FACETS STEREO VISION (FAST VISION) APPLIED TO DIGITAL COLOUR IMAGES

B. Kaiser, B. P. Wrobel
Institute for Photogrammetry and Cartography
University of Technology Darmstadt (Germany)

International Society of Photogrammetry and Remote Sensing
Commission III, WG III/2
XVIII ISPRS Congress Vienna, Austria, 9-19 July 1996

KEYWORDS: Photogrammetry, Matching, Reconstruction, DEM/DTM, Orthoimage, Colour, Experiment

ABSTRACT:
In this paper a modification of the well-known object space based Facets Stereo Vision (FAST Vision) method is introduced. It works with vectors as observations in each of the pictures being processed for surface reconstruction. Former versions of FAST Vision only used scalars for each pixel as input values. The new version makes it possible to exploit the full information of multi-channel imagery. The formal modifications of the method are presented in this paper as well as an example of a surface reconstruction applying the new method to a pair of colour aerial pictures.

ZUSAMMENFASSUNG:
In dieser Arbeit wird eine Modifikation der wohlbekannten objektraumbasierten Methode des Facetten-Stereosehens präsentiert. Sie arbeitet mit Vektoren als Beobachtungen in jedem der Bilder, das für die Oberflächenrekonstruktion herangezogen wird.. Bisherige Versionen von FAST Vision benutzen Skalare als Eingabewerte. Die neue Version macht die volle Ausnutzung von mehrkanaligem Bildmaterial möglich. In dieser Arbeit werden sowohl die formalen Modifikationen der Methode als auch ein Beispiel einer Oberflächenrekonstruktion aus farbigen Luftbildern beschrieben.

1.Introduction

The major difference concerning the exploitation of data between digital and analytical photogrammetry consists in the use of colour information. Whereas the input for analytical photogrammetry usually consists of colour images, the input for digital photogrammetry are usually pictures with 256 (8 bit) grey values. This results in a considerable disadvantage for the digital method. A simple visual comparison of grey value and colour images depicting the same scene illustrates how much information is lost in grey value images. The same loss of information occurs when other channels of a picture, e.g. a infra-red channel is not used as input for surface reconstruction. The reason for the limitation of digital photogrammetry to grey value images consisted in the amount of image data which increases threefold in the case of three colour channels. But nowadays, the price of storage media with large capacity can be neglected. As the data not only has to be stored but also processed, colour images require an improved processing performance of the computer. This performance has not ceased to increase with every new generation of computer processors in recent years and won´t do so in coming years. Therefore the complete exploitation of image information in digital photogrammetry becomes a possibility which should be realised (Weisensee, Wrobel, 1991).

2.Facets Stereo Vision

The method for object reconstruction discussed in this paper is an object space base one: Facet Stereo Vision (Wrobel, 1987). It is based on inverting the process of image formation using a finite element model of the object surface and of the object grey values consisting of patches called facets. Generally, functions T', T''... describe the transformation of image grey values G'(x',y'),G''(x'',y'') ... at pixel positions (x',y'), (x'',y'') to object grey values G(X,Y) at a position (X,Y) in object space.

$$\begin{aligned} T'\{G'(x',y')\} &= G(X,Y) \\ T''\{G''(x'',y'')\} &= G(X,Y) \end{aligned} \qquad (1)$$

If outer and inner orientation and approximate values $Z^0(X^0,Y^0)$ for the object surface are known the grey values in the vicinity of (X^0,Y^0) can be described by a Taylor series.

$$\begin{aligned} G^0(X^0+dX,Y^0+dY)+dG^0(dX,dY) &= G^0(X^0,Y^0) \\ +\frac{\partial G^0(X^0,Y^0)}{\partial X}dX &+ \frac{\partial G^0(X^0,Y^0)}{\partial Y}dY + dG^0(X,Y). \end{aligned} \qquad (2)$$

Inserting (2) into (1) yields:

$$\begin{aligned} G(x',y') = T'^{-1}(G^0(X^0,Y^0) &+ \frac{\partial G^0(X^0,Y^0)}{\partial X}dX \\ &+ \frac{\partial G^0(X^0,Y^0)}{\partial Y}dY + dG^0(X,Y)). \end{aligned} \qquad (3)$$

The image rays passing the point of projection (X'_0,Y'_0,Z'_0) yield a formula linking differences dX and dY with differences dZ:

$$dX = \frac{X^0 - X'_o}{Z^0 - Z'_o} dZ \quad ; \quad dY = \frac{Y^0 - Y'_o}{Z^0 - Z'_o} dZ. \tag{4}$$

Inserting (4) into (3) yields:

$$G(x',y') = T^{-1}(G^0(X^0,Y^0) + \frac{\partial G^0(X^0,Y^0)}{\partial X} \cdot \frac{X^0 - X'_o}{Z^0 - Z'_o} \cdot dZ + \frac{\partial G^0(X^0,Y^0)}{\partial Y} \cdot \frac{Y^0 - Y'_o}{Z^0 - Z_o} \cdot dZ + dG^0(X,Y)). \tag{5}$$

In the Finite Element Method the surface inside a finite element is interpolated by the weighted sum of the grid values of the element:

$$Z(X,Y) \approx \tilde{Z}(X,Y) = \sum_{m=0}^{r}\sum_{n=0}^{s} Z(X_m,Y_n) \cdot a(X - X_m, Y - Y_n) \tag{6}$$

The sum of the weights $a_{m,n}$ is always 1:

$$\sum_{m=0}^{r}\sum_{n=0}^{s} a_{m,n} = 1 \tag{7}$$

In Facets Stereo Vision the finite elements are called facets. They are used for interpolating the surface Z as well as object grey values G(X,Y). Thus, replacing Z by G in (6) yields the formula for interpolating object grey values from the respective finite elements.

Differences of the object surface are interpolated by differences of the grid points of the finite element (replacing Z with G again yields a formula for object grey value differences):

$$d\tilde{Z}(X^0,Y^0) = \sum_{m=0}^{r}\sum_{n=0}^{s} dZ(X_m,Y_n) \cdot a_{m,n} \tag{8}$$

Inserting finite element interpolation for object surface and object grey values into (5) forms the basic equation of Facets Stereo Vision which describes the relationship between image grey values G'(x',y') and object surface and object grey values.

$$G(x',y') = T^{-1}\Big(\sum_{m=0}^{r}\sum_{n=0}^{s} G^0(X_m,Y_n) \cdot \alpha_{m,n} + \sum_{m=0}^{r}\sum_{n=0}^{s} dG(X_m,Y_n) \cdot \alpha_{m,n} + \sum_{m=0}^{k}\sum_{n=0}^{l} G^0(X_m,Y_n) \cdot \left(\frac{\partial \alpha_{m,n}}{\partial X} \cdot \frac{X^0 - X'_o}{Z^0 - Z_o} + \frac{\partial \alpha_{m,n}}{\partial Y} \cdot \frac{Y^0 - Y'_o}{Z^0 - Z_o}\right) \sum_{m=0}^{r}\sum_{n=0}^{s} dZ(X_m,Y_n) \cdot a_{m,n}\Big). \tag{9}$$

Differences of brightness and contrast between image grey values and object grey values are modelled by a linear function T:

$$T\big(G'(x',y')\big) = g'_0 + g'_1 \cdot G'(x',y'). \tag{10}$$

Adding the correction term $v_{G'}$ for the observation G' in (19) the linearization of the right hand side of (10) leads to:

$$g_0'^0 + dg'_0 + \left(g_1'^0 + dg'_1\right) \cdot \left(G'(x',y') + v_{G'}\right) = g_0'^0 + dg'_0 + g_1'^0 \cdot G'(x',y') + dg'_1 \cdot G'(x',y') + g_1'^0 \cdot v_{G'}. \tag{11}$$

We denote the differences between the grey values of the image and their interpolation within the grey value facets as l':

$$l' = g_0^0 + g_1^0 \cdot G(x',y') - \sum_{m=0}^{k}\sum_{n=0}^{l} \alpha_{m,n} \cdot G^0(X_m, Y_n \,. \tag{12}$$

Replacing T in (9) by the right hand side of (11) and l' by the right hand side of (12) yields the following equation:

$$v_{G'} = \frac{1}{g_1'^0} \cdot \Big(\sum_{m=0}^{k}\sum_{n=0}^{l} \alpha_{m,n} \cdot dG\,(X_m,Y_n) + \sum_{m=0}^{k}\sum_{n=0}^{l} G^0(X_m,Y_n) \cdot \left(\frac{\partial \alpha_{m,n}}{\partial X} \cdot \frac{X^0 \quad X'_0}{Z^0 - Z'_0} + \frac{\partial \alpha_{m,n}}{\partial Y} \cdot \frac{Y^0 \quad Y'_0}{Z^0 - Z'_0}\right) \cdot \sum_{m=0}^{r}\sum_{n=0}^{s} a_{m,n} \cdot dZ\,(X_m,Y_n) - dg'_0 - dg'_1 \cdot G'(x',y') - l' \Big) \tag{13}$$

By forming this equation for any pixel in any image which is used as input for FAST Vision a system of linear observation equations can be obtained, in which the unknowns x are the differences of surface heights and surface grey values respectively plus the parameters of radiometric transformation:

$$v = A \cdot x - l.$$

From this system of equations a system of normal equations can be obtained in the well-known way in order to get a least squares solution for the variables x. Starting with reasonable Z-values the approximate values for the object grey values are calculated from one of the images. The shape of the facets is iteratively improved until the images calculated from this model surface are sufficiently similar to those pictures used as input. By taking into account the object surface the method of Facet Stereo Vision has an advantage over image space based methods of surface reconstruction in case of complex surface geometry.

Up to now, the method of Facet Stereo Vision comprises the modelling of object geometry and radiometry by three sets of parameters: Object Z-values describe the object surface, object grey values describe the object radiometry and additional variables for each picture model the inter-image differences of brightness and contrast.

3. Modifications of the FAST Vision approach for colour images

Two of the three sets of parameters in the method of Facets Stereo Vision have to be modified in order to be able to process colour images: Instead of single object grey values there are vectors of object colour values - usually containing three elements. The scalar variables modelling differences of brightness and contrast between the pictures also have to be replaced by vectors. Only the variables modelling surface geometry remain unchanged. All variables in equation (13) containing a capital G have

to be replaced by a variable containing G_i, where i indicates the number of the channel i of the multi-channel image, i.e. i runs from 1 to 3 for colour images. The number of equations in each Z-facet this way multiplies with the number of channels of the picures. The linear function T modelling radiometric differences between the images also has to be replaced by a set of functions, one function for each channel. The scalars g_0' and g_1' become vectors g_{0i}' and g_{0i}'. Thus, not only the amount of input data and the number of observation equations increase threefold, but the amount of output data also more than doubles. The size of the system of normal equations grows with the number of additional unknowns, i.e. with the additional colour value facets. The other features of FAST Vision - adaptive regularisation to overcome the ill-posedness of the problem of image inversion (Wrobel et al. 1992a, Wrobel et al. 1992b), adaptive determination of facet size, the use of images pyramids (Kaiser et al., 1992) etc. - also remain unaltered. Again, processing starts with calculating approximate values for object colour values by using the approximate surface and one of the input pictures.
In contrast to object reconstruction from grey value pictures colour images should be pre-processed: In order to minimize the correlation between the - usually three - colour values of each pixel a transformation of the colour values of the picture takes place before the application of FAST Vision, see next chapter.

4. An example of surface reconstruction from colour images

In order to check the computer program derived from the modifications of FAST Vision described above, the surface reconstruction was tested with the input of computer-generated imagery. These experiments, which are not shown here, proved the reliability of the new method by a comparison of the exactly known surface with the reconstructed. A next step was the application of the modified method to realistic image material which, in this case, consisted of a pair of colour aerial pictures. They were taken from a rural landscape in the southwestern part of Germany. The area is partly covered by buildings and vegetation. The scale of the pictures is 1:6000, the overlap is 60%. The pictures were digitized with three colour channels. Fig 1-6 show the three channels of the left and the right image. The differences of the information contained in the three channels are obvious. Especially the lack of information about vegetation in the blue channel is significant.

Fig.1: Blue channel of the left image.

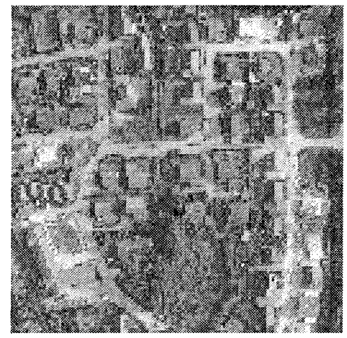

Fig.2: Green channel of the left image.

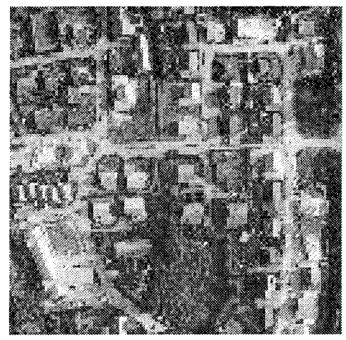

Fig.3: Red channel of the left image.

Fig.4: Blue channel of the right image.

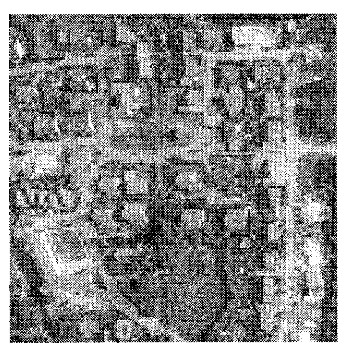

Fig.5: Green channel of the right image.

Fig.6: Red channel of the right image.

As mentioned before, a transformation of image grey values has to take place in order to reduce correlation between observations. It is understood that this should not lead to a loss of information, i.e. that the original grey values have to be recoverable from the transformed ones. Besides the fact that a loss of information is generally not desirable the possibility of computing object grey values for orthoimage generation has to be maintained. One possible transformation is the computation of differences between the channels. If one original channel is used as input for FAST Vision, the differences of the other two channels with the former one can be used instead of the original channels. For the experiments described in this paper these images were used as input in combination with the green channel.

The results of this surface reconstruction using FAST Vision for multi-channel images were compared to those results obtained from using grey value images as input. The purpose of these experiments was not to compare the surface reconstruction with surface values derived by a different method. Besides the fact, that the first method yielded colour orthoimages, the differences of surface reconstruction were not significant. This can be attributed to the fact, that the grey value pictures already contain the necessary information.
Fig. 7 and 8 show an enlarged part (the area entirely covered by vegetation) of the green channels of the left and the right image. This area is displayed here because of the great differences of the three colour channels, which occur especially here, whereas there are good contrasts in all other parts of the three channels. White crosses indicate the position of the grid points of the Z-facets in image space. It is obvious that there are some ambiguities in the colour values which have to be overcome by regularization. Unfortunately, the ambiguities are present in the colour as well as in the grey value pictures.
Each of the 16 surface heights depicted in Fig. 7 and 8 was reconstructed by FAST Vision as the centre of a window containing $8 \cdot 8$ Z-facets. These in turn contained $4 \cdot 4$ colour value facets each ($4 \cdot 4$ grey value facets resp.). The colour value facets had a size of 4m in object space, thus the Z-facets had a size of $16m \cdot 16m$ in object space. Each colour value facet contained approximately $4 \cdot 4$ pixel. So the number of observation equations in each Z-facet was approximately 250 in each grey value image and 750 in each colour image. The number of unknowns in each of the windows containing $8 \cdot 8$ Z-facets, i.e. the window from which one of the surface heights in Fig. 7 and 8 is derived, amounts to 1172 using grey value images as input and 3354 using colour images.
The mean standard deviation of surface heights, i.e. of the unknowns Z of FAST Vision, was 0.2m. This corresponds to less than $0.2^0/_{00}$ of the flying altitude of 1200m from which the pictures were taken. The differences between Z-values derived from colour images and those derived from grey value images was below 0.1m in all cases.

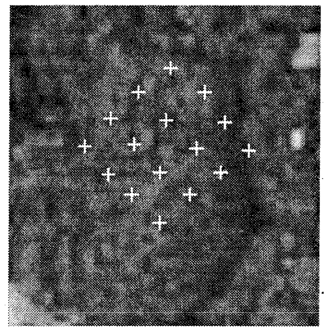

Fig.7: Reconstructed surface heights in the right image (green channel).

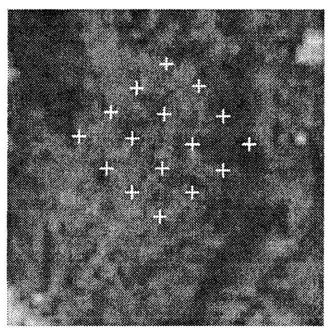

Fig.8: Reconstructed surface height in the left image (green channel).

5. Conclusion

This paper shows the latest modification of FAST Vision: the use of vectors as observations instead of scalars. By taking this step the full exploitation of image information becomes possible, i.e. the use of all channels of multi-channel imagery as input for surface reconstruction. The progress in data processing technology allows to handle the increase of data to be stored and processed which multiplies with the number of image channels being used as input. The experiments shown here using colour aerial pictures as input result in a reconstructed surface which is not significantly different from that reconstructed from grey value images. Further tests with all kinds of image material will probably show cases where the input of multi-channel imagery is a clear advantage. Besides, the other output of Facets Stereo Vision, the orthoimage, is a colour image instead of a grey value picture.
Further tests with the modified Facets Stereo Vision have to be carried out in order to prove if there is real image material where the new approach offers significant advantages. Computer-generated images where this is the case can easily be computed. Furthermore, the possibility

of a pre-processing reducing the number of channels used as input for FAST Vision before applying FAST Vision has to be examined. This pre-processing should be able to extract the maximum information of all channels of a picture on one hand and reduce the number of channels used as input for FAST Vision on the other. The results of the experiments shown in this paper indicate that this will be possible in many cases.

Acknowledgement

This research work was supported by Deutsche Forschungsgemeinschaft (German Research Foundation)

6. References

Kaiser, B., Schmolla, M., Wrobel, B., 1992: *Application of Image Pyramid for Surface Reconstruction with FAST Vision (=Facets Stereo Vision)*, International Archives of Photogrammetry and Remote Sensing, Vol. 29, Part B3, pp. 341-345.

Kempa, M., Wrobel, B., Weisensee, M., 1990: *Convolutional Approaches for Target Location by Digital Photogrammetry*, International Archives of Photogrammetry and Remote Sensing, Vol. 29, Part 3/2, S. 346-357, Wuhan 1990.

Kempa, M., 1995: *Hochaufgelöste Oberflächenbestimmung und Orientierung von Bildern mit dem Facetten-Stereosehen*, Dissertation, Technische Hochschule Darmstadt.

Weisensee, M., Wrobel,B., 1991: *State-of-the-Art of Digital Image Matching for Object Reconstruction*, in: Ebner, Fritsch, Hepike (ed.), Digital Photogrammetric Systems, Herbert Wichmann Verlag, Karlsruhe.

Weisensee, M., 1992: *Modelle und Algorithmen für das Facetten-Stereosehen*, Deutsche Geodätische Kommission, Reihe C, Nr. 374, München.

Wrobel, B. 1987: *Facets Stereo Vision (FAST Vision) - A New Approach to Computer Stereo Vision and to Digital Photogrammetry*, Intercommission Conference of International Society for Photogrammetry and Remote Sensing on Fast Processing of Photogrammetric Data, Interlaken, Switzerland.

Wrobel, B., Kaiser, B., Hausladen, J., 1992a: *Adaptive Regularization - A New Method for Surface Reconstruction from Images*, International Archives of Photogrammetry and Remote Sensing, Vol. 29, Part B3, pp. 824-831.

Wrobel, B., Kaiser, B., Hausladen, J., 1992b: *Adaptive Regularization of Surface Reconstruction by Image Inversion, in: Robust Computer Vision*, Förstner, W., Ruwiedel, S., Herbert Wichmann Verlag, Karlsruhe.