MULTI-IMAGE MATCHING IN OBJECT SPACE ON THE BASIS OF A GENERAL 3-D SURFACE MODEL INSTEAD OF COMMON 2.5-D SURFACE MODELS AND ITS APPLICATION FOR URBAN SCENES

Martin Schlüter*
Bundesamt für Kartographie und Geodäsie
Richard-Strauss-Allee 11, D-60598 Frankfurt am Main, Germany
Phone: +49-69-6333-306
e-mail: martin.schlueter@usa.net
ISPRS Commission IV Symposium, vol. 32, part 4, 545-552, Stuttgart, 1998

KEY WORDS: Surface Reconstruction, Matching, 3-d Surface Models, Object Space, Urban Scene, Occlusion, Facets Stereo Vision

ABSTRACT

A new method for high resolution surface reconstruction by multi-image matching in object space is formulated with regard to the integration of a general 3-d surface model in object space, because common 2.5-d surface models cannot handle arbitrary surfaces in $\mathbb{R}^3$. An adequate geometric surface model is based on an explicitly formulated surface topology which is spread by a spatial TIN (triangular irregular network) in three-dimensional space. This topological description is completed by transfinite interpolants, who are well suited for automatic reconstruction tasks. The self-detection of visibility resp. occlusion of the surface in object space is integrated explicitly. The first part of this paper introduces into the motivation for the new approach and outlines the new surface model as well as the basic observation equation of the matching algorithm.

Applications with respect to urban scenes profit from the new surface reconstruction method, because the reconstruction process can benefit from the complete visible information, which is available within each of the participating images, even if it shows vertical or overhanging parts of buildings. So, vertical parts of buildings have not to be modeled by discontinuities in object space, like many binocular vision algorithms do, but take part in the reconstruction process directly. Since the results of the new approach are only improved, if a certain amount of vertical building parts is visible in at least two images, really meaningfull results can only be expected from a simultaneous multi-image matching procedure. Some results from evaluations on the basis of large scale aerial imagery of built-up areas are given in the second part of this paper. They prove the success of the new method against former formulations using $2\frac{1}{2}$-d surface descriptions like bilinear or bicubic interpolation on a regular grid. The results seem to be a very promising starting point for further scene understanding tasks, especially with regard to automatic building reconstruction for 3-d GIS. Furthermore, the special design of the three-dimensional surface model might be of interest in relation to modeling concepts for GIS, too. Common VRML-descriptions can be easily derived.

1 INTRODUCTION

Surface reconstruction by digital image matching in object space can be formulated in a very general way following the concept of facets stereo vision due to (Wrobel, 1987). Using direct pixel transfer, the geometric relation between every pixel position in image space and the corresponding position on the surface in object space is re-established. Furthermore, the observation equation for a least squares estimation is written for each pixel grey value in image space, while the coefficients of the design matrix are set up completely by numeric quantities in object space. In this context, the gradients of the surface grey values in object space play an important role. This course of action offers many advantages, especially concerning the simple integration of more than 2 images and the simultaneous use of multisensoral, multispectral and multitemporal image data, cf. (Schlüter, 1998).

To reach a high degree of independence concerning the geometric start values required in object space, which describe the approximate run of the surface, one usually falls back on multigrid techniques in object space resp. image pyramids in image space. By this way, first a rough approximation of the resulting surface is gained by a very low vertex density in object space using the images from the highest level of the image pyramid, while afterwards the vertex density in object space is sucessively refined as well as the level of the image pyramid is decreased. If in doubt, we don't hesitate to use even a large number of image pyramid levels, because the computational work at the higher levels can be neglected compared to the computational burden at final resolution. This sucessive way to sculpture the resulting surface step by step finer and finer at each level of the multigrid is of importance for the following extension of the existing methods towards fully three-dimensional surfaces.

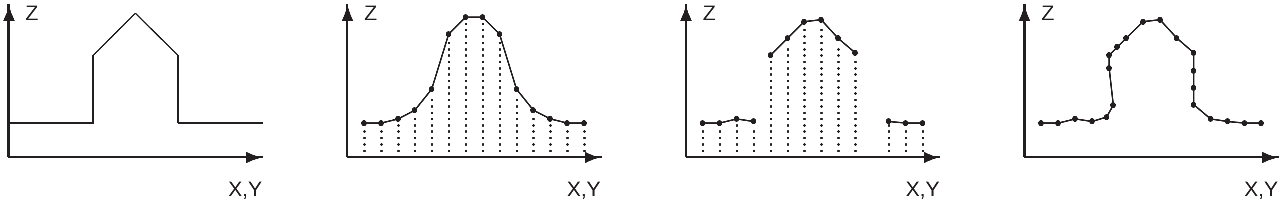
In the past, several $2\frac{1}{2}$-d surface models have been used for surface reconstruction in object space, bilinear and bicubic interpolation of heights upon a regular grid seem to be the most common approaches, cf. (Heipke, 1991), (Weisensee, 1992), (Tsay, 1996) and fig. 1(b). This kind of surface modeling is in principle sufficient, if only two images are used simultaneously for surface reconstruction. In this special case, vertical parts of the surface (vertical with respect to the stereo base) are often visible only in one image. In computer vision, the correspondencies between occlusions and discontinuities are often taken explicitly into account, cf. (Belhumeur, 1993), (Geiger et al., 1995), (Luo and Burkhardt, 1995) and fig. 1(c).

There are many reasons to use more than two images for simultaneous surface reconstruction. The main reason might be the growing robustness of the automatic matching procedure against different sources of disturbances, cf. (Schlüter and Wrobel, 1996). This fact holds also for feature based reconstruction strategies, cf. (Maas and Kersten, 1997) e.g. Here, we focus on the fact that the consequent use of noticeably more than two images implies interesting consequences concerning the local visibility of the surface in object space. Just for simplicity, we concentrate on large scale aerial imagery of urban scenes for this discussion. The more images are at our disposal, the more vertical parts of buildings will be visible in at least two of these images. But using only a $2\frac{1}{2}$-d surface model like $Z = f(X, Y)$ in object space[†], the vertical or even overhanging parts of the surface cause trouble for the reconstruction process, just because they cannot be handled correctly

---

[*] formerly with Institute of Photogrammetry and Cartography, Darmstadt University of Technology, where the development and implementation of this new approach took place

[†] More generally, every surface representation, which can be represented as a graph of a bivariate function, is here refered to be a $2\frac{1}{2}$-d surface representation



(a) (b) (c) (d)

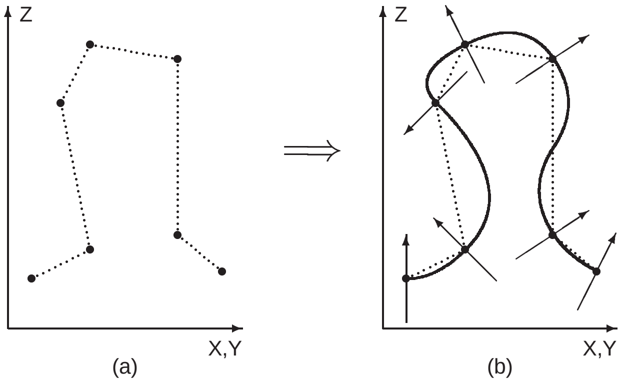
Figure 1: Quality steps of surface representations for image matching in object space, cf. (Schlüter, 1997) . (a): Surface in object space. (b): $2\frac{1}{2}$-d Reconstruction with global smoothness assumption. (c): Reconstruction of continuous areas only. (d): Reconstruction with 3-d surface representation

by the (in this case) inadequate $2\frac{1}{2}$-d surface model. But surface reconstruction should not fail just because of an inadequate surface model. A more general 3-d surface model, refer to fig. 1(d), should lead to improved results of the automatic surface reconstruction procedure. If occlusion is explicitly taken into account, advantage can be taken of the complete grey value information of all participating images.

In the following, the basics of an adequate 3-d surface model are summarized. Subsequently, the integration of this surface model into digital image matching in object space (facets stereo vision) is treated. Finally, some results from large scale aerial images of an urban scene are presented.

2 SET UP OF THE 3-D SURFACE MODEL

In this section the new 3-d surface model is briefly introduced, for details refer to (Schlüter, 1998). It consists mainly of an explicit topological structure between somehow given vertices in $\mathbb{R}^3$, cf. fig. 2(a), and suitable local interpolants without any restriction concerning the run of the surface in $\mathbb{R}^3$, cf. fig. 2(b).

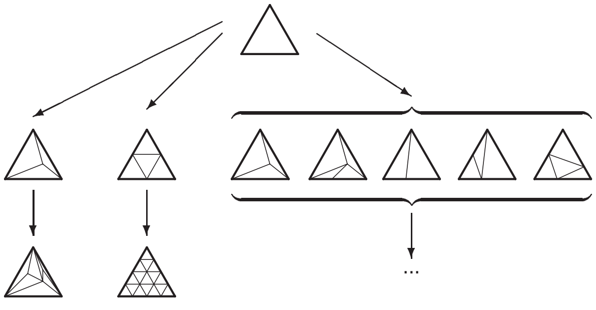


(a) (b)

Figure 2: 3-d surface model for multi-image matching in object space. (a): Explicit topology between vertices (•) in $\mathbb{R}^3$. (b): Completion of the surface model by interpolants with selected continuity properties. The outlined normal vectors are typical boundary constraints for G^1-continuous interpolants.

2.1 Explicit surface topology in $\mathbb{R}^3$

To build up the explicit topology in object space, triangular simplices are used in connection with a *Delaunay*-like triangulation. Working with aerial imagery, it is easy to gather a geometric initialisation, because for the startup geometry usually a $2\frac{1}{2}$-d assumption is valid and sufficient. Thus, often just depending on an intermediate height of flight above ground, vertices can be spread regularly on the startup geometry and be triangulated by a common 2-d approach (here, always the *Delaunay*-criterion is used). The point density of this start triangulation is very low, because it belongs to the upper level of the multigrid procedure in object space. Quaternary subdivision offers the new topology for each new multigrid level which is entered during the reconstruction process, cf. 3(b). It is essential, that a global retriangulation in object space is absolutely avoided during the reconstruction process, because the surface model is intended to pass over from $2\frac{1}{2}$-d to full 3-d characteristics – to obtain a surface triangulation from vertices arbitrarily spread in $\mathbb{R}^3$ is another story, cf. the overview in (Uray, 1996). Nevertheless, it is advantageous to have local operations like vertex insertion or deletion at one's disposal - the required changes of the global triangulation can be realized by local transformations using a *Delaunay*-criterion with respect to the local tangential plane.



(a) (b) (c)

Figure 3: Subdivision schemes for hierarchical triangulations, following (De Floriani and Puppo, 1995). (a) ternary subdivisions, (b) quaternary subdivisions, (c) schemes for heuristic-hierarchical subdivisions.

It seems to be possible and also fascinating to integrate concepts like data dependent triangulation within the photogrammetric surface reconstruction process, which are often discussed in connection with hierarchical triangulations, cf. (De Floriani and Puppo, 1995), because both geometric and grey value information concerning the surface in object space are at hand for this task. But one has to keep in mind that heavy changes of the local vertex densitiy always lead to mal-shaped thin and elongated triangles. For the examples in this paper, the vertex density is aimed to be as constant as possible. For this task *Laplacian smoothing*[‡], cf. (Cavendish et al., 1985), is adopted once per each multigrid level, if required.

2.2 Triangular interpolants in $\mathbb{R}^3$

Every single triangular mesh is completed by a local interpolant. Since former investigations have shown much better convergence rates of the reconstruction process, if C^1- instead of only C^0-continuous surface representations are used, cf. (Tsay, 1996); this continuity constraint shall also hold for the 3-d surface representation. Strictly speaking, only G^1- instead of C^1-continuity can be reached, cf. (Veltkamp, 1992), because a global parameterization is not at hand, but barycentric coordinates are used for each surface triangle separately. To avoid numerical problems, as reported for interpolation in $\mathbb{R}^3$ using triangular *Bézier*-patches by (Pfeifer

---

[‡] Not to mistake for the well-known image processing operator

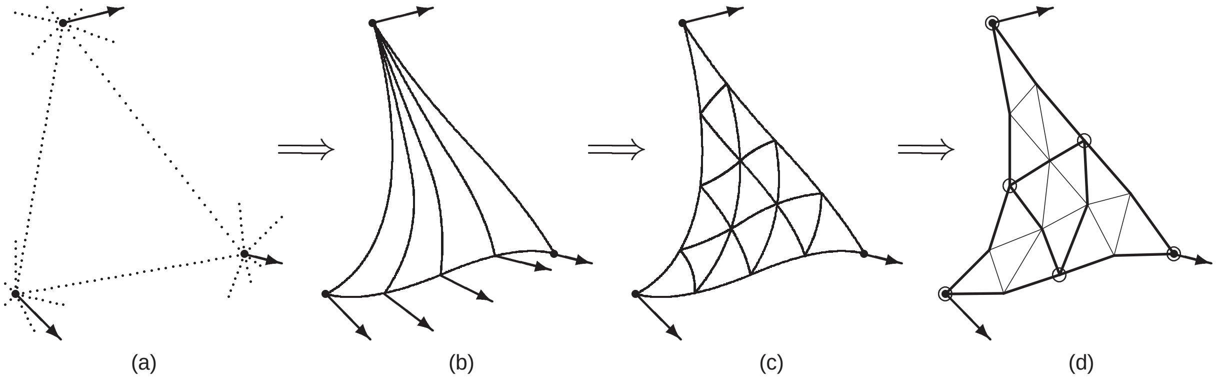


Figure 4: G^1-continuous transfinite interpolation. (a): Approximation of surface normal vectors for the vertices (•) in $\mathbb{R}^3$ on the basis of the topological relations. (b): Definition of boundary curves, some exemplary curves of a single partial interpolant are drawn additionally. (c): Some parameter curves of the complete transfinite interpolant. (d): Subdivision of the geometric interpolant into sixteen small planar patches (for computational reasons) and into four surface grey value facets (○ = vertices of the surface grey value representation).

and Pottmann, 1996), transfinite interpolants are chosen. The interpolation and also the building up of the boundary constraints and the interpolant itself have to be repeated after each iteration of the automatic surface reconstruction process; therefore, we have a little bit different prerequisites as if working with spline patches in an interactive manner. The concept of transfinite interpolations guarantees, that the whole procedure is numerically very insensitive and stable. Transfinite interpolants strictly act locally, therefore, they are a suitable completion of the underlying triangular data structure with respect to computation time.

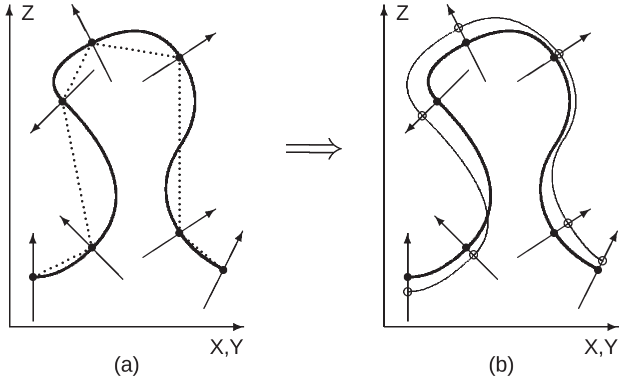
Here, we focus on G^1-continuous transfinite interpolation as introduced by (Nielson, 1987) by the so-called side-vertex-method. The extension towards G^2-continuity is possible, cf. (Hagen and Pottmann, 1989). The basic idea of transfinite interpolation is first to build up a network of boundary curves between given vertices in $\mathbb{R}^3$ and additional boundary constraints within these vertices. Afterwards, the final interpolant results from interpolating the curve network.

To reach G^1-continuity, in a first step the surface normal vectors for each vertex are approximated, cf. fig. 4(a), by calculating the arithmetic mean of the surface normal of all incident triangles, e.g. Weighting strategies depending on the area of each triangle are obsolete, since the above-mentioned strategy concerning a constant vertex density in object space leads to similar areas of adjacent triangles. Based on vertex positions and normal vectors, incident vertices are connected by *Hermite*-splines of degree three. In addition, surface normal vectors along the spline curves are defined, interpolating the normals at the vertices and being locally perpendicular towards the spline curve. Following this course of action again, three partial interpolants can be built by the bunches of curves, connecting each one of the original vertices with the vertices and their local normal vectors of the opposite boundary curve. Fig. 4(b) shows the three boundary curves and exemplary curves of one single partial interpolant. The final G^1-continuous interpolant, cf. fig. 4(c), follows from a linear combination of the partial interpolants, using local weights depending on the barycentric coordinates of the triangle. A proper weight function guarantees the degree of continuity within the triangle and at the border of the triangle; thus the global surface has a continuously varying outward surface normal, which is G^1-continuous.

For computational reasons, the resulting G^1-continuous surface patches are subdivided regularly into small linear patches. This subdivision accelerates both the calculation of the coefficients of facets stereo visions observation equation and the detection of occlusions in image space. The small triangles in fig. 4(d) give an illustration. Furthermore, fig. 4(d) shows a second kind of subdivision into four surface grey value facets. Since the resolution for the grey values in object space is usually finer than for the geometry in object space, it is highly recommended to build up the triangulation of the surface grey value representation in object space directly upon the triangulation of the geometry in object space. This proceeding is again based upon the concepts of hierarchical triangulations, for the examples in section 4 always a regularly subdivision with a ratio of both sixteen grey value triangles and sixteen linear geometry patches per triangle of the geometric representation is used.

3 IMAGE INVERSION IN OBJECT SPACE

Based on the preceding surface representation, cf. fig. 5(a), the photogrammetric surface reconstruction process in object space is reformulated. The unknown geometric parameters, describing the variation of the surface in object space, which are calculated by an iterative least squares estimation, are defined not any more with respect to a global direction, like $\Delta Z = \Delta Z(X,Y)$, but with respect to the direction of the local surface normal vectors $\boldsymbol{N}_{\mathrm{i}}$, cf. fig. 5(b).



(a)

(b)

Figure 5: 3-d surface model for multi-image matching in object space. (a): 3-d surface model, (b): vertex-positions (∘) after one iteration of the photogrammetric surface reconstruction process.

The correspondence condition between image pixel grey values $G'(x',y')$, $G''(x'',y'')$, ... from different overlapping images and the surface grey value $G(t_1,t_2)$ in object space remains to be the basic starting point for the observation equation, cf. (Weisensee, 1992), with residuals $v_{G'}$, $v_{G''}$, ... of the image grey values and the radiometric parameters h_0 and h_1 per image:

$$\begin{aligned} \Big(G'(x',y') + v_{G'}(x',y')\Big)\cdot h_1' + h_0' &= \\ \Big(G''(x'',y'') + v_{G''}(x'',y'')\Big)\cdot h_1'' + h_0'' &= \\ ... &= G(\tilde{t}_1,\tilde{t}_2) \\ &= G(t_1,t_2)\ . \end{aligned} \tag{1}$$

The transition between barycentric coordinates $\tilde{t}_1$,$\tilde{t}_2$ with respect to a triangle of the grey value representation and t_1,t_2, related to a triangle of the geometric surface representation, cf. fig. 4(d) again, can be established easily, if the subdivision of the geometry triangle has been formulated previously on the basis of t_1 and t_2. By this way, the barycentric coordinates t_1,t_2 represent the connecting link between the grey value function in object space and the geometric surface description, which is modeled by triangular, transfinite interpolants with G^1-continuity as introduced above, here abbreviated by

$$\boldsymbol{X}(t_1, t_2) = \boldsymbol{f}_{\boldsymbol{G}^1}[\boldsymbol{X}_{\mathrm{i}}, \boldsymbol{N}_{\mathrm{i}}](t_1, t_2) \ . \tag{2}$$

The contents of the angular brackets in (2) indicate, that the run of the surface depends on the three vertex positions $\boldsymbol{X}_{\mathrm{i}}$ of the triangle in question and the normal unit vectors $\boldsymbol{N}_{\mathrm{i}}$ in these vertices. Presuming known inner and outer orientations of an image, the relation between each pixel position in image space and its ray in object space, resp. its pertaining point of intersection $\boldsymbol{X}(t_1, t_2)$ on the surface in object space, is given by the well-known perspective camera model. We apply *Taylor*-linearization to the grey value function in object space around the approximative known point of intersection $\boldsymbol{X}^\circ(t_1^\circ, t_2^\circ)$ with the ray from a single image pixel[§]. Afterwards, we change from the partial derivatives along the barycentric coordinates, $\partial G^\circ(t_1^\circ, t_2^\circ)/\partial t_1$ and $\partial G^\circ(t_1^\circ, t_2^\circ)/\partial t_2$, to directional derivatives $\mathrm{d}G^\circ(t_1^\circ, t_2^\circ)/\mathrm{d}s_1$, $\mathrm{d}G^\circ(t_1^\circ, t_2^\circ)/\mathrm{d}s_2$ along the corresponding curves on the surface in object space with respect to elements $\mathrm{d}s_i$ of arc length. The equation of the pixels' ray in object space allows to turn from directional elements tangential to the surface to a single directional element perpendicular to the surface. For this purpose, the tangent unit vectors $\boldsymbol{T}_1^\circ(t_1, t_2)$ and $\boldsymbol{T}_2^\circ(t_1, t_2)$ along the parameter curves on the surface are introduced, cf. fig. 6.

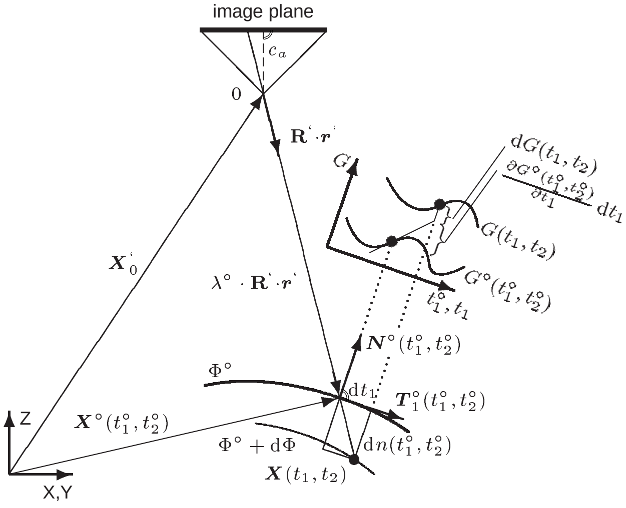


Figure 6: Image inversion: image grey values are projected back to the surface Φ° in object space. The estimation of the geometric change from Φ° to $\Phi^\circ + \mathrm{d}\Phi$ within one iteration of the reconstruction is based on the partial derivatives of the grey value function in object space, $\frac{\partial G^\circ(t_1^\circ, t_2^\circ)}{\partial t_1}$ and $\frac{\partial G^\circ(t_1^\circ, t_2^\circ)}{\partial t_2}$.

Independent of the continuity degree of the surface grey value representation, a linear approach (which is G^0-continuous) seems to be sufficient within our linearized formulation to express the differential change $\mathrm{d}G(t_1, t_2)$ of the surface grey value within a single grey value triangle in dependency of the grey value changes $\mathrm{d}G_{\mathrm{i}}$ at the vertices of the triangle:

$$\begin{aligned} \mathrm{d}G(t_1, t_2) &= t_1 \cdot \mathrm{d}G_1 + t_2 \cdot \mathrm{d}G_2 + (1 - t_1 - t_2) \cdot \mathrm{d}G_3 \\ &= f_{G^0}[\mathrm{d}G_{\mathrm{i}}](t_1, t_2) \ . \end{aligned} \tag{3}$$

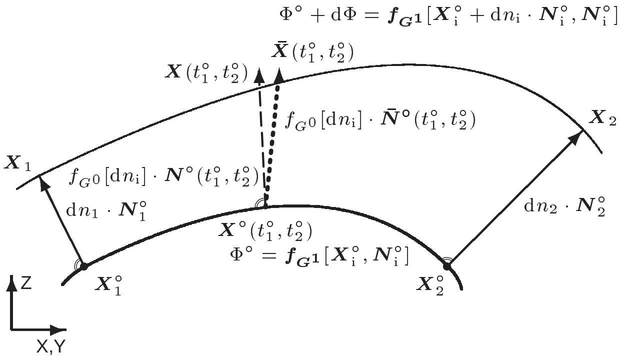


Figure 7: Translation of the surface within one iteration of the photogrammetric reconstruction depending on surface normal vectors $\boldsymbol{N}^\circ(t_1^\circ, t_2^\circ)$ or pseudo normal vectors $\bar{\boldsymbol{N}}^\circ(t_1^\circ, t_2^\circ)$.

For the geometric changes, we follow a similar strategy. Again in principle independent from the actual degree of continuity of the surface representation, though G^1-continuity is recommended, the new run of the surface is found by

$$\begin{aligned} \boldsymbol{X}(t_1^\circ, t_2^\circ) &= \boldsymbol{X}^\circ(t_1^\circ, t_2^\circ) + f_{G^0}[\mathrm{d}n_{\mathrm{i}}](t_1^\circ, t_2^\circ) \cdot \bar{\boldsymbol{N}}^\circ(t_1^\circ, t_2^\circ) \\ &\approx \boldsymbol{f}_{\boldsymbol{G}^1}[\boldsymbol{X}_{\mathrm{i}}^\circ + \mathrm{d}n_{\mathrm{i}} \cdot \boldsymbol{N}_{\mathrm{i}}^\circ, \boldsymbol{N}_{\mathrm{i}}^\circ](t_1, t_2) \ , \end{aligned} \tag{4}$$

with

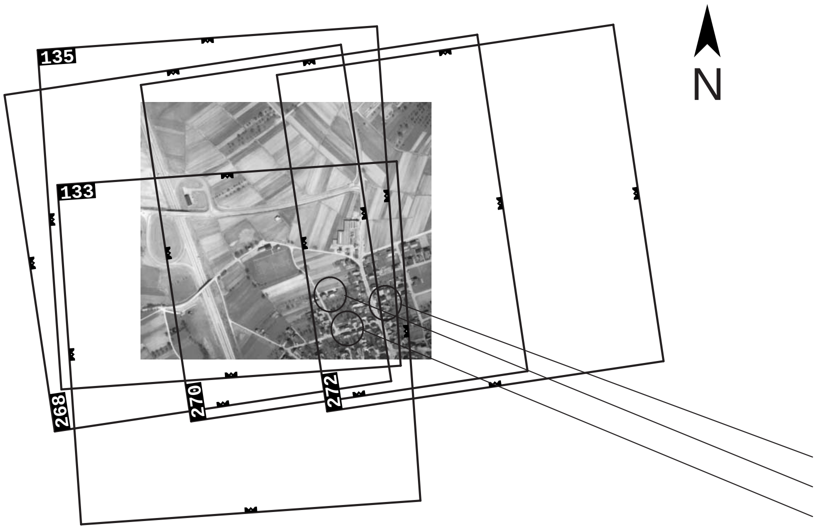
$$\bar{\boldsymbol{N}}^\circ(t_1^\circ, t_2^\circ) = \boldsymbol{f}_{\boldsymbol{G}^0}[\boldsymbol{N}_{\mathrm{i}}^\circ](t_1^\circ, t_2^\circ) \ . \tag{5}$$

Formula (5) explains, that within a single geometric triangle not the geometric normal vectors $\boldsymbol{N}^\circ(t_1^\circ, t_2^\circ)$ are used, but some linearly interpolated pseudo normal vectors $\bar{\boldsymbol{N}}^\circ(t_1^\circ, t_2^\circ)$. As outlined in fig. 7, neither the use of $\boldsymbol{N}^\circ(t_1^\circ, t_2^\circ)$ nor $\bar{\boldsymbol{N}}^\circ(t_1^\circ, t_2^\circ)$ results in a strict relation between Φ° and the new run of the surface $\Phi^\circ + \mathrm{d}\Phi$. But considering all possible cases of the surfaces behaviour, the use of the pseudo normal vectors offers in general the better approximation. So, good rates of convergence can be expected for the whole reconstruction process. Furthermore, the computation of $\bar{\boldsymbol{N}}^\circ(t_1^\circ, t_2^\circ)$ can be done much faster than that of $\boldsymbol{N}^\circ(t_1^\circ, t_2^\circ)$. Since the whole procedure needs to be run iteratively, because only the linear terms of the *Taylor*-series are considered, it can be well-accepted, that the second line of (4) is just a good approximation, but does not hold strictly. Following the handling of the radiometric parameters due to (Tsay, 1996), the linearized observation equation can be written by

$$\begin{aligned} & v_{G^{\backprime}}(x^{\backprime}, y^{\backprime}) + G^{\backprime}(x^{\backprime}, y^{\backprime}) - \frac{G^\circ(\tilde{t}_1^\circ, \tilde{t}_2^\circ) - h_0^{\backprime\circ}}{h_1^{\backprime\circ}} \\ &= \frac{1}{h_1^{\backprime\circ}} \Bigg[-G^{\backprime}(x^{\backprime}, y^{\backprime}) \cdot \mathrm{d}h_1^{\backprime} - \mathrm{d}h_0^{\backprime} + f_{G^0}[\mathrm{d}G_{\mathrm{i}}](\tilde{t}_1^\circ, \tilde{t}_2^\circ) \\ &\quad + \Bigg(\frac{\mathrm{d}G^\circ(\tilde{t}_1^\circ, \tilde{t}_2^\circ)}{\mathrm{d}s_1} \cdot \frac{(\boldsymbol{X}^\circ(t_1^\circ, t_2^\circ) - \boldsymbol{X}_0^{\backprime}) \cdot \boldsymbol{T}_1^\circ(t_1^\circ, t_2^\circ)}{(\boldsymbol{X}^\circ(t_1^\circ, t_2^\circ) - \boldsymbol{X}_0^{\backprime}) \cdot \bar{\boldsymbol{N}}^\circ(t_1^\circ, t_2^\circ)} \\ &\quad + \frac{\mathrm{d}G^\circ(\tilde{t}_1^\circ, \tilde{t}_2^\circ)}{\mathrm{d}s_2} \cdot \frac{(\boldsymbol{X}^\circ(t_1^\circ, t_2^\circ) - \boldsymbol{X}_0^{\backprime}) \cdot \boldsymbol{T}_2^\circ(t_1^\circ, t_2^\circ)}{(\boldsymbol{X}^\circ(t_1^\circ, t_2^\circ) - \boldsymbol{X}_0^{\backprime}) \cdot \bar{\boldsymbol{N}}^\circ(t_1^\circ, t_2^\circ)} \Bigg) \\ &\quad \cdot \ f_{G^0}[\mathrm{d}n_{\mathrm{i}}](t_1^\circ, t_2^\circ) \Bigg] \ . \end{aligned} \tag{6}$$

In (6), the centre of projection is denoted by $\boldsymbol{X}_0^{\backprime}$. It is obvious, that the coefficients of the design matrix concerning the unknown geometry parameters $\mathrm{d}n_{\mathrm{i}}$ depend on the local run of the surface in connection with the direction of ray from the image pixel, and on the direction derivatives of the surface grey values in object space. $\mathrm{d}G^\circ(\tilde{t}_1^\circ, \tilde{t}_2^\circ)/\mathrm{d}s_i$ should be calculated by numerical derivation. According to fig. 4(d), the grey value derivative can be approximated by switching from arc length on the surface to a small appropriate line element. Additional equations for regularization, formulated with respect to the surface geometry as considered by (Wrobel et al., 1992) for the $2\frac{1}{2}$-d surface model, are adapted for the 3-d surface model, cf. (Schlüter, 1998).

[§] The circular index always indicates approximative start values before an iteration step of the reconstruction process



flying altitude:	$600m$ above ground
image scale:	$M_b \approx 1:4000$
camera, film:	ZEISS RMKA, negative b/w
image format:	$23 \times 23cm^2$
camera constant:	$c_a = 153mm$
scanner:	ZEISS PS1
pixel size:	$15 \times 15\mu m^2$
quantisation:	$8bit$ / pixel

Figure 8: Overview of scene *Walddorfhäslach* (Baden-Württemberg, Germany). Image cover with respect to the reconstructed orthoimage of the whole area.

It has been pointed out in the beginning, that the reconstruction procedure is formulated as a multigrid procedure in connection with an image pyramid. Within each level of the multigrid, the whole evaluation area is subdivided by overlapping cubes in object space, containing about some hundred surface triangles each, which are computed separately. This process is organized by an octree structure, which is also used in connection with the detection of occlusion, especially to determine the participating images for each computation window. The final check of occlusion is done on the basis of local distance buffers in image space.

The resulting normal equations for each computation window are stored by a sparse matrix storage scheme (MRS = modified compressed row storage format, cf. (Saad, 1996)), because usually there exists no strictly regular pattern of sparsity. Iterative solvers work efficiently on such schemes, here SSOR $(=$ symmetric successive overrelaxation$)$ and Conjugate Gradients acceleration are used. Since the main part of the solution vector is dominated by additions with high-frequency characteristics, because the parts of the solution with lower frequencies usually were already determined on previous levels of the multigrid, the number of iterations concerning the solution of the linear system is very low. This offers a huge saving of computation time with respect to direct solutions of the normal equations.

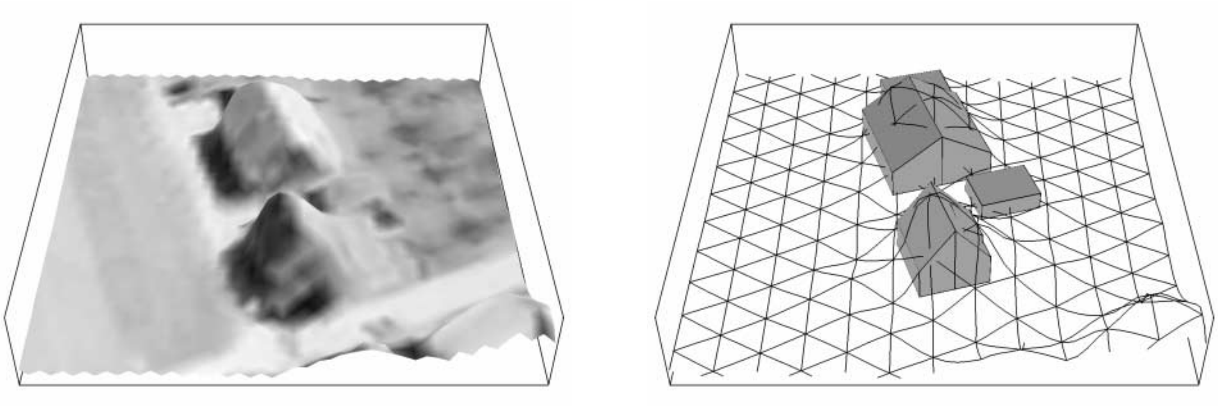
4 RESULTS FROM LARGE SCALE AERIAL IMAGERY

The new approach for surface reconstruction was applied to aerial image data of the scene *Walddorfhäslach* as outlined in fig. 8. Formerly published results from this data, which give interesting insights into typical advantages of multi-image surface reconstruction in object space, were based on the common $2\frac{1}{2}$-d approach, cf. (Schlüter and Wrobel, 1996). Standard deviations of the heights up to $\sigma_Z \approx \pm 3\text{-}4cm$ have been reached and confirmed. Some areas of the already presented results have been reconstructed also by the new 3-d approach. The comparison of both results shows that the height differences are always in the small range, which has to be expected with respect to the corresponding standard deviations. This conformity holds for several local surface types, for flat terrain with low texture as well as in the presence of steep slopes. Furthermore, possible sources of disturbances like moving cars, telegraph poles, etc., usually do not influence the geometric results. It has been proved that the switch from a $2\frac{1}{2}$-d to the mentioned 3-d surface model does not weaken the reconstruction procedure, even though the $2\frac{1}{2}$-d representation would be sufficient. Since also the computation times can be kept similar to each other, there seems to be no reason against a global use of the 3-d surface model.

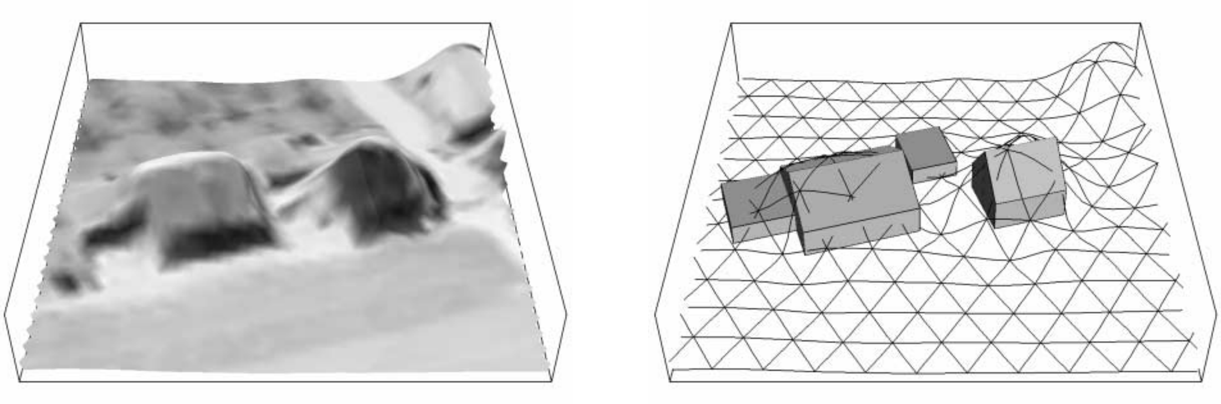
Of course, improved and, therefore, different results of the surface reconstruction using an underlying 3-d surface model are expected and are observed in the presence of buildings. Within the hierarchical reconstruction process, first differences of the intermediate results in dependence of the underlying surface model begin to appear at the fourth level of the image pyramid. Fig. 9 gives an example of such an intermediate result, showing both the surface including the surface grey values (on the left) and only the reconstructed geometric surface triangles (on the right). Manually measured outlines of the buildings are shown in addition for comparison (in grey). Obviously, the surface could be still described by a $2\frac{1}{2}$-d model at this level. The outlined geometry triangles (facets) have an edge length of approximately $4m$. Furthermore it can be noticed in fig. 9(a-d.2), that the geometric surface description is G^1-continuous as described above. A regular subdivision of every geometry triangle into sixteen subtriangles yields the radiometric surface description, here G^0-continuous interpolation is considered to be sufficient due to the high resolution of the grey value representation, which is about four image pixels per image (of the respective pyramid level) per grey value triangle.

Fig. 10 shows the final result of the automatic surface reconstruction, belonging to the same scene as in fig. 9. In addition, figs. 11 and 12 give the results for two different example areas, again rendered from different points of view. All locations are marked in fig. 8. Here, the geometry triangles have an edge length of about $0.5m$, the ratio of approximately four pixels per grey value triangle still remains.

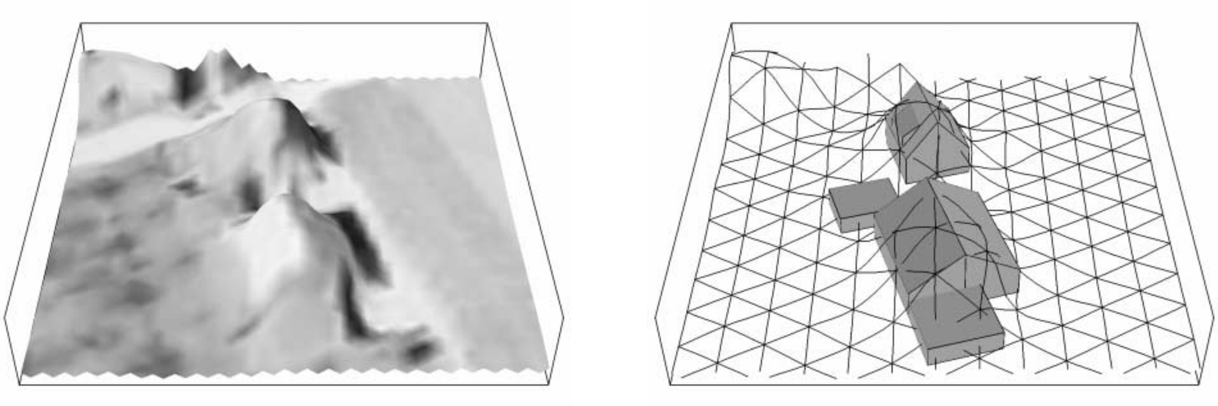
It has to be noticed that mainly the western and northern vertical building side-walls are visible within the aerial images, cf. fig. 8, whereas the southern and eastern walls are often invisible or only visible within a single image, which allows no geometric reconstruction, of course. This image configuration is partly instructive, but in no way optimal. Only a ninefold image overlap, resulting from a photo flight with both overlap and side lap of 70%, using again a wide-angle aerial camera, would guarantee that as many side-walls as possible are visible within at least two or three images. Of course, since several missing information about object space due to occlusion will never be completely avoided, it is also of interest to learn about the new algorithms' behaviour at these regions. Furthermore it has to be pointed out that even the outer regions of an aerial image usually offer a poorer image scale for vertical than for horizontal parts of the surface - this also lowers the reachable local accuracy for the geometry in object space.



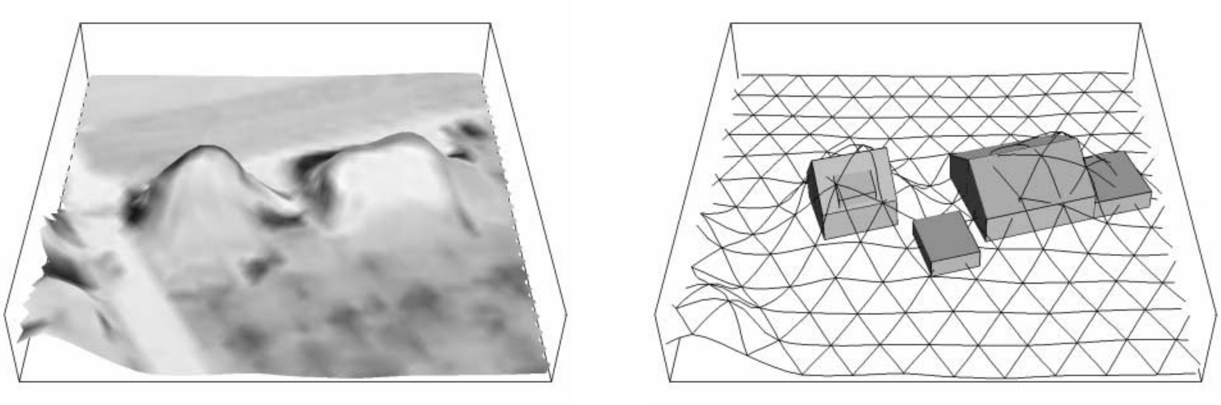
(a.1) View from west (a.2)



(b.1) View from north (b.2)

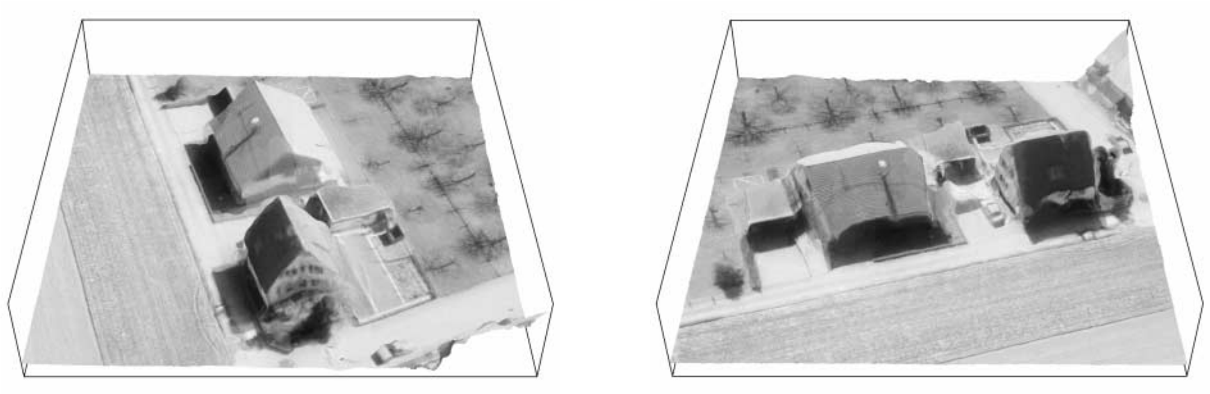


(c.1) View from east (c.2)



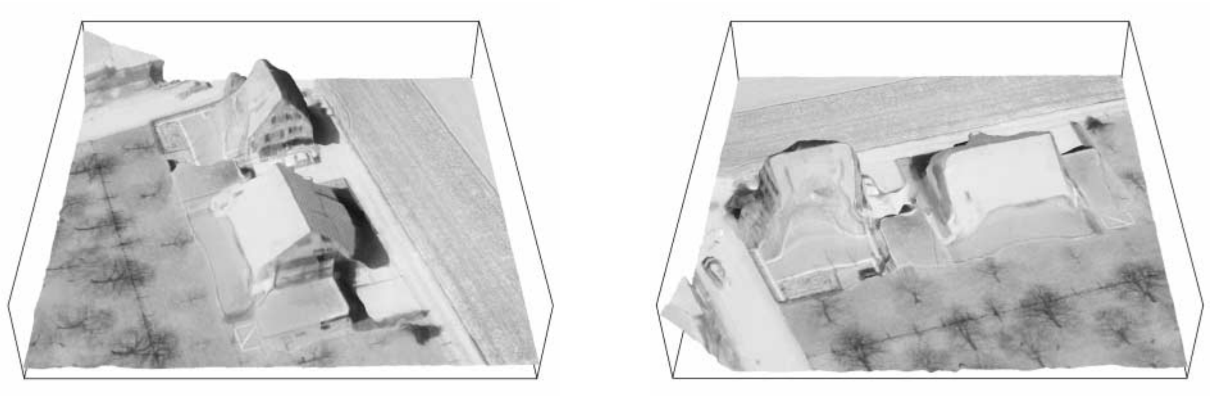
(d.1) View from south (d.2)

Figure 9: Intermediate results (4th level of multigrid), example 1. (a-d.1): Reconstructed surface and surface grey values in object space. (a-d.2): Resulting triangular geometry facets in comparison to manually measured buildings (outlined in gray).



(a) View from west

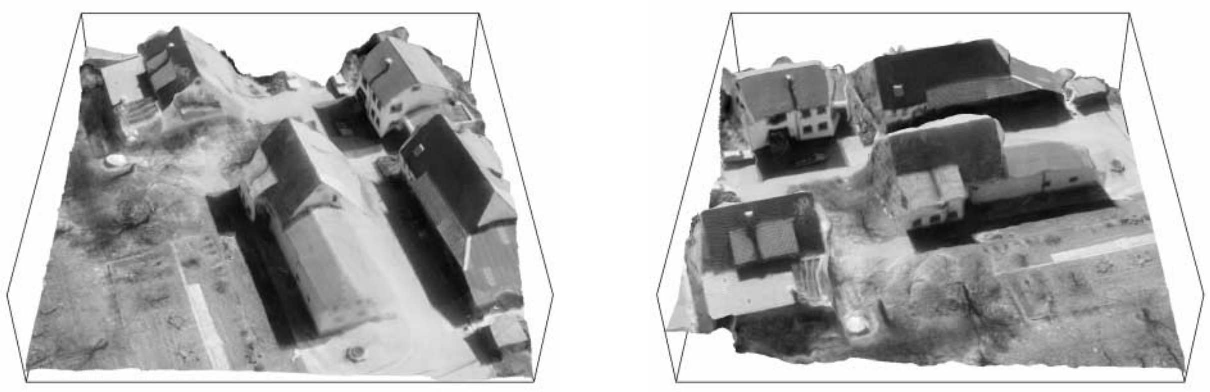
(b) View from north



(c) View from east

(d) View from south

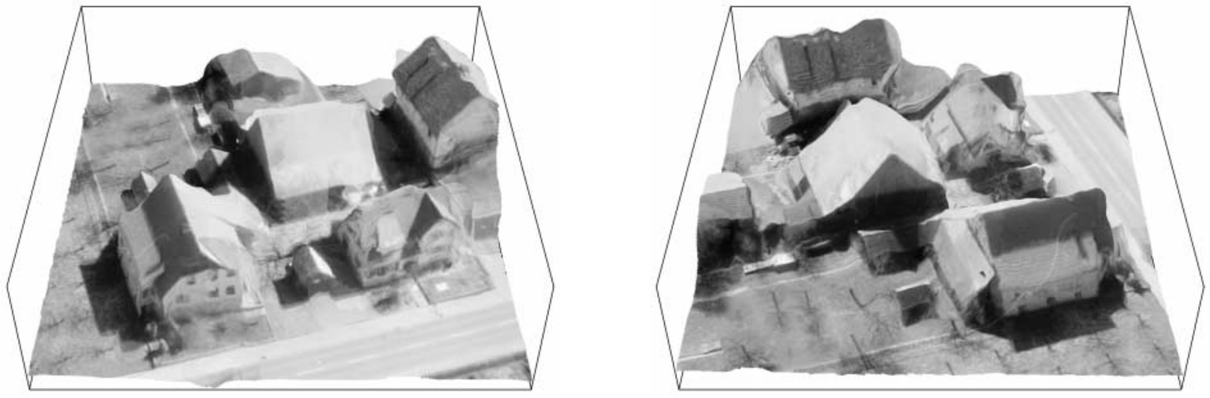
Figure 10: Result of 3-d surface reconstruction (1st level of multigrid), example 1.



(a) View from west

(b) View from north

Figure 11: Result of 3-d surface reconstruction (1st level of multigrid), example 2.



(a) View from west

(b) View from north

Figure 12: Result of 3-d surface reconstruction (1st level of multigrid), example 3.

Consequently, the quality of the reconstructed surface differs locally very much, depending on the local visibility, as the visual impressions of fig. 10 - 12 and also the resulting standard deviations for the surface geometry indicate. For the regions of the invisible building walls, only similar results as obtained by a $2\frac{1}{2}$-d surface model were reached and expected - a smooth approximation with appropriate high standard deviations. But for all the building side-walls, where two or more participating images offer enough grey value information, significantly improved results for the geometry are reached. Due to the chosen approach for regularization - a simple flatness constraint - the resulting surface remains to be some kind of smooth approximation. But, nevertheless, a detailed analysis of the resulting surface geometry shows, cf. (Schlüter, 1998), that most of the northern and eastern side-walls contain wide parts with a strict vertical and sometimes even slightly overhanging run of the surface, proving that the use of the 3-d surface model is successful.

5 CONCLUSIONS

This contribution gives an impressive example of the great variety of possibilities, which is offered by image matching in object space following the photogrammetric concept of facets stereo vision. It has been outlined how the estimation of a general 3-d surface description from multiple images can be carried out succesfully. This point is not only a topic of interest with respect to digital photogrammetry, but also related to the automatic generation of general, three-dimensional surface descriptions in computer aided geometric design.

Building detection and reconstruction should profit from the improved quality of the geometric surface description within urban scenes.

The new approach profits heavily from multi-image data and by this way from flexible conditions while taking images. Thus, it should be of special interest also for close range applications. The resulting triangular data structures can be easily embedded within concepts like VRML (= virtual reality modeling language), which offers standardized ways towards further data conversion or 3-d visualization.

ACKNOWLEDGEMENT

I am very grateful to my supervisor, Prof. Wrobel, for his strong support, his confidence and his precise and constructive criticism. Furthermore I express my sincere thanks to the staff of the *Institute of Photogrammetry and Cartography*, *TU Darmstadt*, for their committed teamwork, and especially to Dr. Düppe, M. Friehl, Dr. M. Kempa, M. Schmolla, M. Schreyer, J.-R. Tsay and Dr. M. Weisensee for helpful discussion and valuable cooperation. The complete underlying digital image data was kindly made available by Dr. Kiefer and Mr. Weiser, *Landesamt für Flurneuordnung und Landentwicklung Baden-Württemberg*. Goodwill by the *Bundesamt für Kartographie und Geodäsie*, *Frankfurt*, made publication of this contribution possible.

REFERENCES

Belhumeur, P. N., 1993. A binocular stereo algorithm for reconstructing sloping, creased, and broken surfaces in the presence of half-occlusion. In: Proc. Fourth International Conference on Computer Vision, IEEE Computer Society Press, Berlin, pp. 431–438.

Cavendish, J. C., Field, D. A. and Frey, W. H., 1985. An approach to automatic three-dimensional finite element mesh generation. IJNME 21, pp. 329–347.

De Floriani, L. and Puppo, E., 1995. Hierarchical triangulation for multiresolution surface description. ACM Transactions on Graphics 14(4), pp. 363–411.

Geiger, D., Ladendorf, B. and Yuille, A., 1995. Occlusions and binocular stereo. IJCV 14, pp. 211–226.

Hagen, H. and Pottmann, H., 1989. Curvature continuous triangular interpolants. In: T. Lyche and L. L. Schumaker (eds), Mathematical Methods in Computer Aided Geometric Design, Academic Press, San Diego, pp. 373–384.

Heipke, C., 1991. Integration von Bildzuordnung, Punktbestimmung, Oberflächenrekonstruktion und Orthoprojektion innerhalb der digitalen Photogrammetrie. Dissertation, TU München, DGK C 366, München.

Luo, A. and Burkhardt, H., 1995. An intensity-based cooperative bidirectional stereo matching with simultaneous detection of discontinuities and occlusions. IJCV 15, pp. 171–188.

Maas, H.-G. and Kersten, T., 1997. Aerotriangulation and DEM/orthophoto generation from high-resolution still-video imagery. PE&RS 63(9), pp. 1079–1084.

Nielson, G. M., 1987. A transfinite, visually continous, triangular interpolant. In: G. E. Farin (ed.), Geometric Modelling: Algorithms and New Trends, SIAM, Philadelphia, pp. 235–246.

Pfeifer, N. and Pottmann, H., 1996. Surface models on the basis of a triangular mesh – surface reconstruction. In: ISPRS 18th Congress, Vol. IAPRS 31(B3) Comm. III, Wien, pp. 638–643.

Saad, Y., 1996. Iterative Methods for Sparse Linear Systems. PWS publishing, New York.

Schlüter, M., 1997. 3D-Modelle für die photogrammetrische Oberflächenrekonstruktion. In: 16. Wissenschaftlich–Technische Jahrestagung der Deutschen Gesellschaft für Photogrammetrie und Fernerkundung (DGPF) 1996, Oldenburg, Publikationen der DGPF, Vol. 5, pp. 271–278.

Schlüter, M., 1998. Von der $2\frac{1}{2}$D-zur 3D-Flächenmodellierung für die photogrammetrische Rekonstruktion im Objektraum. Dissertation, TU Darmstadt. to appear.

Schlüter, M. and Wrobel, B. P., 1996. High resolution surface reconstruction of a landscape from large scale aerial imagery by Facets Stereo Vision – an extended test. In: ISPRS 18th Congress, Vol. IAPRS 31(B3) Comm. III, Wien, pp. 758–763.

Tsay, J.-R., 1996. Wavelets für das Facetten-Stereosehen. Dissertation, TH Darmstadt, DGK C 454, München.

Uray, P., 1996. From 3D Point Clouds to Surfaces and Volumes. Dissertation, TU Graz.

Veltkamp, R. C., 1992. Survey of continuities of curves and surfaces. Computer Graphics Forum 11(2), pp. 93–112.

Weisensee, M., 1992. Modelle und Algorithmen für das Facetten-Stereosehen. Dissertation, TH Darmstadt, DGK C 374, München.

Wrobel, B. P., 1987. Digital image matching by facets using object space models. In: Fourth International Symposium on Optical and Optoelectronic Applied Science and Engineering 'Advances in Image Processing', SPIE, Vol. 804, Den Haag, Niederlande, pp. 325–333.

Wrobel, B. P., Kaiser, B. and Hausladen, J., 1992. Adaptive regularization of surface reconstruction by image inversion. In: W. Förstner and S. Ruwiedel (eds), Robust Computer Vision, Wichmann, Karlsruhe, pp. 351–371.