ON THE APPLICATION OF SCALE SPACE TECHNIQUES IN DIGITAL PHOTOGRAMMETRY

Anthony Stefanidis

Toni Schenk

Department of Geodetic Science and Surveying
The Ohio State University, Columbus, Ohio 43210-1247
USA

Commission III

ABSTRACT

Scale space techniques are widely used in digital photogrammetry. Typical implementations use the scale space as a discrete representation, thus inherently assuming that all features represented in images of similar resolutions belong to the same scale space level. However, this approach ignores differential scale variations that exist between conjugate features in multiple images, or even between different features in a single image. The subject of this paper is an investigation into theoretical and practical aspects associated with the use of scale space techniques in both the image and object space domains. The interrelationship between the scale space representations of these two domains and the effects of differential scale variations in digital photogrammetric operations, such as matching, object space reconstruction, and orthophoto production are also addressed.

1. INTRODUCTION

Physical phenomena in object space occur over a wide variety of spatial extents. Macro-variations of a surface express its major trend, while micro-variations correspond to trends of smaller extent. The concept of macro- and micro-variations is relative and depends on the specific application. What is considered a macro-variation in one application might very well be viewed as a micro-variation in another. In digital images, changes in gray values correspond to object space phenomena, which can also be perceived within areas of different sizes, ranging from few pixels to large regions. However, even region-wise changes occur over an extensive array of region sizes, ranging from as little as a few pixels to as much as a large part of the image. The identification of these changes is essential in decoding the information which inherently exists in an image.

The scale space representation of signals in general, or digital images in particular, is widely used to successfully produce several versions of the same image in which the information content is changing in a systematic and, therefore, easy to exploit fashion [Lindeberg, 1990], [Yuille & Poggio, 1983]. Physical phenomena of various extents can be easily identified through the behavior of their images in different levels of scale space [Lu & Jain, 1989], [Witkin, 1983].

In our paper, we present the basic axioms of scale space, and we analyze the corresponding mathematical aspects, together with the proper selection of scale-generating functions. The effect of differential scale variations on photogrammetric procedures is discussed, and we report how a continuous scale space can be used to bypass the shortcomings of this effect. Finally, the scale space representation of object space and its potential use in photogrammetry are explored.

2. SCALE SPACE

The scale space representation of a signal $f(x,y)$ is a set of signals $\{f_s^n(x,y;n)\}$, representing the original one in various scale levels as function of a scale parameter n. The set of signals $\{f_s^n(x,y;n)\}$ is called the scale space family of $f(x,y)$.

The objective of the scale space representation of any signal is to create a scale space family in a way that information conveyed by this signal will become more explicit. In order for this goal to be met, the generation of scale space family has to follow some basic guidelines [Lindeberg, 1990]:

- The scale space family has to be generated by the convolution of the original signal with a single scale-generating function $s(x,y;n)$

$$f_s^n(x,y;n) = s(x,y;n) * f(x,y) \tag{1}$$

- The scale-generating function should be selected in a proper manner, such that larger values of n would create coarser versions of the original signal through elimination of the finer details which correspond to higher frequency phenomena. We want to be able to identify large trends in lower resolutions and include spatially limited details in finer levels. For $n = 0$, at the finest resolution of scale space, we have the original signal itself

$$f_s^0(x,y;0) = f(x,y) \tag{2}$$

which is obviously the upper limit as far as fine resolution is concerned.

A Gaussian filter is mathematically expressed as a function

$$g(x,y) = ke^{-\frac{x^2+y^2}{2\sigma^2}} \tag{3}$$

where σ is the associated standard deviation. In applications, the multiplicative factor k may receive various values, creating a large array of Gaussian filters which are essentially scaled variations of the core function, e.g.,

$$G(x_g, y_g; \sigma) = \frac{1}{2\pi\sigma} e^{-\frac{x_g^2+y_g^2}{2\sigma^2}} \quad (4)$$

attempts to preserve the output within a prespecified range [Agouris et al., 1989]. The use of a Gaussian filter, with standard deviation σ as the associated scale parameter, as a scale-generating function satisfies the above set criteria [Babaud et al., 1986]. Therefore, the scale space family of a signal $f(x,y)$ can be created as

$$f_s^n(x, y; \sigma) = G(x_g, y_g; \sigma) * f(x, y) \quad (5)$$

A digital image is a two-dimensional discrete signal $I(x,y)$. Its convolution with the Gaussian kernel

$$I_\sigma^n(x, y; \sigma) = \sum_{x_g=-\infty}^{\infty} \sum_{y_g=-\infty}^{\infty} \frac{1}{2\pi\sigma} e^{-\frac{x_g^2+y_g^2}{2\sigma^2}} I(x - x_g, y - y_g) \quad (6)$$

can be used to construct its scale space family. Members of the scale space family may have the same dimensions as the original image, or, more commonly, their dimensions may decline in coarser resolutions. Assuming the original image $I(x,y)$ to have dimensions 4096 × 4096 pixels, we can form its scale space family by creating m versions of the image (all of dimensions 4096 × 4096 pixels), each one by convolving $I(x,y)$ with a Gaussian kernel of different scale parameter σ.

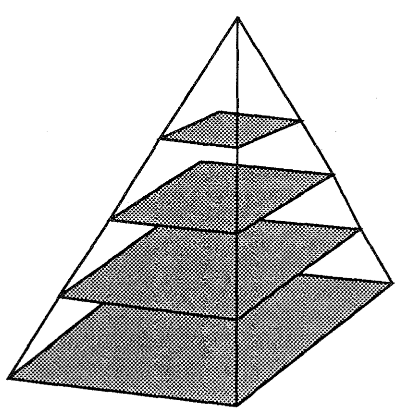

Figure 1: An image pyramid as a representation of discrete scale space

However, in most applications coarser levels of scale space are represented by images of smaller dimensions. By convolving the image with a Gaussian kernel and resampling every n^{th} pixel we can create a lower resolution copy of size $4096/n \times 4096/n$. A scale space family in which lower resolution members are represented by smaller size images is called an image pyramid [Fig. 1]. Various members of the image pyramid can be perceived as images of the same object scene in various geometric scales. For practical reasons the dimensions of the members of the scale space family are integer powers of two. Typically, the image pyramid of an original image of 4096 × 4096 pixels includes versions of the image in dimensions of 2048 × 2048, 1024 × 1024 and 512 × 512 pixels. Fig. 2 shows two windows of equal dimensions, one from the 512 × 512 pixel member of an image pyramid and the other from the 2048 × 2048 pixel version to demonstrate the associated differences in resolution. Both images were obtained by the convolution of the original 4096 × 4096 image with a Gaussian kernel, and by proper resampling.

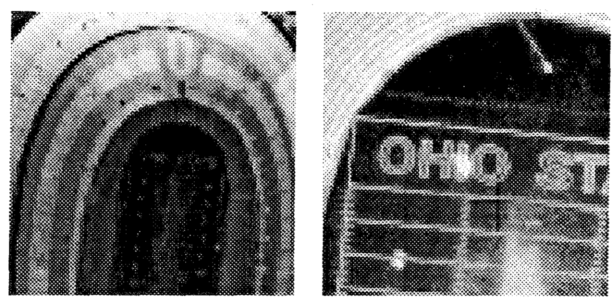

Figure 2: Two windows of equal size in pixels, one in 512 × 512 resolution (left) and the other in 2048 × 2048 resolution (right).

The use of a Gaussian kernel as a scale-generating function offers certain advantages, most notably exploited when combining smoothing with edge detection. Edges are identified as discontinuities in the image function, and therefore correspond to zero-crossings of the twice-differentiated image. The orientation independent second derivative of a two-dimensional function is obtained through a Laplacian operator

$$\nabla^2 = \frac{\partial}{\partial x^2} + \frac{\partial}{\partial y^2} \quad (7)$$

The associative property of convolution allows the combination of scale space generation with a Gaussian function $G(x_g, y_g)$ and differentiation with a Laplacian operator, thus substituting two convolutions by a single one

$$\nabla^2[G(x_g, y_g; \sigma) * I(x, y)] = [\nabla^2 G(x_g, y_g; \sigma)] * I(x, y) \quad (8)$$

Instead of scaling the image with $G(x_g, y_g)$ and then looking for edges in the smoothed image, we simultaneously smooth the image and extract its orientation-independent second derivative in a single convolution by the Laplacian of Gaussian (LoG) function

$$\nabla^2 G(x_g, y_g; \sigma) = [2 - \frac{x_g^2 + y_g^2}{\sigma^2}] e^{-\frac{x_g^2+y_g^2}{2\sigma^2}} \quad (9)$$

The size of the LoG operator is determined by the value of σ or altenatively, by the diameter w of its positive central region, which is related to σ through the equation

$$w = 2\sqrt{2}\sigma \quad (10)$$

Scale space family generation and edge detection can thus be succesfully combined. By using the Gaussian kernel for scaling we ensure that in any scale level fewer edges occur than in finer resolutions and more than in coarser ones, thus performing proper scale space generation. This property has a qualitative aspect in addition to its obvious quantitative meaning. Edges detected in coarser levels using large σ (or w) values will also appear in finer levels. The same edge can be traced through various resolutions, since its images display a certain degree of geometric similarity, with the degree of localization (closeness to the true edge) increasing with resolution [Lu & Jain, 1989], [Witkin, 1983]. This is demonstrated in Fig. 4 and Fig. 5 which show edges of the original image (shown in Fig. 3) produced by its convolution with a

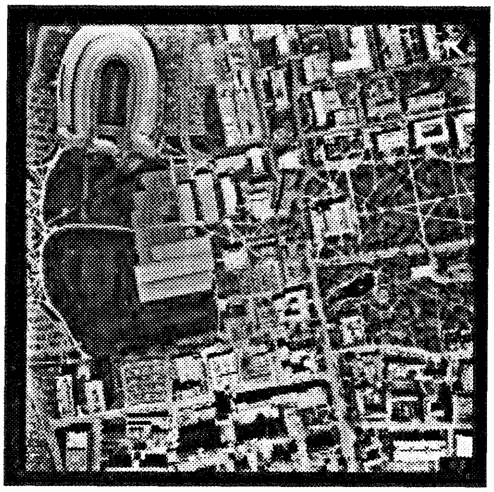

Figure 3: The original image

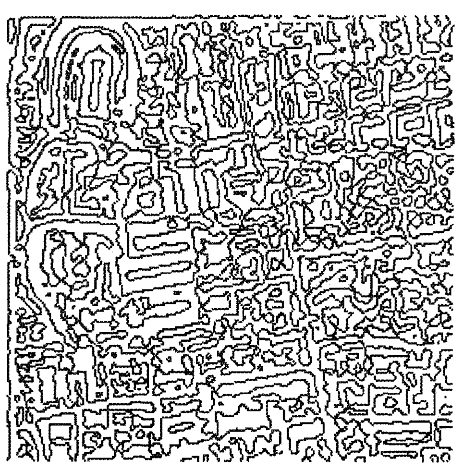

Figure 4: Edges detected with a fine LoG operator ($w = 10$)

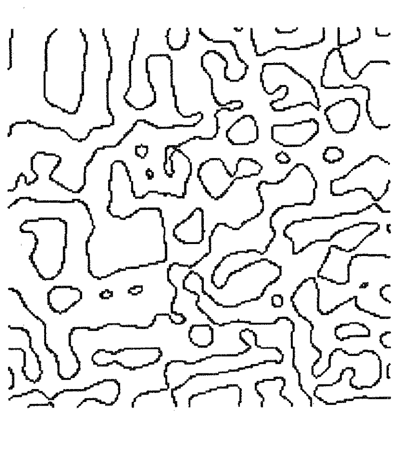

Figure 5: Edges detected with a coarse LoG operator ($w = 30$)

fine ($w = 10$) and a coarse ($w = 30$) LoG operator respectively. In addition, the traces of edges in various resolutions offer a complete representation of the original signal, thus allowing its reconstruction [Yuille & Poggio, 1983].

3. DIFFERENTIAL SCALE VARIATIONS

When representing the scale space family of a digital image as a pyramid, we create a number of discrete representations of the original image with each representation corresponding to a specific scale level. However, unless the image-generating projection is parallel, the exposure vertical and the object surface planar, features within the same image pyramid level will not have the same geometric scale, expressed as

$$S^A = \frac{A'}{A} \qquad (11)$$

with A' the image of a feature A of the object space. For the projective transformation governing the image formation process, the scale factor S^i at a point (x^i, y^i) of the image, corresponding to a point (X^i, Y^i, Z^i) in the object space will be given through the formula

$$\begin{bmatrix} x^i \\ y^i \\ -c \end{bmatrix} = S^i R \begin{bmatrix} X^i - X_o \\ Y^i - Y_o \\ Z^i - Z_o \end{bmatrix} \qquad (12)$$

where R is the rotation matrix and (X_o, Y_o, Z_o) the exposure station coordinates of the photo. It is apparent that different features in the same image will have different scale factors. In addition, the images of the same object space feature in two or more different images will have different scales, particularly when the exposure conditions (rotations, exposure stations) differ significantly (e.g., converging photography) or the object space surface displays high variations. In the extreme case, the scale becomes 0 and occlusions occur.

Assuming each image pyramid level i to correspond to an average scale S_i, features within this image will thus appear in scales

$$S_i + dS_i, \quad dS_i = f(X, Y, Z, \omega, \phi, \kappa, \frac{\partial Z}{\partial X}, \frac{\partial Z}{\partial Y}) \qquad (13)$$

which in general will not coincide with any of the discrete scales represented by the image pyramid. Image pyramids though are discrete representations of the scale space which itself is continuous. While the discrete representation is obtained using only a number of values of the scale parameter σ of the Gaussian kernel used to convolve the image, a continuous representation is the outcome of the same convolution allowing σ to receive any allowable real value.

Scale variations between members of stereopairs become apparent in digital photogrammetric operations, with matching serving as a good example. In least squares matching, we attempt to match windows of pixels by minimizing their radiometric differences. This is achieved by forming one observation equation for every pair of conjugate pixels within a pair of approximately conjugate image windows $g_L(x_L, y_L)$ and $g_R(x_R, y_R)$ in the left and right image respectively

$$g_L(x_L, y_L) - g_R(x_R, y_R) = e(x, y) \qquad (14)$$

The solution is obtained by allowing one of the two windows to be geometrically reshaped according to an affine transformation and by resampling gray values for this newly defined window. Differences in scale are accommodated by the two scale factors assumed in the six-parameter affine transformation

$$x_R = a_1 + a_2x_L + a_3y_L \tag{15}$$

and

$$y_R = b_1 + b_2x_L + b_3y_L \tag{16}$$

Updating the above affine transformation parameters by the solution of the linearized observation equations

$$\begin{aligned} g_L(x_L, y_L) - e(x, y) &= g_R^o(x_R^o, y_R^o) + g_{R_x}da_1 + g_{R_x}x_Lda_2 \\ &+ g_{R_x}y_Lda_3 + g_{R_y}db_1 + g_{R_y}x_Ldb_2 \\ &+ g_{R_x}y_Ldb_3 \end{aligned} \tag{17}$$

we define a new window in the right image within which we resample the gray values.

Scale variations will affect this procedure in various stages. When two image patches are represented in two different scale levels in a stereopair, their scale difference will be both geometric and radiometric. When resampling the gray values $g_R(x_R, y_R)$ we use the original image, spreading or shrinking its gray values over a new area, according to the updated affine transformation parameters. As a result we produce a new window in the right image which might belong to the same geometric level of scale space as its conjugate left image template $g_L(x_L, y_L)$ but will still differ from it in the radiometric scale space. This will have obvious effects on the observation equations, since we use gray level differences as observations. The same problem occurs during digital image warping or rectification for orthophoto production [Doorn, 1991],[Novak, 1992]. Conjugate patches in two overlapping orthophotos are brought to the same scale level geometrically, using as a reference a digital elevation model of the object space. Radiometrically though, these patches remain unequal to the same degree that the corresponding windows in the original stereopair were unequal. This causes conjugate patches in overlapping orthophotos to differ radiometrically, even when their gray level histograms are adjusted for average and standard deviation differences.

To accommodate for the problem of different scales, the scale concept has to be introduced into the matching process itself. This will be conceptually performed by the alteration of the observation equations to accommodate for scale as

$$g_L(x_L, y_L; s_L) - g_R(x_R, y_R; s_R) = e(x, y) \tag{18}$$

which would correspond to a matching process adapting itself into various scales. The above equation may be linearized with respect to x, y and s, essentially adding to the previously mentioned (eq. 17) linearized observation equations one term

$$\begin{aligned} g_L(x_L, y_L, s_L) - e(x, y) &= g_R^o(x_R^o, y_R^o, s_R^o) + g_{R_x}dx_R + \\ &+ g_{R_y}dy_R + g_{R_S}ds_R \end{aligned} \tag{19}$$

The added term g_{R_S} expresses how gray levels change at a point whenever the scale level of the window within which this point is located changes within the continuous scale space. The term s has conceptual meaning and may be substituted by the σ of the Gaussian filter or any other quantity sufficiently describing scale.

The introduction of a scale parameter in least squares matching may introduce linear dependency. The terms g_{R_x} and g_{R_y} also express gray level gradients, but are different than the term g_{R_S} in that they are highly localized and obviously orientation dependent. Even in the case that high dependency exists, matching may be implemented in two distinct sets, properly constraining some of the parameters to realistic estimated values. To assure succesful implementation, matching has to be performed in the highest possible common resolution of the two conjugate patches. That will obviously be the resolution of the coarser patch, and therefore the finer patch has to be transferred into another scale level using a Gaussian filter.

4. SCALE SPACE REPRESENTATION OF OBJECT SPACE

Object space can be described by the combination of two two-dimensional continuous signals, one ($Z(X, Y)$) expressing its geometric and another ($R(X, Y)$) expressing its radiometric properties. Discretized, these signals are represented by a Digital Elevation Model and a Digital Radiometry Model which can be together referred to as DERM.

Each of the signals can be individually expressed in a scale space representation using the Gaussian kernel, thus preserving the scale space family properties that we presented in section 2. The scale space family of the DEM will consist of DEM of lower resolutions, with each lower resolution level representing a smoothed version of the original signal. Taking advantage of the self-reciprocity of the Gaussian function which states that the Fourier transform of a Gausian is another Gaussian

$$\mathcal{F}[G(x)] = G(w) \tag{20}$$

we see that convolution with a Gaussian function in the space domain is equivalent to a filtering with a filter of the same shape in the frequency domain [Weaver, 1983]. Therefore, Gaussian convolution can be perceived as filtering with a low-pass filter, the cut-off frequency of which is determined by the scale parameter σ. Coarse scale representations of the DEM preserve the major geometric trends of the surface, corresponding to the lower frequencies of its frequency domain equivalent. In finer resolutions, frequencies of higher order are introduced. Edge detection, with the application of an LoG function to the DEM signal, will locate breaklines [Chakreyavanich, 1991]. Breakline detection can be applied hierarchically, similarly to edge detection in images. In coarse levels of scale space (large w parameter) we detect major breaklines in the topographic surface, while moving to finer resolutions we not only improve the spatial accuracy of these breaklines, but we also identify breaklines of smaller spatial extent.

In a similar fashion, the Digital Radiometry Model (DRM) of the surface can be processed with a Gaussian filter for the generation of its scale space family. Edges in the DRM will correspond to positions where the radiometric properties of the surface present discontinuities.

The recorded image gray values represent the DRM as altered due to the geometric properties of the object space. In the scale space family of DRM there will exist a member which most closely corresponds to the image depicting this DRM. For a DERM with no geometric variations, the edges detected in the image function would correspond to discontinuities in DRM. In realistic situations though, DEM is not flat and the image edges reflect the combined effect of geometric and radiometric discontinuities. Taking advantage of this we can distinguish edges created by geometric and radiometric discontinuities in the object space, by comparing the scale space of the image to the scale spaces of the object space.

5. COMMENTS

Scale space can be used to represent two dimensional signals in various resolutions. This representation can thus be used for images as well as for radiometric and/or geometric descriptions of the object space. It is structured and explorable and it can offer valuable assistance in various photogrammetric processes.

The concept of scale space provides the theoretical foundation for hierarchical implementation of digital photogrammetric tasks, allowing otherwise cumbersome and time consuming modules to be performed quickly and effectively. For instance, automatic stereopair orientation can be performed using digital image pyramids to effectively lead the results to continuously improving accuracies [Schenk et al., 1991].

However, besides implementing some modules in a hierarchical fashion, scale space theory can also be used to refine the performance of well-established processes, such as least squares matching and orthophoto production. By investigating the differential scale variations which exist between conjugate features in different images, we can deduce a scale-adapting matching process aiming at the optimization of least squares matching. In orthophoto production, we can bring features to the same radiometric and geometric level of scale space, thus eliminating discrepancies and improving its overall performance.

In general, the advantage of using scale space theory to represent the object space is twofold. Signals describing the object space can be stored in a compact yet efficient way by recording their discontinuities through scale space and in addition, image and object space can be directly compared and semantic information can be extracted from this comparison.

References

[1] Agouris P., Schenk A.F. & Stefanidis A. (1989) *Zero-Crossings for Edge Detection*, Proceedings 1989 ASPRS Fall Convention, Cleveland, pp. 91-99.

[2] Babaud J., Witkin A., Baudin M. & Duda R.O. (1986) *Uniqueness of the Gaussian Kernel for Scale-Space Filtering*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 8, No. 1, pp. 26-33.

[3] Bergholm F. (1987) *Edge Focusing*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 9, No. 6, pp.726-741.

[4] Chakreyavanich U. (1991) *Regular DEM Data Compression by Using Zero Crossings: The Automatic Breakline Detection Method*, Report No. 412, Dept. of Geodetic Science, The Ohio State University.

[5] Doorn B.D. (1991) *Multi-Scale Surface Reconstruction in the Object Space*, Report No. 413, Dept. of Geodetic Science, The Ohio State University.

[6] Lindeberg T. (1990) *Scale-Space for Discrete Signals*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 12, No. 3, pp. 234-254.

[7] Lu Y. & Jain R.C. (1989) *Behavior of Edges in Scale Space*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 11, No. 4, pp. 337-356.

[8] Novak K. (1992) *Rectification of Digital Imagery*, Photogrammetric Engineering & Remote Sensing, Vol. 58, No. 3, pp. 339-344.

[9] Schenk A.F., Li J.C. & Toth C. (1991) *Towards an Autonomous System for Orienting Digital Stereopairs*, Photogrammetric Engineering & Remote Sensing, Vol. 57, No. 8, pp. 1057-1064.

[10] Weaver J.H. (1983) *Applications of Discrete and Continuous Fourier Analysis*, John Wiley & Sons.

[11] Witkin A.P. (1983) *Scale-Space Filtering*, Proceedings 7th International Joint Conference on Artificial Intelligence, Karlsruhe, pp. 1019-1022.

[12] Yuille A.L. & Poggio T. (1983) *Fingerprints Theorems for Zero-Crossings*, A.I. Memo 730, Artificial Intelligence Laboratory, Massachusetts Institute of Technology.