Proceedings of the 1991 IEEE
International Conference on Robotics and Automation
Sacramento, California - April 1991

Determining Optimal Scales for Edge Detection Using Regularization[1]

Hong Jeong[2], Chang-Ik Kim, and Woon-Tack Woo

Pohang Institute of Science and Technology

Abstract

This paper suggests a regularization method for determining scales of edge detector adaptively for each part of the image. We extend the concept of optimal filter by Poggio, et al. and scale space by Witkin to an adaptive scale parameter.

We first introduce an energy function defined over continuous scale space. Natural constraints for edge detection are incorporated into the energy function. To obtain a set of optimal scales which can minimize the energy function, a parallel relaxation algorithm is introduced. Experiments for synthetic and natural scenes show the advantages of the new algorithm over the edge detectors by Marr and Hildreth or Canny.

I Introduction

In this paper, we consider edge detection as the problem of locating and measuring changes of light intensity in the image. The higher goal to detect and locate physical edges in the three dimensional surfaces being imaged is beyond the scope of this paper.

The size of the filter with which to perform edge detection has always been an unresolved issue in computer vision. This problem is also known to be ill-posed in the sense of Hadamard [8]. Marr and Hildreth [6, 4] have argued that the Gaussian is an optimal filter for edge detection due to its localization properties in both the spatial and frequency domains. Since then, the Gaussian has been used in a number of edge detectors: e.g., the algorithms by Canny [2] and Bergholm [1].

Poggio, et al. [9] proved that the Gaussian is indeed an optimal filter for estimating ideal data from incomplete data defined over a discrete grid. Using the regularization method, Poggio, et al. [9, 10] defined the edge detection problem: find g that minimizes

$$\int\int[(Sg-f)^2+\lambda(\nabla^2 g)^2]dxdy, \qquad (1)$$

where the data on image intensity ($f = f(x,y)$) are given on a discrete lattice, the operator S is the sampling operator on the continuous distribution g to be recovered, and λ is a stabilizing constant. The purpose of this scheme is first to find an estimate of the original image which generated the given degraded image. After then, we take a suitable derivative of the original image to detect the intensity variations.

They showed that (a) the solution g of this problem can be obtained by convolving the data $f(x,y)$ with a convolution filter $R(x,y)$, and (b) that the filter $R(x,y)$ is a cubic spline with a shape very close to a Gaussian $G(x,y)$ and a size controlled by the regularization parameter λ which is equivalent to the role of σ for the Gaussian ($\lambda \approx \sigma^4$). Differentiation can then be accomplished by convolving the data with an appropriate derivative of this filter.

Although the shape of optimal filter has been derived, it is not yet known how to determine an optimal size in terms of λ (equivalently σ for Gaussian) when errors on the data or on the smoothness conditions are not known in advance [8].

To remove noise and keep good localization, edge detection must be performed at different scales of resolution. The basic strategy for this problem is that different parts of the image must be blurred differently, depending on the degree of noisiness, before further processing is involved.

Marr and Poggio [5], Marr and Hildreth [6, 4] have

[1]This research has been supported by the grants from the Ministry of Education, the Ministry of Science and Technology and the Research Institute of Science and Technology during the period of 1990-91, in Korea

[2]Department of EE, Pohang Institute of Science and Technology, P.O.Box 125, Pohang 790-330, South Korea, Tel: (0562) 79-2223, FAX:(0562) 79-2903, E-MAIL:hjeong@vision.postech.ac.kr@RELAY.CS.NET

used several sizes of filters in order to perform edge detection. However, there arise problems how to integrate the output of different scales. The basic strategy of the Canny's feature synthesis [3] is using several filters of discrete scales and combining the edge maps using a predefined rule. The underlying problem is that there is no known optimal method to integrate the edge maps.

In order to choose the optimal scale, Witkin [11] proposed a method that combines output from different scale levels by treating scale as being continuous - essentially filtering across a continuous scale value. The problem is that finding the natural scale of the filter, as suggested by Witkin, is based on heuristics. To avoid this problem and the matching problem in approaches using logarithmically spaced, independent filters, Bergholm [1] introduced a method in which the edges are tracked from coarse to fine resolution. This rule-based approach, however, requires some complicated rules for detecting local features and for tracking and predicting possible scale parameter.

In fact, combining the output from scales of different sizes, either continuous or discrete, is an ill-posed problem, for the solution does not exist or is not unique or does not depend continuously on the data. The following sections show that the ill-posed problem can be regularized by introducing suitable stabilizing functional.

II Determining Optimal Scales

In order to determine optimal scales for a Gaussian filter, we could define an energy function which determines quantitatively the usefulness of the underlying edge map. For any edge detection method which uses uniform blurring and localization of maximum derivative, the basic conflict between noise and localization remains. To bypass the conflict, we discard the concept of uniform blurring and instead use adaptive filters for noise reduction which use different scales for different parts of the image.

If the noise in an image can be assumed to be random and ergodic, the average taken over a neighborhood goes to zero as the extent of the neighborhood expands. On the other hand, a large filter tends to have influence on the location of the intensity discontinuities. To achieve the conflicting goals, i.e., suppressing false detection and localizing accurate changes of light intensity, it is certain that the size of filters must be determined adaptively fir different parts according to some context of the image.

Under this assumption, the edge detection problem reduces to the following formulation. If we assume $\sigma(x,y)$ which is the size parameter of a Gaussian filter, and that $E(\sigma)$ denotes an energy defined over the function σ, then the edge detection problem reduces to the following problem:

$$\left\{ \begin{array}{ll} \text{Find} & \sigma(x,y) \\ \text{such that} & \min_\sigma \quad E(\sigma) \end{array} \right. . \tag{2}$$

Since the energy is a functional of $\sigma(x,y)$, the underlying natural constraints must be expressed in terms of σ. Also the energy function must be bounded below so that its attractors are located at the states of the desired edge configuration.

It is important to notice that the Gaussian filter $G(\sigma(x,y),x,y)$ is a functional of the variable $\sigma(x,y)$ which is in turn a function of the coordinates (x,y). In fact, this adaptive filter is space variant and noncausal. Therefore, we can hypothesize that a better estimate of the ideal image $g(x,y)$ is $G(\sigma,x,y) * f(x,y)$, where $*$ means convolution. In addition, we assume a constraint on the variation of $\sigma(x,y)$. For example, $\sigma(x,y)$ should not change abruptly from pixel to pixel to discourage fragmented edges due to noise. Then, the regularization of the energy consists of two terms.

$$\begin{aligned} E(\sigma) &= \int\int (f - G*f)^2 + \lambda |\nabla \sigma(x,y)|^2 dxdy \\ &= \int\int (f - [\int\int \frac{1}{2\pi\sigma^2} \exp(\frac{-(\alpha^2+\beta^2)}{2\sigma^2}) \\ &\quad f(x-\alpha, y-\beta) d\alpha d\beta])^2 \\ &+ \lambda[(\frac{\partial\sigma}{\partial x})^2 + (\frac{\partial\sigma}{\partial y})^2] dxdy. \end{aligned} \tag{3}$$

A set of $\sigma(x,y)$ minimizing this energy will satisfy the conditions: (1) $\sigma(x,y)$ must be large at uniform intensity area, thus smoothing out random noise, (2) $\sigma(x,y)$ must be small at change of intensities, thus recovering edges very accurately, and (3) $\sigma(x,y)$ shouldn't change abruptly from pixel to pixel to prevent broken edges due to the random noise. The first and the second properties correspond to the first term of Eq. (3), while the third property corresponds to the second term of this equation.

An alternative method for finding $\sigma(x,y)$ which minimizes Eq. (3) is converting this integrodifferential equation into Euler-Lagrange equation. Even so, we cannot easily obtain a closed form of $\sigma(x,y)$. A more practical approach is the use of numerical calculation as we will see in the next section.

III A Relaxation Scheme for Minimizing Energy Function

To minimize the energy by a relaxation algorithm, we must first find a discrete form of Eq. (3). Replacing the continuous coordinates (x, y) by the discrete coordinates (i, j) and leaving (α, β) untouched for convenience, we get

$$E = \sum_{ij}\{(f_{ij} - [\sum_{\alpha\beta} \frac{1}{2\pi\sigma_{ij}^2} \exp(\frac{-(\alpha^2+\beta^2)}{2\sigma_{ij}^2}) f_{i-\alpha j-\beta}])^2 + \lambda[(\sigma_{i+1j} - \sigma_{ij})^2 + (\sigma_{ij+1} - \sigma_{ij})^2]\}. \quad (4)$$

Note that the intensity $f(x, y)$ is replaced by f_{ij} and the parameter $\sigma(x, y)$ by σ_{ij}. Also, the domain of (i, j) and (α, β) is the $N \times M$ image plane.

Taking a derivative of this equation in terms of the function σ_{ij}, we obtain

$$\frac{\partial E}{\partial \sigma_{ij}} = -2(f_{ij} - \sum_{\alpha\beta} \frac{1}{2\pi\sigma_{ij}^2} \exp(\frac{-(\alpha^2+\beta^2)}{2\sigma_{ij}^2}) f_{i-\alpha j-\beta}) \times \sum_{\alpha\beta}[\frac{\alpha^2+\beta^2-2\sigma_{ij}^2}{2\pi\sigma_{ij}^5} \exp(\frac{-(\alpha^2+\beta^2)}{2\sigma_{ij}^2}) f_{i-\alpha j-\beta}] + 2\lambda(4\sigma_{ij} - \sigma_{i+1j} - \sigma_{ij+1} - \sigma_{i-1j} - \sigma_{ij-1}). \quad (5)$$

Here, $i \in \{0, 1, \ldots, N-1\}$ and $j \in \{0, 1, \ldots, M-1\}$. An optimal solution σ_{ij} that minimizes E must satisfy the condition:

$$\frac{\partial E}{\partial \sigma_{ij}} = 0. \quad (6)$$

To find a minimum of this equation by iterative calculations, we can use the gradient descent method:

$$\sigma_{ij}^{n+1} = \sigma_{ij}^n - \eta \frac{\partial E}{\partial \sigma_{ij}} \Delta\sigma + m\underline{\sigma}^{n-1}, \quad (7)$$

where n (≥ 0) represents the iteration number, $\Delta\sigma$ is the unit step size for each iteration, η is a parameter to be determined heuristically (usually $\eta = 1.0$), and m is a momentum parameter (usually $m = 0$).

The state of this system is defined by $\underline{\sigma} = \{\sigma_{ij}\}$. With this notation, the dynamics of this system is represented by

$$\underline{\sigma}^{n+1} = \underline{\sigma}^n - \nabla_\sigma E(\underline{\sigma}^n)\Delta\sigma, \quad (8)$$

where $\underline{\sigma}^0$ is an initial state, and ∇_σ is a gradient in the direction of $\underline{\sigma}$. The trajectory of this dynamical equation is chosen in such a way that the energy $E(\underline{\sigma}^n)$ decreases steadily along the path. Hence, $E(\underline{\sigma}^n)$ is a Lyapunov function for this dynamical equation.

Note that this algorithm converges to a minimum point nearest to the initial state. In general, the gradient search method has the property of converging to one of the local minima. Thus the initial value of σ must be chosen optimally(in this experiment, $\sigma^0 = 4.0$). We assume that the energy is analytic and also that the energy is bounded below, i.e., $E \geq 0$.

A complete form of the relaxation equation is given by Eq. (9).

$$\sigma_{ij}^{n+1} = \sigma_{ij}^n - \{-2(f_{ij} - \sum_{\alpha\beta} \frac{1}{2\pi\sigma_{ij}^2} \exp(\frac{-(\alpha^2+\beta^2)}{2\sigma_{ij}^2}) f_{i-\alpha j-\beta}) \times \sum_{\alpha\beta}[\frac{\alpha^2+\beta^2-2\sigma_{ij}^2}{2\pi\sigma_{ij}^5} \exp(\frac{-(\alpha^2+\beta^2)}{2\sigma_{ij}^2}) \times f_{i-\alpha j-\beta}] + 2\lambda(4\sigma_{ij} - \sigma_{i+1j} - \sigma_{ij+1} - \sigma_{i-1j} - \sigma_{ij-1})\}\Delta\sigma. \quad (9)$$

In the above equation we must be careful in assigning the value of $\Delta\sigma$ and λ. If $\Delta\sigma$ is smaller than predefined value, the convergence speed becomes slow. Conversely, if $\Delta\sigma$ is larger, σ may diverge. Also, the stabilizing constant λ is used to control the smoothness. In this experiment, we assigned 0.0001 to $\Delta\sigma$ and 1000 to λ. This equation can be computed by an array processor. A processing element in this array stores and updates the state σ_{ij}^n by using information coming from nearby processors, together with its previous state. Each path delivers the state as well as the image intensity of the nearby processor. Hence, the states $\underline{\sigma}^n$ of this system can be updated simultaneously (also synchronously) according to Eq. (9).

Either Jacobi or Gauss-Seidel method can be used for simulating this equation. To terminate the iteration, we can define in advance either the maximum number of iterations or a lower bound of the change of σ in successive steps.

IV Results

Our algorithm has been tested for both the Marr and Hildreth edge detector and the Canny edge detector. In both cases, the first step is to compute $\underline{\sigma}$ as stated above. Using these size specifications, the discrete supports for $G(x, y)$, having $\sigma(x, y)$ as its size parameter and defined over a small rectangular region, are computed. Then, these filters have been directly used for the remaining parts of the Marr-Hildreth or the Canny edge detector.

The Marr-Hildreth edge detector [6] can be easily modified so that the adptive Gaussian filters may be used for it. For a given gray scale image $f(x, y)$, we obtain an edge map by the following method: 1) find $\sigma(x, y)$ using Eq. (9), 2) convolve the image $f(x, y)$ with $\nabla^2 G(\sigma(x, y), x, y)$, and 3) find the zero crossings of $(\nabla^2 G) * f$.

Similarly, Canny algorithm [3] can be easily modified. For a given gray scale image $f(x, y)$, we do the following steps: 1) find $\sigma(x, y)$ using Eq. (9), 2) convolve the image $f(x,y)$ with $G(\sigma(x,y),x,y)$, 3) compute the directional derivatives, $\nabla(G(\sigma(x,y),x,y) * f(x,y))$, 4) apply the non-maximum suppression algorithm, and finally 5) perform thresholding.

To examine the performance of the new algorithms, we used three examples of images (256 x 256 with 256 gray levels): two synthesized images and one natural image. The first example, as shown in Fig. 1, is a synthesized image containing a vertical object boundary. A half plane with uniform intensity is first produced

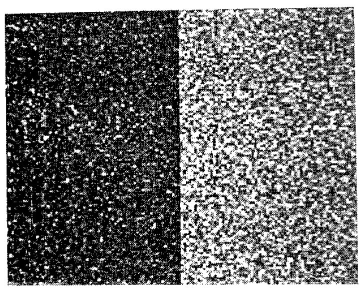

Figure 1: A vertical edge with additive Gaussian noise.

and then Gaussian noise with zero mean and $\sigma = 1$ has been added to the binary image. The primary purpose of this experiment is to observe the behavior of the dynamical equation around the vertical boundaries especially in term of the adaptation capability of the size parameter.

The results are shown in Fig. 2. In this figure, (a) must be compared with (b), while (c) must be compared with (d). The edge map (a) is the result of Marr-Hildreth and (c) of Canny. The results of the new algorithms are shown in (b) and (d). For the image pairs being compared: (a) and (b), and (c) and (d), the same threshold parameters have been maintained. For (a) and (c), $\sigma = 1$ and for (c) and (d), the same single threshold was used to compare under the same condition. It is easy to see that (b) is better than (a) and (d) is better than (c). The new algorithms tend to remove the noise noticeably across the uniform intensity areas while detecting the vertical edges very accurately.

Fig. 3 shows a horizontal crosssection of the image $f(x,y)$ and the computed $\sigma(x,y)$ for this part of image. Note that, as expected, $\sigma(x,y)$ is large on the left and the right areas where change of intensity is minimal, while $\sigma(x,y)$ is small, i.e., $\sigma \approx 1.0$, where change of

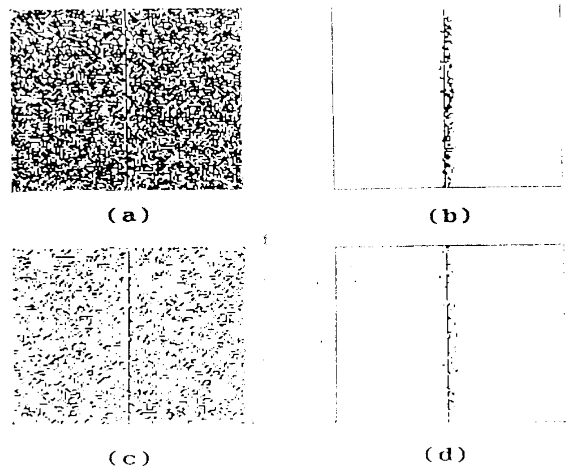

Figure 2: Edge maps: (a) Marr-Hildreth(σ = 1.0), (b) Marr-Hildreth with adaptive scale, (c) Canny(σ = 1.0) and (d) Canny with adaptive scale.

intensity is maximal, i.e., $x \approx 62$. Also note that the filter size varies slowly across the object boundaries.

To see the performance of the edge detectors around the sharp object corners, we synthesized an image as shown in Fig. 4. Analogously to the previous experiment, Gaussian noise has been added to the binary image which contains a rectangular object with uniform brightness.

Fig. 5 contains edge maps. The four pictures are obtained and arranged in the fashion stated above. As before, we can observe that (b) and (d) are respectively better than (a) and (c). In particular, the sharp corners of the rectagular object have been successfully detected by the new algorithms.

The final example is a natural scene as shown in Fig. 6. The corresponding edge maps are shown in Fig. 7. In (c), we used the threshold with hysteresis but in (d), used a single threshold. After a number of trials, the best threshold parameters for hysteresis were chosen and used in (c). Although the original

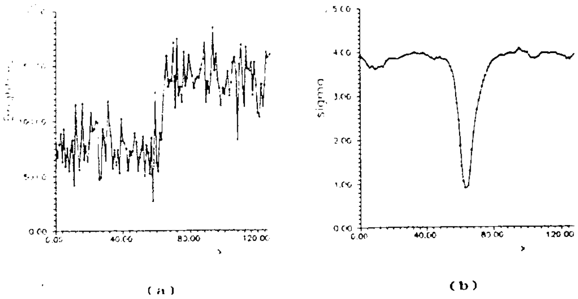

Figure 3: (a) A horizontal crosssection of the image and (b) the corresponding $\sigma(x, y)$.

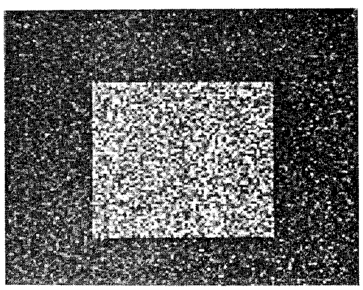

Figure 4: A rectangle object with additive Gaussian noise.

image is complicated because of the edges, shadows, and noise, the new algorithms tend to remove noise and detect edges very remarkably. We can observe that (b) and (d) are respectively better than (a) and (c).

In summary, it is certain that the size of filter, when it is determined by the dynamical equation, Eq. 9, can vary in an adaptive manner according to the context contained in the image. The problem of this algorithm is its computational complexity. An approach to cope with this problem is using several discrete scales instead of continuous scale as has been used in this algorithm. Then the relaxation equation must perform a combinatoral optimization.

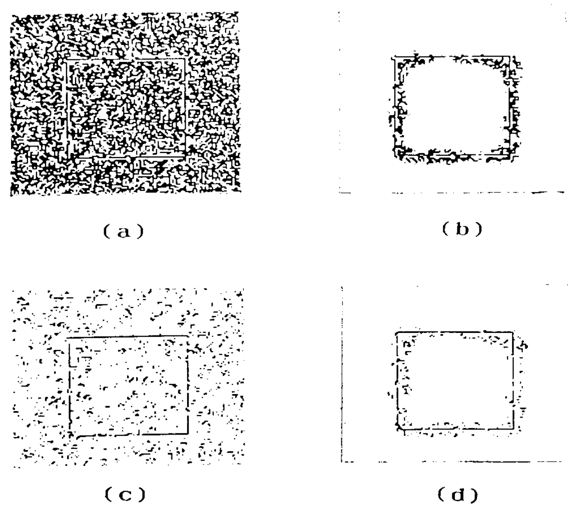

Figure 5: Edge maps: (a) Marr-Hildreth($\sigma = 1.0$), (b) Marr-Hildreth with adaptive scale, (c) Canny($\sigma = 1.0$) and (d) Canny with adaptive scale.

Conclusion

As a method by which to choose the optimal scale for edge detection, this paper analized the limitations of the optimal filtering by Poggio, et al. [7], the scale space filtering by Witkin [11], and the feature synthesis by Canny [3]. In order to overcome the limitations of these methods, we introduced an alternative scheme which can determine optimal scales for each part of the image in an adaptive manner.

An energy function, which contains the natural constraints for optimal edge detection, is suggested. To compute the optimal scales by minimizing the energy function, we introduced a parallel relaxation scheme which can be realized on an array processor. Compared with the edge detectors by Marr and Hildreth [6] or Canny [3], it is experimentally shown that our algorithm has very excellent properties in detecting and localizing edges.

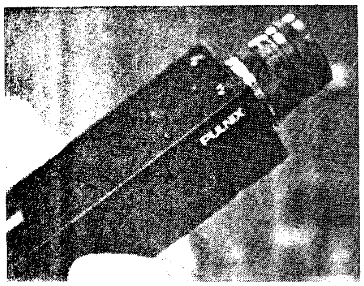

Figure 6: A natural scene.

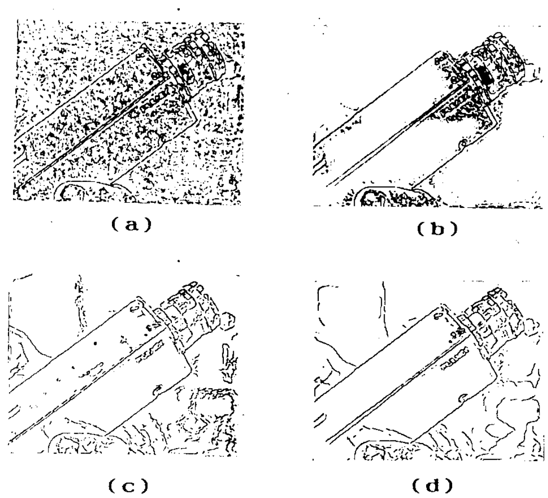

Figure 7: Edge maps: (a) Marr-Hildreth(σ = 1.0), (b) Marr-Hildreth with adaptive scale, (c) Canny with hysteresis(σ = 1.0) and (d) Canny with adaptive scale.

References

[1] F. Bergholm, "Edge focusing," *IEEE Trans. on Patt. Analys. Machine Intell.*, vol. PAMI-9, no. 6, Nov., 1987, pp. 726-741.

[2] J.F. Canny, "Finding edges and lines in images," AI-TR 720, AI Lab, MIT, 1983.

[3] J. F. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, Nov. 1986.

[4] E.C. Hildreth, "Implementation of a theory of edge detection," AI-TR-579, MIT AI Lab, 1980.

[5] D. Marr and T. Poggio, "A computational theory of human stereo vision," in *Proc. Roy. Soc. London B*, vol. 204, 1979, pp. 301-328. (Also in AI Memo 451, MIT AI Lab., 1977.)

[6] D. Marr and E. Hildreth, "Theory of edge detection," *Proc. Roy. Soc. London*, vol. B-207, 1980, pp. 187-217.

[7] T. Poggio, H. Voorhees and A. Yuille, "Regularizing Edge Detection," *AI Memo*, no. 776, AI Lab., MIT, Cambridge, 1984.

[8] T. Poggio, H. Voorhees and A. Yuille, "A regularized solution to edge detection," AI Memo, MIT AI Lab, May 1985.

[9] T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory," *Nature*, vol. 317, n. 6035, Sep. 1985, pp.314-319.

[10] V. Torre and T. Poggio, "On edge detection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 2, March 1986, pp. 147-163.

[11] A. Witkin, "Scale-Space Filtering," *Proc. IJCAI*, 1983, pp. 1019-1021.