

Derivatives in Scale Space*

Ge Cong and SongDe Ma

National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences
PO.Box 2728, Beijing, 100080, P.R.China
email: congge@prlsun1.ia.ac.cn

Abstract. Various derivatives(1st to 4th order) describe different signal characters in scale space: the first derivative represents signal variation which can be used for coding; the second derivative characterizes the size of signal local structure and can be used for adaptive smoothing; the third and fourth order derivatives relate zero-crossing(or other features) contour in scale space to a differential equation which can be used for feature tracking and make quantitative measurements of zero-crossing properties possible. In the case that different derivatives are calculated by a special class of derivative filters, all the derivative signals can be treated in the same way.

1 Introduction

A common problem in early vision is to select the effective scale(s) to perform signal processing. Since we have no *a priori* knowledge about the original signal available, a multiscale strategy in which the signal was smoothed at many different scales was used[17]. Since Witkin[19] and Koenderink[6] introduced the *Scale Space* in 1983 and 1984 respectively, the research in this field became very popular. In their scale space, the signal was convoluted with Gaussian filter with its scale parameter changing continuously from zero to infinite, and thus a parameter family of the derived signals was formed. However, since scale space is highly information redundant, To find a more efficient representation, it becomes important to study the relationship between signals at different scales. So far, the most popular method for this topics is the *coarse-to-fine tracking*: the features at the coarsest scale, which usually represent globe information and are relatively easy to be detected, are first detected. And then we track along the feature contours down to the finest scale to obtain their precise locations.

The idea of using multiple scales in image processing can be traced as early as to 1971, when Rosenfeld and Thurston used operators of different size to perform edge detection. Klinger[5] introduced maybe the first multiscale representation of image, the quad tree. Then Burt[1] and Crowley[3] developed the low-pass pyramid representation. In addition to Witkin and J.J.Koenderink, Eklundh[9],

* This work is supported by the Chinese National Natural Science Foundation and Sino-France Co-operation Project

Florack[4], Lindeberg[9][10][11] had also made great contributions to the field. The work by Mallat[14][15][16] and Meyer who considered multiscale signal processing by wavelets also contributed to this subject.

Most of the further works in this field, however, were concentrated on finding appropriate kernels for scale space construction[2][20][6]. There is relatively less work done to analyze the relationship between signals at different scales. The relationship is important, however, if we note that the purpose of using scale space is to find better methods to performing feature extraction in early vision systems. Derivatives play a dominant role in the extraction because, first, derivatives with respect to x are needed to extracted features at current scale; second, derivative with respect to σ , the scale, is essential to relate the features at different scales. Since scale space is the solution of the so called *Heat Equation*[2][6][18], the derivative in σ equals to the second derivative in x . Then the analysis between scales can be carried out by studying the higher order derivatives in current scale.

In this paper, we show that different derivatives (1st to 4th) represent different signal properties in scale space; The first derivative represents signal variation and can be used for coding; the second derivative determines the signal local structure size so it is a essential descriptor for signal measurement in adaptive smoothing; When the third derivative equals to zero, the zero-crossing curve in scale space terminates and if the fourth derivative equals to zero, the zero-crossing location is preserved across scale. So the third and fourth derivatives which describe the behaviour of the zero-crossing passing from finer to coarser scale give us a method to measure its properties quantitatively . And then, we show if the derivatives are calculated by a special class of derivative filters, the higher order derivative can be decomposed into the linear combination of the lower order derivatives at different scales and then they can be used conveniently in image processing.

In this paper, following notation is used:

Gaussian function: $g(x, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} : (\mathbb{R} \times \mathbb{R}^+) \rightarrow \mathbb{R}$, and there is:
 $\frac{\partial g}{\partial \sigma} = \sigma \frac{\partial^2 g}{\partial x^2}$.

Scale space: $L(x, \sigma) = f(x) * g(x, \sigma) : (\mathbb{R} \times \mathbb{R}^+) \rightarrow \mathbb{R}$, where $f(x)$ is the original signal, $f : \mathbb{R} \rightarrow \mathbb{R}$, and $*$ is the convolution operator.

The m -th derivative of one dimensional signal: $f^{(m)}(x) = \frac{d^m f}{dx^m}$

2 The first derivative: signal variation

The first derivative:

$$L^{(1)} = \frac{dL}{dx} \quad (1)$$

describes signal variation and is used for signal coding, *e.g.* DPCM, since it is variation, but not constant signal, that carries information. By calculating the first derivative, we eliminate most of the correlation between points in the signal and then $L^{(1)}$ is much more compact than $L(x)$ itself. In nonlinear diffusion, $L^{(1)}$

is used for calculating diffusion coefficients due to its ability to measure signal local variations[18].

In scale space, there is:

$$L^{(1)} = f^{(1)} * g(x, \sigma) \quad (2)$$

Just like the signal itself, $L^{(1)}$ also becomes smoother and smoother, which means less and less information is contained in the derived signal. At a certain point, there is:

$$\frac{\partial L^{(1)}}{\partial \sigma} = f^{(1)} * \frac{\partial g}{\partial \sigma} = f^{(1)} * \sigma \frac{\partial^2 g}{\partial x^2} = \sigma L^{(3)} \quad (3)$$

The information is reduced at the rate of the magnitude of $L^{(3)}$.

3 The second derivative, signal local structure size

In adaptive smoothing and other image processing tasks, what we are concerned with is how to select an appropriate scale smoothing by which we can suppress noise while preserve local structure. We should then define a variable, say, \mathcal{S} , to describe the size of signal local structure: large structure has large \mathcal{S} , while small structure has small \mathcal{S} . The local scale is then selected according to \mathcal{S} : for big \mathcal{S} , we select large scale so to suppress noise; for tiny \mathcal{S} , we select small scale so to avoid too much deformation. However, different from signal variation, this variable should not depend on $L^{(1)}$.

Let's see an example. Considering the signal:

$$f(x) = ax + b \quad (4)$$

It is easy to show that:

$$f(x) * g(x, \sigma) = f(x) \quad (5)$$

The smoothing result is independent of the smoothing scale, which means \mathcal{S} should also be independent of σ, a, b and it should be very large to suppress noise. However, $L^{(1)}$ is determined by a, b and the only common thing for signal with different a, b is $L^{(2)} = 0$.

In general, if

$$L^{(2)}(x, \sigma) = 0 \quad (6)$$

holds at current scale σ , it should hold in a certain range of scale $[\sigma - \xi, \sigma + \xi]$ around σ . We have:

$$\frac{\partial L}{\partial \sigma} = 0, \quad \sigma \in [\sigma - \xi, \sigma + \xi] \quad (7)$$

The smoothing result is independent of the scale, so the size of the structure should be independent of the signal. However, in this situation, $L^{(1)}$ can be of anything, the only common thing is $L^{(2)} = 0$.

In this paper, \mathcal{S} is defined as:

$$\mathcal{S} = \frac{\partial L}{\partial \sigma} \quad (8)$$

In another word, the local structure is of large size if it remains the same after the scale varies a little and is of smaller size if the structure changes sharply . This definition is obviously appropriate for adaptive smoothing and other image processing tasks. We have:

$$S = \frac{\partial L}{\partial \sigma} = f * \frac{\partial g}{\partial \sigma} = f * \sigma \frac{\partial^2 g}{\partial x^2} = \sigma L^{(2)} \quad (9)$$

So S is defined by $L^{(2)}$.

Liking $L^{(1)}$, there is:

$$L^{(2)} = f^{(2)} * g(x, \sigma) \quad (10)$$

so the signal local structure also becomes smoother and smoother with the increase of the scale. At a certain point, there is:

$$\frac{\partial L^{(2)}}{\partial \sigma} = f^{(2)} * \frac{\partial g}{\partial \sigma} = f^{(2)} * \sigma \frac{\partial^2 g}{\partial x^2} = \sigma L^{(4)} \quad (11)$$

The structure is smoothed at the rate of the magnitude of $L^{(4)}$.

4 The third and the fourth derivatives, zero-crossing properties

In this section, we use ZC to indicate the zero-crossing of $L^{(2)}$.

See Figure 1, ZC at different scale form many curves in scale space. It has been proved [13][14] that the ZC curves always disappear in pairs, so two ZC curves intersect only at the terminating scale, and the intersection point is smooth. The pair of curves terminating together form the so called *ZC contour*[2][22]. Different ZC contours never intersect with each other. For a certain ZC curve, it is a implicit function of x in $\sigma: x = x(\sigma)$. ZC can be formulated as (notice x is the ZC location at different scales):

$$L^{(2)}(x, \sigma) = 0 \quad (12)$$

On both sides of (10), calculate the derivative in σ :

$$\frac{dL^{(2)}}{d\sigma} = \frac{\partial L^{(2)}}{\partial \sigma} + \frac{\partial L^{(2)}}{\partial x} \frac{dx}{d\sigma} = 0 \quad (13)$$

And then:

$$\frac{dx}{d\sigma} = -\frac{\sigma L^{(4)}}{L^{(3)}} \quad (14)$$

We can see from (12) if $L^{(3)}(x, \sigma) = 0$, then $dx/d\sigma = \infty$, so the curve terminates at the scale. Since $dx/d\sigma$ is unique defined, two ZC curves intersect with each other only when they terminate. In another word, the ZC curves always disappear in pairs at the point where $L^{(3)}(x, \sigma) = 0$. In figure 2, the ZC and the zero-crossing of $L^{(3)}$ are both curved, we can see that the ZC curves always terminate where they intersect with the zero-crossing curves of $L^{(3)}$.

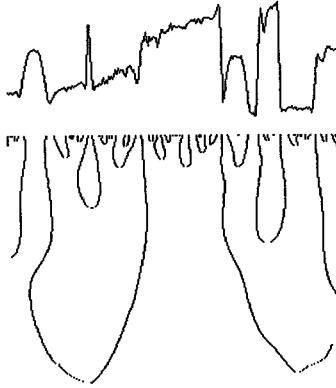


Fig. 1. ZC in scale space

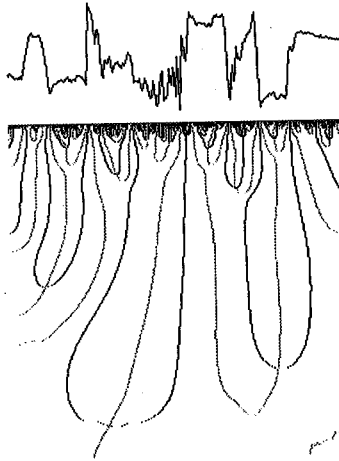


Fig. 2. ZC terminates where $L^{(3)} = 0$. Black curve, ZC curve; Dash curve, the curve zero-crossing of $L^{(3)}$.

If $L^{(4)}(x, \sigma) = 0$, then $dx/d\sigma = 0$, the ZC curve does not shift across scale; Otherwise, it shifts. Figure 3 presents the curves of both ZC and zero-crossing of $L^{(4)}$, we can see that those ZC with their $L^{(4)}$ equaling to 0 (or being very small) go much further to coarser scale than other ZC. The straighter the curve is, the longer it lives. The range of scale in which the ZC curve lives is called the lifetime of the ZC curve. So, $L^{(4)}$ can be used to measure the strength and predict the lifetime of the ZC curve. The fact that $L^{(4)}$ is a very important descriptor of ZC can be understood in this way: if $L^{(4)}(x, \sigma) = 0$, then x is not only a edge point in the signal, but also a edge point in $L^{(2)}$, so x is a sharper zero-crossing in $L^{(2)}$ and obviously stronger than other ZC.

In general, at a edge point where $L^{(2)}(x, \sigma) = 0$, $L^{(3)}, L^{(4)}$ can be of any values. As we can see, the ZC with stronger $L^{(3)}$ and weaker $L^{(4)}$ is more important than the ZC with weaker $L^{(3)}$ but stronger $L^{(4)}$. Another very important conclusion can be derived from the analysis above is that there should not be $L^{(3)}(x, \sigma) = 0$ and $L^{(4)}(x, \sigma) = 0$ at the ZC at the same time except for $f(x) = \text{constant}$.

It is useful to analyze the variation of some variables, A , *e.g.*, the derivatives, along the ZC curve: $\frac{\partial A}{\partial s}$, where s is the path parameter along the curve. Calculating the directional derivative along the curve's tangent direction, we have:

$$\frac{\partial A}{\partial s} = \frac{\partial A}{\partial x} \cos \alpha + \frac{\partial A}{\partial \sigma} \sin \alpha \quad (15)$$

where α is the angle between the tangent and the x axis. Now, suppose that at the edge point, $L^{(1)} > 0$, then according to Perona[18] there is $L^{(3)} < 0$, and note $0 \leq \alpha \leq \pi$, therefore

$$\cos \alpha = \frac{\sigma L^{(4)}}{\sqrt{L^{(3)2} + \sigma^2 L^{(4)2}}}, \quad \sin \alpha = -\frac{L^{(3)}}{\sqrt{L^{(3)2} + \sigma^2 L^{(4)2}}} \quad (16)$$

For example, if $A = L^{(1)}$, then (note $L^{(2)} = 0$):

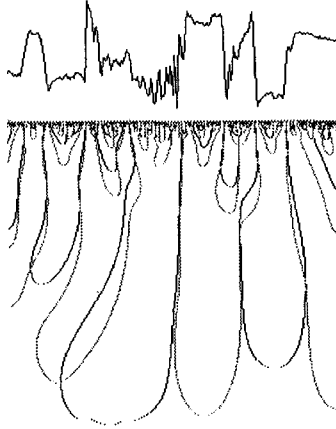


Fig. 3. ZC curve and its relationship with $L^{(4)}$. Black curve, ZC curve; Dash curve, the zero-crossing of $L^{(4)}$.

$$\begin{aligned} \frac{\partial L^{(1)}}{\partial s} &= \frac{\partial L^{(1)}}{\partial x} \frac{\sigma L^{(4)}}{\sqrt{L^{(3)2} + \sigma^2 L^{(4)2}}} - \frac{\partial L^{(1)}}{\partial \sigma} \frac{L^{(3)}}{\sqrt{L^{(3)2} + \sigma^2 L^{(4)2}}} \\ &= \frac{-\sigma L^{(3)} L^{(3)}}{\sqrt{L^{(3)2} + \sigma^2 L^{(4)2}}} < 0 \end{aligned} \quad (17)$$

We can see, as the scale increases, the magnitude of $L^{(1)}$ of the edge point always decreases along the ZC curve.

5 Higher order derivatives

It has been proved that a $2k$ -th derivative filter can be designed as:

$$h_{2k}(x) = \sum_{j=0}^k \alpha_j g(x, \sigma_j) \quad (18)$$

where the coefficients α_j satisfy the moment conditions[7][8], one of which is:

$$\sum_{j=0}^k \alpha_j = 0 \quad (19)$$

Define:

$$h_2\left(\frac{x}{r}\right) = g(x, r) - g(x, cr) \quad (20)$$

where r is the scale of h_2 and we select:

$$\sigma_{j+1} = c\sigma_j, \quad c > 1 \quad (21)$$

Then easily we obtain:

$$h_{2k}(x) = \sum_{j=0}^{k-1} \beta_j h_2\left(\frac{x}{c^j \sigma_0}\right) \quad (22)$$

where

$$\beta_0 = \alpha_0, \quad \beta_{j+1} - \beta_j = \alpha_{j+1}, \quad j = 1, \dots, k-2, \quad \beta_{k-1} = -\alpha_k \quad (23)$$

A $2k$ -th derivative filter can be decompressed into the second derivatives at different scales. Similarly, a $(2k+1)$ -th derivative filter can be designed as[7][8]:

$$h_{2k+1}(x) = \sum_{j=0}^k \alpha_j g^{(1)}(x, \sigma_j) \quad (24)$$

and one of the moment conditions is:

$$\sum_{j=0}^k \alpha_j \sigma_j = 0 \quad (25)$$

Define:

$$h_3\left(\frac{x}{r}\right) = cg^{(1)}(x, r) - g^{(1)}(x, cr) \quad (26)$$

where r is the scale of h_3 and we select:

$$\sigma_{j+1} = c\sigma_j, \quad c > 1 \quad (27)$$

then:

$$h_{2k+1}(x) = \sum_{j=0}^{k-1} \beta_j h_3\left(\frac{x}{c^j \sigma_0}\right) \quad (28)$$

where

$$c\beta_0 = \alpha_0, \quad c\beta_{j+1} - \beta_j = \alpha_{i+1}, \quad j = 1, \dots, k-2, \quad \beta_{k-1} = -\alpha_k \quad (29)$$

A $(2k+1)$ -th derivative filter can be decompressed into the third derivatives at different scales. Thus we show that the *Even* and *Odd* kernels used for derivative filter construction in [7][8] can be replaced by even and odd derivatives as in (20)(24) respectively so the integrals of all the kernels are zero.

6 Conclusion

To use the information obtained from scale space more efficiently, we have to find the relationship between the features at different scales. We show that derivatives, with respect to both x and σ , play a dominant role in finding the relationship and different derivatives in scale space represent different characters of the smoothed signal.

References

1. P.J.Burt and E.H.Adelson, The Laplacian Pyramid as a Compact Image Code, *IEEE Trans. on Commun.*, vol.9, pp532-540, 1983.
2. J.Baraud, A.P.Witkin, M.Baudin, and R.O.Duda, Uniqueness of the Gaussian Kernel for Scale-Space Filtering, *IEEE Trans. Pattern Anal. Machine Intell.*, vol.8, no.1, pp26-33, 1986.
3. J.L.Crowley and A.C.Parker, A Representation for Shape Based on Peaks and Ridges in the Difference of Low-Pass Transform, *IEEE Trans. Pattern Anal. Machine Intell.*, vol.6, no.2, pp 156-170,1984.
4. L.M.J.Florack, B.M.ter Haar Romeny, J.J.Koenderink and M.A.Viergever, Cartesian Differential Invariants in Scale-space, *J.of Mathematical Imaging and Vision*, vol.1.3, no.4, pp327-348, 1993.
5. A.Klinger, Pattern and Search Statistics, *Optimizing Method in Statistics*(J.S.Rustagi, ed.), (New York), Acaden Press, 1971.
6. J.J.Koenderink, The Structure of Images, *Biol.Cybern.*, vol.50, pp363-370,1984
7. B.C.Li and S.D.Ma, Multiscale-Filtering Based Computation of High Order Derivatives, *SPIE. on Visual Communication and Image Processing*, Chicago, USA, 1994.
8. B.C.Li and S.D.Ma, Difference-Polynomial-Based Implementation of Image Filtering, *SPIE's OE/Technology on Intelligent Robots and Computer Vision XIII*, Boston, USA, 1994.
9. T.Lindeberg and J.-O.Eklundh, The Scale-space Primal Sketch: Construction and Experiments, *Image and Vision Computing*, vol.10, pp3-18, 1992.
10. T.Lindeberg, Scale-space Theory: A Basic Tool for Analyzing Structures at Different Scales, *J. of Applied Statistics*, vol.1.21, no.2, pp225-270,1994.

11. T.Lindeberg, Effective Scales: A Natural Unit for Measuring Scale-Space Lifetime, *IEEE Trans. Pattern Anal. Machine Intell.*, vol.15, no.10, pp1068-1074, 1993.
12. Y.Lu and R.C.Jain, Behavior of Edges in Scale Space , *IEEE Trans. Pattern Anal. Machine Intell.*, vol.11, no.4, pp337-356, 1989.
13. Y.Lu and R.C.Jain, Reasoning about Edges in Scale Space , *IEEE Trans. Pattern Anal. Machine Intell.*, vol.14, no.4, pp450-468, 1992.
14. S.G.Mallat, A Theory for Multiresolution Signal Decomposition: The Wavelet Representation, *IEEE Trans. Pattern Anal. Machine Intell.*, vol.11, no.7, pp674-694, 1989.
15. S.G.Mallat and S.Zhong, Characterization of Signal from Multi-scale Edges, *IEEE Trans. on Pattern Anal. Machine Intell.*, vol.14, no.7, pp710-723, 1992.
16. S.G.Mallat, Multifrequency Channel Decompositions of Images and Wavelet Models, *IEEE Trans. Acoust Speech and Signal Processing*, vol.37, pp2091-2110, 1989.
17. D.Marr and H.Hildreth, Theory of Edge Detection, *Proc. Roy. Soc, London*, vol.B207, pp187-217, 1980.
18. P.Perona and J.Malik, Scale-Space and Edge Detection Using Anisotropic Diffusion, *IEEE Trans. Pattern Anal. Machine Intell.*, vol.12, no.7, pp629-639, 1990
19. A.P.Witkin, Scale-space Filtering, *Proc.8th Int. Joint Conf. Art. Intel.*,(Karlsruhe, West Germany), pp1019-1022, 1983.
20. A.L.Yuille and T.A.Poggio, Scaling Theorems for Zero Crossing, *IEEE Trans. Pattern Anal. Machine Intell.*, vol.8, no.1, pp15-25, 1986.