9] P Campisi and G. Scarano, "A Multiresolution Approach for Texture Synthesis Using the Circular Harmonic Functions," *IEEE Trans. Image Processing,* vol. 11, pp. 37-51, 2002.

[10] M.N. Do and M. Vetterli, "Rotation Invariant Texture Characterization and Retrieval Using Steerable Wavelet-Domain Hidden Markov Models," *IEEE Trans. Multimedia,* pp. 146-158, 2002.
[11] E.P. Simoncelli and W. T. Freeman, "The Steerable Pyramid: A Flexible Architecture for Multi-Scale Derivative Computation," *Proc. Int'l Conf. Image Processing,* pp. 444-447, 1995.
[12] A.F. Frangi, W.J. Niessen, K.L. Vincken, and M.A. Viergever, "Multiscale Vessel Enhancement Filtering," *Medical Image Computing and Computer-Assisted Intervention,* vol. 1496, pp. 130-137, 1998.
[13] M. Jacob, T. Blu, and M. Unser, "3D Reconstruction of DNA Filaments from Stereo Cryo-Electron Micrographs," *Proc. IEEE Int'l Symp. Biomedical Imaging: Macro to Nano,* vol. 2, pp. 597-600, 2002.
[14] D. Eberly, R. Gardner, B. Morse, S. Pizer, and C Scharlah, "Ridges for Image Analysis," *J. Math. Imaging and Vision,* pp. 353-373, 1994.
[15] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes in C.* Cambridge Univ. Press 1997.
[16] M.A. Ruzon and C. Tomasi, "Color Edge Detection with the Compass Operator," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* vol. 2, pp. 160-166, 1999.
[17] R. Deriche and G. Giraudon, "A Computational Approach for Corner and Vertex Detection," *The Int'l J. Computer Vision,* vol. 10, no. 2, pp. 101-124, 1993.
[18] L. Kitchen and A. Rosenfield, "Gray Level Corner Detection," *Pattern Recognition Letters,* pp. 95-102, 1982.
[19] Z. Zheng, H. Wang, and E.K. Teoh, "Analysis of Gray Level Corner Detection," *Pattern Recognition Letters,* vol. 20, pp. 149-162, 1999.
[20] H. Wang and M. Brady, "Real Time Corner Detection Algorithm for Motion Estimation," *Image and Vision Computing,* vol. 13, no. 9, pp. 695-703, 1995.
[21] R. Deriche and O. Faugeras, "2D Curve Matching Using High Curvature Points: Application to Stereo Vision," *Proc. 10th Int'l Conf. Pattern Recognition,* 1990 vol. 1, pp. 240-242, 1990.
[22] A.A. Bharath, "Steerable Filters from Erlang Functions," *Proc. British Machine Vision Conf.,* pp. 144-153, 1998.
[23] D. Demigny, "On Optimal Linear Filtering for Edge Detection," *IEEE Trans. Image Processing,* vol. 11, no. 7, pp. 728-737, 2002.
[24] W. Rasband, Image J, http://rsb.info.nih.gov/ij/. 2004.

Mathews Jacob received the ME and PhD degrees from the Indian Institute of Science, Bangalore in January 1999 and the Swiss Federal Institute of Technology, Lausanne, in September 2003, respectively. He is currently a Beckman fellow at the University of Illinois at Urbana Champaign. His research interests include biomedical shape processing, inverse problems in imaging, and sampling theory. He is a member of the IEEE.

Michael Unser (M'89-SM'94-F'99) received the MS (summa cum laude) and PhD degrees in electrical engineering in 1981 and 1984, respectively, from the Swiss Federal Institute of Technology in Lausanne (EPFL), Switzerland. From 1985 to 1997, he was with the Biomedical Engineering and Instrumentation Program, National Institutes of Health, Bethesda, Maryland. He is now a professor and the head of the Biomedical Imaging Group at the EPFL. His main research area is biomedical image processing. He has a strong interest in sampling theories, multiresolution algorithms, wavelets, and the use of splines for image processing. He is the author of more than 100 published journal papers in these areas. Dr. Unser is the associate editor-in-chief for the *IEEE Transactions on Medical Imaging*. He is on the editorial boards of several other journals, including *IEEE Signal Processing Magazine*, *Signal Processing*, *IEEE Transactions on Image Processing* (1992-1995), and *IEEE Signal Processing Letters* (1994-1998). He serves as regular chair for the SPIE conference on wavelets, held annually since 1993. He was general cochair of the First IEEE International Symposium on Biomedical Imaging, held in Washington, DC, 2002. He received the 1995 best paper award and the 2000 Magazine Award from the IEEE Signal Processing Society. In January 1999, he was elected a fellow of the IEEE with the citation: "for contributions to the theory and practice of splines in signal processing."