SCALE DIFFERENCE CONSIDERATIONS IN CONJUGATE FEATURE MATCHING

Anthony Stefanidis
Peggy Agouris

Department of Spatial Information Science and Engineering
and the National Center for Geographic Information Science and Analysis
University of Maine
5711 Boardman Hall, Rm. 348
Orono, ME 04469-5711
Tel: (207) 581 2180, Fax: (207) 581 2206, e-mail: {tony, peggy}@spatial.maine.edu

Commission III, Working Group 2

KEY WORDS: Softcopy, Image Matching, Scale Space

ABSTRACT

This paper addresses the problem of matching features which have been recorded in two spatially overlapping images at substantially different scales. This phenomenon may be associated with foreshortening, in which case the scale differences are feature- and direction-dependent, or simply with the simultaneous processing of images of different scales, in which case the scale variations are obviously bidirectional and global in nature. We approach this problem by employing principles of scale space theory, which deals with the formalization and classification of signal contents and trends by examining the behavior of signals in various resolutions. Coarse resolutions convey only the dominant trends of a signal (corresponding to low-frequency information), while in finer resolutions information details (high-frequencies) are also included. When matching features recorded in substantially different scales in digital imagery, we are actually attempting to establish correspondences among different scale representations of the same object space scene. Typical matching techniques fail or perform poorly in terms of accuracy in such cases, because they do not consider that beyond geometric, scale differences are also of radiometric nature. The methodology presented in this paper proceeds by identifying scale differences among conjugate features, identifying proper image pyramid levels at which matching should be performed, and only then precisely matching conjugate features. The analysis of the matching results permits the transformation of matching uncertainties through scale space, and the derivation of realistic accuracy estimates.

1. INTRODUCTION

Matching, the task of identifying similar features in two or more spatially overlapping images, is a dominant research issue in digital image analysis, as it is a fundamental operation, involved in practically all photogrammetric applications. Despite the great advancements made in digital image matching, and the numerous algorithms and strategies developed employing geometric and radiometric similarity criteria to identify conjugate features, there still exist problematic cases, where matching fails to produce reliable results. The lack of sufficient radiometric variations is a typical example of such a case. These problems are, to a certain extent, adversely affecting the role of digital image matching for geoinformation generation, thus delaying the much anticipated full automation of the mapping process.

Among the cases where matching performs poorly, producing unreliable or even no results at all, is the case of features which have been recorded in two spatially overlapping images at substantially different scales. This phenomenon can be associated with isolated features within a pair of images of otherwise similar scales, or with the processing of images of overall different scales. The first case is rather object-oriented and its occurrence is dependent on specific image capturing and object shape combinations. The latter is an issue which is expected to receive much higher attention in the near future, as it is inherently associated with three-line sensor imagery (e.g. MOMS) which is becoming more widely available [Schneider & Hahn, 1992], while research also moves towards the fusion of aerial and satellite digital imagery for geoinformation extraction [Gruen et al., 1995], or the integration of digital imagery within geographic information systems [Agouris et al., 1996], whereby digital imagery of various scales is combined during the performance of complex digital image analysis processes.

In this paper we examine the problems occurring when attempting to match conjugate features whose images differ in scale. The presented method employs scale space concepts for the identification and accommodation of scale differences in matching.

2. SCALE SPACE THEORY

The information content of a signal is encoded in its values and their variations. These variations occur over a wide range of spatial extents, with macro-variations expressing major signal trends, and micro-variations expressing highly localized trends, manifesting themselves within spatially limited areas. The visual perception and distinction of macro- and micro-variations in images is an intricate human cognitive process, involving perception, reasoning and often intuition. As such, this task is fundamentally complex to be algorithmically duplicated and functionally mimicked by machine-supported operations.

The concept of examining the behavior of signals in multiple scales can be traced back to the seventies with research in hierarchical information structures [e.g. Tanimoto & Pavlidis, 1975]. However, scale space theory has been formally introduced and developed in the signal processing community only during the previous decade, with the papers of Witkin credited as introducing the concept [Witkin, 1983; Witkin 1986]. It deals with the identification and classification of trends encoded in the values of signals by analyzing the behavior of those signals in various resolutions. The scale space of an m-dimensional signal defined in the space spanned by $(x_1, x_2, \ldots x_m)$ is the (m+1)-dimensional space $(x_1, x_2, \ldots x_m, s)$ if and only if the additional parameter s expresses the resolution of the signal. Digital images are two-dimensional discrete intensity functions defined in the (x, y) space, and therefore their scale space is the three-dimensional (x, y, s) space. A discrete representation of the continuous in s scale space of a signal $f(x, y)$, comprising a set of n derivative signals $\{f_s^n(x, y, s_n)\}$ representing the original one in various resolutions (termed scale levels), corresponding to n distinct values $(s_0, s_1, \ldots s_{n-1})$ of the scale parameter s, is an n-order scale space family of the original signal. Figure 1 shows a scale space family and demonstrates how the original signal is decomposed at coarser scale levels.

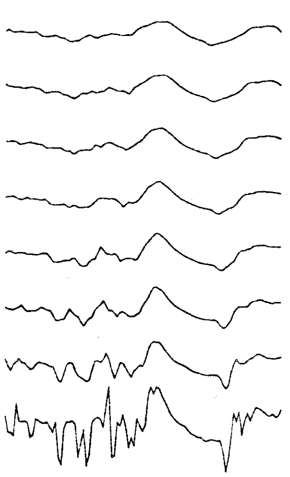

Fig. 1: A scale space family of a signal. The original signal is at the bottom and resolution decreases upwards.

For different scale parameter values, different scale space families of an original signal can be generated. This generation is performed through the numerical manipulation of the original signal. The aim of generating scale space families of signals is to provide representations in which the information content of a signal changes in a systematic and therefore exploitable manner. In order for this goal to be met, scale space generation has to follow certain rules [Lindeberg, 1990; Lindeberg 1994]:

- The scale space is generated through the convolution of the original signal with a single scale-generating function (or its discrete kernel) $k(x,y,s)$

$$f_s^n(x, y, s_n) = k(x, y, s_n) * f(x, y) \qquad \text{Eq. 1}$$

- The scale generating function has to be selected in such manner that, through its application, signal resolution will change monotonically for respective changes of the scale parameter s.

Both rules aim at the optimization of the interpetation potential of the generated scale space: the use of more than one scale-generating function (e.g. different functions for different scale parameter ranges) would make practically impossible the comparison of different scale space versions of a signal. The non-monotonic change of resolution would have similar implications.

Scale generating functions have to possess certain properties, in order to satisfy the above rules [Burt, 1981; Babaud et al., 1986; Meer et al., 1987], among which the most important are:

- symmetry, in order for direction independance to be satisfied,
- normalization, for ensuring the (essential in terms of data handling and processing) compatibility in value range of the multiresolution versions of a signal,
- unimodality, to avoid semantic distortions due to the disproportionate participation of distant information during scale space generation, and
- separability, for the alleviation of the computational requirements associated with scale space generation and manipulation.

Considering two-dimensionality, as is the case for digital imagery, the separability property of a scale generating kernel $k(x,y)$ allows its decomposition into two one-dimensional signals

$$k(x, y) = [k_x(x)]^T k_y(y) \qquad \text{Eq. 2}$$

and thus permits the use of different scale values in x and y, effectively allowing us to consider the scale space of images as a four-dimensional one. Actually, even for m-dimensional signals we could, in the same manner, consider the scale space as a $2m$-dimensional space. Scale space generation applied on digital imagery leads to the generation of digital image pyramids [Burt, 1984; Meer et al., 1987].

Arguably, the most important operation associated with scale space is to link the information of all scale space members together. This is achieved through feature tracing, which can be defined as the problem of identifying global features out of local signal properties, and of tracing the position and behavior of these features through various levels of the signal's scale space. Features are typically identified at the coarser signal levels, where overlaying high frequency phenomena and their interferences have been removed. Subsequently, these features are traced back to the finer levels, where their precise spatial positions are determined, free of the positional distortions which are introduced in coarser levels by the convolution with the scale generating kernels. This analysis of signals allows the identification and classification of major signal trends, thus making explicit the information which is inherently contained in the signal values [Lu & Jain, 1992].

3. SCALE VARIATIONS OF CONJUGATE FEATURES WITHIN A STEREOPAIR

The scale differences between conjugate features in a pair of spatially overlapping images can be:

- one-dimensional, associated with the foreshortening problem, and
- two-dimensional, associated with images which differ substantially in their orientation parameters.

Problems of the first type are highly localized and object-dependent. They occur only for certain features within a pair of images of otherwise similar scales, and will be the main focus of this paper. The extension of the presented methodology to two-dimensional is quite simple when taking into account the separability of two-dimensional scale generating kernels.

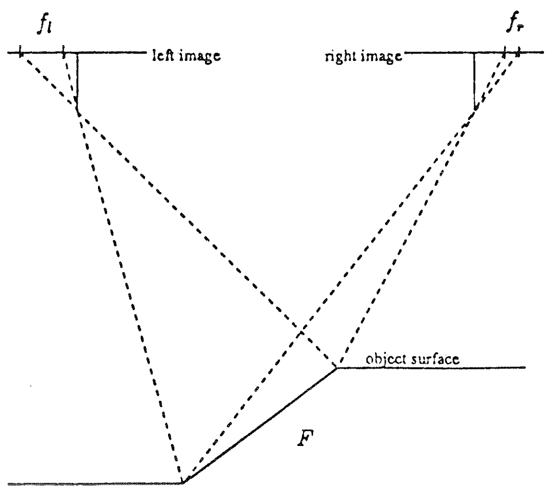

Fig. 2: Scale differences of conjugate features due to the foreshortening problem.

Figure 2 shows the foreshortening problem for a pair of photographs and a feature in the object space (ramp F at the center) for which the angle between the vertical and the surface normal is substantially different than 0. As it can easily be observed, the feature's inclination causes its image f_l in the left photo to be substantially larger than its image f_r in the right photo. In this manner, foreshortening causes the images of certain objects to be recorded at different scales in two stereomates. The geometric difference is accompanied by differences in radiometric scales.

Assuming error-free gray value registration, the recorded gray values are quantized expressions of the amount of energy incident to the light sensitive material at the corresponding sensor location, expressed by image irradiance [Wrobel, 1991]. According to the $\cos^4$ law of irradiance, image irradiance is proportional to a combined measure of object space surface reflectance characteristics and illumination conditions [Horn, 1986; Alvertos et al., 1989].

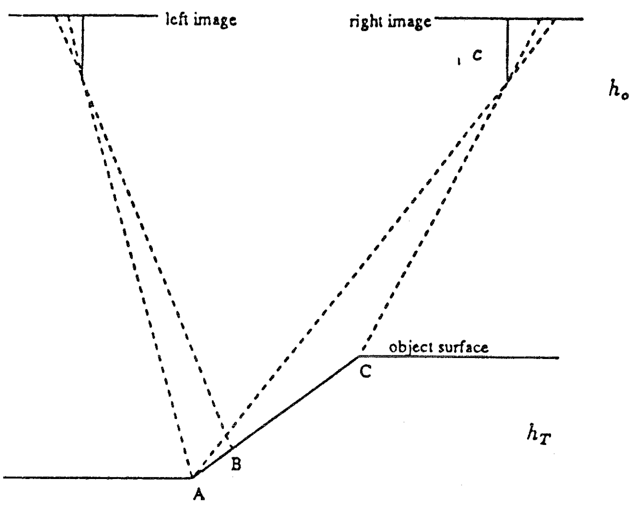

Fig. 3: Objel differences in conjugate feature registration in a stereopair.

The gray value recorded in a pixel is essentially expressing the irradiance of the object area which is the projection of this pixel in object space, hereinafter referred to as object equivalent pixel, or in short *objel*. In other words, the objel expresses the object space area which is imaged in a single digital image pixel. While all pixels of a sensor have the same size, their object equivalents vary according to the shape variations of the object space, as shown in Fig. 3.

The averaging operation associated with sensor charging and subsequent gray value assignment can be considered an operation equivalent to scale space generation. A series of images of an object space scene from various, constantly increasing heights, is actually forming a scale space family of the radiometric content of this scene. Images from higher exposure stations correspond to larger objel sizes and, consequently coarser scale levels than images captured from exposure stations closer to the actual object space. Conjugate pixel groups are actually scale space representations of their equivalent object space area, with image orientation (and by this we refer to both rotations and exposure station position), object space shapes, and sensor characteristics being the parameters defining the scale generation process. In a stereopair, the same sensor is used and, considering the excellent performance of metric quality cameras, it can be assumed that the effects of sensor characteristics during image formation are similar for conjugate features in stereopairs. The remaining combined effect of exposure orientation and terrain shape make image capturing through central projection unique in terms of scale space generation: scale might actually vary within an image, with various features belonging to different levels of the terrain scale space as it would have been obtained had an orthogonal projection been used.

4. EFFECTS ON LEAST SQUARES MATCHING

Matching features whose images are distorted due to foreshortening within a stereopair, using classic least squares techniques, will lead to failure when scale differences are sufficiently large [Stefanidis, 1993]. The reasons for this failure are:

- erroneous pixel correspondences (and consequently observation equations) are formed due to the dissimilarities of the initially selected conjugate patches, and this problem cannot be corrected during the iterative solution;
- negligence to directly access the radiometric scale differences of conjugate patches renders the mathematical model (which considers solely geometric relationships) inadequate; and
- violation of the flat terrain assumptions which are inherent in the geometric model used to relate conjugate patches (affine transformation).

Failure to bring conjugate features at comparable scales prior to matching will in essence result into matching by comparing non-conjugate gray values, and will therefore produce observations inconsistent with the geometric model used to relate conjugate image windows. In this case, two types of errors can occur (in direct analogy to errors in statistical decision making):

- truly conjugate pairs of features may be rejected by the matching solution due to the contradictory information provided by the comparison of non-conjugate gray values, or
- non-conjugate pairs can be matched by the adjustment solution due to the contamination of the matching process by erroneous observation equations.

5. MATCHING THROUGH SCALE SPACES

To overcome the previously described scale-pertinent problems, we can proceed in the following manner:

1. For every pair of matching candidates, the scale space behavior of the feature to which these matching candidates belong is examined, and large scale variations between them are identified.
2. The scales which are most proper for matching are determined.
3. Only then is precise matching performed, with the participation of radiometric scale parameters in a classic least squares matching manner.

By applying scale space techniques, we can substitute a stereopair by two image pyramids (a stereopyramid). It has been shown [Babaud et al., 1986] that the two-dimensional Gaussian function

$$g(x,y,s) = ke^{-\frac{x^2+y^2}{2s}} \qquad \text{Eq. 3}$$

is most appropriate for scale space generation using digital images. Typical photogrammetric multiresolutional techniques proceed by comparing similar pyramid levels of two stereomates (e.g. the 512×512 pixel version of the left stereomate is compared to the 512×512 pixel version of the right stereomate, the 4096×4096 left to the 4096×4096 right etc.). However, as we discussed in the previous two sections individual elements within these images may actually belong to dissimilar scale levels. Our first objective is to identify within the stereopyramid those scale levels of the two stereomates at which the specific feature currently processed is represented at comparable scales. In this manner, it is possible to establish stereo-correspondences by matching for example a feature at the 512 pyramid level of the left stereomate to its conjugate at the 1024 pyramid level of the right stereomate. Thus, matching is performed in the four-dimensional space (x,y,s_x,s_y), with s_x and s_y being the scale parameters in the x and y directions respectively.

Given approximate conjugate positions in a stereopair, we can examine the scale space differences of the features to which these points belong. The separability property of the two-dimensional Gaussian function allows it to be substituted by the product of two one-dimensional Gaussian functions

$$g_x(x,s_x) = k_1 e^{-\frac{x^2}{2s_x}} \quad \text{and} \quad g_y(y,s_y) = k_2 e^{-\frac{y^2}{2s_y}} \qquad \text{Eq. 4}$$

and thus permits us to substitute a two-directional search by two one-dimensional ones, resulting in great computational gains.

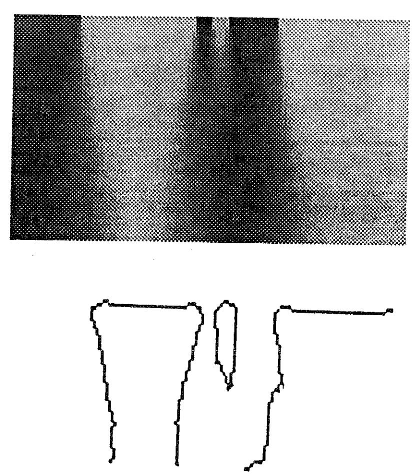

Fig. 4: A profile scale space image (top) and traces of features in it, detected as edges (bottom).

For searching in scale space, we introduce the concept of *profile scale space images*. As the name implies, a profile scale space image is the scale space representation of an image profile. It is stored and processed as any digital image file, but in this case, while one direction (columns) corresponds to image coordinates along the profile, rows are discrete representations of the continuous scale space, corresponding to discrete values of the scale parameter. The number of these discrete levels is user-defined, and its selection depends on the amount of information conveyed by the profile and the relevant storage and computational requirements. Figure 4 (top) shows such a profile scale space image, with the top row containing the original profile gray values, and rows underneath that containing increasingly smoother versions of the original signal, corresponding to gradually larger values of the scale parameter s. Thus, such an image will have a coordinate system (p, s_p), with p being the distance along the profile direction, and s_p being the scale parameter.

The use of digital image files (rather than simple signal values as is the case in typical scale analysis applications) to express the scale behavior of signals has great advantages, as it permits us to employ digital image analysis algorithms. To identify scale differences among conjugate features, we can match their corresponding profile scale space images. Matching proceeds similarly to least squares matching, but this time a shift in the s direction denotes a difference in scale among conjugate profiles. A shift in the profile direction p corresponds to a refinement of the initially available conjugate locations. By performing this matching process along the two directions (which for practical reasons are the base direction and its perpendicular), we can identify the exact correspondence in the stereopyramid

$$(x_l, y_l) \Leftrightarrow (x_r, y_r, s^r{}_x, s^r{}_y) \qquad \text{Eq. 5}$$

for comparing a specific feature.

This procedure can be enhanced when combined with edge detection. Figure 4 (bottom) shows the edges in a profile scale space image, which actually show how the various objects intersected by this profile (variations in top row gray values) behave in scale space. The extracted feature outlines describe not only the behavior of a single feature, but also its interaction with its surroundings. Robust features are remaining evident throughout the profile's scale space, while ephemeral ones disappear fast. The feature to which the given approximation belongs is the one which surrounds the available approximation. We can easily examine whether the given approximations lay on a robust or ephemeral feature. Points on robust features are better matching candidates. Furthermore, we can examine whether the given approximations lie on the same feature by comparing the major radiometric characteristics (absolute gray values, gradients) of the features to which the approximate points belong. This check can help us avoid gross matching errors which are associated with erroneous approximations.

Once scale space correspondences are established, assigning to a feature at a specific scale level in a stereomate its proper conjugate at the other stereomate's scale space, we can proceed with subsequent precise matching. Radiometric parameters can be introduced in it, to fully express the remaining radiometric differences between conjugate patches. By taking advantage of the diffusion equation of the Gaussian function, according to which

$$\frac{\partial g(x, s_x)}{\partial s_x} = \frac{1}{2}\frac{\partial^2 g(x, s_x)}{\partial x^2} \qquad \text{Eq. 6}$$

the derivative with respect to the scale parameter is equivalent to the second derivative with respect to the spatial coordinate, allowing us thus to directly introduce it in the linearized least squares matching observation equations (with the second derivatives of gray values as corresponding coefficients in the Jacobian matrix) [Stefanidis, 1993].

6. EXPERIMENTS

The mathematical models and matching procedure presented in the previous sections were tested in several experiments using both synthetic and real images. Synthetic data were generated by creating a DEM with substantial local inclinations (ramps, tall buildings etc.), assigning radiometric values to it and projecting back to fictitious exposure stations. By varying scale space inclinations, variations in scale differences among conjugate features were generated. The criteria by which the performance of the technique was judged were pull-in range in scale differences and positional accuracy of the obtained matching results. For ramp structures (like the one in Fig. 2) it was found that, even with excellent approximations typical least squares matching failed when the scale differences exceeded 20-30%. This range of scales is due to variations in the local radiometric content. Using the above described method we managed to match images of the ramp which differed by arbitrary amounts in scale. The identification of sufficient initial correspondences between features was the only limit. This task is indeed becoming less trivial as scale differences increase. When certain features were significantly different (in gray values) from their surroundings, we were even able to identify cases of occlusions and tag them as such. In terms of positional accuracy, our results were comparable to typical least squares matching results (on the order of 0.1 pixel). This should be considered quite successful when considering that these matching accuracies refer to cases where typical matching methods failed to produce any results. The reader is referred to [Stefanidis, 1993] for a more detailed description and evaluation of experiments.

7. COMMENTS

The presented technique addresses the problem of matching under the presence of extreme scale variations. The technique proceeds by identifying and taking into account such variations, and subsequently performing precise matching. Considering the automation potential of matching, this technique is viewed functioning as a module within a general matching strategy, complementing matching results in areas in which regular matching has failed. Of course it can function as a stand-alone matching module, but it would be computationally cumbersome to perform a detailed scale space analysis for every single patch to be matched. The developed concept of profile scale space images opens a new direction for scale space analysis. Not only do these images offer great visualization potential, allowing an operator to check the process, but they also have the great advantage of being, by design, compatible with digital image processing and analysis algorithms and software. This makes their complete integration in an existing general matching strategy very easy. They can be effectively combined with edge detection for automated, fast, and reliable scale space feature tracking, showing great promise for use towards image understanding.

REFERENCES

❑ Agouris P. & A. Stefanidis (1996) *Integration of Photogrammetric and Geographic Databases,* International Archives of Photogrammetry & Remote Sensing, Vol. XXXI, Part B4 (in print).

❑ Alvertos N., D. Brzakovic & R.C. Gonzalez (1989) *Camera Geometries for Image Matching in 3-D Machine Vision,* IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 11, No. 9, pp. 897-915.

❑ Babaud J. A. Witkin, M. Baudin & R.O. Duda (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.

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

❑ Burt P.J. (1981) *Fast Filter Transforms for Image Processing,* Computer Graphics Image Processing, Vol. 16, pp. 20-51.

❑ Burt P.J. (1984) *The Pyramid as a Structure for Efficient Computation,* in `Multiresolution Image Processing and Analysis' (A. Rosenfeld ed.), Springer-Verlag, New York, NY, pp. 6-35.

❑ Chin F., A. Choi & Y. Luo (1992) *Optimal Generating Kernels for Image Pyramids by Piecewise Fitting,* IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14, No. 12, pp. 1190-1198.

❑ Gruen A., O. Kuebler & P. Agouris (eds.) (1995) *Automatic Extraction of Man-Made Objects from Aerial and Space Images,* Birkhaeuser Verlag, Basel, Switzerland.

❑ Hahn M. (1990) *Estimation of the Width of the Point Spread Function within Image Matching,* International Archives of Photogrammetry & Remote Sensing, Vol. 28-3/2, pp. 246-267.

❑ Horn B.K.P. (1986) *Robot Vision,* MIT Press, Cambridge, MA.

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

❑ Lindeberg T. (1994) *Scale-Space Theory: A Basic Tool for Analyzing Structures at Different Scales,* Journal of Applied Statistics, Vol. 21, No. 2, pp. 224-270.

❑ Lindeberg T. (1994) *Scale-Space Theory in Computer Vision,* Kluwer Academic Publishers, Dordrecht, Netherlands.

❑ Lu Y. & R.C. Jain (1992) *Reasoning about Edges in Scale Space,* IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14, No. 4, pp. 450-468.

❑ Meer P., E. Baugher & A. Rosenfeld (1987) *Frequency Domain Analysis of Image Pyramid Generating Kernels,* IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 9, No. 4, pp. 512-522.

❑ Schneider P. & M. Hahn (1992) *Matching Images of Different Geometric Scale,* International Archives of Photogrammetry & Remote Sensing, Vol. XXIX, Part B3, pp. 295-302.

❑ Stefanidis A. (1993) *Using Scale Space Techniques to Eliminate Scale Differences Across Images,* Ph.D. Dissertation, Dept. of Geodetic Science & Surveying, The Ohio State University.

❑ Tanimoto S. & T. Pavlidis (1975) *A Hierarchical Structure for Picture Processing,* Computer Vision, Graphics and Image Processing, Vol. 4, No. 2, pp. 104-119.

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

❑ Witkin A.P. (1986) *Scale Space Filtering,* in `From Pixels to Predicates' (A.P. Pentland ed.), Ablex Publishing Co. Norwood, NJ. pp. 5-19.

❑ Wrobel B.P. (1991) *The Evolution of Digital Photogrammetry from Analytical Photogrammetry,* Photogrammetric Record, Vol. 13, No. 77, pp. 765-776.