

Coarse and fast modelling of urban areas from high resolution stereo satellite images

Thomas Krauß, Peter Reinartz, Manfred Lehner
Remote Sensing Technology Institute
German Aerospace Center (DLR)
PO Box 1116, 82230 Weßling, Germany
thomas.krauss@dlr.de

Uwe Stilla
Photogrammetry and Remote Sensing
Technische Universität München
Arcisstrasse 21, 80333 München, Germany

Abstract—With the availability of very high resolution (VHR) satellite images fast and relatively cheap investigations of large urban areas in comparison to aerial photography are possible. In urban areas a predominant use of satellite data is the generation of city models for applications like mobile phone signal propagation or flooding and catastrophe simulation. Since more and more large and quick growing cities emerge in developing countries, monitoring and modeling of these areas from satellite will be the cheapest if not the only possibility. Most of the actual methods used for the generation of city models depend on mainly manual work. A method for automatic derivation of – in a first step very coarse – models of urban environments will be of great use. In this paper a production chain and the methods used for such a automatic modeling is presented. The method is based on stereo images from VHR satellite stereo imagery provided, e.g., by IKONOS or QuickBird. In a first step a digital surface model (DSM) is derived from the stereo data. Subsequently a digital terrain model (DTM) and ortho images are created. Based on the local height differences between DSM and DTM and the normalized difference vegetation index (NDVI) a coarse classification can be made. Upon this classification object models can be selected and object parameters can be adapted to create an object-based representation of the satellite image scene. The method used is evaluated and the results are discussed.

I. INTRODUCTION

Models of urban areas become important for more and more applications. Beside the creation and updating of maps from sprawling urban settlements also models for simulation and planning in case of catastrophic events like flooding, earth quakes or civil defense applications are requested. Also many companies, e.g. from telecommunication sector, rely on high resolution three-dimensional city models for determining antenna positions and wave propagation.

In cases of low-resolution imagery like those from Landsat, SPOT or comparable satellites a detailed urban modeling is not possible since the spatial resolution is too low. On the other side high resolution aerial photography provide images with a ground resolution in the ranges of tens of centimeters. For this very high resolutions exist already a broad variety of methods and also of companies which create high resolution city models from such aerial photographs.

But since the appearance of high resolution satellites like Ikonos, QuickBird or the new French, Indian and Japanese missions with ground resolutions of one meter or below a coarse urban modeling from satellites becomes possible.

With this potential for the first time the modeling of large and quick growing cities in developing countries comes into reach. Such models can give the administration of such cities a much better base for decisions especially in cases of catastrophes.

So the intention of this work is the generation of three-dimensional models of urban areas from high-resolution satellite imagery. Since this limitation implies also the generation of a high-resolution digital surface model (DSM) of the urban area stereo images are needed. Furthermore due to the large coverage of about 12 km × 12 km per image for example for Ikonos, the whole process should be implemented in a mostly automatic processing chain.

The generation process of a three-dimensional city model can be divided in the following steps: After preprocessing the satellite data a DSM has to be generated from these images. The next steps contains the generation of true ortho images and classification. Based hereon objects for modeling can be selected, fitted to the data and exported in a standard modeling format.

II. STATE OF THE ART

Currently there exist a broad variety of methods for city modeling. These methods are mostly based on cadastral data, aerial images, aerial and terrestrial laser scanner data, terrestrial photographs and much more. The modeling integrates data from several of these sources in mostly intense manual work for the urban models [3], [1].

For example CyberCity uses a semi-automatic extraction of 3D point clouds with a photogrammetric workstation. From these points the roofs are generated and walls are built by projection of the roofs to the digital terrain model or directly from cadastral data. The textures of roofs and walls are finally extracted by projection of the models into the aerial photos. As stated by CyberCity a good operator is able to extract up to 500 roof parts per day.

2007 Urban Remote Sensing Joint Event

In our approach a big limitation is the missing availability of several data sources. Creating models for any selected urban area of the world in a short time is often only possible by obtaining very few high-resolution satellite images of the area.

Subject of this work is to develop a system, which creates urban models only from a minimum of two high-resolution stereo satellite images. Such images are provided at the moment, e.g., by Space Imaging (Ikonos, [13]) with a ground resolution of about 1 meter panchromatic and 4 m multispectral or Digital Globe (QuickBird, 60 cm panchromatic and 2.4 m multispectral).

In the near future many more high-resolution optical satellites are due to go into operation like WorldView I (2006) and II (2008) offering half-meter panchromatic and 1.4 to 1.8 m multispectral resolution ([4]).

III. DATA

The method relies on stereo image pairs which are best acquired in the same orbit with same illumination conditions, two distinct viewing angles and known internal and external orientation of the satellite (e.g. orbital positions and look angles or RPCs).

The following investigations were done mainly with two distinct Ikonos stereo image pairs:

- Athens scene: acquired 2004-07-24, 9:24 GMT, ground resolution 88 cm, viewing angles -19.99° and $+13.17^\circ$, level 1B image: full sensor corrected standard stereo product in epipolar geometry (Figure 1.)
- Munich scene: acquired 2005-07-15, 10:28 GMT, ground resolution 83 cm, viewing angles $+9.25^\circ$ and -4.45° , level 1A image: only corrected for sensor orientation and radiometry (Figure 2.)



Figure 1. Section $500\text{ m} \times 400\text{ m}$ from the Athens scene, top: left stereo image, bottom: right stereo image



Figure 2. Section $500\text{ m} \times 400\text{ m}$ from the Munich scene (area of technical university), top: left stereo image, bottom: right stereo image

IV. PROCESSING CHAIN

The herein proposed, implemented and tested processing chain consists of the following steps which are explained in detail in the following sections:

- Preprocessing of the raw imagery
- Creating a high-resolution digital surface model (DSM)
- Derivation of a digital terrain model (DTM) from the surface model (DSM)
- Calculating pansharpened images
- Projecting these pansharpened images to the DSM and creating high-resolution true ortho photos
- Classifying these ortho images to simple base classes
- Extracting distinct objects
- Selecting applicable models for all extracted objects
- Fitting model parameters for these objects
- Representing the object models through geometric primitives
- Texturing the geometric primitives
- Exporting geometric primitives into a suitable 3D-file-format

A. Preprocessing of the raw imagery

High-resolution Ikonos satellite imagery is accompanied by rational polynomial coefficients (RPCs) for every image. These coefficients are used to transform the geographical coordinates longitude X , latitude Y and ellipsoid height Z to image coordinates $(x,y)^T$ by division of two polynoms with 20 coefficients each [8], [6].

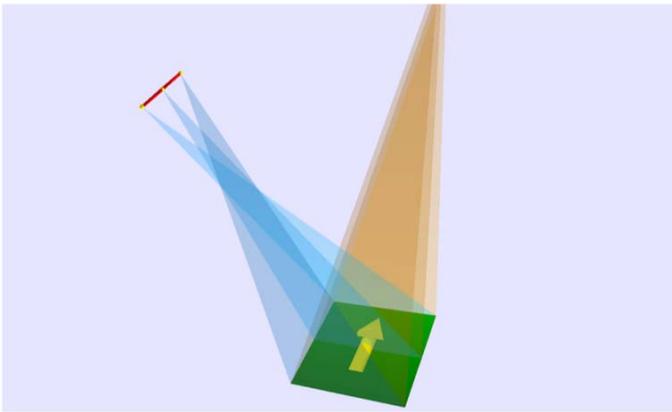


Figure 3. Ikonos acquisition geometry for the Munich-scene calculated from the provided RPCs – first image taken from north (brown), second in reverse mode from west (blue)

Based on these polynomials an image coordinate is calculated for every object point and vice versa longitude and latitude for a given height by recursive fitting from the image coordinates. In Figure 3. the satellite positions for the two images of the Munich-scene are calculated in this matter for three selected image scanlines (orbit height reduced by a factor of 20, longitude and latitude of satellite by a factor of 3).

This geometric calculations are used throughout the generation of the DSM and further for absolute georeferencing and ortho image generation.

Preprocessing consists mainly of the import of the delivered data from the satellite operating company to suitable formats and the extraction of all needed metadata like orbital and navigational data, internal and external orientation, rational polynomial coefficients (RPCs) etc.

In a further preprocessing step the multispectral channels with much coarser resolution (one pixel is about four meter) than the panchromatic channel are pansharpened. Therefore every stereo multispectral scene will be transformed in color space to a intensity channel and channels describing the color information. These images will be scaled to the resolution of the corresponding panchromatic image. A back transformation in color space using the pan channel instead of the calculated intensity channel results in the pansharpened multispectral image.

B. Creating the digital surface model (DSM)

The first complex processing step is the generation of a digital surface model from the image stereo pair. For creating a “depth map” from two stereo images a broad variety of methods exists. All methods use a kind of matching of features found in the first image to corresponding features in the second image. The relative displacement of this features give the disparity vectors and so a global depth map of the scene.

Two different approaches can be distinguished: Area based matching compares small windows between the images. Feature based matching compares extracted objects like lines or areas between the images.

Analyzing the different approaches in respect to the given imagery (“high-resolution” of about 1 m, steep walls, many occlusions, most features in size of resolution) leads to the following results:

- Pixel-, area-, intensity-based methods:
 - need a small correlation window to find correlating areas in the stereo image pair
 - steep features as common in urban areas smooth out on the size of this window
 - problems occur with occlusion that are common in urban scenes
- Object-, line-based methods:
 - “high-resolution” satellite imagery has a too coarse resolution for line extraction
 - most of the features, which can be found in urban images, are of the size of the resolution
 - so lines and features found in one image mostly have no corresponding features in the other image

Therefore a third approach besides pure area based and pure object based matching has to be used which depends not on large features and needs only small windows of about one pixel. If the geometry of the images is fairly good known pixels in one of the images can be projected to absolute object coordinates by using a given height. So projecting corresponding pixels from both images with the same given height fit to the same absolute object coordinates.

This approach comes to a simple solution if epipolar images can be used. A good overview of a selection of methods applicable for such simple epipolar geometry can be found on the Stereo Vision Research Page of the Middlebury College maintained from Daniel Scharstein, Middlebury College and Richard Szeliski, Microsoft Research [12].

Besides these an existing intensity based method and another epipolar based method developed at the German Aerospace Center (DLR) by [10] and [9] where investigated.

In the first – classical – method interest points are searched via the usage of the Förstner interest operator [5] in one image of the stereo pair. Based on a set of six conjugate points searched manually in images scaled down in an image pyramid by a factor of 32 an affine transformation is estimated. The interest points found in the first image are transformed via this transformation to the second image of the stereo pair and an area based matching of image windows around these points is performed. Subpixel refinement via local last square fit, propagating up the image pyramid and performing forward-backward-matching delivers a good and stable set of conjugate points in the full scale stereo pair.

Applying a region growing [11] gives a dense quantity of points and correlated disparities. Transforming this information to object space using ephemeris and pointing data and applying a triangulation to fill a grid delivers the digital surface model.

The second approach named “dynamic line warping” was developed applying a method known in speech recognition as “dynamic time warping” line by line for stereo images. This algorithm warps one linear array of values onto another. In speech recognition these arrays are voice input and recorded speech that are compared. In stereo image matching the two arrays are the corresponding lines of the two images. This approach requires epipolar images.

The two input arrays are correlated using a dynamic programming approach. The required cost function for image matching may be a pixel-by-pixel gray value distance or the cost function used by Birchfield and Tomasi [2]. Based on this distance a cost matrix $M_{i,j}$ can be calculated as

$$M_{i,j} = |I_{x+i,y} - I'_{x',y'}|$$

with $x' = \text{rpc}_x(x+i, y, h)$ and $y' = \text{rpc}_y(x+i, y, h)$ representing functions which calculate the coordinates (x', y') in image I' from the pixel position (x, y) in the first stereo image I and the disparity $h = i - j$.

The dynamic programming approach calculates a minimum distance between the two linear input arrays. In speech recognition only this distance for finding best matching words is of interest. But the matrix delivered by applying the dynamic programming delivers beneath the “minimum distance” as the bottom rightmost value also a so called “minimal path” from the left top to the right bottom corner.

Calculating the minimal path starts with the calculation of the cost matrix M and applying the dynamic programming approach for simplifying an exhaustive recursion. In this step the rows and columns of the cost matrix are cumulated to a matrix D filling the first line and column according to

$$D_{1,j} = \sum_{k=1}^j M_{1,k} \qquad D_{i,1} = \sum_{k=1}^i M_{k,1}$$

and the remaining of the matrix $D_{i,j} (i,j > 1)$ according to

$$D_{i,j} = M_{i,j} + \min \{D_{i-1,j}, D_{i,j}, D_{i-1,j-1}\}$$

Taking two gray value profiles from corresponding lines in the stereo images, calculating the matrix $D_{i,j}$ and marking this minimal path creates the path shown in Figure 5. (i, j run in this figure from the bottom left to the upper right corner).

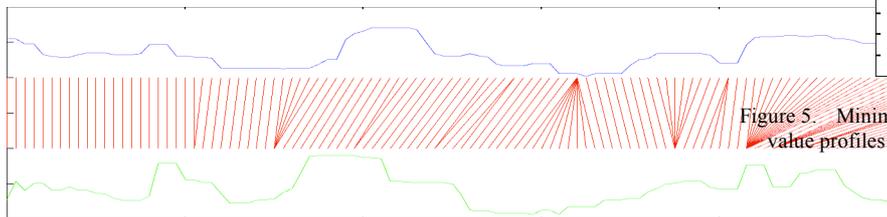


Figure 4. Found correlations between the two input gray value profiles (top and bottom of graphic)

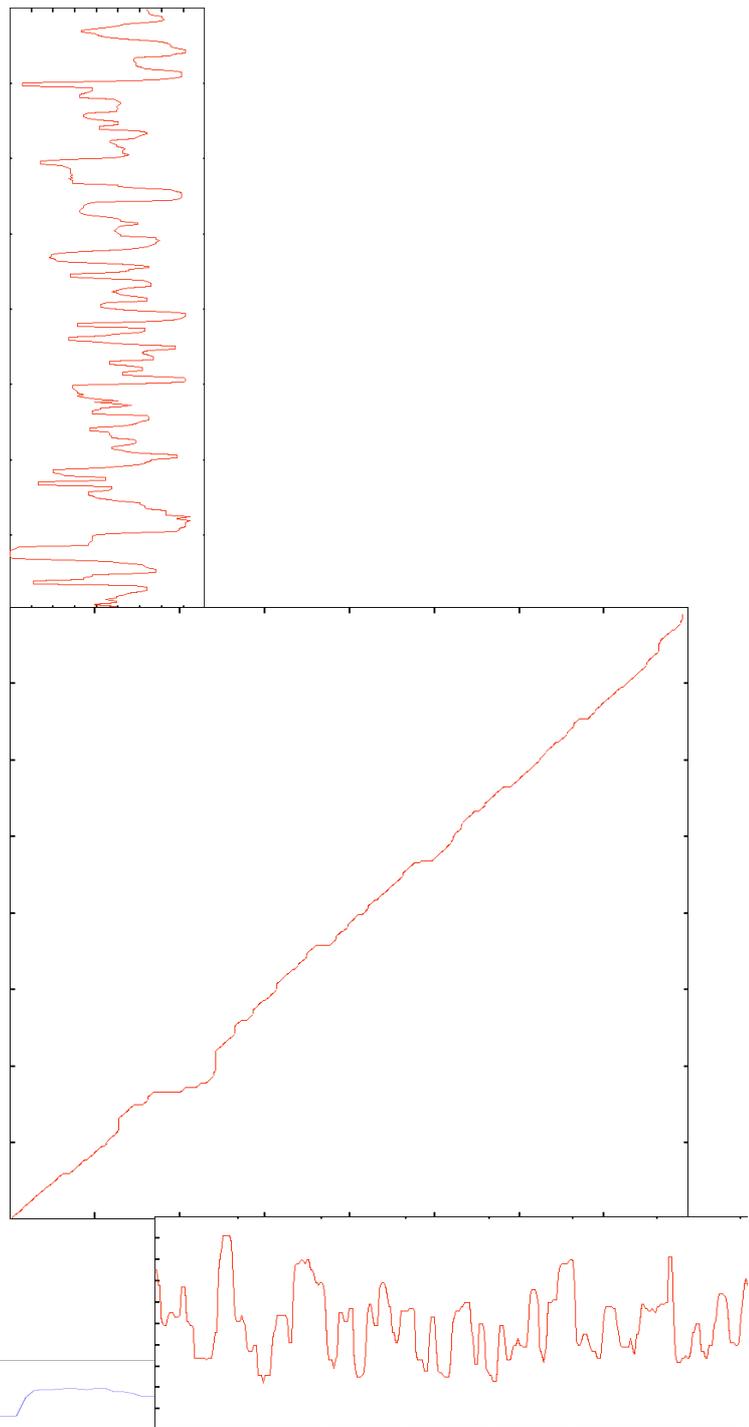


Figure 5. Minimal path created by the line warping algorithm with the grey value profiles as inputs (shown on the left and bottom for better visual correlation)

Picking a small area of this result and showing the extracted correlations between the input gray value profiles yields Figure 4. Areas with different widths in the profiles are mapped correctly to each other as can be seen well.

2007 Urban Remote Sensing Joint Event

A perfect diagonal line as minimal path represents no relative displacements between the two profiles. All deviations from this line indicate the searched parallaxes. If this minimal path is represented by pairs of coordinates $\mu_k = (i,j)_k$ it is possible to fit a parabola through $D_{i-1,j+1}$, $D_{i,j}$ and $D_{i+1,j-1}$ to calculate the parallaxes as the minimum of this parabola in subpixel accuracy.

Using small window areas of diameter w around every line position for calculating the distance $M_{i,j}$ gives better results but introduces on the other hand smoothing effects as known from area-based methods. The distance $M_{i,j}$ of line y with images $I_{x,y}$ and $I'_{x,y}$ and window size w is calculated as

$$M_{i,j} = \sum_{\lambda=-w/2}^{w/2} \sum_{\mu=-w/2}^{w/2} |I_{i+\lambda,y+\mu} - I'_{j+\lambda,y+\mu}|$$

This means that over a window of size $w \times w$ all absolute values of gray value distances between the corresponding pixels in the two images are summed up.

Afterwards a vertical median filter is applied to reduce line streaking and blunders resulting from the line by line correlation. For reduction of processing time a maximum correlation window size can be chosen. Outside of this distance from the main diagonal the matrix M simply gets filled with a maximum distance value instead of the calculated distance.

For the final calculation of the DSM the above resulting disparity image is back-transformed from epipolar image space to original image space. Forward intersection using RPCs and subsequent interpolation produce a regular DSM.

The advantage of dynamic line warping in respect to an area based matching approach is that the used „window“ can be as small as only one pixel and so no smearing at steep edges should occur. But disadvantages of this method are:

- the method needs epipolar images
- streaking occurs since there is no interconnection of the epipolar lines
- errors occur in water and grass areas

One solution for avoiding the streaking between neighbouring epipolar lines is the multidirectional approach „semi-global matching“ described by [7].

In this similar dynamic programming approach the summation occurs not only epipolar line for epipolar line but parallel for all epipolar lines up to a given maximum disparity. The disparity matrix is in this case a three-dimensional matrix of size image width \times image height \times \pm maximum disparity. Beside this the used cost function can be calculated after Birchfield/Tomasi or using mutual information.

The three investigated methods for the calculation of the DSM are in summary:

- Classical, area based hierarchical matching (Figure 6.)
- Matching epipolar lines using dynamic line warping (Figure 7.)

- Matching epipolar lines using semi-global matching (Figure 8.)

The results can be enhanced by applying a small median filter. So sharp edges are retained but outliers and parts of extreme streaking effects are filtered out (Figure 9.).

Also the cost function can be expanded from only one pixel to a small window of size $-w/2 \dots w/2$ as described above which creates better results but introduces on the other hand a smoothing effect on steep ridges. In all following calculations a window and median size of 3 pixels were used.

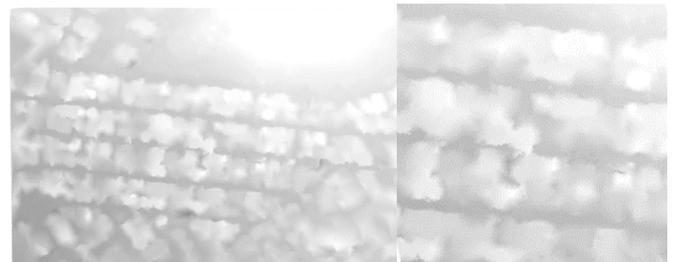


Figure 6. DSM generated by classical area based matching (Athens, clips 640 m \times 400 m and 250 m \times 250 m)

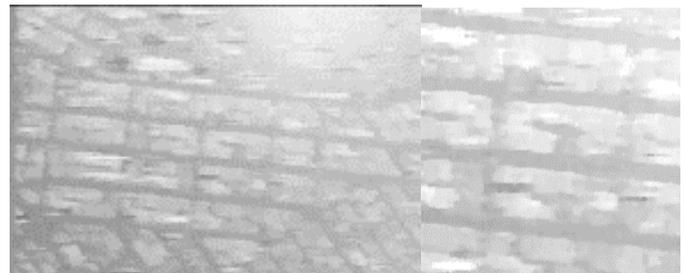


Figure 7. DSM generated by dynamical line warping (Athens, clips 640 m \times 400 m and 250 m \times 250 m)

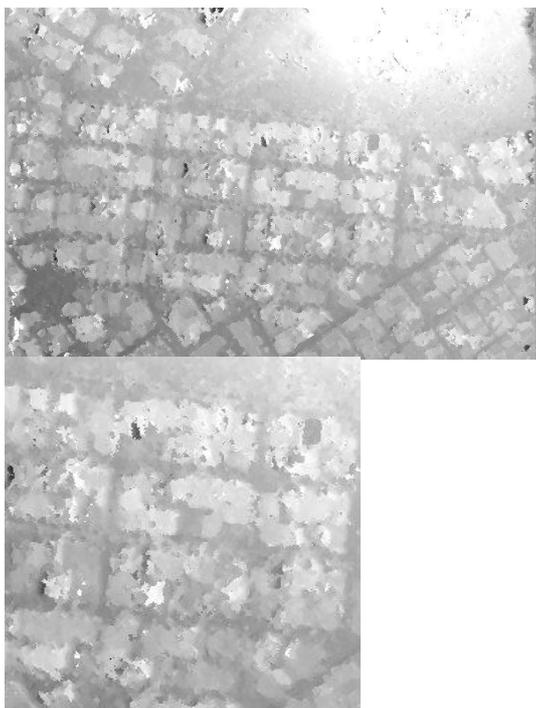


Figure 8. DSM generated by semi-global matching (Athens, clips 640 m × 400 m and 250 m × 250 m)

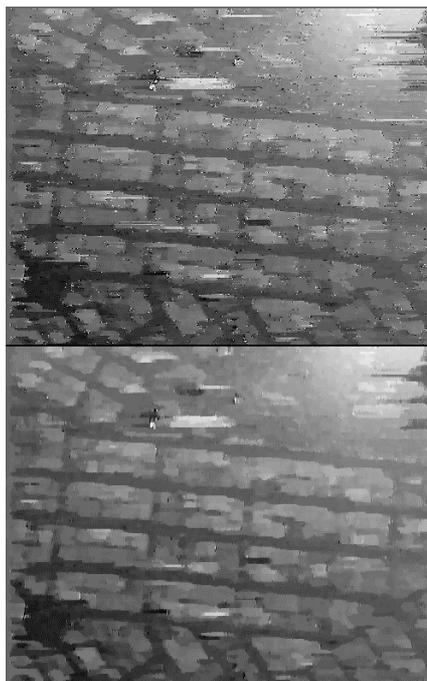


Figure 10. Same DSMs as in figure above but both with a window size of 5 pixels

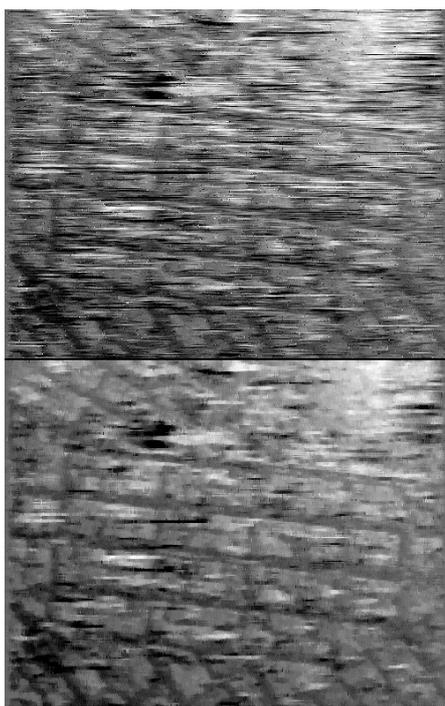


Figure 9. DSMs from Athens scene generated by dynamical line warping, left with median of size 1, right with median of size 5, both clips 500 m × 400 m

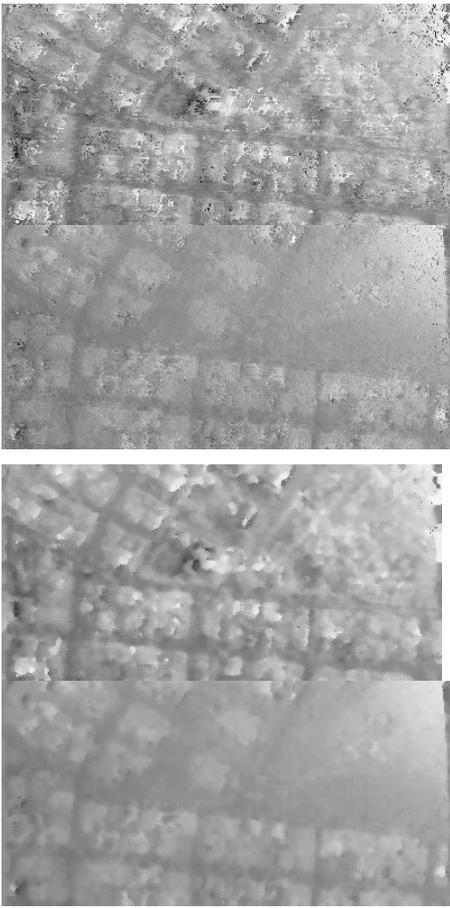


Figure 11. DSMs from Athens scene generated by semi global matching, left with intensity-based cost function (Birchfield-Tomasi), right with mutual information, upper with median of size 1, lower with median size of 5, clips 640 m x 400 m

Using mutual information in the calculation of costs (semi-global matching) instead of simple gray value differences or Birchfield/Tomasi also gives some remarkable improvements as shown in Figure 11. .

C. Extracting the digital terrain model

The generation of the digital terrain model (DTM) is based on the calculated DSM. First a maximum of the smallest horizontal cross sections of all buildings has to be estimated. For a normal urban area this is about 20 m wide. In industrial areas the buildings are normally much larger so lengths up to 100 m may be necessary.

Applying a minimum filter of this diameter to the DSM results in a height image with every pixel representing the minimum height in this distance around the pixel. Problems may arise at steep slopes since the slope will be raised. For calculating the DTM in reality a median filter will be applied, which does not return the middle value after sorting but rather a low value. If the lowest value will be returned it's the same as a minimum filter or a morphological erosion but we take a value of about 1 to 5 %. So single outliers with too small values does not dominate the generated DTM.

After applying this “minimum filter” an averaging using the same filter size is applied to obtain a smoother DTM.

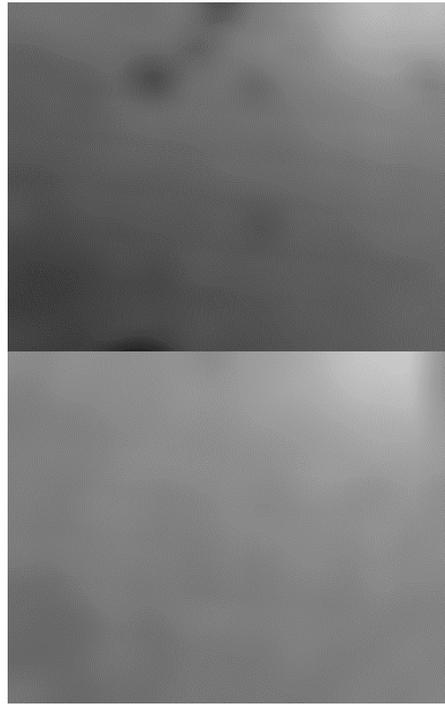


Figure 12. DTMs calculated from the DSMs resulting from the digital line warping (left) and semi global matching (right)

D. Creating true ortho images

Based on the detailed DSM and the multi-spectral channels from the stereo images it is possible to create so called true ortho images. In simple ortho images a much coarser DSM is used. In distinction to simple ortho images in true ortho images also buildings will be orthorectified and not only squeezed onto ground.

Projecting the pansharpened images using the known RPC for every image of the stereo pair and the calculated disparities from the DSM vertically down to a given ground a true ortho image will be generated.



Figure 13. Orthoimage Athens scene, channels red, green, blue



Figure 14. Orthoimage Athens scene, channels near infrared, red, green

Figure 13. and Figure 14. show the calculated ortho images from the Athens scene as RGB- and IRG-image respectively. The used DSM was calculated with the line-warping method.

E. Classification

Based on the pansharpened images and the beforehand calculated DSM and DTM a coarse classification is made. This classification is based on two binary masks.

The first is a mask of “high objects”. This mask is calculated by subtracting the DTM from the DSM and taking into account all heights greater than a given height-threshold as shown in Figure 15. .

The second mask is the vegetation mask. First the NDVI (normalized difference vegetation index) defined as a pixelwise normalized difference from the near infrared and red channels of the pansharpened multispectral true ortho image:

$$NDVI = (NIR - Red) / (NIR + Red)$$



Figure 15. Height masks calculated from the line-warping-DSM (left) and the semi-global-DSM (right) (all images 500 m × 400 m)

Applying also a threshold (in the examples always half of the full intensity) a binary vegetation mask can be calculated (see Figure 16.).

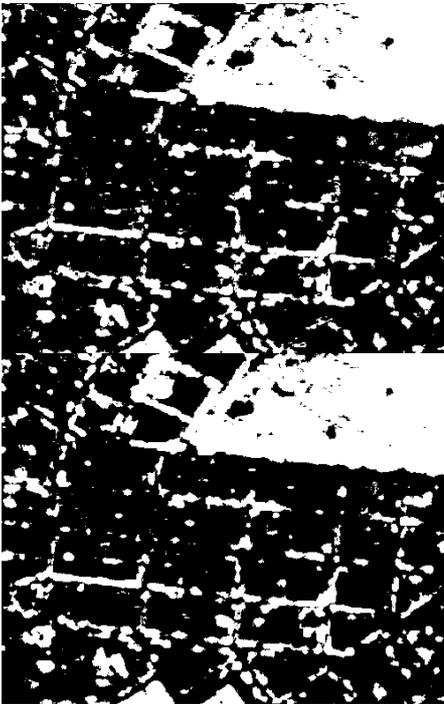


Figure 16. Vegetation masks calculated from the ortho image based on the line-warping DSM (left) and the semi-global DSM (right) (all images 500 m × 400 m)

Combining the two masks (“high objects” + “Vegetation”) leads to four classes:

- low and no vegetation: streets, plain soil, . . .
- high and no vegetation: buildings, . . .
- low and vegetation: meadows, grass, . . .
- high and vegetation: trees, bushes, . . .

Figure 17. and Figure 18. show the simple classifications calculated by this approach for two different DSM generation methods respectively.

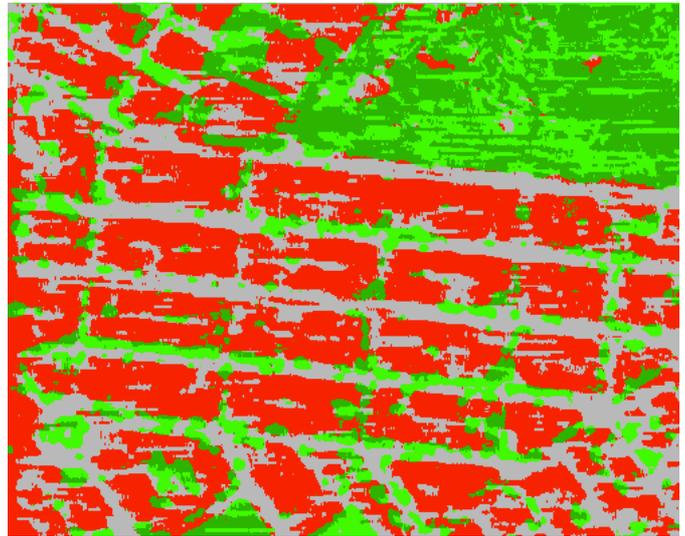
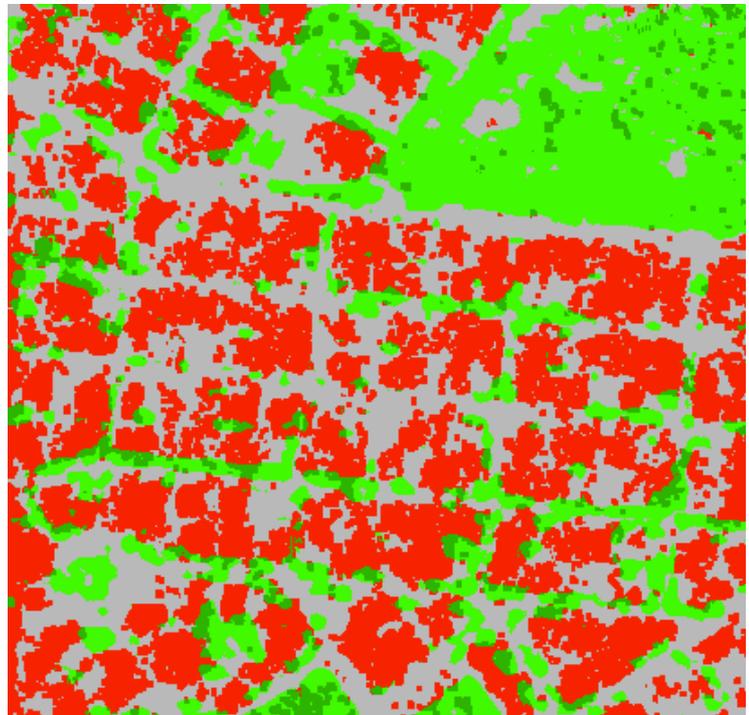


Figure 17. Classification of the Athens scene using the DEM and DTM generated by dynamic line warping



2007 Urban Remote Sensing Joint Event

Figure 18. Classification of the Athens scene using the DEM and DTM generated by semi-global matching

F. Object extraction

Based on the classification objects can be extracted. Using the mask for high objects combined with the vegetation mask two images containing (i) high vegetation and (ii) high non-vegetation can be extracted from the DSM. In image (i) separated small round objects up to a certain diameter can be classified as solitary trees whereas larger areas get classified as forest and modeled coarse as a bunch of overlapping single trees.

Classifying objects from image (ii) is more challenging since such objects will be of more interest in the resulting models. In a first approach sketched out below following assumptions are made:

- all high, non-vegetation objects are buildings (in the scene are no flyovers or steep rocks; and in contrast to (i) no roof gardens or roofs with grass)
- the buildings are all designed rectangular (in the first approach no round, triangular or other buildings are supported)

A section from the Munich test scene will be used for describing the method of building extraction. Figure 19. show the Technical University of Munich in the center of the image with four identified objects used in the following description.

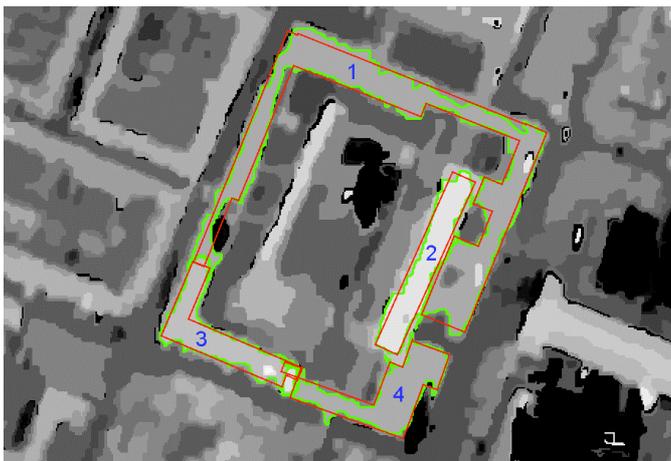


Figure 19. Four selected outlines for following investigations in Munich scene (image size: 500 m × 400 m)

In the first step the height image (difference of DSM and DTM) will be masked with a logical AND of the height- and the non-vegetation-mask. In this resulting image neighboring heights of small distance are combined to height classes.

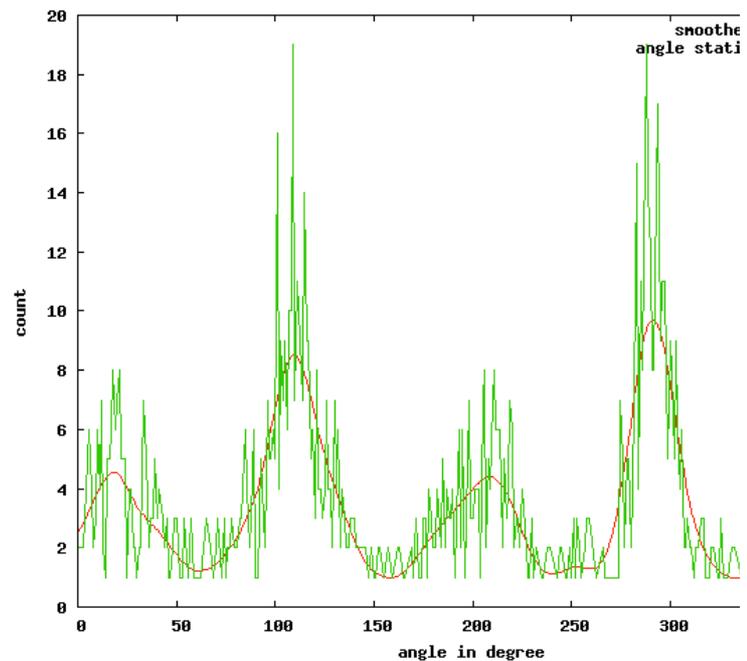
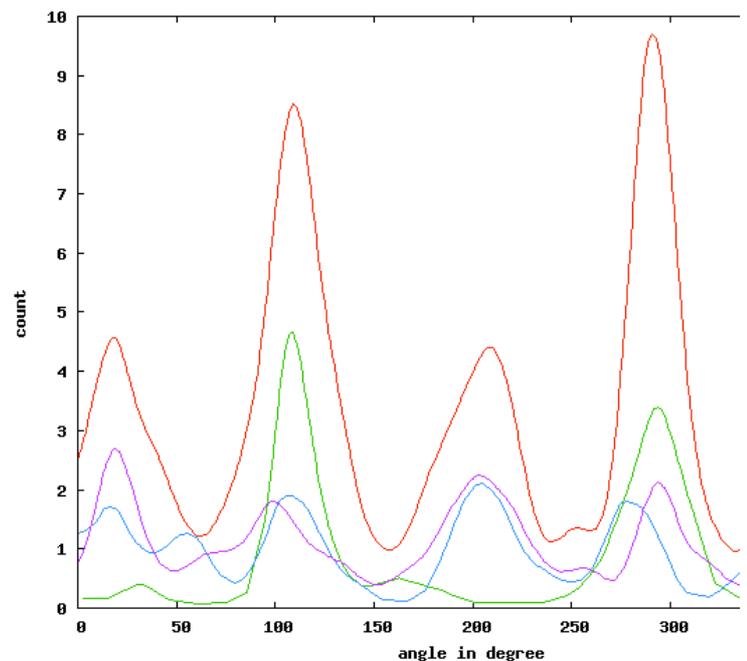


Figure 20. Statistics of measured angles for first polygon (up: count, right: angle in degree)

The borders of these classes are traced and converted to polygons. Figure 20. shows the statistics for the depicted green polygon „1“ in Figure 19. . For the statistics at every point of the polygon the weighted average direction from this point to any of the seven previous or following points is calculated. The direction statistics is averaged afterward applying a filter of size 10.

Figure 21. shows the smoothed statistics of angles for all four polygons. Selecting prominent maxima of these distributions and forcing perpendicularity leads to TABLE I. .



2007 Urban Remote Sensing Joint Event

Figure 21. Smoothed statistics of all angles in the depicted polygons smoothed on weighted ± 7 polygon points (up: count, right: angle in degree)

TABLE I. EXTRACTED MAXIMA FROM THE ANGULAR STATISTICS GIVE THE MAIN DIRECTIONS OF THE RECTANGULAR OBJECT

Polygon	Number of points	Found maxima
1	1167	4: 22°, 112°, 202°, 292°
2	292	4: 24°, 114°, 204°, 294°
3	349	4: 23°, 113°, 203°, 293°
4	400	4: 20°, 110°, 200°, 290°

Filtering and grouping angles to the found main directions found above from the angular statistics leads for the first polygon to the results in Figure 22. .

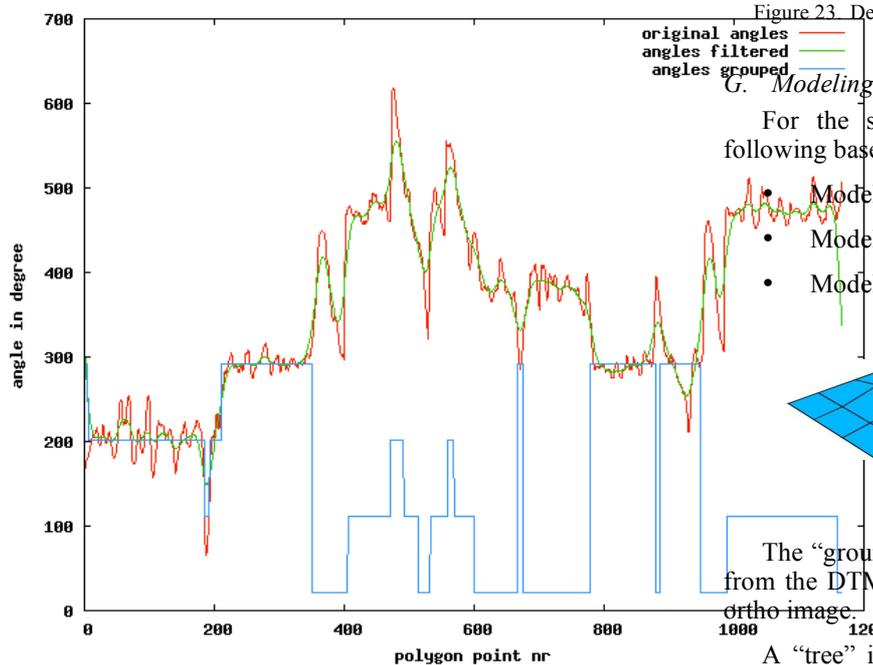


Figure 22. Angles for first polygon (up angle in degree, right pixel number) measured in polygon, smoothed and grouped together after identifying main directions

Figure 23. shows lines through all above by means of main directions grouped polygon points. In a last step consecutive line segments are intersected and the closed red polygons shown in Figure 19. above are generated.

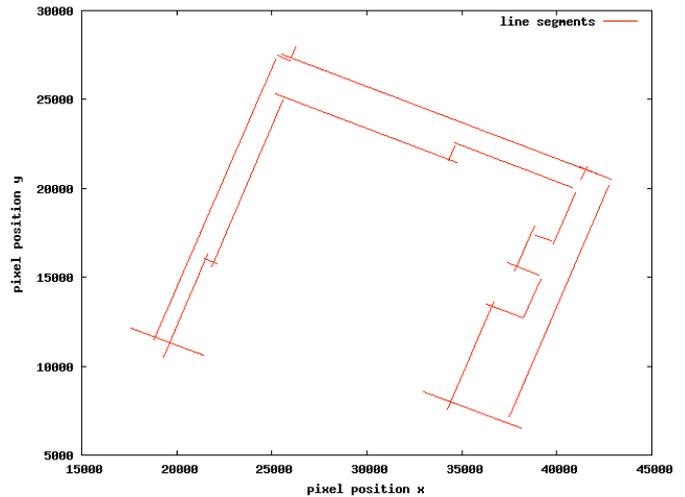


Figure 23. Derived line segments of polygon 1 after grouping of angles

G. Modeling

For the subsequent modeling of the extracted objects following base models are used:

- Model “ground” (class “low”, any type of vegetation)
- Model “tree” (class “high” and “vegetation”)
- Model “building” (class “high” and “no vegetation”)

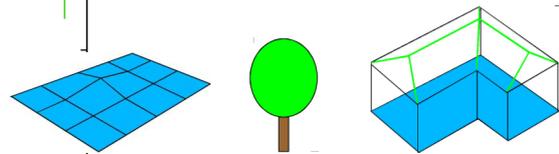


Figure 24. Simple models used

The “ground” is simply inserted as a height field extracted from the DTM with an optional texture directly from the true ortho image.

A “tree” is described by a crown diameter and a treetop height extracted directly from the classification (by single trees) and the DSM respectively.

A surrounding rectangular polygon describing a “building” is extracted as described in the previous chapter. In a first version a coarse modeling can be done by simply generating a polygonal model with the polygon as circumference and a flat top with an average height extracted from the DSM.

In a required more detailed version the polygonal circumference has to be divided into rectangles. An averaged section across these rectangles may give a hint if it's a gabled or a flat roof and lead to a more detailed model.

H. Representing the object models through geometric primitives

The models will be represented through following geometric primitives:

2007 Urban Remote Sensing Joint Event

- “ground”: as height field derived directly from the DTM – one ground object is therefore sufficient for the whole scene
- “tree”: trunk as cylinder and crown as ellipsoid from with extracted diameter
- “building”: with rectangular vertical walls following the extracted circumference and a horizontal polygonal roof in the first version. In the next version (split up to cuboids): cuboid representing inner and outer walls topped with symmetrical roof planes (gable and eaves height extracted from the averaged sections along the identified cuboid).

I. Texturing the geometric primitives

The “ground” object(s) derive their texture directly from the true ortho image. For texturing the “trees” a color statistics from the corresponding area in the true ortho image has to be made and the predefined texture of the ellipsoid fed with the identified main color components.

“Buildings”: For every polygon in the building a visibility in the two stereo images has to be calculated using the RPCs. (i) A average color or (ii) the full texture can be extracted from these images by transforming the polygon edges to the corresponding image and warping the identified area back onto the polygon.

J. Exporting geometric objects into a suitable 3D-format

After all these steps the optionally textured geometric primitives have to be exported into a suitable 3D vector format. Figure 25. shows the area around the technical university from the Munich scene as shown by a VRML viewer.



Figure 25. Simple 3D view generated from the Munich scene, size 640 m × 400 m, center: Technical University of Munich, right Old Pinacotheca

V. SUMMARY AND OUTLOOK

In this paper a processing chain for the automatic extraction of three-dimensional city models directly from high-resolution stereo satellite images is suggested. The processing chain is still in development. So the DSMs generated are not satisfying enough up to now. Also the automatic extraction of objects from the classified images exists only in a first evaluation version. The texturing is still missing up to now. But the results gained from each step are encouraging enough to follow the path and refine every step of the chain to receive a newly fully automatic system for generating coarse three-dimensional urban models from stereo satellite imagery in a short time.

REFERENCES

- [1] 3D Geo. <http://www.landexplorer.net/> (accessed 02/2006).
- [2] Birchfield, S. and Tomasi, C., 1998. Depth discontinuities by pixel-to-pixel stereo. Proceedings of the 1998 IEEE International Conference on Computer Vision, Bombay, India, pp. 1073–1080.
- [3] CyberCity. <http://www.cybercity.tv/>. (accessed 02/2006).
- [4] DigitalGlobe. <http://www.digitalglobe.com/about/imaging.shtml>. (accessed 02/2006).
- [5] Förstner, W. and Gülch, E., 1987. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In: ISPRS Intercommission Workshop, Interlaken.
- [6] Grodecki, J., Dial, G. and Lutes, J., 2004. Mathematical model for 3D feature extraction from multiple satellite images described by RPCs. In: ASPRS Annual Conference Proceedings, Denver, Colorado.
- [7] Hirschmüller, H., 2005. Accurate and efficient stereo processing by semi-global matching and mutual information. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
- [8] Jacobsen, K., Büyüksalih, G. and Topan, H., 2005. Geometric models for the orientation of high resolution optical satellite sensors. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 36 (1/W3). ISPRS Workshop, Hannover.
- [9] Krauß, T., Reinartz, P., Lehner, M., Schroeder, M. and Stilla, U., 2005. DEM generation from very high resolution stereo satellite data in urban areas using dynamic programming. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 36 (1/W3). ISPRS Workshop, Hannover.
- [10] Lehner, M. and Gill, R., 1992. Semi-automatic derivation of digital elevation models from stereoscopic 3-line scanner data. ISPRS, 29 (B4), pp. 68–75.
- [11] Otto, G. and Chau, T., 1989. Region growing algorithm for matching of terrain images. Image and vision computing (7) 2, pp. 83–94.
- [12] Scharstein, D. and Szeliski, R. Middlebury stereo vision page. <http://cat.middlebury.edu/stereo>. (accessed 01/2007).
- [13] SpaceImaging. <http://www.spaceimaging.com/>. (accessed 05/2006).