OPTIMAL STEERABLE FILTERS FOR FEATURE DETECTION

Mathews Jacob and Michael Unser

Biomedical Imaging Group,
Swiss Federal Institute of Technology Lausanne,
CH-1015, Switzerland.

ABSTRACT

We present a new approach for the design of optimal steerable 2-D templates for feature detection. As opposed to classical schemes where the optimal 1-D template is derived and extended to 2-D, we directly obtain the 2-D template. We choose the template from a class of steerable functions based on the analytic optimization of a Canny-like criterion. Our approach gives more orientation selective templates that have simple closed form expression. We illustrate the method with the design of operators for edge and ridge detection and demonstrate their performance improvement in practical applications.

1. INTRODUCTION

In his seminal paper on computational edge detection, Canny proposed an approach to derive optimal operators for the detection of image features [1]. His results on optimal edge detection, in particular, had a great impact on the field and stimulated further developments using different optimality criteria and design strategies [1–3]. All these authors considered the derivation of 1-D operators and extended the detectors to 2-D by applying the optimal 1-D operator in the direction orthogonal to the feature boundary while smoothing in the perpendicular direction (along the boundary). As the optimal 1-D templates did not have explicit formulae, they were typically approximated by simple first or second order differentials of a Gaussian that could be applied inexpensively in a directional manner via the computation of smoothed image gradients or Hessians.

An alternative to these differential approaches to rotation-independant feature detection is provided by the elegant work of Freeman and Adelson on steerable filters [6]. The underlying principle there 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 and Simoncelli used this framework to approximate/design orientation-selective feature templates [4, 5]. In this paper, we are proposing to reconcile the two methodologies—computational approach and steerable filterbanks—by presenting a general strategy for the design of 2-D steerable feature detectors. Our filter optimization is in 2-D as opposed to 1-D as was the case with the computational approaches of Canny and others. Moreover, in contrast with the work of Perona, we do not approximate a given template within a steerable solution space, but we search for the filter that gives the best response according to an optimality criterion. The latter is specified to provide the best compromize in terms of signal-to-noise ratio, false detections and localization.

THIS WORK WAS SUPPORTED BY THE SWISS NATIONAL SCIENCE FOUNDATION UNDER GRANT 2100-053540

2. FEATURE DETECTION

2.1. Rotating matched filtering

Suppose our task is to detect some feature in an image $f(\mathbf{x})$ 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 and rotated versions of a 2-D feature template $h(-\mathbf{x})$ at every point in the image. A high magnitude of the inner-product indicates the presence of the feature; the angle of the corresponding template gives the orientation. Mathematically, the estimation algorithm is

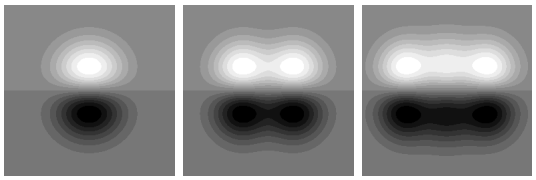
$$\theta^*(\mathbf{x}) = \arg\max_{\theta} \left(f(\mathbf{x}) * h(\mathbf{R}_\theta \mathbf{x})\right) \quad (1)$$

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

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. This algorithm is optimal in the maximum likelihood sense (for additive white Gaussian noise model) but it is not very practical for it requires the implementation of a large number of filters.

2.2. Steerable filters

To cut down on the computational load, we select our detector within the class of steerable filters defined by Freeman et. al [6]. These filters can be rotated very efficiently by



(a) (b) (c)

Fig. 1. Edge Detectors for different parameters (a) Canny's detector (b) $M = 3$; $\mu = 0.09$ (c) M=5; $\mu = 0.15$

taking a suitable linear combination of a small number of filters. Specifically, we consider templates of the form

$$h(x,y) = \sum_{k=1}^{M} \sum_{i=0}^{k} \alpha_{k,i}\, g_{k,i}(x,y), \qquad (3)$$

where $g_{k,i}(x,y) = \frac{\partial^{k-i}}{\partial x^{k-i}} \frac{\partial^{i}}{\partial y^{i}}\, g(x,y)$ and $g(x,y)$ is an arbitrary isotropic window function. We can prove that such an M^{th} order filter is steerable. This means that the convolution of a signal $f(x,y)$ with any rotated version of $h(x,y)$ can be evaluated as

$$f(\mathbf{x}) * h(\mathbf{R}_\theta \mathbf{x}) = \sum_{k=1}^{M} \sum_{i=0}^{k} b_{k,i}(\theta)\, f_{k,i}(\mathbf{x}), \qquad (4)$$

where $f_{k,i} = f * g_{k,i}$ and where the weights $b_{k,i}(\theta)$ are polynomials of degree M in $\cos(\theta)$ and $\sin(\theta)$. Since the number of partial differentials in (4) for a general M^{th} order template is $L = M(M+3)/2$, $h(x,y)$ is steerable in terms of as many individual functions.

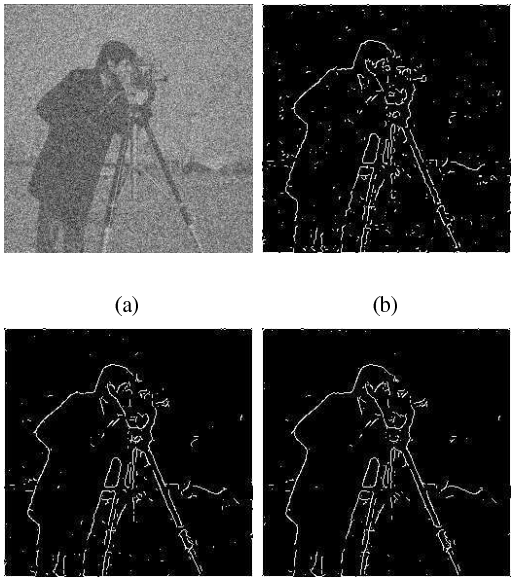
Interestingly, when $g(\mathbf{x})$ is a Gaussian, the family is equivalent to the class of moment filters discussed in [6], but the filters are not identical.

3. DESIGN OF STEERABLE FILTERS

The widely-used contour extraction algorithm [1] has three steps: (a) feature detection, (b) non-maximum suppression, and, (c) thresholding. We now present a general strategy for the design of steerable filters for feature detection, 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 in (3)—for that particular class of steerable functions.

3.1. Optimality Criterion

We now review Canny's criterion and modify it slightly to enable analytical optimization. To derive the optimal 2-D operator, we assume that the feature (edge/ridge) is oriented in some direction 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 3 different terms in Canny's criterion are as follows:



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

Fig. 2. (a) Noisy Image (b) Classical gradient-based detector (Canny) (c) $M = 3$; $\mu = 0.09$ (d) $M = 5$; $\mu = 0.15$

3.1.1. Signal-to-Noise Ratio

The response of a filter $h(\mathbf{x})$ to a particular signal $f_0(\mathbf{x})$ (eg. idealized edge) centered at the origin is given by

$$S = \int_{\mathcal{R}^2} f_0(x,y)\, h(-x,-y)\, dx\, dy \qquad (5)$$

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:

$$\text{Noise} = \int_{\mathcal{R}^2} |h(x,y)|^2\, dx\, dy \qquad (6)$$

We desire to have a large S for a given value of Noise.

3.1.2. Localization

Since the estimated feature position corresponds to the location of the local maximum of the response, the presence of noise can cause an undesirable shift in the estimated feature location. To reduce this shift, we maximize the second derivative of the response, orthogonal to the boundary, at the origin

$$\text{Loc} = -\int_{\mathcal{R}^2} f_0(x,y)\, h_{yy}(-x,-y)\, dx\, dy$$

A high second derivative of the response implies a narrow peak, thus resulting in a low shift-variance.

3.1.3. Elimination of false maximas

In 2-D, we desire that the response be relatively free of false maximas orthogonal to the feature boundary. The average distance between the false maximas can be maximized by penalizing the oscillation term :

$$R_o = \int_{\mathcal{R}^2} |h_{yy}(x,y)|^2 \, dx \, dy \tag{7}$$

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

$$R_p = \int_{\mathcal{R}^2} |h_{xx}(x,y)|^2 \, dx \, dy \tag{8}$$

These terms will force the filter to be smooth; the response is less oscillatory leading to fewer false detections.

3.2. Derivation of the optimal detector

We combine the individual terms to obtain a single criterion

$$C = S \cdot \mathrm{Loc} - \mu \underbrace{(R_o + R_p)}_{R} \tag{9}$$

The filter in the class (3) which maximizes this criterion, subject to the constraint $\mathrm{Noise} = 1$, is the optimal detector. The free parameter $\mu > 0$ controls the smoothness of the filter. To make the solution invariant to the dilation of the window (by a factor σ), we weigh each of the terms in (9) with an appropriate power of σ. For the ease of notation, we collect the component functions of (3) into a function vector $\mathbf{g}$ of length L; an arbitrary function in the family is represented in a compact form as

$$h(\mathbf{x}) = \mathbf{a}^{\mathrm{T}} \mathbf{g}(\mathbf{x}) \tag{10}$$

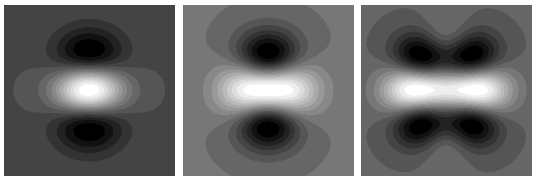
where $\mathbf{a}$ is the vector containing the $\alpha_{i,k}$'s in (3). We now express the terms of the criterion in a matrix form as $S = \mathbf{a}^{\mathrm{T}}\mathbf{s}$, $\mathrm{Loc} = \mathbf{a}^{\mathrm{T}}\mathbf{q}$, $\mathrm{Noise} = \mathbf{a}^{\mathrm{T}}\mathbf{P}\,\mathbf{a}$ and $R = \mathbf{a}^{\mathrm{T}}\mathbf{R}\,\mathbf{a}$, where

$$[\mathbf{s}]_i = \langle f_0(\mathbf{x}), [\mathbf{g}(-\mathbf{x})]_i \rangle_{L_2} \tag{11}$$

$$[\mathbf{q}]_i = \left\langle f_0(\mathbf{x}), ([\mathbf{g}(-\mathbf{x})]_i)_{yy} \right\rangle_{L_2} \tag{12}$$

$$[\mathbf{P}]_{i,j} = \left\langle [\mathbf{g}]_i, [\mathbf{g}]_j \right\rangle_{L_2} \tag{13}$$

$$[\mathbf{R}]_{i,j} = \left\langle ([\mathbf{g}]_i)_{yy}, \left([\mathbf{g}]_j\right)_{yy} \right\rangle_{L_2} + \left\langle ([\mathbf{g}]_i)_{xx}, \left([\mathbf{g}]_j\right)_{xx} \right\rangle_{L_2} \tag{14}$$



(a) (b) (c)

Fig. 3. Ridge Detectors for different parameters (a) $M = 2; \mu = 2$, (b) $M = 2; \mu = 0$, (c) $M = 4; \mu = 0.25$

Thus, (9) can be expressed in the matrix form as

$$C = \mathbf{a}^{\mathrm{T}} [\mathbf{Q} - \mu \mathbf{R}] \, \mathbf{a}, \tag{15}$$

where $\mathbf{Q} = \mathbf{s}\,\mathbf{q}^{\mathrm{T}}$. Since all the terms are quadratic, the optimal parameters can be found analytically by using Lagrange's multiplier method. Maximizing (15) subject to the constraint, we get

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

Rearranging the terms and using the fact that $\mathbf{P}$ is invertible, we get

$$\mathbf{P}^{-1} [\mathbf{Q} - \mu \mathbf{R}] \;\; \mathbf{a} = -\lambda \, \mathbf{a} \tag{16}$$

which implies that λ is an eigenvalue of the matrix $-\mathbf{P}^{-1}[\mathbf{Q} - \mu\mathbf{R}]$. The corresponding L 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

$$\bar{\mathbf{a}} = \max \left\{ \mathbf{a}_{\lambda_i}^{\mathrm{T}} [\mathbf{Q} - \mu \mathbf{R}] \, \mathbf{a}_{\lambda_i}; \; i = 0 ... L \right\} \tag{17}$$

3.3. Feature detection by local optimization

The optimal angle θ^* in (1) is obtained as the solution of $\frac{\partial}{\partial \theta} (f(\mathbf{x}) * h(\mathbf{R}_{\theta^*}\mathbf{x})) = 0$. Using (4), we get

$$\sum_{k=1}^{M} \sum_{i=0}^{k} f_{k,i}(x,y) \underbrace{\frac{\partial}{\partial \theta} (b_{k,i}(\theta)) \, |_{\theta=\theta^*}}_{c_{k,i}(\theta^*)} = 0 \tag{18}$$

(18) is a polynomial of order M (in $\cos(\theta)$ and $\sin(\theta)$). If $h(x,y)$ has only odd/even order partial derivatives, (18) can be simplified to a polynomial in only one variable—$\tan(\theta)$; we have an analytic solution when $M <= 3$; otherwise it is obtained using iterative refinement of an initial approximate solution from a lower order detector.

4. EXAMPLES & APPLICATIONS

We now design operators optimized for the detection of different 2-D features. We chose the window function to be a Gaussian $g(\mathbf{x}; \sigma)$, where σ is the standard deviation.

4.1. Edge Detection

As model for the edge, we choose the Heaviside step function; ie. $f_0(x,y) = U(y)$. Since it is an odd function of y, the even order derivatives do not contribute to the signal energy; we therefore ignore them in (3). When $M = 1$, we find that the optimal template is equivalent to the well known Canny's edge detector. For higher M, we obtain a family of solutions that are increasingly smooth when μ goes up. The impulse responses of some optimal templates are shown in Fig. 1. The higher order templates are more elongated thus having higher SNR and Localization . We have an analytic solution for for the optimal angle in (1) or (18) when $M = 1$ and $M = 3$. For $M > 3$, we refine the $M = 3$ solution using a golden search.

To illustrate the algorithm, we perform edge detection on the cameraman image corrupted with additive white noise (c.f. Fig. 2-a). The size of the Gaussian window is the same in all the experiments. The detected edges after non-maximum suppression and thresholding are presented in Fig. 2. The Canny's edge detector gives a lot of false detections and wiggly contours. The new detectors have significantly lower false detections and better localization, thus confirming the theoretical improvement.

4.2. Ridge Detection

We chose the idealized line model as $f_0(x,y) = \delta(y)$, where δ denotes the Dirac's delta function. 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.

Some examples of optimal templates are shown in Fig. 3. Interestingly, the template Fig. 3-b is more directional than the classical one in Fig. 3-a (2nd derivative of a Gaussian) which has the same computational complexity. For $M = 2$, the optimal direction and ridge magnitude can be computed with the eigen-decomposition of $\mathbf{H}_{\text{mod}}$. For fourth order detectors, we initially compute the $M = 2$ solution and refine the criterion around the current solution.

An interesting application is the detection of DNA filaments (cf Fig. 4-a) from cryo-electron micrographs. The micrographs are extremely noisy because they are exposed to a low electron dose to avoid the degradation of the specimen. The results (Fig. 4-b - 4-d) correspond to the output of ridge detection algorithm followed by non-maximum 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.

5. CONCLUSIONS

We proposed a general approach to derive optimal 2-D operators from a family of steerable functions for feature detection. We proposed an analytical optimization scheme based on a slight modification of Canny's criterion. We used this method to derive optimal operator for edge and ridge profiles and demonstrated their utility in practical applications; the results are promising.

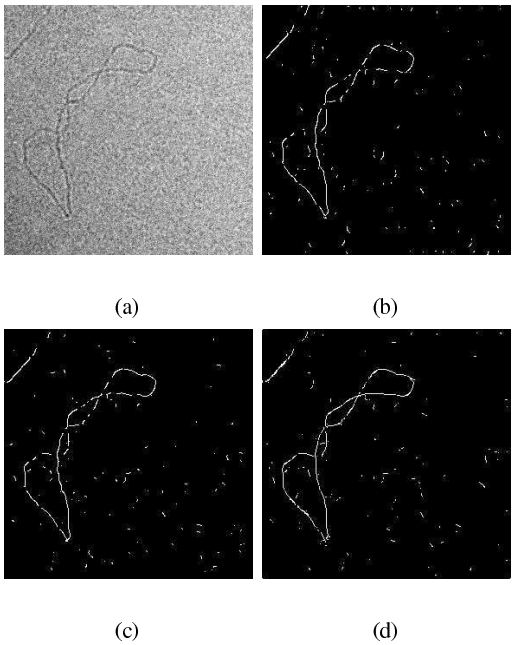


Fig. 4. (a) Electron micrograph of DNA (b) Hessian based classical detector (c) $M = 2; \mu = 0$ (d) $M = 4; \mu = 0.15$

6. REFERENCES

[1] J. Canny, "A computational approach to edge detection," *IEEE Transactions on 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 Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 11, pp. 1154–1170, 1991.

[3] R. Deriche, "Fast algorithms for low-level vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 78–87, 1990.

[4] P. Perona, "Deformable kernels for early vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 5, pp. 488 –499, 1991.

[5] E.P. Simoncelli and H. Farid, "Steerable wedge filters for local orientation analysis," *IEEE Transactions on Image Processing*, vol. 5, no. 9, pp. 1377–1382, 1996.

[6] W.T. Freeman and E.H. Adelson, "The design and use of steerable filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 9, pp. 891–906, 1991.