MATCHING IMAGES OF DIFFERENT GEOMETRIC SCALE

Franz Schneider and Michael Hahn
Institute for Photogrammetry, University of Stuttgart
Stuttgart, FRG

Commission III

0 ABSTRACT

In typical photogrammetric image processing applications scale differences between the images are usually small. There are exceptions, however. In close range applications it can not be avoided that at least the image scales of some recorded objects differ considerably from image to image. The same problem we meet in the stereo-image recordings of the MOMS-02 digital space-camera. The pixel size related to the terrain surface is 4.4 m in the high resolution nadir channel and 13.2 m in the backward and forward locking channels. For many digital photogrammetric tasks like point transfer, orientation, DTM reconstruction, etc., the problem of image or feature matching has to be solved under such conditions.
The paper presents theoretical and experimental investigations into the matching problem with significantly different geometric image scales. The special objective is to arrive at the highest possible precision by guiding the feature based matching by the high resolution image. This implies multilevel and focusing techniques. The quality of the matching is assessed by the experimental investigations with simulated and real images.

Keywords: Algorithm, Image Matching, Focusing, Resolution, Stereoscopic

1 INTRODUCTION

The reason for this paper is a very practical one: Within the MOMS (modular electrooptical multispectral scanner) - project exists the problem of matching images which are significantly different in geometric scale. Images of the terrain surface are recorded by a three line scanner. Because of limitations of the real-time recording capacity it has been decided to establish different resolution channels to solve the stereo tasks of photogrammetry and mapping. Therefore in the panchromatric, high-resolution stereomodule of STEREO-MOMS the pixel size in both, the backward and forward channel is 13.2 m, whilst in the downward channel the size of one pixel measures linearly 4.4 m on ground (Ackermann et al., 1989). So naturally the questions arises on how to exploit the more of information gained by the downward channel in combination with the lower resolved other two channels. At first the scale difference of factor 3 between the channels has consequences for all automatic measurement processes. This implies the selection of features in each channel as well as the establishment of feature correspondences. The typical steps of photogrammetric data evaluation like point transfer, orientation and point determination (Ebner and Kornus, 1991), DTM reconstruction (Hahn and Schneider, 1991), which are investigated in the MOMS-project, depend all directly or indirectly on the measurement process. Because the quality of the measurement process propagates to all subsequent evaluation processes it is of primer interest to investigate possibilities on matching images of different scale. Similar problems we meet in photogrammetric close range applications or in navigation with image sequences. The scale differences of objects imaged from different places may amount to 100%, in navigation distinctly more in some tasks. In the latter case the appearance of an object in an image sequence when approaching e.g. by a mo

ving vehicle, may evolve from a small blob to a detailed thistle bush. This famous example is presented by Bobick and Bolles (1989) to explain, that recognition and tracking of objects needs knowledge based systems and a suited representation space for the description of an object at different scales. Sester (1990) poses the question of treating the representation and reasoning problem in a multiresolution of a single image.

The closeness in considering two images recorded with different geometric scale and two level of pyramidal representation of an image are at hand. Assuming lowpass filtering of an image with a smoothing radius of σ pixels and resampling a "new" image is generated. The recorded and the derived "new" image has a scale ratio of $\sigma:1$. The most simple procedure for the measurement process would be to transform the higher resolved image to the scale of the lower one by lowpass filtering. Matching can then performed with other images of similar scale. Of coarse by proceeding in this way the high quality information in the large scaled image is lost because of the smoothing operation.

The procedure for intensity based matching of images presented by Hahn (1990) allows to estimate geometry and radiometry transformation parameters, and furthermore the difference in the smoothness between the two images. The parameter used to measure smoothness differences is the Gaussian scale parameter σ. This scale parameter of Gaussian smoothing is added to a set of other parameters, i.e., it is also estimated by the procedure. If the coarser resolved image is resampled to the sampling density of the higher resolution image the effect of smoothing as mentioned above is still present. By application of the matching procedure proposed by Hahn the smooth image and a stereo partner are matched. This implies, that the transformation parameters between two images of different geometric scale are

estimated.

In this paper we explore the matching problem based on features. For feature extraction we use the point operator proposed by Förstner (1987, 1989). The points selected in the different images have to be matched e.g. to solve the problem of point transfer or the problem of DTM reconstruction. Today image analysis and matching are based on pyramids and some strategies, which guide the process usually from coarse to fine resolution (for examples cf e.g. Ackermann and Hahn, 1991). For the analysis of the single image, tracking of features through scale space is proposed by Bergholm (1987).The scale space Witkin (1983) is a kind of an image pyramid which consists of an infinite number of smoothed image levels. The characteristic of this representation is that the scale (smoothness) parameter is continuous which is of benefit for tracking. In the matching case, e.g. in DTM reconstruction, it is usual to work with standard image pyramids, almost with a Gaussian image pyramid. The smoothing levels of this pyramids are fixed and the spatial resolution between two consecutive levels decreases from the bottom to the top of the pyramid. In this case not the tracking idea is dominant but the questions due to approximate values, efficiency of the algorithm and reliability of matching are addressed.

The concept of this investigations and according to this the organization of the paper is as follows: (1) We want to find out characteristics of the point operator in scale space. This implies questions due to the tracking of the point location of the interest point from fine to coarse and vice versa. Moreover the significance of the selected features in scale is important. The example used in section 2 is a synthetic image. (2) The scale space tracking of features in real images is discussed in section 3. Influences due to physical (illumination, etc.) and geometric (perspective projection) aspects can be observed. Mainly the consequences for the image location of the features and for the stereo displacements are of interest. For the synthetic image as well as for the real images we want to restrict ourself to one dimension. The one-dimensional real signals are taken from a epipolar stereo pair.

1.1 RELATED WORK

Related work which has not been addressed up to now mainly concerns the representation and reasoning about features in scale space. Since the early days of computer vision it was quite clear that high-level processes need and have to use a lot of different knowledge for reasoning. For low-level processes such as edge detection a common belief was that they are simply data driven without use of explicit knowledge. This assessment today changes. A lot of operators for edge detection have been proposed in parts but research has clearly demonstrated that the edge detected by these techniques do not give satisfying results (Lu and Jain, 1992; Bergholm, 1987). Because of this the role of reasoning in low level processing comes into the center of interest. As one of the first Witkin (1983) analyzed thoroughly the behavior of edges in scale space. He reflected work of Marr (1982), who argued "that physical processes act on their own intrinsic scales". A scale-structured representation, called the interval tree, he introduced to describe contours over scale. This organization characterizes the information over a broad range of scale, that means, between a coarse resolution level with a small number of edges and a fine resolution level with usually significantly more edges. This organization is expected to be useful Witkin for matching or object reconstruction tasks. The sym-

bolic image description over scale is generated by the zero-crossings of a Laplacian of Gaussian (LoG) convolved image, in which the Gaussian σ is addressed as scale. Three typical edge behaviors in Gaussian scale space were observed by researchers: (1) The locations of edges in filtered images using different scale parameters can (and in general will) be different. (2) in scale space zero crossing occurring at finer scales can disappear at coarser scales. (3) Spurious edges are those that occur at a coarser scale but have no corresponding edges at a finer scale. For more details cf. Lu and Jain (1992). This two authors presented the most sophisticated algorithm up to now, called RESS, which stands for reasoning about edges in scale space. The knowledge about edge behavior in scale space is explicitly formulated in 35 rules and is used in RESS to select proper scale parameters, to correct dislocation of edges (1), to recover missing edges (2), and to eliminate noise or false edges (3). The separation of significant edge information from noise mainly has been also the aim of Bergholm (1987) in his multiscale tracking procedure, called edge focusing, as well as Canny (1986), who proposed a multiscale edge detector.

Finally we want to address the work of Heikkilä (1989), which has some similarity to our investigation because he used also the Förstner point operator. In the one-dimensional case the point operator coincides with an edge detector. The estimated point position locates the edges with subpixel accuracy, at least in theory. Therefore the estimation of the edge location in scale space will be interesting. In the paper of Heikkilä the properties of the operator by varying the scale parameter of the integrating window are investigated. The integration works on the squared gradient image. Consequently the interest operator is a nonlinear edge operator. So far this is presumably the main difference to other edge operators like the linear LoG operator mentioned before. Our interest is not the problem of a varying window size. We investigate the behavior of the operator applied to a series of Gaussian smoothed images. In the 2 D case the operator can be formulated as

$$[\nabla(G_{\sigma_1} * f(x,y)) \ \nabla^T(G_{\sigma_1} * f(x,y))] * G_{\sigma_2} \qquad (1)$$

With σ_1 the scale space image of $f(x,y)$ is generated, whilst σ_2 is responsible for the size of the weighting in the window. the dyadic product $\nabla\nabla^T$ indicates the nonlinearity of the operator. If σ_1 is constant and σ_2 varies mainly the following characteristic for edges can be observed:

With increasing σ_2 the number of edges decreases, i.e. the edges fuse or disappear with coarser scale σ_2. Just invert is the situation when the scale space parameter σ_1 varies and σ_2 is constant:

296

Even though on the finer scale more edges are detected which disappear at coarse levels, it is to observe that in tracking from coarse to fine contour elements merge together.

In the next sections we want to deepen insight mainly from an experimental point of view.

2 EDGE DETECTION IN MULTISCALE IMAGES

The surface obtained by convolving the image with a Gaussian kernel with varying scale parameter is called the scale space image. In practice only a series of a finite number of multiscale representations can be generated. For our experiments we therefore approximate the Gaussian kernel by binomial kernels which are recursively obtained by convolution with $\frac{1}{2}(1\ 1)$.

The interest operator proposed by Förstner for point detection is described in detail in (Förstner and Gülch 1987). In the two dimensional formulation the operator is designed to detect, classify and precisely locate corners, circular features and other isotrop textures. The 2x2 matrix (e.q. 1), written in terms of convolutions and matrix multiplication, is the so-called normal equation matrix used within the point location step. In the one-dimensional case this matrix reduces to a scalar quantity and the point operator is just an edge operator. Denoting the signal with $f(x)$ and the scale space image generated from the signal by convolution with a Gaussian kernel by according to

$$g(x,\sigma) = f(x) * G_\sigma,$$

the formalism for edge location with the 1D-interest operator reads as

$$(p * g_x^2)\ \hat{x} = px * g_x^2 \qquad (2)$$

As outlined above p can also be a Gaussian, i.e. $p = G_{\sigma_1}$. The scale σ_2 stands for the width of the convolution window and by this also for the window size. Varying σ_2 as investigated by Heikkilä (see above) produces scale space of the squared gradient image. In this investigation we keep σ_2 and choose σ_2 throughout all the experiments. In the discrete approximation of the Gaussian kernel by a binomial kernel this fits to a window size of 5 pixels. The second operator in (2) denoted by px can be computed according to $px(i) = p(i) \cdot x(i)$ for all pixels i within the window. For convenience the coordinates x are centered to the mean of the window, so that px is symmetric and shift invariant. Thus $\hat{x}$ gives a field of edge positions, in which for each pixel the corresponding edge location is estimated. The quantity $w = p * g_x^2$ also found for each pixel is sometimes called interest value or weight or response of the window operator. It measures the roughness of sale space image. In the experiments presented below the roughness w and edge locations $x+\hat{x}$ are computed for all pixels x of the multiscale image in a series of discrete scale levels. For calculation of the gradients the image is convolved with (-1 1).

The edge detection scheme is based on the roughness image. The local maxima in w are considered to indicate edges. The window size for extracting local maxima (or suppressing non-

maxima) is chosen in accordance with size of the convolving operator p, i.e. in the experiments also 5 pixel are used for the non-maxima window size. Though a larger non-maxima window may suggests that the remaining edges are more distinct, in our opinion this is not recommendable. A quite not rare observation in scale space is that with increasing scale parameter the weight of a local maxima may decrease considerably faster than that of a neighboring edge. One common aspect in the strategies behind edge focusing (Bergholm, 1987) or RESS (Lu and Jain, 1992), but also the work of Lindeberg and Eklundh (1990) on scale space blobs is that those features are of interest which achieve high weights on all levels or on the coarser scales.

The rest of the paper is devoted to experiments. Even the first experiences with simulated and real image data have shown that for example the very nice rules for reasoning in the linear LoG scale space presented by Lu and Jain can not simply be applied to the nonlinear edge detector (eq. 2). The characteristics of the zero crossings in the LoG scale space we have outlined above. Moreover a further property we have to include which occurs just in the case of nonlinear operator: new edges may be generated as the scale parameter increases (Yuille and Poggio, 1986).

3 TRACKING IN A SYNTHETIC IMAGE

In this section we would like to illustrate the tracking problem at the example of an idealized synthetic image. The signal shown together with the multiscale representation in fig. 3.1 has just two intensity levels. It consists of a series of pulses whose width is chosen randomly. The number of pixels plotted is approximate 200.
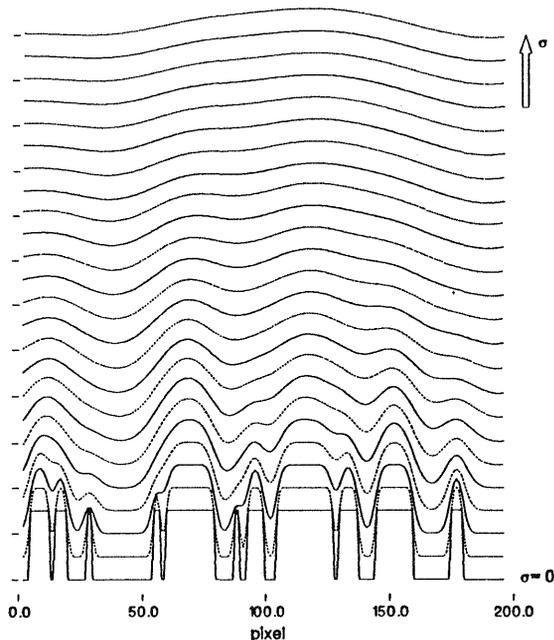

Figure 3.1 Multiscale image with 23 levels and step size $\Delta\sigma = 1$

The following three figures show the roughness of the signal in scale space. The logarithm ln $w(x,\sigma)$ is plotted. The partitioning of the scale range ($\sigma = 0-23$) in three parts mainly intents to improve the visual impression from details of the plots. The global

297

view of the figures shows dramatic changes on true scale levels (figure 3.2) and significant generalization on the coarser levels (figure 3.4).
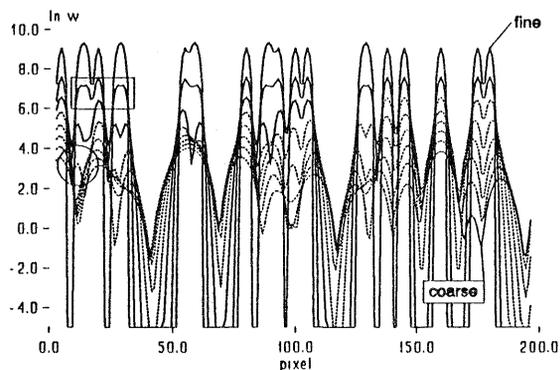

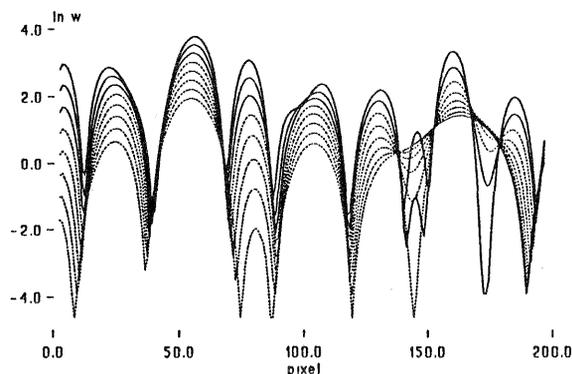Figure 3.2 Roughness of the signal in scale space (σ varies from 0 to 7, $\Delta\sigma$= 1)

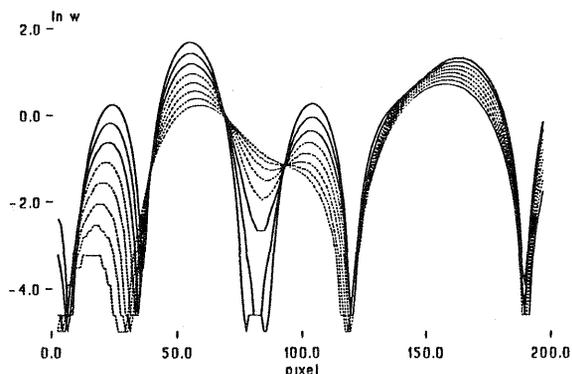Figure 3.3 Roughness for the scale levels σ= 8-15, $\Delta\sigma$= 1

Figure 3.4 Roughness for the scale levels σ= 16-23, $\Delta\sigma$= 1

Further in figure 3.4 two nodes (at x = 70 and x = 90) are easy to recognize. A local maxima in w (at x= 160) could also be noticed. It is verged on local minima and which are far away from the edge.

A closer look to the data shows some of the scale space phenomena mentioned in section 1.1. For this purpose the range $0 \leq x \leq 35$ is zoomed in figures 3.5 - 3.9. The $\Delta\sigma$ intervals in the multiscale image are chosen smaller with step sizes $\Delta\sigma^2$= 0.5. The area marked by a rectangle in figure 3.2 (cf. also figure 3.5) shows that an edge is splitted into two edges increases as the scale parameters.
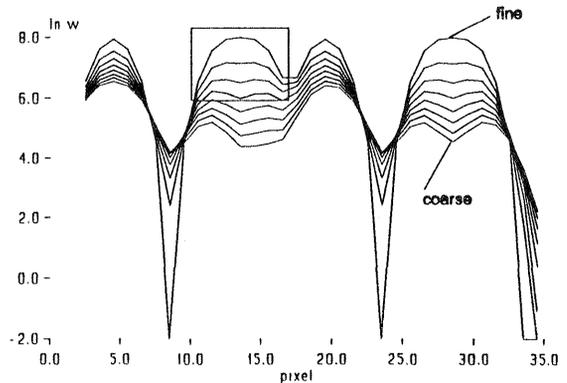
Figure 3.5 Roughness of the scale space images for high resolution levels (σ= 0-2, $\Delta\sigma^2$= 0.5)

The area marked by a circle in figure 3.2 indicates a local minima which switches to a local maxima. That means the edge is detected up to a certain level and then it is lost. This effect in figure 3.6 (at x= 11 and x= 27) can be visually tracked very well.
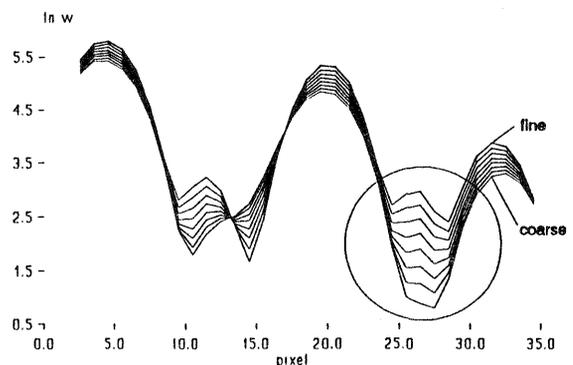

Figure 3.6 Roughness for the scale range (σ= 3.1-3.5, $\Delta\sigma^2$= 0.5)

The node to the right (x= 13) seems to have a certain meaning for this process. For a lot of other nodes which are stable in their position over a certain σ range similar processes are to be recognized: local maxima which are weaker and lie in the neighborhood of such a node will be absorbed, i.e. the edge is lost at this point. The comparison of figures 3.6 and 3.7 gives further hints to the stability of nodes.
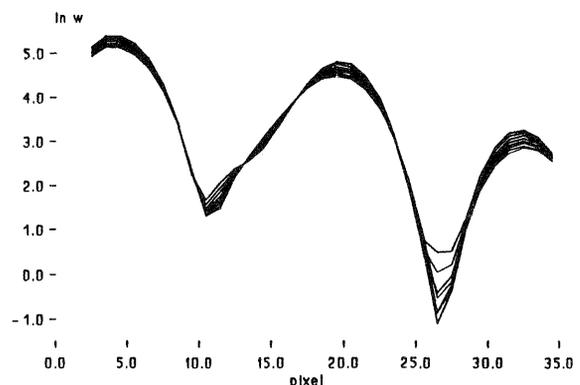

Figure 3.7 Roughness for the scale range (σ= 3.6-4.0, $\Delta\sigma^2$= 0.5)

In figure 3.8 the estimated location $\hat{x}$ (eq. 2) are plotted for all pixels. The underlying multiscale image is that of figure 3.1. At $\sigma = 0$ the position of the pixels without point estimation are marked for reference. It is interesting to note that the order of the estimated location $\hat{x}$ was never changed, i.e. for all pixels i the relation $x(i) \leq x(i+1)$ produces $\hat{x}(i) \leq \hat{x}(i+1)$. The equal sign was only observed at $\sigma = 0$. The dislocation of edges over scale is easy to see at the dominant edge in figure 3.8.
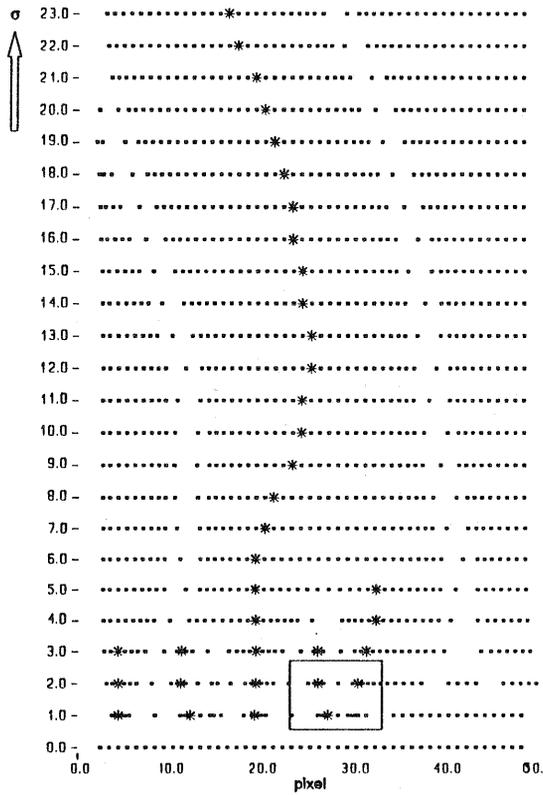
Figure 3.8 Location of edge candidates in scale space. The detected edges are marked by *

Again, figure 3.9 gives a detailed view of the locations x estimated with subpixel position. The tracking of the edges from coarse to fine seems not a severe problem with these small differences between the scale levels. Between $\sigma = 2$ and $\sigma = 3$ we left a gap. The missing information caused by this gap does not rise any reasoning problems in this example.
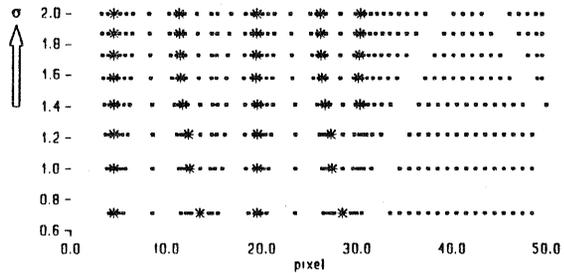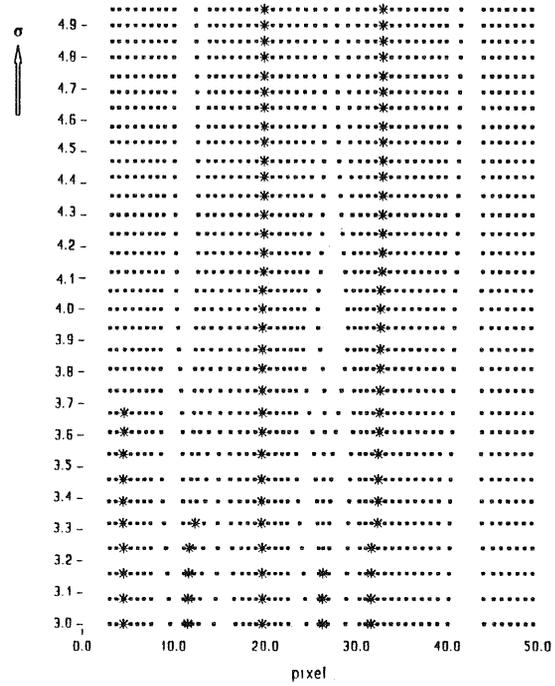
Figure 3.9 Detailed view on the estimated locations $\hat{x}$ in scale space (σ range is 0.7-5.0, $\Delta\sigma^2 = 0.5$)
The splitting process of an edge (at x = 30) with increasing scale can be observed at the level $\sigma = 1.4$.

4 TRACKING AND MATCHING IN REAL IMAGES

The task of tracking images through scale space and the consequences for matching will be demonstrated with examples of real images. The real images are taken from Hahn and Förstner (1988). The stereo pair is given in epipolar geometry. The real data example used consists of corresponding rows from these images. The scale space images are visualized in figures 4.1l and 4.1r. The appended l and r denote the left and right stereo partners, respectively.
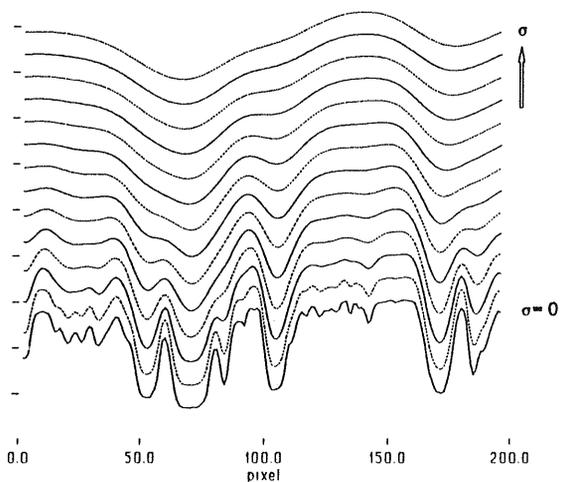
299

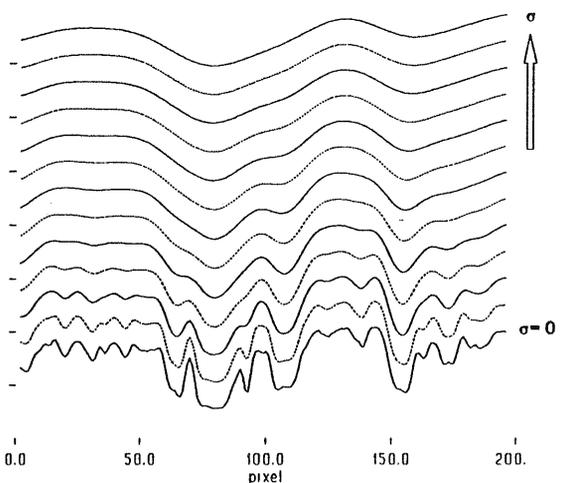Figure 4.1l Scale space image with 13 levels and step size Δσ= 1 (left stereo partner)

Figure 4.1r Scale space image with 12 levels and step size Δσ=1 (right stereo partner)

In the fine scale levels striking differences between two signals are recognized resulting from geometric and radiometric scale differences between two images. These influences are also easily observable on the coarser scale. The roughness images of the corresponding image pair can be seen in figures 4.2l and 4.2r for the finer scale levels and in figures 4.3l and 4.3r for the coarser scale levels σ= 8-13.
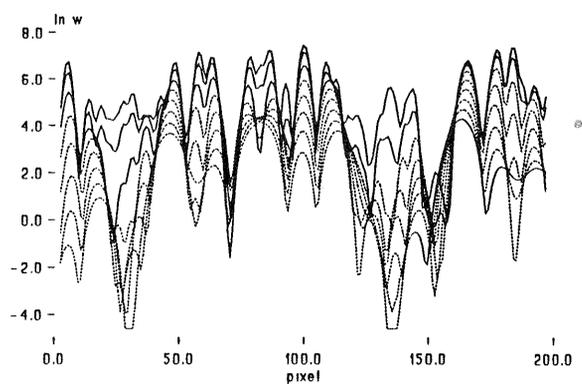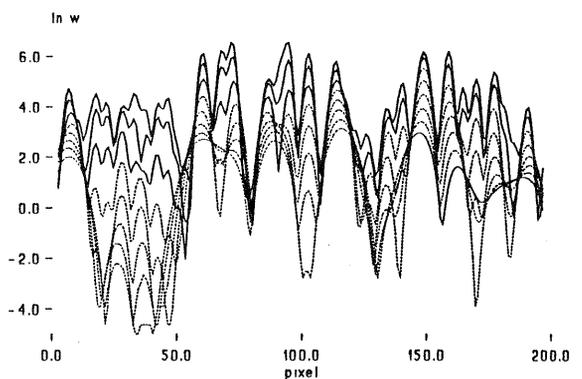

Figure 4.2l Roughness of the signal in scale space (σ varies from 0 to 7, Δσ= 1)

4.2r Roughness for the signal in scale space. (σ varies from 0 to 7, Δσ= 1)
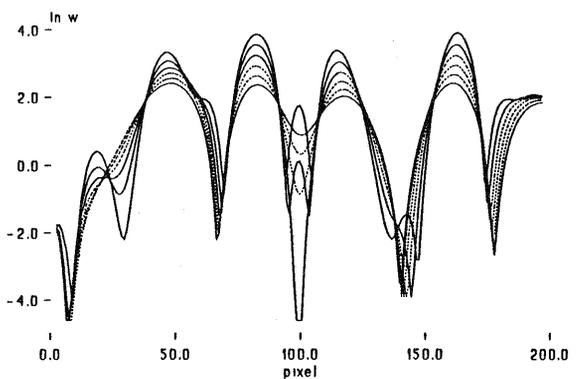
Figure 4.3l Roughness for the scale levels σ= 8-13, Δσ= 1

In w

2.0 –

0.0 –

– 2.0 –

– 4.0 –

    | | | | |
   0.0   50.0   100.0   150.0   200.0
                   pixel
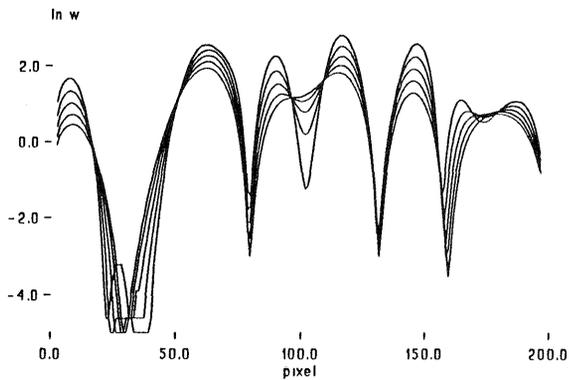
Figure 4.3r Roughness for the scale levels σ= 8-12, Δσ= 1

In the case of detailed manual analysis the direct comparison of the roughness images on the finer scales (figures 4.2) is a really severe problem. The job of solving this problem on the coarser (figures 4.3) levels is much simpler. Similarity in this generalization allows to establish corresponding edges. We do not want to recall the statements observed in section 3 with the simulated image, though they can also be observed in these real images. But one characteristic feature which can be seen in figure 4.2l is not discussed up to now. In the area marked by a circle (cf. figure 4.2l) a spurious edge is generated (at x= 26, in level σ= 3.0). But this edge is only present on a small scale range (moves to x= 28 in level σ= 4.2).

The contours of the estimated edge locations in scale space for this stereo pair are drawn in figures 4.4l and 4.4r. The corresponding edges are identified and marked. In figure 4.4r two dominant edges are present. Unfortunately the left of both is just outside the region shown in figure 4.4l.
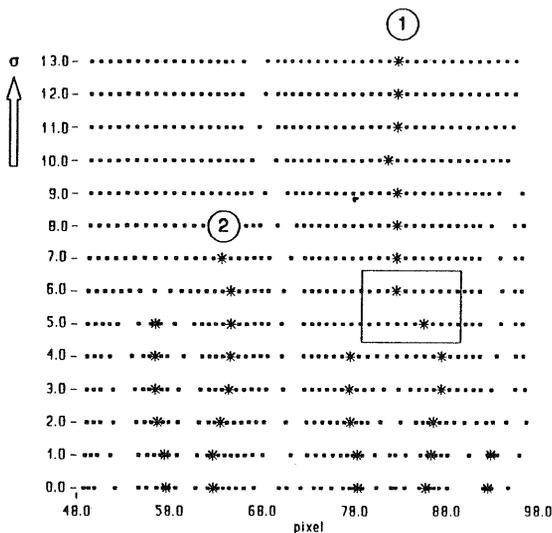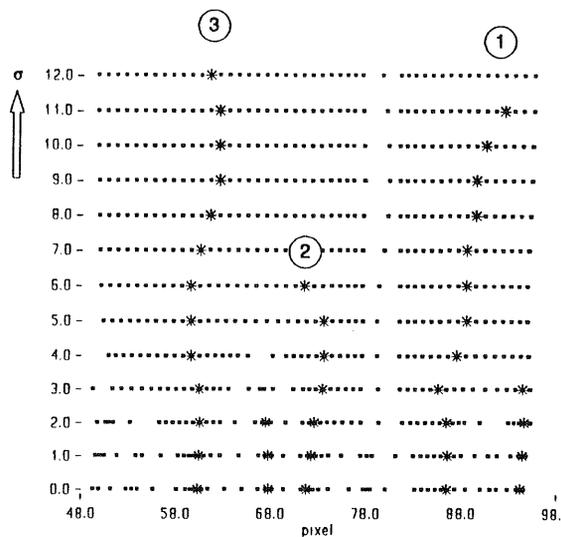
Figure 4.4r Location of edge candidates in scale space. The detected edges are marked by * (right stereo partner)

Finally we address some consequences of tracking edges through scale for matching. The correspondence of edges is established manually and with signal control. We start (a) the tracking at level σ= 3 and (b) at level σ= 12. The results are listed in table 1.

Table 1 Influences of tracking to point location and matching

	image left Δx (pixel)		image right Δx (pixel)		left - right Δpx (pixel)	
	mad	mse	mad	mse	mad	mse
σ_{start} = 3 N= 15 points	1,03	0,62	0,54	0,77		
σ_{start} = 12 N= 8 points	1,05	0,55	0,31	0,39	0,89	0,68

Figure 4.4l Location of edge candidates in scale space. The detected edges are marked by * (left stereo partner)

Δx measures the dislocation of the edges between the start level (σstart) and the level σ= 0, i.e. the unsmoothed image. Δpx measure the difference in these dislocations, i.e. it gives the resulting systematic effect for the parallaxes. mad stands for mean about differences and mse denotes the mean square error. This statistic shows that in theses cases a dislocation of about 1 pixel is observed on the average. The largest shift between two levels (Δσ= 1) amounts to 2.9 pixels. The area in which this shift occurs is marked by the rectangle in figure 4.4l. The systematic effect for the parallax nearly reaches about 1 pixel on the average. Because such a parallax error directly propagates on the calculated heights, we have to investigate this effect in more detail in future.

301

5 CONCLUSION AND OUTLOOK

It is quite clear that reasoning in scale space will be a key to come to a satisfying solution for edge detection and location. To approach the reasoning we analyzed simulated and real image data to gain insight into problems of tracking and matching in multiscale images. Aspects due to the tracking of edges located by the Förstner Operator are observed and discussed. Those edges which can be tracked through a large scale range are most important for solving the matching process. They may serve for the elimination of spurious edges and give approximate information for matching edges which are missed after smoothing with a small scale parameter. Further we expect that other knowledge like smoothness of the displacement field will be a complementary and helpful information, suited to support the tracking and matching in combination.

REFERENCES

Ackermann, F., et al. (1989): MOMS-02-D2 Ein multispektrales Stereo-Bildaufnahmesystem für die zweite deutsche Spacelab-Mission D2. In: Geoinformationssysteme 3, S. 5-11.

Ackermann, F., Hahn, M. (1991): Image Pyramids for Digital Photogrammetry. In: Ebner, H. et al.: Digital Photogrammetric Systems, Wichmann-Verlag, Karlsruhe.

Sester, M. (1990): Multiscale representation for knowledge Based recognition and Tracking of Objects in Image Sequences. Int. Arch. of Photogrammetry and Remote Sensing Vol 28 3/2, pp. 868-888.

Bergholm, F. (1986): Edge Focusing. IEEE PAMI 9, No 6, pp 726-741.

Canny, J. (1986): A Computational Approach to Edge Detec tion, IEEE PAMI Vol 8, No. 6, pp 679-689.

Förstner, W., Gülch E. (1987): A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centres of Circular Features. Proc. of the Intercom. Conf. on Fast Processing of Photogrammetric Data, Interlaken.

Förstner, W. (1988): "Statistische Verfahren für die automati sche Bildanalyse und ihre Bewertung bei der Objekterkennung und Vermessung", Habilitationsschrift, Uni Stuttgart.

Hahn, M. (1990): Estimation of the Width of the Point Spread Function within Image Matching. Int. Arch. of Photogrammetry and Remote Sensing, Vol 28-3/2, Wuhan, Peoples Rep. of China, pp 246-267.

Hahn, M., Schneider, F. (1991): Feature Based Surface Recon struction - A Hierarchical Approach Developed for MOMS-02 Imagery. International Geosience and Remote Sensing Symposium, Vol 3, pp 1747-1750, Helsinki, University of Technology Espoo, Finland.

Hahn, M., Förstner, W. (1988): The Applicability of a Feature Based Matching and a Least Squares Matching Algorithm for DEM-Acquisition. Int. Arch. of Photogram-metry and Remote Sensing, Vol 27, B9, pp 321-331.

Heikkilä, J. (1989): Multiscale Representation with Förstner Operator. The Photogrammetric Journal of Finnland. Vol 11, No. 2, pp 40-54.

Lindeberg, T., Eklund, J-O. (1990): Scale Detection and Region Extraction from Scale-Space Primal Sketch. Proc. 3rd Int. Conf. on Comp. Vision, Osaka, pp. 416-426.

Lu, Y., Jain R. L. (1992): Reasoning about Edges in Scale Space. IEEE PAMI 14, No. 4, pp. 450-468.

Marr, D. (1982): Vision. San Francisco: W. H. Freeman.

Witkin, A. (1983): Scale Space filtering. Proc. Int. Joint Conf. Artif. Intell., Karlsruhe, pp. 1019-1021.

Yuille, A., Poggio, T. (1986): Scaling Theorems for Zero Crossings, IEEE PAMI, Vol 8, No. 1, pp. 15-25.