Experiments with Facets Stereo Vision (FAST - Vision) for Object Surface Computations

Th. Korten, B. Wrobel, M. Franek and M. Weisensee
Technical University Darmstadt
Petersenstrasse 13
D 6100 Darmstadt
Federal Republic of Germany

Presented paper to the 16th International Congress of ISPRS Kyoto, Japan, 1988
Commission III

Abstract

The fundamentals of facets stereo vision (FAST-Vision) have been presented in a separate paper at this congress, see (Wrobel, 1988).

Due to the well-known complexity of computer-stereo-vision the detailed features of FAST-Vision were investigated by computer simulation. Results were obtained by computing linear images and calculating an object elevation profile as well as an object density model. In this paper aspects of accurracy and convergence will be discussed and other properties which can be derived from verying object surface textures and the number of facets for height and object density. The influence of a third image on object surface resolution is dealt with. The computational approach is deterministic. The picture scale is 1:12 000 . All results are given for the linear version of FAST-Vision. Their general validity has to be tested yet.

1. Introduction

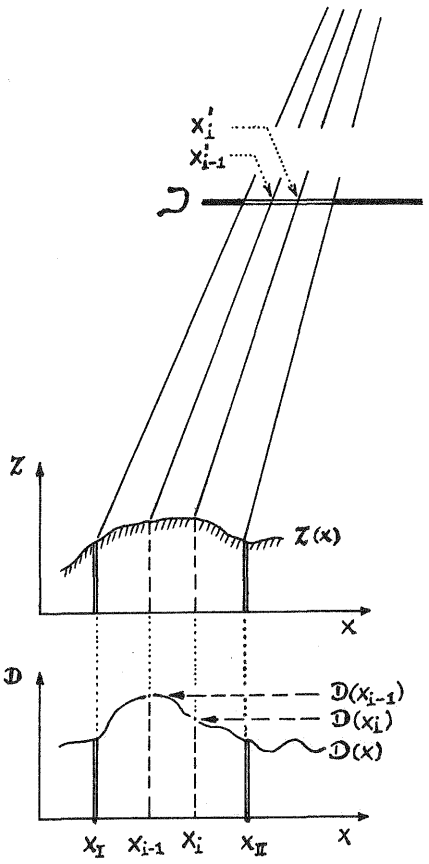
To examine the properties of FAST-Vision a linear version has been developed, which is expected to give results valid also for two dimensional calculations, as texture orthogonal to the base line does obviously not affect the method, neither accuracy nor convergence (Wamser,1987). In this context we will regard 'texture' as the change in object density, i.e. the gradient of the object surface grey value function.

The ' linear FAST-Vision program system ' consists of two parts: one to calculate linear images, the other to compute an object elevation profile. In spite of the general approach, which includes a simple linear illumination model, the mathematical formulation is restricted to unknown parameters in object height and surface density. Corresponding to this aspect there is no illumination model involved when creating linear images. Furthermore the influence of noise is reduced to round-off errors resulting from the fact that grey levels are treated as integers. The following chapter gives a brief introduction to both programs.

2. Programs and Mathematics

2.1 Generating linear Images

The data evaluated by program 'FAST' are distinct integer values within a predefined gray-scale range. A picture is formed by an array of pixels of a certain length specification. Pixel length is fixed to 20 µm. The object gray-value function is defined by parameters of a cosine wave, i.e. amplitude, frequency and phase. The input data is transformed, so that for a specified range (x_{min} , x_{max}) the frequency 1.00 means, that one period of a cosine wave fits exactly into this range. The discrete grey values are orthogonally projected to the surface of an elevation model, which may be defined



in a similar way. From there they are transformed by standard photogrammetric equations to the picture coordinate system, which in this case for reason of simplicity coincides with the pixel index system but for a scale factor and a constant. The object surface is sampled at a rate that leads to discrete values in the linear image at distances of less than 5 µm. The final grey value is obtained by integrating over pixel length. As just one picture will be processed at a time, the number of pictures is unlimited.

Let x_I and x_{II} be the border of pixel j of image $\mathcal{D}$ in object space. The object length $\Delta x_j = (x_j - x_{j-1})$ and its equivalent $\Delta x'_j$ in image space are defined by fig. 2/1. The image density D'_j is given by:

$$D'_j = INT(D'^*_j + 0.5) \quad ,$$

$$D'^*_j = \frac{1}{\Delta p'} \sum_{i=2}^{n} \Delta x'_i \cdot (D'(x'_i) + D'(x'_{i-1}))/2 \quad .$$

$$\Delta p' = \sum_{i=2}^{n} \Delta x_i \quad \text{is the pixel length.}$$

Fig. 2/1 : Calculating image density D'_j of pixel j of one of the linear images. The lower sketch represents the object density function D(x).

2.2 Calculating linear FAST-Vision

The mathematical formulation of linear FAST-Vision is given by:

$$D'(x') + v_D(x') = \{D^o(x_A)\frac{\partial \alpha_A(x^o)}{\partial x} + D^o(x_B)\frac{\partial \alpha_B(x^o)}{\partial x}\} \frac{x^o - x'_o}{z^o - z'_o}\{a_I(x^o)\,dz(x_I) + a_{II}(x^o)\,dz(x_{II})\}$$

$$+ \alpha_A(x^o) \cdot D(x_A) + \alpha_B(x^o) \cdot D(x_B) \qquad (2.01)$$

and

$$a_I(x^o) = 1 - x^o_z \qquad \alpha_A(x^o) = 1 - x^o_d$$

$$a_{II}(x^o) = x^o_z \qquad \alpha_B(x^o) = x^o_d$$

$$x^o_z = (x^o - x_I)/(x_{II} - x_I) \qquad x^o_d = (x^o - x_A)/(x_B - x_A)$$

$x^o = f(z^o_I, z^o_{II})$ is the point of intersection of the linear elevation profile and an image ray defined by the centre of a pixel and the focal point.

Fig. 2/2 illustrates the parameters involved in equation (2.01). The dotted lines represent the real object surface or object density as they are approximated by z- and d-facets. Within a certain window equation (2.01) is evaluated for every pixel. The pixel grey values are regarded as original observations. A direct pixel transfer is chosen, i.e. the center of a pixel is projected onto the object surface, represented by an interpolation function. Linear FAST-Vision leads to normal equations of sparse matrices form, especially a band width of 2 and a border of $n_z \cdot n_d$. A modified Cholesky algorithm was used to reduce CPU time. All calculations were performed at

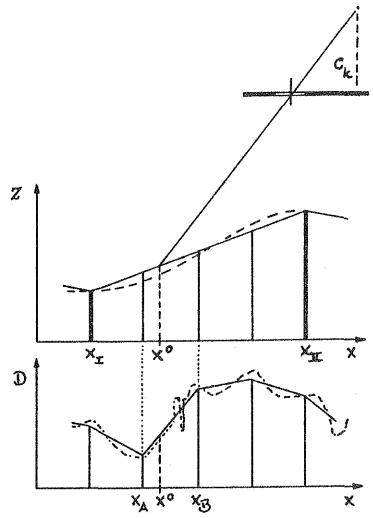


Fig. 2/2: Parameters of equation (2.01)

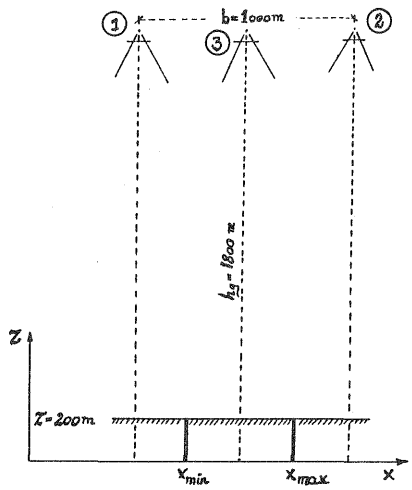


Fig. 2/3 : Configuration of images.

single precision except those rendering index positions or points of intersection. Numerical tests were confined to proove, that within the range of convergence different approximate values lead to identical results.

Three images are introduced into these studies and symmetrically arranged along the object of interest. See fig. 2/3. The third image is located in the middle of the scene and provides redundant information only.

Within a predefined range marked by x_{min} and x_{max} a number n_z of z-facets is declared. They form an equidistant grid, which is constant during program execution. A second grid is formed by the number of d-facets n_d , which are expected to allow linear interpolation of grey values in between. Therefore a dense grid is necessary which will be one aspect of the following tests. This grid is kept constant as well.

Within the object range a window is chosen by centre and length. All calculations are performed for this window only. As normal equations increase with the product of n_z and n_d window extension affects computing time - an optimal window length is subject of present studies.

To differentiate the influence of z- and d-parameters on the solution a horizontal elevation profile was chosen, i.e. z=const.

3. Experiments & Results

3.1 Parameters involved in FAST-Vision

FAST-Vision contains several parameters which have basic influence on the solution and which in genaral are dependend on one another. They are:

- pixel size — fixed
- geometric parameters, calibration of camera, orientation data — fixed
- object surface function — fixed
- parameters of illumination model — none
- grey level range and contrast
- signal to noise ratio (SNR)
- wave-length components of object density λ_d
- number and ratio of n_z and n_d
- length of window
- approximate values

All results will be given for the linear version. Only some parameters were included in these investigations. This paper will concentrate on the last six topics.

3.2 Choice of Parameters

The fact that all parameters affect FAST-Vision solutions at the same time necessitates to alter some of them and study them while other parameters are fixed. In the end some statements will refer to those fixed parameters as far as general conclusions can be drawn from the experiments.

Important parameters are the object density function and the signal to noise ratio (SNR) which condition normal equations and the accuracy of the method. Accuracy is also affected by the n_z to n_d ratio. Window length is important for practical performance and convergence requires approximate values within a certain range. Except for synthetic textures projected onto the object surface (Ackermann, 1984) only the very last parameters can be influenced by operator control. All experiments are characterized by following items:

- pixel length 20 µm
- image scale 1 : 12 000
- parallel linear images
- baseline to height ratio ϑ_1 = 1 : 1.8 for two images and ϑ_2 = 1 : 3.6 for three.
- focal length 150 mm
- object range 500 m
- object elevation z = const.

The signal to noise ratio is supposed to be

$$SNR = \gamma / 0.3 \qquad (3.01)$$

Herein the noise is assumed to be 0.3 because of the fact that it is caused by round-off errors only which can be regarded as being of equal likelihood. In case of real images the noise can be estimated from calculating the difference of two densitometer measurements (Göpfert, 1987). γ will be defined as the actual gray-value range.

3.3 Accuracy of Height Estimation

3.3.1 SNR and Height Precision

First of all the question shall be answered which precision can be expected by linear FAST-Vision when SNR decreases to a minimum. This is performed by reducing γ in (3.01). The results are shown in figure (3/1). This reveals that a very high precision can be reached even under low SNR. All other tests were performed with γ < 128. A binary picture (0,1), however, is not to be processed sufficiently.

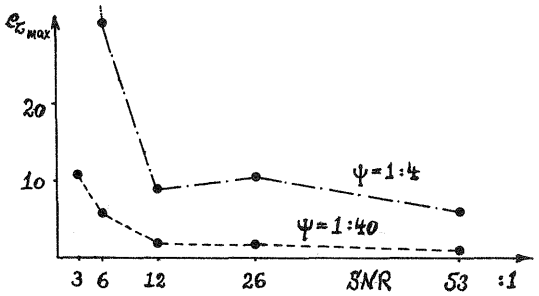


Fig. 3/1 :
Maximum error increases by reducing SNR. The precision depends on the ratio of n_z and n_d (ψ). Remember that 0.18 m is 0.1 °/oo h_g .

3.3.2 Number and Ratio of n_z and n_d

As the object surface function z(x) is kept constant for all tests, no empirical results will be presented concerning the number of z-facets n_z . We will assume that there are enough parameters to approximate the object surface and its gradient. For z(x)=const. this assumption is always satisfied.

The ratio $\psi = n_z : n_d$ is limited by equation (2.01) to a maximum of 1:1 . There is no such limit for a minimum and it turns out, that for flat objects precision increases with decreasing ψ , see figure (3/2). Furthermore there is a direct influence in σ_o which in case of optimal approximation should be 0.3, i.e the noise.

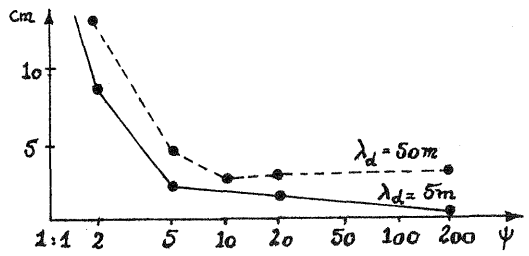


Figure 3/2: Increasing precision by decreasing ψ .

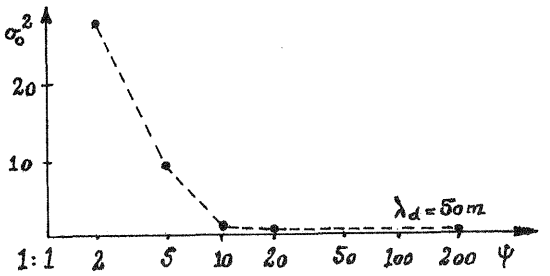


Figure 3/3 : Decreasing σ_o caused by decreasing ψ.

λ_d is the wave-length of a grey value cosine wave and in general a single wave-length is involved in the calculation. For all tests λ_d is connected with a constant γ in the following relation (if not told otherwise):

λ_d	>50	10	5	1	.75	.50	m in object space
γ	124	64	64	64	32	10	grey values

3.3.3 *A Ratio between λ_d and n_d*

As FAST-Vision parameter estimation is based on the gradient of object grey values, d-facets must be numerous enough to approximate this gradient and to allow linear interpolation of object density without loss of accuracy. In this case it means that $\sigma_o > 0.3$ should be indicating false approximation. This theory is supported by results of the studies, which can't be given in detail. We will also show that λ_d requires a minimum of 4 d-facets to avoid gross systematic errors.

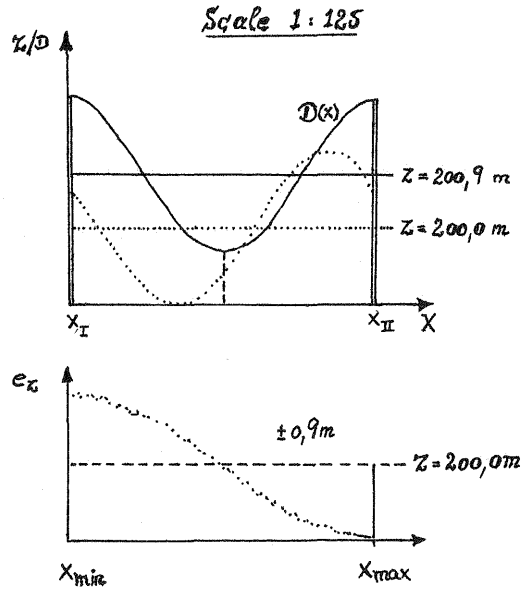


Fig. 3/5 : Phase dependend solution for $n_d = 2$ and a phase of 45°.

For $\lambda_d = 5$ m = $x_{II} - x_I$ and $\gamma = 64$ tests on the minimum number of d-facets per wave-length λ_d led to results as follows :

$n_d = 1$: No convergence, parameters for object density alternated about mean gray level by +/-1. Z-parameters alternated in a way to compensate remaining residuals in density.

$n_d = 2$: At first look the results show no obvious gross errors. Deeper investigations, however, revealed dependences between the phase of λ_d and the accuracy of the profil. This property is illustrated in the opposite figure. The dotted lines in the upper sketch represent the functions for height and object density within one z-facet as originally given for generating linear images. x_I and x_{II}

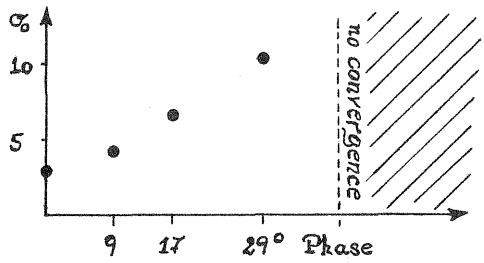


Fig. 3/6: Increasing σ_o under non-zero phase. In comparison to these figures: Introducing $n_d = 20$ leads to $\sigma_o = 0.3$.

are situated near x_{min}. Observation of the whole range (x_{min}, x_{max}) leads to half a cosine wave for the error function (lower sketch):

$$e_z = z_{calc.} - 200.00 \text{ m},$$

where 100 z-parameters have been equally spaced over the range in 5 m intervals. Different calculations led to convergence upto 45° phase and to a ratio of 2 cm height error per 1° phase. As σ_o increases with e_z this effect is most evident. See fig. 3/6.

$n_d = 3$: Precision becomes independent from phase, i.e. there is no linear relation any more. With regard to a phase of 45° the following results were obtained: There is one maximum error of about 0.40 m on the window border. Neglecting this gross error the true mean error becomes 0.07 m, whereas $\sigma_o = 6.0$ is still very high.

$n_d = 4$: No gross errors, true mean square error of z-parameters less then 0.03 m. $\sigma_o = 2.9$ is half the value of $n_d = 3$. Still there are systematical errors by insufficient linear approximation of the true object density function. Although σ_o decreases by expanding the number of d-facets there is no reasonable gain in precision any further.

From these experiments we shall draw the conclusion that the number of d-facets per λ_d must be greater then or equal to 4 to become independent from the phase of λ_d . In other words: The shortest wave-length $\lambda_{d.min}$ of the object density function, that linear FAST-Vision is able to approximate, is four times the length Δd of a d-facet. As Δd_{min} is limited to pixel length in object space, the shortest wave-length is provided by the sampling rate on image digitization (Rosenfeld/Kak, 1982).

3.3.5 The Choice of D-facets

To sum up the results of the previous chapters we will give a rule for the number of d-facets which is supposed to be fairly independend from z-parameters. We will introduce so many d-facets that yields:

$$\Delta d \Longrightarrow m_b \cdot \Delta p' \quad , \qquad (3.02)$$

where m_b is the image scale and $\Delta p'$ is pixel length.
This is for two reasons. First we regard the increasing precision for $\psi \Rightarrow 0$. As the number of z-facets is chosen according to object surface conditions the minimum of z-facets is not optional. So the minimum of ψ is always given by (3.02). Secondly we have to approximate the object density function most properly. This will be an important postulation when trying to interpret the residuals of FAST-Vision for detection of surface ambiguities and hidden spots.

3.3.6 The Influence of a Third Image

To study the influence of a third image almost every test was calculated twice. The results do not reveal any influence of a third image, neither in convergence nor in precision. The standard deviation of z-parameters was not reduced. The values turned out to be identical, although redundancy increased significantly. Obviously this is due to the fact that systematic errors are dominant. Where they are not, the minimum of σ_o is given by (3.01). A third image is expected to become important for real scenery. The theoretical maximum $\psi = 1:1$ is due to a third image.

3.4 Aspects of Convergence

3.4.1 Approximate Values and the Uniqueness of a Solution

FAST-Vision is based on taylorisation of non-linear equations and therefore dependent on approximate values, which have to be very close to the true values because the solution of FAST-Vision is not unique. Besides a differential change in scale another solution exists one wave-length apart of the true one, assumed that only one wave-length is introduced. The range of convergence is therefore given by approximately half a wave-length δp in object space which is turned into an elevation intervall dz by:

$$dz = \frac{h}{b}g \cdot \delta p = \vartheta^{-1} \cdot \delta p \quad , \qquad \delta p = \lambda_d/2 \tag{3.03}$$

where ϑ is the baseline to height ratio.

For the dataset in fig. 3/7 the theoretical range of convergence was computed and compared with that of FAST-Vision simulations, see fig. 3/8. The scale of the ordinate axis is one wave-length, so that 0.5 is the theoretical maximum. However, results show that approximate z^o values should be better than 0.4 parts of a wave-length. Approximate values within the hachured range lead to unpredictible results (different parts of a window may converge towards different solutions). Furthermore the range of convergence is diminished by decreasing ψ. As large numbers of iterations have to be excepted, methods to speed up convergence will be required.

λ_d	1	2	5	10	50	m
SNR	200	200	200	200	400	:1
dz	.9	1.8	4.5	9	45	m

Fig. 3/7: Parameters for which dz was computed.

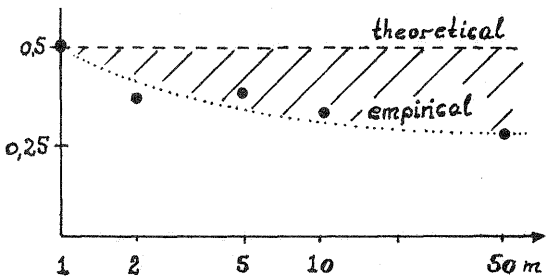


Fig. 3/8: Range of convergence in parts of one wave-length λ_d.

Further calculations reveal, that the shortest wave-length is generally limiting the range of convergence. If the amplitude is small against the more powerful density waves, the range of convergence is extended. Convergence and reliability of FAST-Vision solutions depend on strategies to obtain accurate approximate values.

3.4.2 Methods to Improve Convergence

This chapter will discuss some approches towards improving convergence. Three have been tested in linear FAST-Vision only, whereas the second topic is allready used to speed up convergence calculating 3-D-objects.

Smoothing the gradient is one method already tested in algorythms that use the gradient of the gray value function to estimate z-components. This is usualy done in image space (Ackermann, 1984), whereas the gradient in FAST-Vision is determined by neighbouring density parameters in object space. To obtain approximate grey values $D^o(x)$ the first approximation for elevation parameters $z^o(x)$ is used to interpolate the height of those points for which the density parameters will be calculated. $D^o(x)$ is calculated as the mean of all correspondent grey values in all different images connected with these points in image space. This means that the gradient is obtained from pixel grey values picked out by chance. Therefore taking the average of some pixels in every image should smoothen the gradient. However, this approach did not satisfy the expectations. No explanation can be offered yet.

Least squares adjustment for d-parameters only during the first iteration of FAST-Vision was performed by neglecting the results for z-parameters. This method, already used successfully for the 3-D-version did not speed up convergence. This may be due to the high SNR in synthetic images. For sampled photographs indeed this pre-adjustment may reduce the unpredictible effects of noise.

Least squares adjustment for z-parameters only. In this case the parameters for object density were ignored by eliminating them from normal equations. After every updating of z-parameters the d-parameters were calculated like approximate values. This method reduced the number of iterations by factor 2, whereas the range of convergence was extended but slightly.

The disadvantage is a loss in theoretical precision, which can be compensated by evaluating both groups of parameters together in the very last step. As every iteration is very time-costing this approach seems to be most promising.

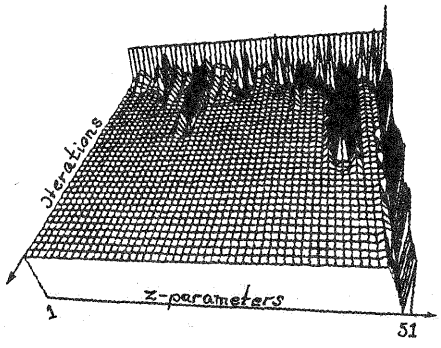


Fig. 3/9 : Convergence of a window without speed up methods.

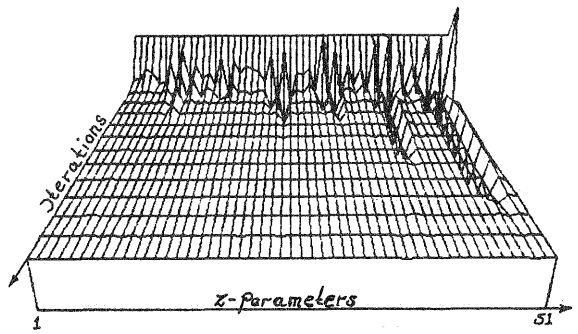


Fig. 3/10: Convergence of a window LSA - z-parameters only.

Computing a **frequency related average (FRA)** of pixels in image space is applied to extend the range of convergence. As stated in the previous chapter this range is so small that in case of poor approximate values z^o no convergence will be obtained at all. If approximate values need to be achieved by intensive preparations the advantages of digital matching vanish. FRA, however, is able to extend the range for approximate values from about 1 °/oo h_g to about 30 °/oo and more, which means about 50 m and more for these examples. FRA is performed according to the information included in the images: The number of pixels comprised is chosen according to frequency analysis of the gray values. The higher frequencies are low-pass filtered by enlarging pixellength and by replacing the original pixels by their average grey value. The fact that the number of parameters and observations, i.e. the pixels, is reduced by a factor of at least 2 or 3 leads to a decreasing effort of at least 4 or 9 for the 3-D-version. The range of convergence depends on powerful low frequencies. Detailed studies on these subjects have to be applied to the 3-D-version.

An example is calculated for the following data set:

λ_d :	∞	1000.0	250.0	142.8	83.3	40.0	2.5	1.0	m in object space
Amplitude :	127.5	38.0	29.0	22.0	17.0	10.0	5.0	2.0	grey levels

To suppress the influence of the high frequencies the factors of FRA should be chosen by following rule. Pixel length in object space is 0.24 m. This means 4 pixels per shortest wave-length, 10 per 2.5 m and 160 per 40.0 m. So 4 and 10 are introduced, whereas 160 would reduce redundancy so much that it is neglected. The theoretical range of convergence (λ_d =1.0, dz= 0.9 m) is extended to 2.25 m because of the more powerful frequency λ_d = 2.5 m . FRA with factors of 10 and 20 lead to results of 200 +/- 1.8 m for equally spaced z-parameters 50 m apart. The number of iterations is about 10. After one act of comprising data the updated approximate values are good enough for

non-comprised data evaluation. The ratio ψ is chosen by the rule (3.02). Other suggestions were made for choosing the FRA factor (Helava, 1988) but haven't been prooved yet.

4. Conclusions

This paper discusses the high accuracy of FAST-Vision as already reported by (Wamser, 1987; Wrobel and Weisensee, 1987). A rule for the number of d-facets is presented, derived from computer simulations.

It is shown that the shortest wave-length of the object density function, that can be properly reconstructed by FAST-Vision, is four times the pixel length in object space. This wave-length also limits the range of convergence. This may be due to the high SNR typical for synthetic images.

Furthermore a hierarchical concept for improving convergence and the reliability of FAST-Vision solutions have been given, which are important for establishing ecconomic operation and for solving more complicated problems like detection of surface ambiguities or hidden spots. (Boochs, 1984; Helava, 1988).

5. References

Ackermann, F.: Digital Image Correlation: Performance and Potential Application in Photogrammetry, Paper presented at the 1984 Thompson Symposium, Birmingham, U.K.

Boochs, F.: Ein Verfahren zur Herstellung digitaler Höhenmodelle aus photogrammetrischen Stereomodellen mit Hilfe der flächenhaften Korrelation in digitalen Bildern. Dissertation Bonn. DGK Reihe C. München 1984.

Ebner H. et al.: Integration von Bildzuordnung und Objektrekonstruktion innerhalb der digitalen Photogrammetrie, BuL Heft 5/87, S. 194-203.

Förstner, W.: On the Geometric Precision of Digital Correlation, Proceedings of ISPRS Comm. III, Symposium Helsinky 1982.

Helava, U. V.: Object Space Least Squares Correlation, HAI, Southfield, Michigan, USA. ACSM-ASPRS Convention, St. Louis, March 1988, p. 13-18.

Göpfert, W.: Raumbezogene Informationssysteme, H. Wichmann Verlag, Karlsruhe 1987

Grepel, U.: Effiziente Rechenverfahren für umfangreiche geodätische Parameterschätzungen, Dissertation, Mitt. aus den Geodätischen Inst. der Rh. Fr. W. Universität, Nr. 75, Bonn 1987.

Koch, K. R.: Parameterschätzung und Hypothesentests in linearen Modellen, Ferd. Dümmlers Verlag, Bonn 1980.

Rosenfeld, A. and Kak, A.C.: Digital Picture Processing, Second Edition, 2 Vol., Academic Press, New York 1982.

Wamser, H.: Entwurf und Erprobung eines Programms zur digitalen Bildzuordnung nach der Facettenmethode, unveröffentlichte Diplomarbeit, Darmstadt 1987.

Woytowicz, D.: Digitale Korrelation von Meßbildern eines Braunkohlentagebaues, unveröffentlichte Diplomarbeit, Bonn 1987

Wrobel, B.: Facets Stereo Vision (FAST-Vision) - A New Approach to Computer Stereo Vision and to Digital Photogrammetry, ISPRS Intercommission Conference on 'Fast Processing of Photogrammetric Data', Interlaken 1987.

Wrobel, B. and Weisensee, M.: Implementation Aspects of Facets Stereo Vision with some Applications, ISPRS Intercommission Conference on 'Fast Processing of Photogrammetric Data', Interlaken 1987.

Wrobel, B.: The Fundamentals of Facets Stereo Vision (FAST-Vision) - A General Least Squares Procedure for Object Surface Computation from Digital Image Data. IAPRS 27, Com. III, Kyoto 1988.

Programs developed on ATARI 1040 ST. Computations on SIEMENS PC MX 2.