

A NETWORK FOR MULTISCALE IMAGE SEGMENTATION

Pietro Perona

Jitendra Malik

Department of Electrical Engineering and Computer Science
University of California at Berkeley - CA 94720 - U.S.A.
e-mail to: perona@united.Berkeley.edu.arpa

ABSTRACT

Detecting edges of objects in their images is a basic problem in computational vision. The scale-space technique introduced by Witkin [11] provides means of using local and global reasoning in locating edges. This approach has a major drawback: it is difficult to obtain accurately the locations of the ‘semantically meaningful’ edges. We have refined the definition of scale-space, and introduced a class of algorithms for implementing it based on using anisotropic diffusion [9]. The algorithms involves simple, local operations replicated over the image making parallel hardware implementation feasible. In this paper we present the major ideas behind the use of scale space, and anisotropic diffusion for edge detection, we show that anisotropic diffusion can enhance edges, we suggest a network implementation of anisotropic diffusion, and provide design criteria for obtaining networks performing scale space, and edge detection. The results of a software implementation are shown.

1 INTRODUCTION

Detecting edges of objects in their images is a basic problem in computational vision and much effort has been devoted to it so far. Surfaces in nature usually [7,6] have a hierarchical organization composed of a small discrete number of levels. At the finest level, a tree is composed of leaves with an intricate structure of veins. At the next level, each leaf is replaced by a single region, and at the highest level there is a single blob corresponding to the treetop. There is a natural range of resolutions (intervals of the scale-space parameter) corresponding to each of these levels of description. Furthermore at each level of description, the regions (leaves, treetops or forests) have well-defined boundaries.

Noise, finer detail, and the structure of the edges themselves, makes local criteria (i.e. decision rules, or operators, based on small fixed-size neighbourhoods) insufficient for determining the existence and location of the edges at different scales of resolution. On the other hand global criteria prove to be hard to formulate, difficult to implement, and computationally expensive. The scale-space technique provides means for taking advantage of both local and global reasoning on images.

The importance of multi-scale descriptions of images has been recognized from the early days of computer vision e.g. Rosenfeld and Thurston [10]. A clean formalism for this problem is the idea of scale-space filtering introduced by Witkin [11] and further developed in Koenderink [6], Babaud, Duda and Witkin [1], Yuille and Poggio [12], and Hummel [4].

The essential idea of this approach is quite simple: embed the original image in a family of derived images $I(x, y, t)$ obtained by convolving the original image with lowpass kernels of aperture $\frac{1}{t}$.

Larger values of t , the scale-space parameter, correspond to images at coarser resolutions. See fig. 1.

Koenderink motivates the choice for a particular lowpass kernel by stating two criteria :

1. *Causality* : Any feature at a coarse level of resolution is required to possess a (not necessarily unique) “cause” at a finer level of resolution although the reverse need not be true. In other words, no

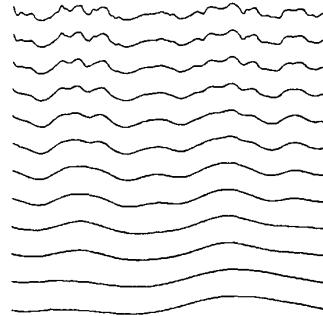


Figure 1: A family of 1D images $I(x,t)$ obtained by convolving the original one (top) with gaussian kernels whose variance increases from top to bottom. (Adapted from Witkin).

spurious detail should be generated when the resolution is diminished.

2. *Homogeneity and Isotropy* : The blurring is required to be space invariant.

It turns out that only the gaussian kernel generates a scale space satisfying these requirements. As pointed out by Koenderink [6], and Hummel [4] the one parameter family of derived images may equivalently be viewed as the solution of the heat conduction or diffusion equation

$$I_t = c(I_{xx} + I_{yy})$$

with the initial condition $I(x, y, 0) = I_0(x, y)$, the original image.

Unfortunately, in the scale-space thus derived the true location of the boundaries at coarse scales is not directly available in the coarse scale images. The edges at the coarse level t_1 are shifted from their true locations due to the interaction of the different brightness “blobs” during diffusion. The only way to obtain the true position of the edges that have been detected at a coarse scale is by tracking across the scale space, to their locations in the original image.

The reason for this spatial distortion is quite obvious—Gaussian blurring does not ‘respect’ the natural boundaries of objects. Suppose we have the picture of a treetop with the sky as background. The gaussian blurring process would result in the green of the leaves getting ‘mixed’ with the blue of the sky, long before the treetop emerges as a feature (after the leaves have been blurred together).

Figure 1 shows a sequence of coarser images obtained by gaussian blurring which illustrates this phenomenon. It may also be noted that the region boundaries are generally quite diffuse instead of being sharp.

With this as motivation, we have enunciated the criteria which we believe any candidate paradigm for generating multi-scale ‘semantically meaningful’ descriptions of images must satisfy.

- Causality** : As pointed out by Witkin and Koenderink, a scale-space representation should have the property that no spurious detail should be generated.
- Immediate Localization** : At each resolution, the region boundaries should be sharp and coincide with the semantically meaningful boundaries at that resolution.
- Piecewise Smoothing**: At all scales, intra-region smoothing should occur preferentially over inter-region smoothing. In the tree example mentioned earlier, the leaf regions should be collapsed to a treetop *before* being merged with the sky background.

In the next section we will outline a framework for achieving these objectives.

2 ANISOTROPIC DIFFUSION

Consider the anisotropic diffusion equation

$$I_t = \operatorname{div}(c(x, y, t) \nabla I) = c(x, y, t) \Delta I + \nabla c \cdot \nabla I$$

where we indicate with div the divergence operator, and with ∇ , and Δ the gradient, and laplacian operators with respect to the space variables. It reduces to the isotropic heat diffusion equation $I_t = c \Delta I$ if $c(x, y, t)$ is a constant. Suppose at the time (scale) t , we knew the locations of the region boundaries appropriate for that scale. We would want to encourage smoothing *within* a region in preference to smoothing *across* the boundaries. This could be achieved by setting the conduction coefficient to be 1 in the interior of each region and 0 at the boundaries. The blurring would then take place separately in each region with no interaction between regions. The region boundaries would remain sharp.

Of course, we do *not* know in advance the region boundaries at each scale (if we did the problem would already have been solved!). What can be computed is a current best estimate of the location of the boundaries (edges) appropriate to that scale. Let $E(x, y, t)$ be a vector valued function defined on the image with the following ideal properties:

- $E(x, y, t) = \mathbf{0}$ in the interior of each region.
- $E(x, y, t) = K e(x, y, t)$ at each edge point, where e is a unit vector tangent to the edge at the point, and K is the local contrast (difference in the image intensities on the left and right) of the edge.

Note that the word *edge* as used above has not been formally defined—we mean here the perceptual subjective notion of an edge as a region boundary. A completely satisfactory formal definition is likely to be part of the solution, rather than the problem definition!

Once an estimate of $E(x, y, t)$ is available, $c(x, y, t)$ is chosen to be a function $c = g(\|E\|)$ of the magnitude of E . According to the previously stated strategy $g(\cdot)$ has to be a nonnegative monotonically decreasing function with $g(0) = 1$. This way the diffusion process will mainly take place in the interior of regions, and it will not affect the region boundaries where the magnitude of E is large. There are many possible choices for $g(\cdot)$ the simplest being a binary valued function. We will also briefly discuss the fact that the function $g(\cdot)$ need not be time invariant, and actually sometimes it should vary with time.

The way the conduction coefficient is chosen guarantees that the diffusion process satisfies the third requirement for generating a scale-space, namely piecewise smoothing. The analysis performed in section 3, and the experimentation (see fig. 4 and fig. 5), will show that the second criterion—immediate localization—is also satisfied. It is important to know which possible choices of the conduction coefficient $c(\cdot)$, if any, respect the first criterion, causality.

For studying causality we make use of a classical result of the theory of partial differential equations : the maximum principle. This theorem guarantees that the maxima and minima of the solution of the anisotropic diffusion equation belong to the initial condition (the original image in our case), and to the boundaries of the domain of interest

(the boundaries of the picture), provided that the conduction coefficient $c(x, y, t)$ is differentiable and non-negative.

This theorem is relevant to our discussion because if we identify “features” in the images with “blobs” of the brightness function $I(x, y, t)$ for different values of the scale parameter, then the maximum principle guarantees that no new features are created for increasing t . In fact the birth of a new “blob” would imply the creation of a maximum, or a minimum, that would have to belong to the interior of the domain of interest in the scale space.

A more precise statement of this theorem, and a proof that anisotropic diffusion satisfies it may be found in [9], or may be derived using techniques in [5].

3 EDGE ENHANCEMENT

A big problem with conventional low-pass filtering and diffusion, is that the price payed for eliminating the noise, and for performing scale space, is the blurring of edges. This causes their detection, and localization to be difficult.

Hummel [3] has suggested enhancing edges by running backwards the diffusion equation. This is an ill-posed problem, and gives rise to numerically unstable computational methods, unless proper regularization is employed.

We will show here briefly that a proper choice of the conduction coefficient makes the anisotropic diffusion enhance edges while running *forward* in time, and still being numerically stable.

Let our model of an edge be a mollified version of a step , for example a step function convolved with a gaussian kernel. A change of coordinates will align the step with the y axis, thus reducing the 2-D problem to 1-D. Thus $I_t = \operatorname{div}(c(x, y, t) \nabla I) = \frac{\partial}{\partial x}(c \cdot I_x)$.

Take c to be locally determined, just a function of the gradient of I : $c(x, y, t) = g(I_x(x, y, t))$. Call $\varphi(\cdot)$ the product $c \cdot I_x$: $\varphi(I_x) \equiv g(I_x) \cdot I_x$.

We are interested in looking at the slope of the edge while this is being diffused. Consider then : $\frac{\partial}{\partial t}(I_x)$. Since the functions that we are dealing with are smooth the order of differentiation may be inverted obtaining :

$$\begin{aligned} \frac{\partial}{\partial t}(I_x) &= \frac{\partial}{\partial x}(I_t) = \frac{\partial}{\partial x}\left(\frac{\partial}{\partial x}(c \cdot I_x)\right) = \\ &= \frac{\partial}{\partial x}\left(\frac{\partial}{\partial x}\varphi(I_x)\right) = \varphi'' \cdot I_{xx}^2 + \varphi' \cdot I_{xxx} \end{aligned}$$

Suppose the edge is oriented in such a way that $I_x > 0$ (if not, reverse the orientation of the x axis). At the point of inflection $I_{xx} = 0$, and $I_{xxx} < 0$ since the point of inflection corresponds, or is very close to, the point with maximum slope. Then in a neighbourhood of the point of inflection $\frac{\partial}{\partial t}(I_x)$ has sign opposite to $\varphi'(I_x)$. If $\varphi'(I_x) > 0$ the slope of the edge will decrease; if, on the contrary $\varphi'(I_x) < 0$ the slope will increase.

Notice that this increase in slope cannot be caused by a scaling of the edge, because this would violate the maximum principle. The edge becomes sharper.

Observe that such a function $\varphi(\cdot)$ is easy to obtain : take, for example, $c(x, y, t) = g(I_x) = \frac{C}{1+(I_x/K)^{1+\alpha}}$ with $\alpha > 0$. Then there exists a certain threshold value related to K and α , below which $\varphi(\cdot)$ is monotonically increasing, and beyond which $\varphi(\cdot)$ is monotonically decreasing, giving the desirable result of blurring small discontinuities, and sharpening edges.

Therefore with passive diffusion, and local reasoning we may obtain *edge enhancement* if we choose the conductivity function $g(\cdot)$ appropriately. Increasing progressively the value of the constant K we may obtain scale- space, the more and more edges disappearing from the picture.

4 NETWORK IMPLEMENTATION

In section 2 we have introduced a class of diffusion equations that generate families of images satisfying the 3 scale-space criteria that we have enunciated in section 1. Section 2 also hints on ways to choose $g(\cdot)$, and E and bounds are put on the possible choices of c . In this section

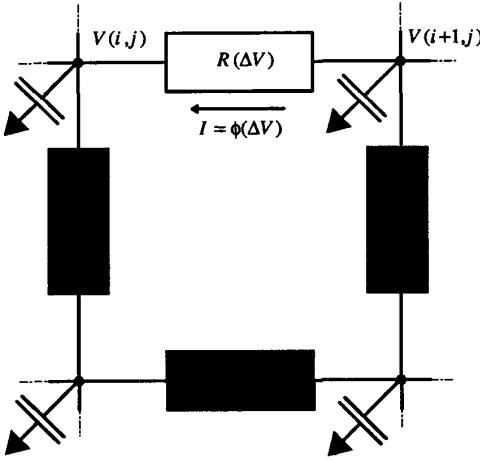


Figure 2: The structure of a network implementing anisotropic diffusion.

we will indicate the structure of a simple network that implements the anisotropic diffusion equation, and provide design criteria for obtaining scale space, and edge detection.

Consider the following simple discretization of the second term of the anisotropic diffusion equation :

$$\begin{aligned} I_x(x, y) &\approx \frac{1}{\Delta x}(I(x + \frac{\Delta x}{2}, y) - I(x - \frac{\Delta x}{2}, y)) \\ I_y(x, y) &\approx \frac{1}{\Delta y}(I(x, y + \frac{\Delta y}{2}) - I(x, y - \frac{\Delta y}{2})) \\ \text{div}(c(x, y, t)\nabla I) &\approx \\ &\approx \frac{1}{\Delta x^2}(c(x + \frac{\Delta x}{2}, y) \cdot (I(x + \Delta x, y) - I(x, y)) - \\ &\quad c(x - \frac{\Delta x}{2}, y) \cdot (I(x, y) - I(x - \Delta x, y))) + \\ &\quad + \frac{1}{\Delta y^2}(c(x, y + \frac{\Delta y}{2}) \cdot (I(x, y + \Delta y) - I(x, y)) - \\ &\quad c(x, y - \frac{\Delta y}{2}) \cdot (I(x, y) - I(x, y - \Delta y))) \end{aligned}$$

Introducing the notation:

$$\begin{aligned} c_N(x, y) &\equiv c(x, y + \frac{\Delta y}{2}) \cdot \frac{1}{\Delta x^2} & c_S(x, y) &\equiv c(x, y - \frac{\Delta y}{2}) \cdot \frac{1}{\Delta x^2} \\ c_E(x, y) &\equiv c(x + \frac{\Delta x}{2}, y) \cdot \frac{1}{\Delta x^2} & c_W(x, y) &\equiv c(x - \frac{\Delta x}{2}, y) \cdot \frac{1}{\Delta x^2} \end{aligned}$$

$$\nabla_N I(x, y) \equiv I(x, y + \Delta y) - I(x, y)$$

$$\nabla_S I(x, y) \equiv I(x, y - \Delta y) - I(x, y)$$

etc. we obtain :

$$I_t = c_N \cdot \nabla_N I + c_S \cdot \nabla_S I + c_E \cdot \nabla_E I + c_W \cdot \nabla_W I$$

This corresponds to diffusion of a quantity I associated to the nodes of a lattice with arcs of conductivity c . An electrical network with the topology of a square lattice, charges with initial value equal to the image brightness I associated to the nodes, and resistors with conductance c associated to the arcs does the job.

We may perform the same analysis as in the previous section using the discretized equations obtaining the same qualitative result: a lattice

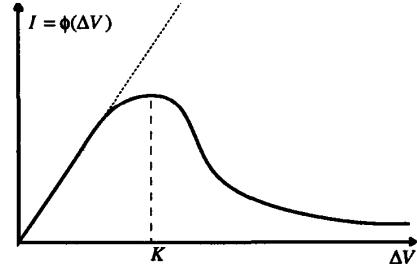


Figure 3: The qualitative behavior of a nonlinear resistor needed to realize scale space (solid line) compared with the linear resistor (dotted line).

network whose nodes are only nearest-neighbour connected, and whose arc elements have only the couple of associated nodes as inputs can perform scale-space and edge detection, including the edge enhancement effect, if the arc elements are chosen to have a nonlinear conductance as indicated in section 4. See also fig. 3.

5 EXPERIMENTAL RESULTS

We have simulated in software the anisotropic diffusion scale space and edge detection network that we propose.

The computational setup had a lattice structure, with brightness values associated to the vertices, and conduction coefficients to the arcs. A constant value of 1 for the conductivity led to gaussian blurring.

In a first set of experiments we used the absolute value of the gradient as the estimate for $\|\mathbf{E}\|$. Different functions were used for $g(\cdot)$ giving perceptually similar results. The image in fig. 4 [2] was obtained using

$$g(\nabla I) = \frac{C}{1 + (\frac{\nabla I}{K})^2}$$

Notice that the smoothed image has a Mondrian-like appearance. This is the effect of the piecewise smoothing, and makes edge detection extremely simple: it is sufficient to threshold the magnitude of the gradient of the brightness. Notice the behaviour of the edge detector on the water, the pavement, and the architectural details.

There is an additional remark to be made: if the function $g(\cdot)$ approaches zero very quickly for increasing contrast values the edges may be too impermeable i.e. the image is transformed to a piecewise-constant Mondrian, and the diffusion is very slow in getting rid of the fine detail. This may be corrected by letting the value of $g(\cdot)$ increase in time so that the weaker edges become unstable and are diffused. If this is done slowly compared with the speed of the diffusion process, the criteria that we have enunciated, namely 'localization of boundaries', and 'piecewise smoothing' are respected, while the computation becomes more efficient. There are other reasons for allowing the function $g(\cdot)$ vary with time. One of these is the presence of high levels of noise in the image. In this case the conduction coefficient has to start off with a high value in order to let the diffusion average the noise, and decrease in the course of the diffusion process. We discuss this in detail in [8].

References

- [1] J. Babaud, A. Witkin, Baudin, and R.Duda. Uniqueness of the gaussian kernel for scale-space filtering. *IEEE Trans. on PAMI*, 8, January 1986.
- [2] Canaletto. View in Venice. Circa 1740.



Figure 4: The effect of anisotropic diffusion on an image. From top to bottom: a) The original image . b) The same image at a coarser resolution.

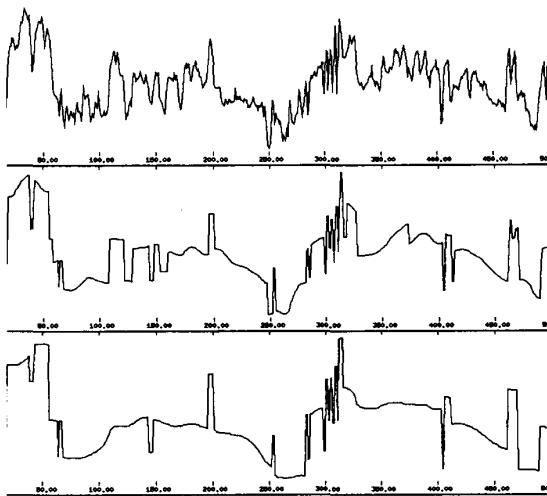


Figure 5: The brightness pattern along a horizontal line crossing the lower part of the image in fig.4 . The top plot represents the brightness of the original image, while the bottom one the brightness of the coarse resolution image 4.b. The middle one corresponds to an image at an intermediate resolution. Observe that the position of the edges does not vary across the scale space.

- [3] R.A. Hummel. *Deblurring Gaussian Blur*. Technical Report , New York University, Courant Institute of Mathematical Sciences.
- [4] R.A. Hummel. *Representations based on Zero-crossings in Scale-Space*. Technical Report 225, New York University, Courant Institute of Mathematical Sciences, 1986.
- [5] Fritz John. *Partial Differential Equations*. Springer Verlag , New York, 1982.
- [6] J.J. Koenderink. The structure of images. *Biological Cybernetics*, 50:363–370, 1984.
- [7] David Marr. *Vision*. W.H.Freeman & Co., 1982.
- [8] Pietro Perona and Jitendra Malik. Technical Report, Deptt. of EECS, U.C.Berkeley. In preparation.
- [9] Pietro Perona and Jitendra Malik. Scale space, and edge detection using anisotropic diffusion. In *IEEE Computer Society Workshop on Computer Vision - Miami*, 1987.
- [10] A. Rosenfeld and M. Thurston. Edge and curve detection for visual scene analysis. *IEEE Trans. on Computers*, C-20:562–569, May 1971.
- [11] A.P. Witkin. Scale-space filtering. In *International Joint Conference on Artificial Intelligence*, pages 1019–1021, 1983. Karlsruhe.
- [12] A. Yuille and T. Poggio. Scaling theorems for zero crossings. *IEEE Transactions on Pattern Analysis, and Machine Intelligence*, 8, January 1986.

Acknowledgements

This work has been supported by grant 82-11-008 of the Semiconductor Research Corporation , and an IBM faculty development award.