IAPRS vol. 31, part B3, pp. 758-763.

HIGH RESOLUTION SURFACE RECONSTRUCTION OF A LANDSCAPE FROM LARGE SCALE AERIAL IMAGERY BY FACETS STEREO VISION – AN EXTENDED TEST

M. Schlüter and B. P. Wrobel
Institute of Photogrammetry and Cartography
University of Technology Darmstadt
Petersenstr. 13, D-64287 Darmstadt, Germany
wrobel@gauss.phgr.verm.th-darmstadt.de
ms@gauss.phgr.verm.th-darmstadt.de

International Society of Photogrammetry and Remote Sensing, Commission III, Working Group 2
XVIIIth ISPRS Congress, Vienna, Austria, July 1996

KEY WORDS: Photogrammetry, Vision Sciences, Matching, Orthoimage, DEM/DTM, Geometric Surface Reconstruction, Stereoscopic Aerial Imagery, Facets Stereo Vision

ABSTRACT

We present results of high resolution surface reconstruction in object space by Facets Stereo Vision, using large scale aerial images. After some references to the basics of Facets Stereo Vision, we try to classify possible sources of difficulties for the surface reconstruction. Consequently, we choose appropriate examplary areas in object space with different topographical character. For these areas, the results of Facets Stereo Vision are given in form of orthoimages and unsmoothed contours. Further on, corresponding quality statements like residual images and images of standard deviations are presented as supplied by Facets Stereo Vision.

Topics like the occurrence of image blunders, moving cars in object space during the exposure interval, changing illumination during the exposure interval, terrain noise and discontinuities in object space are covered. We discuss their treatment as well as the reached accuracy.

The results seem to be promising for further scene understanding tasks.

1 ABOUT FACETS STEREO VISION

Facets Stereo Vision is a method for surface reconstruction in object space, cf. [6]. DTM-heights, orthoimage grey values and radiometric parameters are directly introduced as unknown parameters, and are estimated simultaneously by least squares adjustment from two or more images containing one or more spectral bands.

For the examples in this paper, the relationship between image space and object space is described by the perspective camera model. Both, DTM-heights and orthoimage grey values, are represented as $2\frac{1}{2}$-D functions over regular grids (called facets), using bilinear interpolation between grid points. Regularization, as discussed in cf. [7], is required to overcome ill-posedness of the image inversion and to bridge areas, in which the gradients of the image grey value signal are low in relation to image noise. As least squares adjustment requires Taylor-linearization, approximate start values of the parameters are needed. Therefore, the whole reconstruction process is formulated as a multigrid procedure: Step by step, a finer resolved DTM is calculated for each level of the image pyramid of the input data. For details see [5]. A consistency check with respect to occlusions is useful to avoid the matching of pixels to hidden parts of the surface.

Our aim is not only to reach a surface reconstruction fitting tightly to the real world surface, but also to gain correct quality statements for our results as well.

2 WHICH RESULTS CAN BE EXPECTED?

Contrary to an operator with an analytical plotter, who measures single point positions and maybe using explicit knowledge for point selection, Facets Stereo Vision works area based, assuming e.g. that the real world's surface is a Lambert reflector and can be approximated by a continuous $2\frac{1}{2}$-D function over regularly spaced grid points, etc. . Keeping this in mind, it is clear that problems of the surface reconstruction have to expected at those areas, where the used mathematical model does not fit properly to the real world.

For a brief discussion of error avoidance or detection and for considerations concerning further developments of postprocessing algorithms, it seems to be helpful to distinguish three classes of possible error sources, even if they cannot exactly be kept distinct from each other:

First, let us have a look at disturbances, which only occur locally in a minority of the used images: E.g. image blunders, objects which move in object space during the exposure interval or local light spots by total reflection. We prove by two examples (cf. fig. 2, 3) that simply using more than two images simultaneously can avoid a lot of trouble.

A second group of errors is caused by terrain noise: Surface elements, which would require a finer resolution of surface facets than used in the reconstruction process can not be approximated properly. Image resolution sets the bound to the resolution in object space. So this error cannot be circumvented easily. But the quality criteria supplied by Facets Stereo Vision offer powerful possibilities for detection and further semantic analysis.

The third class of errors is caused by a deficiency of the $2\frac{1}{2}$-D surface model: Vertical surface areas can just be approximated by steep facets. A 3-D surface model should remedy this situation, cf. [4].

Last but not least it should be mentioned, that Facets Stereo Vision does not use any explicit knowledge to eliminate parts of the surface – of course the outer surfaces of buildings, leaved trees, parking cars, etc. become part of the reconstructed surface.

The examples presented in this paper deal with these problems and show the surface reconstruction results of appropriate areas.

3 SCENE AND DATA

We use 4 black and white aerial images for our experiments, cf. fig. 1. They are taken by the aerial camera ZEISS RMKA with an image format of $23 \times 23cm^2$ and a calibrated focal length $c_k = 153mm$. The flying altitude of $600m$ above ground corresponds to an image scale $m_B \sim 1 : 4000$. The exposures were taken in the south of Germany at the beginning of springtime, so the vegation is still leafless. The exposure interval between the images 133/135 and 268/270 is about $20min$.

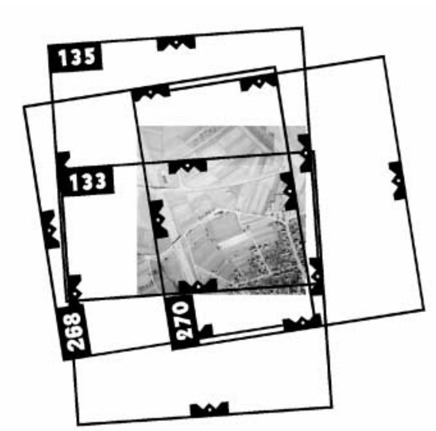


Figure 1: Overlap of aerial images ($m_B \sim 1 : 4000$) in relation to the reconstructed orthophoto

The images were scanned by the photogrammetric scanner ZEISS PS1, with $8bits$ per pixel and a pixelsize of $15 \times 15 \mu m^2$. The mathematical model of Facets Stereo Vision deals with noisy image data – straightforward, neither geometric nor radiometric preprocessing of the images has been applied.

The scene contains different degrees of difficulty of topographical surface types: Relatively flat agricultural areas, steep slopes, different kinds of leaved and unleaved vegetation and man-made objects like buildings and highway bridges.

In our experience, difficulties of surface reconstruction usually grow within the multigrid process by growing image scale. So we choose an image scale as large as possible to get in contact with all of the problems of terrain noise and discontinuities in object space. As distortions like image blunders usually disappear at higher levels of the image pyramid, the performance of Facets Stereo Vision concerning this point, too, can be shown to be best at finest image resolution. Consequently, in this paper we only present results gained with the original pixel values and resolution from scanning.

4 PARAMETER SETTINGS AND SOME SPECIAL HINTS

We started the multigrid process at the 9th level of the image pyramid. The corresponding resolutions are given in tab. 1. Start values for the heights were obtained by simply bilinear interpolating the coordinates of four outlying tiepoints from the bundle block adjustment.

level:	pixel/facet:	size:
9	image pixel	$4 \times 4mm^2$ in image space
	orthoimage facet	$32 \times 32m^2$ in object space
		$\sim 2 \times 2$ image pixels
	height facet	$128 \times 128m^2$ in object space
		$\sim 8 \times 8$ image pixels
1	image pixel	$15 \times 15\mu m^2$ in image space
	orthoimage facet	$12.5 \times 12.5cm^2$ in object space
		$\sim 2 \times 2$ image pixels
	height facet	$50.0 \times 50.0cm^2$ in object space
		$\sim 8 \times 8$ image pixels

Table 1: Facet parameters and image pixel sizes for highest and lowest multigrid level

The test area covered $\sim 600 \times 600m^2$ in object space, corresponding to $\sim 1.4 \cdot 10^6$ estimated height parameters and $\sim 22.4 \cdot 10^6$ estimated orthoimage grey value parameters. Because of the high image noise of $\sigma_0 \sim \pm 6 - 8$ grey values we choose non adaptive curvature minimization as regularization procedure with a weight factor of $\lambda = 1 \cdot 10^5$. The relativly large size of the heigt facets in relation to the pixel size causes some implicit regularization, too.

In principle, the long time delay (cf. section 3) between the two flight strips can be taken into account by simultaneously estimating one separate set of orthoimage parameters for the images of each strip. The basic idea of this proceeding is very similar to the treatment of color images, as explained by [1]. By that way, the different surface texture caused by different shadow locations in the images of the different strips can be taken into account precisely. In this paper we treat all images as isochronous exposures. So, small errors in those regions, were shadow gives texture, have to be expected.

The contours plotted into the orthoimages on the following pages of this paper exactly reproduce the results obtained by Facets Stereo Vision by bilinearly interpolating within each height facet for every orthopixel position. Please note, that our goal for this paper is to document the original reconstruction results of Facets Stereo Vision, but not the results of any additional fine contour smoothing algorithm!

We believe that it might be useful not only to offer height and orthoimage data, but also the accompanying quality criteria to a further semantic analysis. This in mind, the decision whether to build a 'good looking' smooth result or not should be dependent on the aim of the following data procession steps.

5 ACCURACY CHECKS

We obtained the parameters of the outer orientation of images by bundle block adjustment based on image coordinates measured in the analogue images. The bundle block adjustment reached an accuracy of $\pm 3cm$ for the height component of the tiepoints in object space. To ensure that the set of our transformation parameters is really correct for the digital images, we compared a set of 10 signalised control points with the corresponding interpolated heights of Facets Stereo Visions surface reconstruction: The absolute height differences all were less than $|\Delta Z_{max}| < 10cm$, with a mean of $\overline{|\Delta Z|} = 4cm$. This result is one indicator that the accuracy of Facets Stereo Vision is comparable to the accuracy which can be reached by an analytical plotter (~ 0.09 $^0\!/_{00}$ $\cdot$ *flight altitude*), as reported earlier for close-range imagery by [3]. Also, comparing the results numerically to those gained from the analytical plotter as well as carrying out a visual stereoscopic check of correctness by reprojecting the estimated 3-D object coordinates into the digital images with subpixel accuracy, yield similar results.

The following examples are intended to give an impression of morphological correctness and of the consequences of terrain noise or model deficits, respectively.

6 DISCUSSION OF OUR TEST RESULTS

In fig. 2 'Meadow' we show the reconstruction results of an area with only low terrain noise. The shaky course of the contours results from the very low equidistance of $\Delta = 20cm$ in comparison to the standard deviations of the heights which are in the range of $\sigma_Z \sim \pm 3cm - \pm 15cm$. The obtained contours are another indicator that σ_Z seems to be very realistic.

Low textured areas beside the little brook are well recorded, the little brook itself seems to be modeled in a morphologically correct way. Fig. 2c allows a closer look at the treatment of image disturbances. Even by using only 3 images for the reconstruction it can be observed, that image blunders mainly affect the residual image which belongs to the disturbed image – by that way, the orthoimage is not disturbed! This effect was already reported for parts of a reseau cross in one image by [2].

Example 'Highway', fig. 3, shows the effect of another disturbance which appears only in one image: In principle, the effect of the driving car is not different from a blunder in one image, except that it covers a larger area, cf. fig. 3c. The more images take part in Facets Stereo Vision, the better are the results. Of course, parking cars which appear at the same position in all images, become part of the reconstructed surface, cf. fig. 4a. The strips on the highway show very nicely, how large grey value gradients in object space result in low standard deviations of the estimated heights, cf. fig. 3b. Again topographic details like the ditch and the slope of the highway are reconstructed in a morphologically correct way.

Contrary to the first two examples, where we treated disturbances in single images, the last example (fig 4) shows two interesting situations where the used $2\frac{1}{2}$-D surface model does not fit properly to reality. First, the building is an example for the occurrence of discontinuities in object space, cf. fig. 4a. Vertical parts of the surface can not be well approximated by a $2\frac{1}{2}$-D surface model – a better way could be the use of a 3-D surface model as proposed by [4]. Second, the tree in the upper left of fig. 4a is an example for the presence of terrain noise, typical for unleaved trees. While the shadow in nearly isochronous exposures gives a fine texture and a good surface reconstruction result on the ground, small height disturbances can appear near the centre of a tree.

Depending on the aim of the reconstruction, one might want to detect and to correct the observed gross errors. Both, σ_0 and σ_Z are valuable indicators for automatic detection and manual correction. σ_G of the orthoimage grey values reflects at first order the number of pixels per orthoimage grey value facet and seems to be less important.

Automatic correction algorithms should be based on a semantic analysis of the observed errors. The resulting residual images could be a helpful tool, e.g. for the calculation of local σ_0's or for a texture analysis to select errors caused by unleaved vegetation.

7 CONCLUSION

We present examplary results (heights, orthoimages and quality criteria) of fully automatic surface reconstruction by Facets Stereo Vision with a very large resolution in object space. Simply using more than 2 images simultaneously for surface reconstruction can avoid to meet several difficulties. Further error avoidance or detection can be based upon Facets Stereo Visions quality statements.

The results including their quality statements are a promising starting position for further semi- or fully automatic scene understanding tasks.

8 ACKNOWLEGEMENT

The project was supported by *Deutsche Forschungsgemeinschaft (DFG)* under WR 18/1-4.

We are very grateful to Dr. Kiefer and Mr. Weiser from the *Landesamt für Flurneuordnung und Landentwicklung Baden-Württemberg*, who made available to us the complete underlying data, including digital and analogue images.

REFERENCES

[1] **Kaiser B. and Wrobel B. 1996.** *Facets Stereo Vision (FAST Vision) Applied to digital Colour Images.* XVIIIth ISPRS Congress, Comm. III, WG 2, Vienna.

[2] **Heipke C. 1989.** *An Integral Approach to Digital Image Matching and Object Surface Reconstruction.* In: Grün A., Kahmen H. (eds): Optical 3 (three) D Measurement Techniques. Wichmann Verlag, Karlsruhe: 347–359.

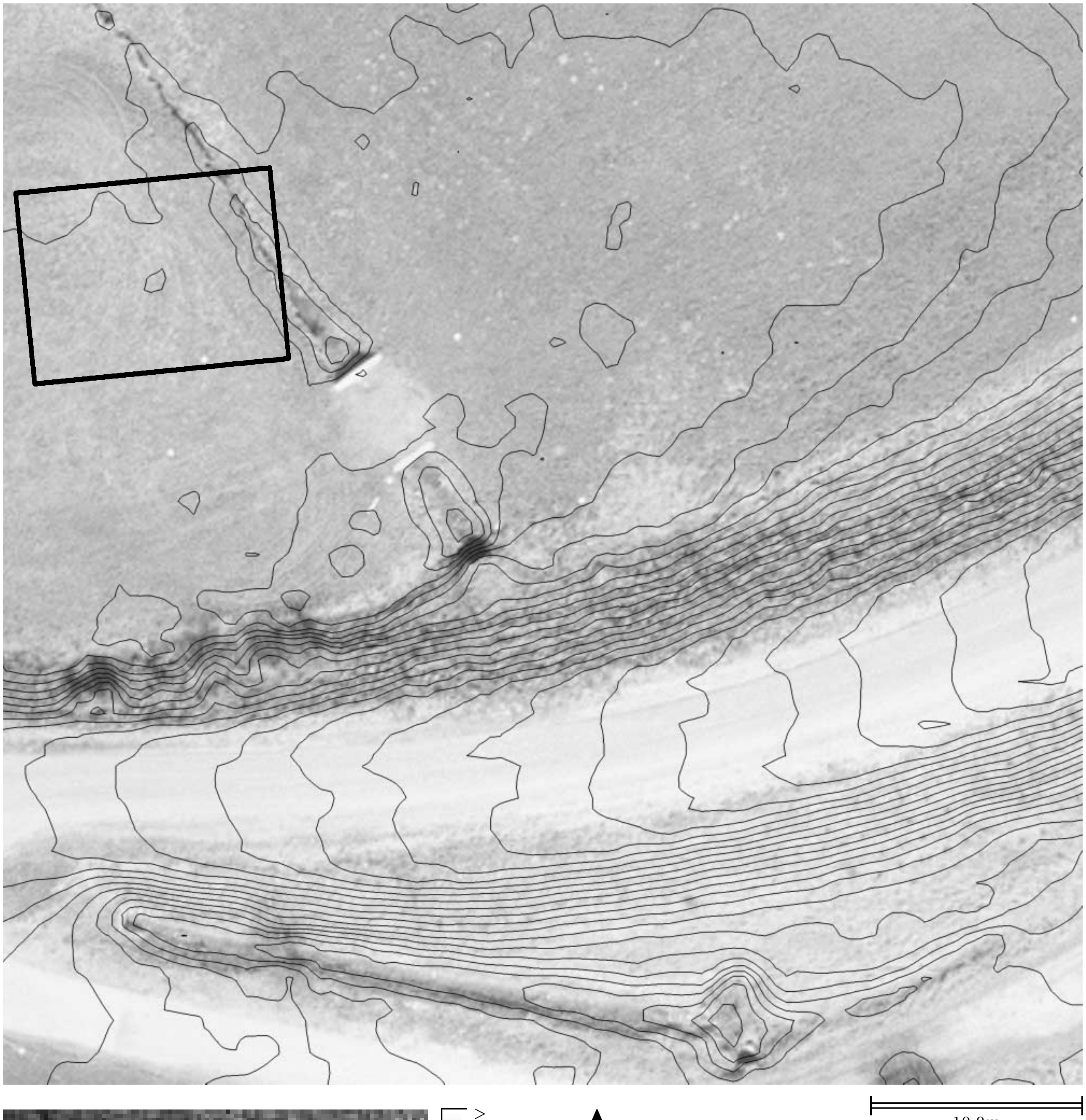
[3] **Kempa M. and Schlüter M. 1993.** *DEM Evaluation by an Operator and Facets Stereo Vision: A Comparison Based on Close-Range Imagery.* 2nd Conference on Optical 3-D Measurement Techniques, Zurich, Switzerland, 502–509

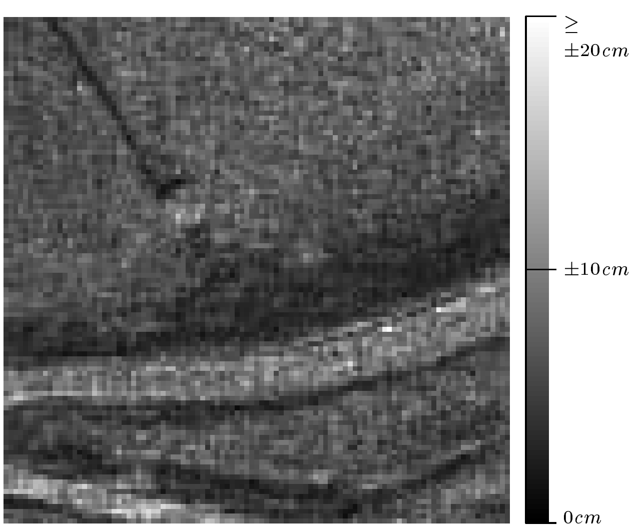
[4] **Schlüter M. 1994** *Object Space Based Surface Reconstruction with Discontinuities – An Approach.* ISPRS Comm. III Symposium 'Spatial Information from Digital Photogrammetry and Computer Vision', vol. 30, part 3/2, 737–744, Munich, Germany

[5] **Weisensee M., 1992.** *Algorithmen und Modelle zum Facetten-Stereosehen.* Deutsche Geodätische Kommision C 374, München.

[6] **Wrobel B. 1987.** *Digital Image Matching by Facets Using Object Space Models.* 4th Int. Symp. on Optical and Optoelectr. Appl. Science and Engineering, The Hague, Netherlands, SPIE 804, 325–333.

[7] **Wrobel B., Kaiser B., Hausladen J., 1992.** *Adaptive Regularization Of Surface Reconstruction By Image Inversion.* In: Förstner W., Ruwiedel St. (eds): Robust Computer Vision. Wichmann Verlag, Karlsruhe: 351–371.





▲ Fig. 2a: Orthoimage and contours,
surface reconstruction using 3 images
orthoimage scale: $1:300$
equidistance of contours: $0.2m$

◀ Fig. 2b: Standard deviations σ_Z of DTM, fig. 2a

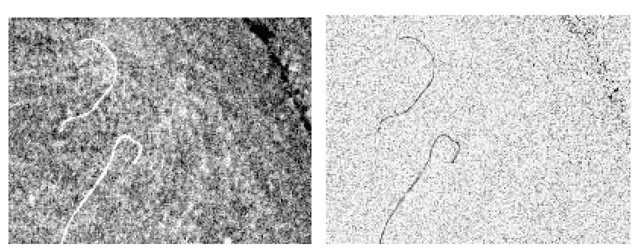
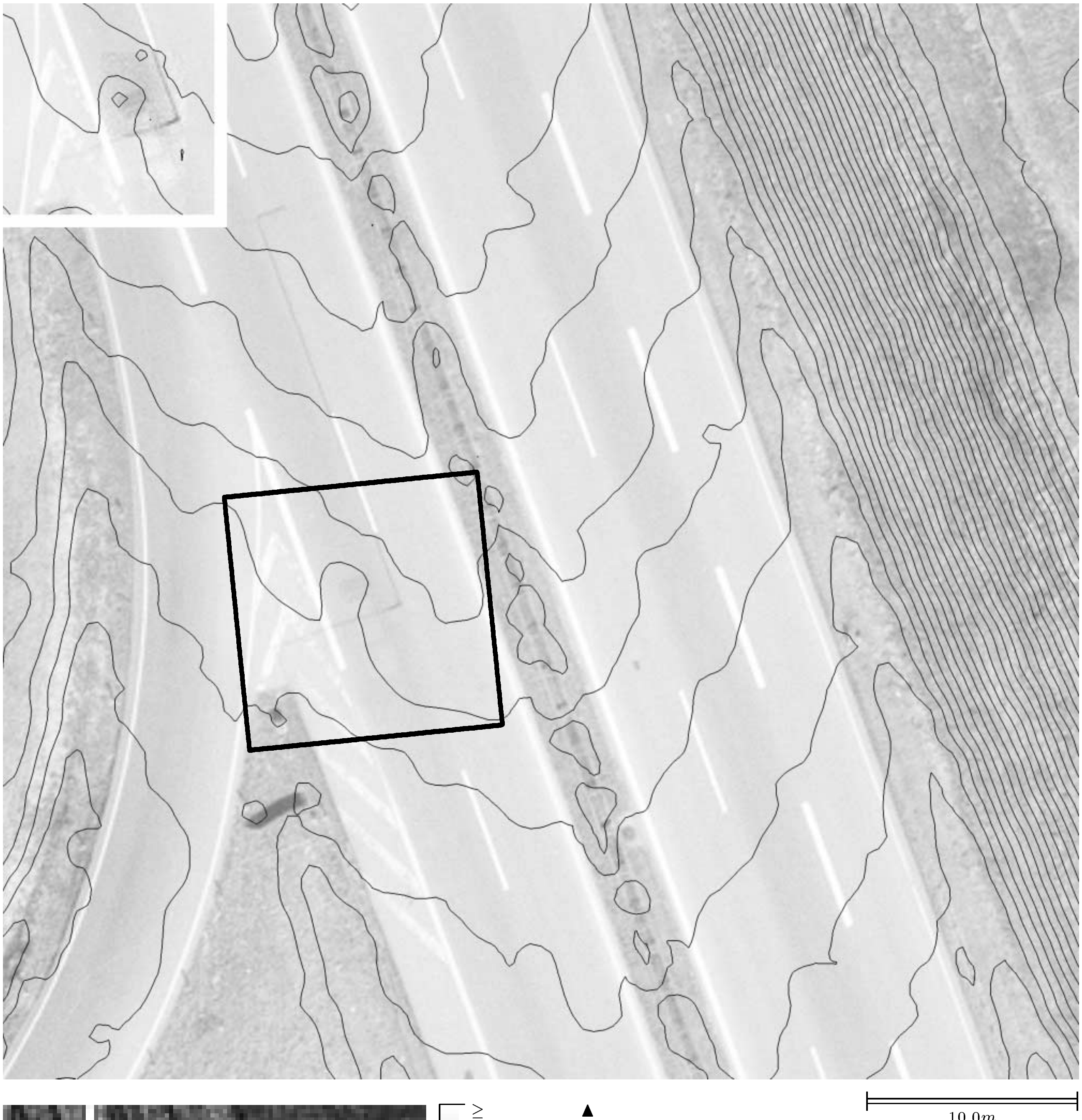
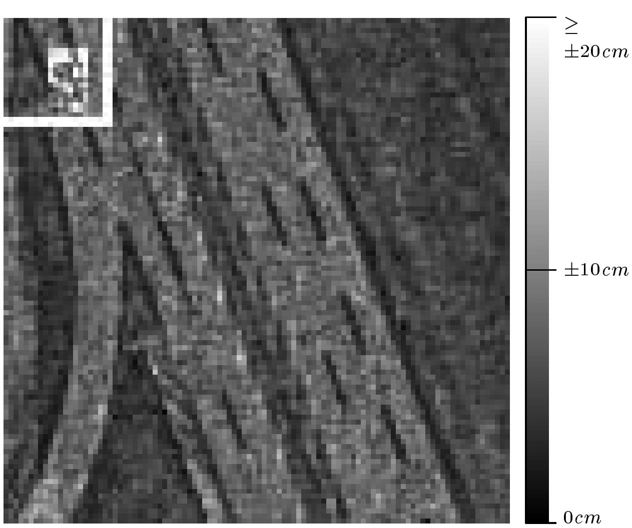


Fig. 2c: Window of original aerial image 135, according to the marked area in fig. 2a, and corresponding residual image

Figure 2: Example *'Meadow'*



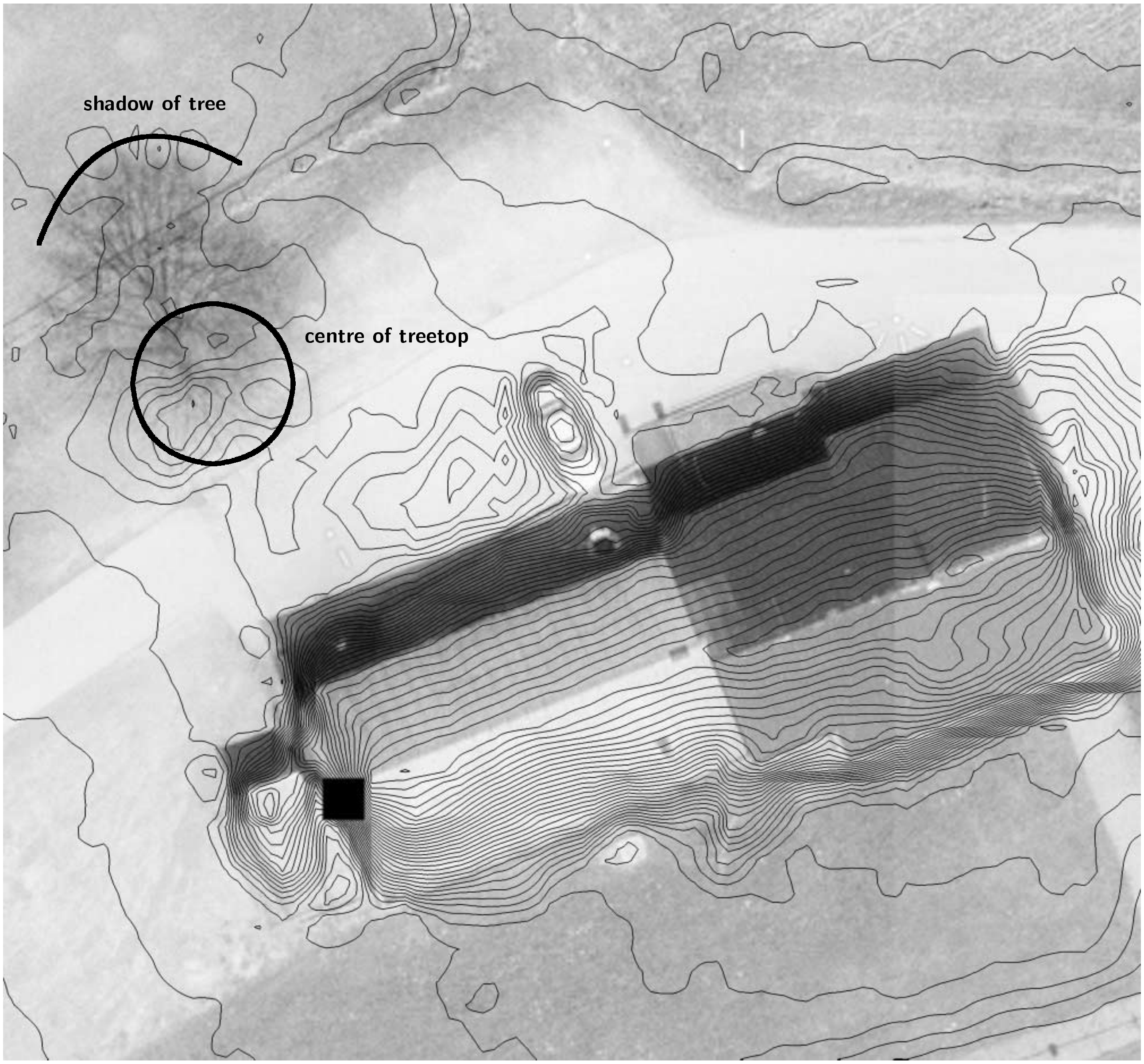
▲ Fig. 3a: Orthoimage and contours,
surface reconstruction using all 4 images
(upper left: surface reconstruction by 3 images only!)
orthoimage scale: 1 : 300
equidistance of contours: $0.2m$



◀ Fig. 3b: Standard deviations σ_Z of DTM, fig. 3a

Fig. 3c: Windows (200×200 $pixels$, contrast enhanced) of the original aerial images, corresponding to the marked area in fig. 3a

Figure 3: Example *'Highway'*



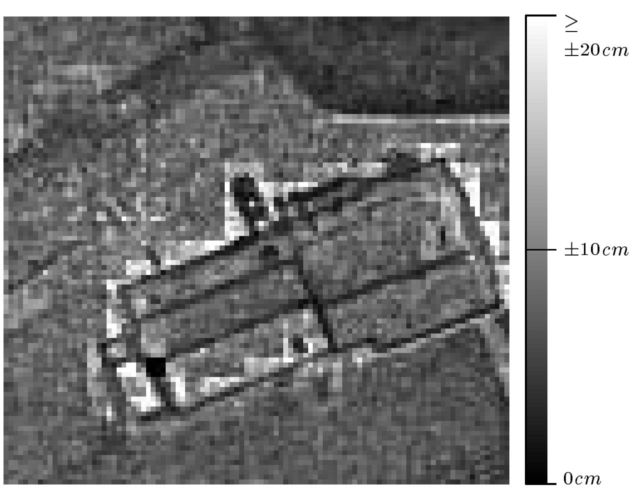
▲ Fig. 4a: Orthoimage and contours, surface reconstruction using all 4 images

orthoimage scale: $1:300$

equidistance of contours: $0.2m$

$10.0m$

▼ Fig. 4b: Standard deviations σ_Z of DTM, fig. 4a



▼ Fig. 4c: Standard errors of unit weight σ_0

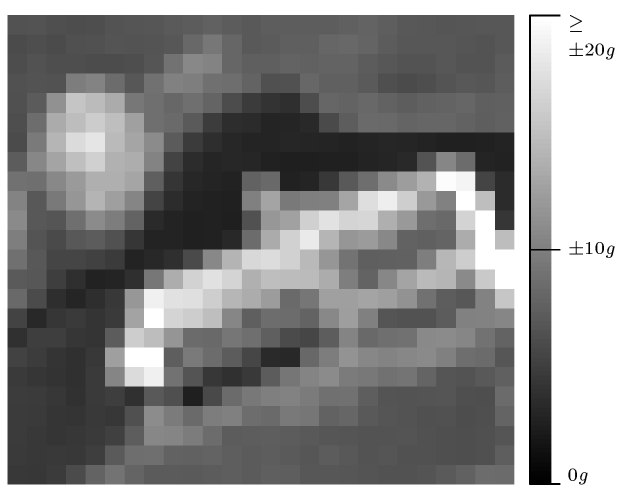


Figure 4: Example *'Building & Tree'*