Preprint

Leonardis Ales et al. (eds.) Confluence of Computer Vision and Computer Graphics.
NATO Advanced Research Workshop (ARW) held at University of Ljubljana, Slovenia, 29-31 August 1999. Published 2000 Kluwer Academic Publishers, NATO Sceince Series, pp 201-214.

URBAN SITE MODELS: ACCURATE, DETAILED, RAPID AND INEXPENSIVE

Franz W. Leberl[1], Konrad Karner, Markus Maresch
Vexcel Corporation, Boulder Colorado (USA)

ABSTRACT

Three-dimensional computer models of urban areas have become the latest topic of discussion in photogrammetric circles, although the production of such models has long been a standard offering of photogrammetric data providers. Initial "killer applications" of such data have been defense organizations to support military operations in urban terrain (MOUT), and the Telecom industry in their optimization of certain communications networks that depend on line-of-sight analyses in urban environments. Other applications are trailing behind these trailblazing needs.

The issues today are cost, accuracy, detail, throughput and manageability of data. This paper therefore ignores the discussion of principles of image information extraction. Instead we are concerned with work flow in the production of such data as they concern the actual production of data sets with perhaps half a million buildings of one metropolitan area. The challenge is to create this at modest budgets, at tight schedules, and with verified accuracy and detail.

1. INTRODUCTION

The primary role of terrain data and maps of the land have not changed with the transition from the strictly 2-dimensional representation on paper maps and in the initial digital geographic data bases to the 3-dimensional renderings on computer monitors. The primary purpose was and is the support of the human navigator, planner, user of the land and explorer of its riches.

The creation of such maps and data has always included the third dimension; had it not, the maps and data would have been distorted by the effect of the vertical dimension on the planimetric representation and positioning of objects. Photogrammetry has, since its inception at the turn of this century, seen the world with its 3 dimensions, and always modeled, to this day, its objects in their 3, sometimes even 4 dimensions. The most radical effect of the transition from the 2-d maps and GIS to 3-d data has not been in the modeling of the objects or the creation of the data, but in the storing and rendering of such data.

The idea of 3-d models of urban areas was initially discussed in the context of Military Operations in Urban Terrain (also denoted as "MOUT", see publications in the context of the Image Understanding Workshops, for example Strat, 1997; Lukes, 1999) and in academic environments (see for example Gruber et al., 1995; Fuchs, 1997; Lang and Förstner, 1996; Henricsson et al., 1996). Very recent telecommunications developments in Broadband Wireless Access need line-of-sight

[1] Also at Institute for Computer Graphics and Vision, Technical University Graz, Austria

analyses from one building top to another, and this has created a market for 3-d building models (Figure 1; see Loffet,1996; Siebe, 1996).

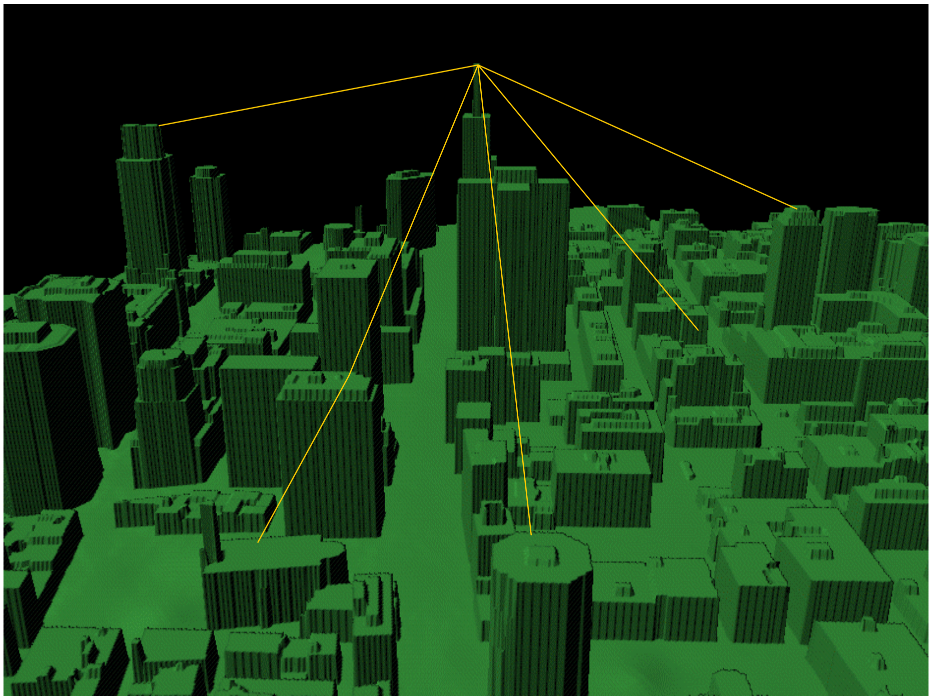

Figure 1: Example of a Telecom-type urban site model for line-of-sight analyses for the so-called Local Multipoint Distribution System LMDS. The example is from downtown Montreal, showing Place-Ville-Marie. The suggested "lines-of-sight" represent broadband wireless access communications.

From a consideration of photogrammetric technology it is self-evident that urban 3-d data have been produced since a long time, using traditional photogrammetric means. In fact, architectural applications were at the root of the creation of the field of photogrammetry, and even the word "photogrammetry" was coined in Berlin by M. Meydenbauer, an architect at the end of the 1800's and user of 3-d building data.

New are thus the computer graphics aspects that have created innovations which now support the use and application of vast quantities of 3-dimensional urban building data (see for example Kofler et al., 1996; Gruber et al., 1997). As a result, the focus of discussions may preferably be on dealing with, and creating an application for, large quantities of such data. Key concepts are work flow, quality control, efficiency and automation to achieve cost advantages, data retrieval and visualization using varying levels of detail and organizational concepts, and trade-offs between geometric detail versus information contained in photographic textures.

2. SOURCE DATA

2.1 Aerial Photography

Aerial photography is the overwhelmingly used source of urban site models (Grün et al., 1995 and 1997; Leberl et al., 1998). This is the straight forward extension of traditional photogrammetry. The scales of standard photography, its overlaps, the use and role of color versus black&white material are at times items for discussions. However, it is the traditional standard of 60% forward and 30% sidelap photography that serve as primary input to urban modeling.

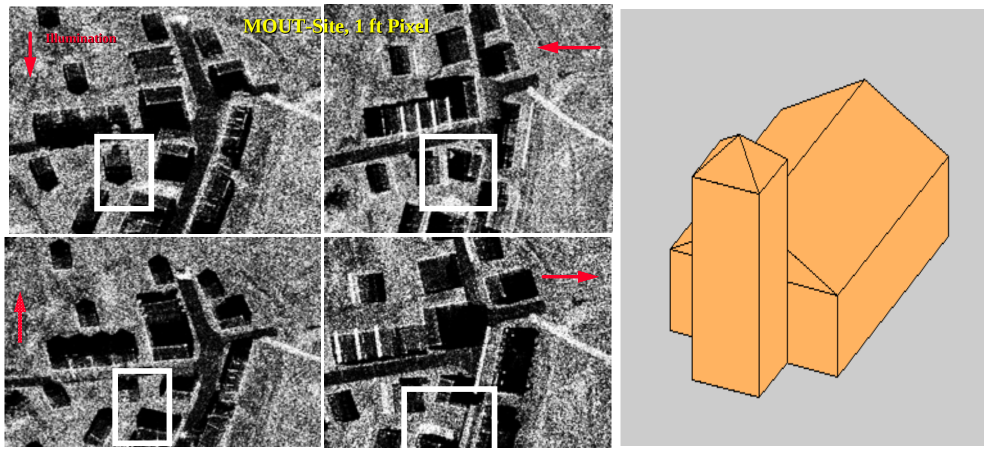

Figure 2: Example of a building reconstruction based on high-resolution SAR imagery, taken from 4 flight lines (left). Shown is the church in the village, as marked in the four images. The SAR images are at a resolution of 30 cm per pixel, and the reconstruction (right) has an uncertainty of about 1 meter.

2.2 Scanning Laser

Laser scanning is currently considered as an alternative to aerial photography (Haala, 1994; Maitre, 1996). There is no other consideration than that of cost that leads one to deviate from photos and propose the use of lasers. A scanning laser and video camera combine into a single pass 3-dimensional data collection and photo texturing system. Generally, however, the advantages of this approach have not yet been demonstrated with sufficient clarity to make it replace the use of imagery. And as soon as aerial photography needs to be collected anyway, one may question the added value received in return for the added cost for the scanning laser data.

2.3 Interferometric Synthetic Aperture Radar

Interferometric radar imagery has begun to get consideration as a primary source for building models. At image resolutions of 30 cm to 10 cm per radar image pixel, and with the benefits of the single pass, autonomous measurements of 3-D shape by means of interferometry, radar may gain in importance for this application (Figure 2). At this time, however, there is hardly any research to develop such autonomous modeling and texturing systems.

2.4 Data Collection from the Ground and Inside a Building

At issue is of course also the detail of buildings both concerning their outside features such as skylights, chimneys, windows, doors or other structures, as well as their inside. This has led to various ideas for efficient source data collections with linear array and digital frame cameras, with indoor imaging and laser scanning systems, again combined with video cameras for autonomous 3D and phototexturing.

3. WORK FLOW ISSUES

3.1 Three Major Processing Function Blocks

A system to build urban site models consists of three major components, each of which consumes about an equal part of the overall resources required:

Aerial photogrammetry subsystem;
Bulk processing for digital canopy elevation modeling;
Manual refinements for building geometry extraction.

All three components are very well established capabilities with a high degree of technological maturity. Many hundreds of commercial companies are currently able technologically to apply these components to create urban site models. At issue is thus not whether one can create such data, but how inexpensively, how accurately, how reliably.

This is where automation is at issue.

However, automation is already a standard in many of the work flow elements. Aerial photography gets scanned and submitted to aerial triangulation, producing the geometric foundation on which the urban site models are built. Stereo matching then follows as a mature technology, and very little gain is being expected from additional automation in the creation of canopy Digital Elevation Models (DEMs).

3.2 Refinement of Building Models

The aspects of actually producing the geometric building models is a "last frontier" at which the opportunity exists for significant gains in productivity and thus in a reduction of cost. However, this still addresses no more than a third of the entire work flow. Manual interaction with the digital sources addresses both the initial creation of building polygons as well as the control of the quality of the resulting data. Many authors have consistently argued that the human operator just be in the loop in this data extraction and quality assurance component of the work flow, and as a result have developed strategies for data extraction that have the computer support the human, instead of having a fully autonomous data processing strategy followed by a manual editing effort.

3.3 Building Details

The inside of buildings, and the detail of building sides, have their separate technological challenges. Ground based imagery is much less well structured as aerial imagery is, and there is no photogrammetric tradition to deal with such ground based data sources. As a result the work flow for the use of such images is less well developed and more opportunities exist to develop and study new methods for data extraction from such images, as will be discussed later.

3.4 Human Stereo versus Monoplotting

Traditionally, photogrammetry has made a strong cases in favor of consistent stereoscopic viewing and measuring on specialized stereo equipment. Recently this view has been challenged, for example by Englert and Gülch (1996) or Spradley and Welch (1999). A much more "scalable" interactive process results when it is based on monocular viewing, but a machine supports the augmentation of the monocularly collected 2-d data into the 3-d domain. Interesting work flow issues result: what really is the human contribution to the image analysis task? Is the man-machine system capable of creating an equally effective product if the human operates monocularly as when the human operates in stereo?

These questions currently have no clear answer, however, in the Urban Site Modeling debate, monocular human interaction has found a valuable role that reduces cost, creates scalability of procedures, and in the process does not compromise the result.

4. THE USE OF IMAGE MATCHING

4.1 Canopy DEM

Matching is the classical approach to create a so-called "canopy DEM". The reflective surface as seen in aerial photography will be modeled by simply finding as dense a set of surface points as is reasonable, and using for that either regular stereo overlaps, or exploiting multiple overlaps involving more than 2 stereo photographs.

4.2 Monocular Man-Machine System

Matching has a second role to play by assigning a third dimension to the 2-dimensional data collected by a human in a monocularly interactive man-machine environment.

By identifying the buildings in one image (2-d polygons) one of the most critical parts of computer vision, the segmentation, is solved by the human operator. This information combines with a rough bald Earth or so-called canopy DEM and the exterior orientation of the aerial images to calculate the 3-d roof line of buildings in the following way:

> *Assuming that the roof line of a building lies in the horizontal plane, we can calculate a mask (see figure 3), to show the areas used for image matching in white areas. The mask is calculated by using the outline of the building and a given inflation and deflation factor. By employing several tests on different data sets, we optimized the inflation and deflation factors, separately for large and small polygons.*
>
> *Using the position of a currently considered building in the image in which it was monocularly defined (thus still a 2-d polygon) and the DEM we get an estimate of the elevation of a building. Furthermore, we use the information of a maximum building height in an area of interest or in a town -- which is easy to find out -- and a tolerance depending on the quality of the DEM, we can restrict the search area for a stereo match by means of the minimum and maximum height and thus the search areas are limited in all images that cover a building (see figure 4).*

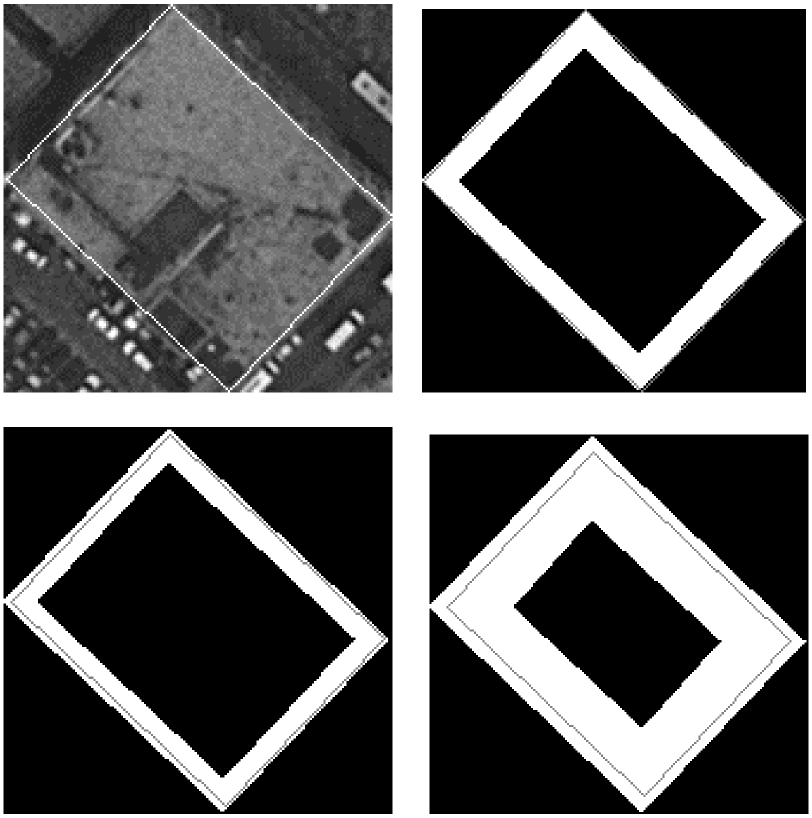

Figure 3: Illustrating the matching areas to assign an elevation to a polygon collected in one image, and the role of inflation and deflation parameters. The manually collected input is shown in the upper left image; the upper right image present one assumed parameter set for the inflation and deflation values, and other parameters are shown in the two lower images.

The process starts with the largest polygons and proceeds down to the smallest ones, using inside/outside tests to decide whether we are considering polygons in an urban or downtown area. This knowledge enables us to restrict the search area significantly and thus decrease the ambiguity of false matches.

Depending on the overlap of the aerial images, polygons are often visible in more than two images. Thus, by employing the above mentioned process to all stereo pairs and a post processing 3-d clustering, one is further able to decrease the ambiguity.

5. SETTING UP LARGE BLOCKS OF IMAGERY

5.1 Classical Aerial Photogrammetry

Mature procedures exist to deal with many 1,000 aerial photographs and set them up geometrically so that individual overlaps can then be exploited for optimum accuracy. Since such images taken from the air follow some very strict rules of flight lines and standardized overlaps, it has become a standard to set up such blocks of images with a great degree of automation.

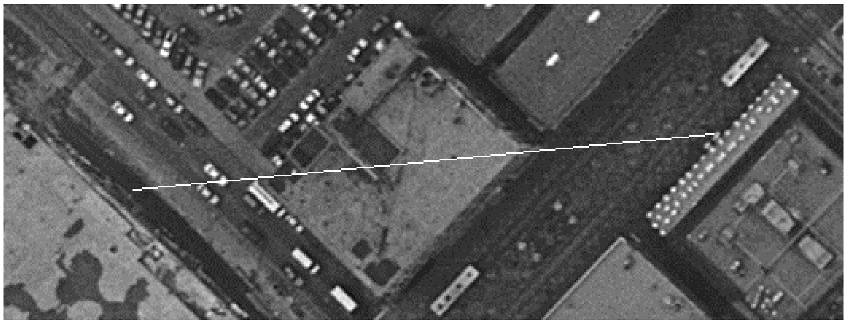

Figure 4: Illustrating the search line in a second image, assuming that we start from the centroid of a manually collected polygon as shown in Figure 3. The length of the line segment is a result of the assumed minimum and maximum building heights.

The determination of relative orientations between aerial images, the identification of common points and finally the calculation of the exterior orientation is a well-known process for aerial images. Furthermore a fairly simple geometry and global positioning systems support these procedures.

5.2 Inside Buildings and Ground Based Images

Modeling the inside of a building may based on 10,000 digital images that need to be geometrically linked together and put into a consistent geometric framework. Because ground based imaging, and the inside of buildings, are subject to vagaries absent in aerial imaging, the task is harder to automate.

An approach has to be chosen for ground based images which are used for façade modeling and extraction of detailed features not visible by aerial images. A generally higher depth variance in images, oblique views, repetitive patterns (windows, man-made structures) and therefore ambiguity make it a more challenging task for automation than is the case in aerial photogrammetry.

We have developed procedures to automate the analytic phase for ground based façade images and are able reduce human effort to about 5 %, compared with the prior fully manual process. Our method is based on 'video-like' image sequences with high overlap (<= 75%), see also Maresch, 1998a. This requirement seems to be justified with the recent dramatic drop in camera and disc storage prices. Some of the frames are used for point transfer only.

The process is based on standard cross correlation based image matching as described in standard texts on image processing, for example by Gonzalez and Woods (1993), and supported by high and low level feature extraction (see Maresch, 1998b, and Smith and Brady, 1997), and relative orientation algorithms to verify and qualify match candidates (Hartley and Richard, 1997; Kraus, 1993).

The simultaneous employment of completely different strategies in most sub-tasks of this process and the use of matching and the calculation of the relative orientation yield a reliable tool. Human effort is reduced to quality control and the throughput for the analytic phase was increased by a factor of 10 to 20 using this method over a purely manual process. About 1 minute of processing time is budgeted per image pair.

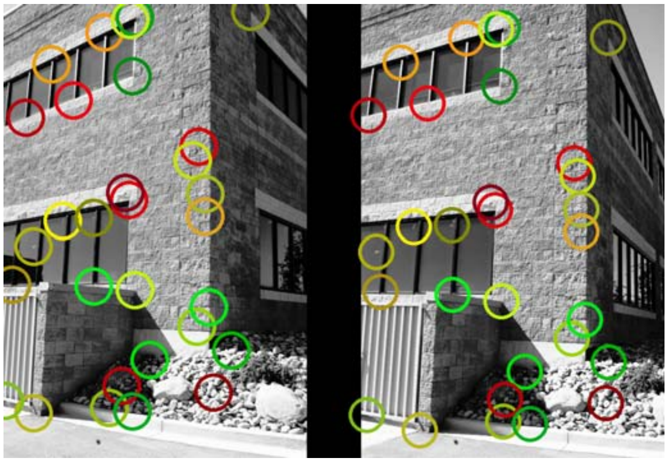

Figure 5: Example of an automatic relative orientation of a pair of overlapping terrestrial photographs using a large number of candidate points in each of two images, and pruning the candidates for correspondence to fit a legitimate relative orientation.

6. EXTRACTING DETAIL OF BUILDING FACADES

So far, building models were simply understood to be “lego-boxes”. This, of course, is entirely satisfactory for many applications that deal with line-of-sight issues such as those in Telecom. However, when buildings themselves are the object of interest, then one is concerned with doors, windows, escape routes, chimneys and so forth. Such detail is costly to collect by hand and efforts are being made to produce such building details with machine support.

The source of such information is ground-based imagery. We are using an approach for modeling 3-d building facades that is similar to the one explained in sub-section 4.2. The human operator again is working monoscopically in one image only and the 3-d information is calculated automatically by image matching employing geometry-based restrictions (e.g. conics) using other images, showing the same part of the object. In the modeling process pre-calculated features like corners, lines, conics and splines are used. The user is able to select and edit these features as well as to define new features, which might not be found in the preprocessing due to weak edges or peculiarities. Thus again, the critical part is the segmentation and it is solved by the human operator by only selecting different kinds of features and combining them to surface facets.

We use well known matching techniques for different kinds of features. The point matching algorithm works similar to the one explained in sub-section 4.2. For line features we are using a slightly modified approach of Schmid and Zisserman (1997). Conics are calculated without calculating cross correlation. We use the approach of Quan (1996), with an additional 3-d clustering, however only if more than 2 images are available, so that one reduces ambiguities. A result is illustrated in Figures 6 and 7.

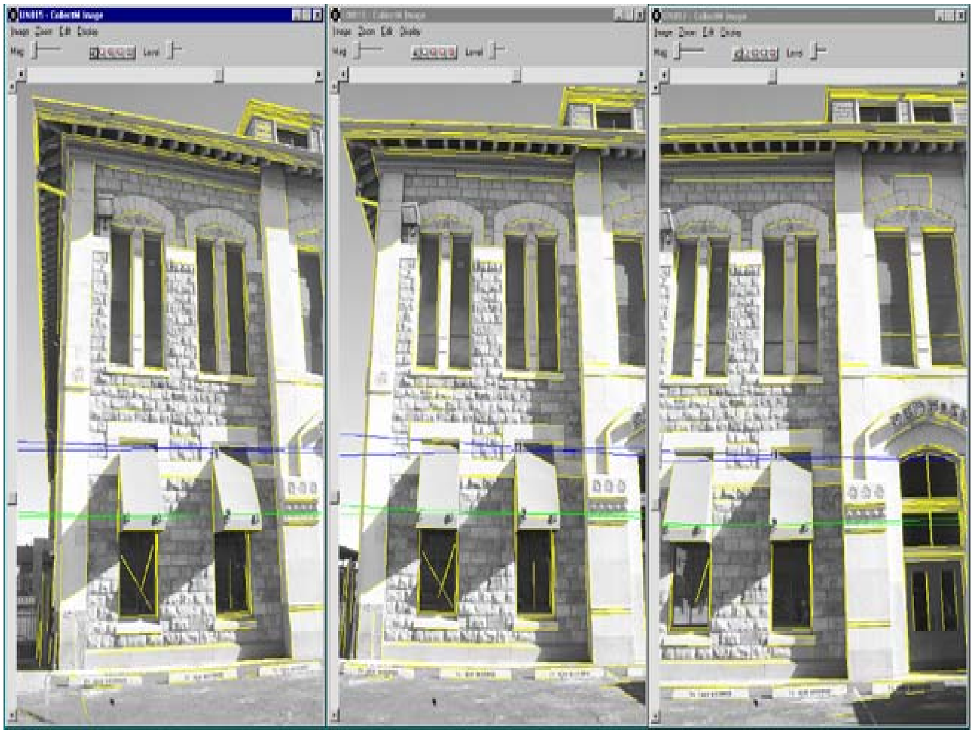

Figure 6: Example of a façade shown in three photographs, in which various features have been found.

7. WHERE WILL THIS LEAD?

The Telecom industry has charged ahead with its requirement for line-of sight-analyses, and has embarked on massive projects to model all major metropolitan areas on the globe. Specifications vary, and such data may not be reusable unless they satisfy specifications for applications other than the planning for hubs of broadband wireless access systems. In Canada, Telecom employs data of all buildings, elsewhere the focus is on business and commercial buildings only. The Canadian concept leads to data sets, say for Montreal, with up to 1,000 buildings per sqkm, and about 0.5 million for the area of interest. Unfortunately, Telecom's interest is only for the densest regions of any city, and therefore its data sets may have large holes in the less densely populated segments of urban areas. As a result, the Telecom-inspired efforts may have to be augmented to be useful for other applications.

There is little doubt that at the current cost of producing such data, the 3-dimensional building data will become as ubiquitous as the 2-d GIS currently is. There are no cost reasons not to go forward with this. A large city like Montreal may have 400,000 buildings scattered over a core area of 400 sqkm. At a cost of US$ 200 per sqkm, it would be a modest expense of US$ 80,000.-- to model all of the buildings of Montreal in 2-d!

Figure 7: Vectors extracted from the overlapping terrestrial photographs in Figure 6. These 3-d vectors have been computed automatically and can now be submitted to a machine-supported manual editing and clean-up procedure.

The texturing of the buildings is another issue. While the aerial photography will automatically, and almost at no additional cost, produce a texture for all horizontal or near-horizontal surfaces, this is not true for the vertical walls of buildings. These textures will need to be separately produced. Once this has been accomplished, the issue may arise concerning the geometric detail of each building versus the information contained in the photographic texture. These issues will move into the foreground once the basic data sets have come into existence.

REFERENCES

Englert R., Gülch E. (1996) *One Eye Stereo System for the Acquisition of Complex 3D Building Description*, GIS, Vol. 4, 1996.

Fuchs, C. (1997) *OEEPE Survey on 3D City Models*, Inst. für Photogrammetrie, Univ. Bonn, Bonn 97.

Gonzalez R., R. Woods (1993) *'Digital Image Processing'*, Chapter: Matching by correlation, pp. 583, Addison Wesley.

Gruber M., S. Meissl, R. Böhm (1995) *Das dreidimensionale digitale Stadtmodell Wien. Erfahrungen aus einer Vorstudie*. VGI (Austrian J. of Surveying and Geoinformation), pp. 29-36.

Gruber M., Kofler M., Leberl F. (1997) *Managing large 3D Urban Data Base Contents supporting Phototexture and Levels of Detail*, Proceedings of the Ascona Workshop '97, Automatic Extraction of Man-Made Objects from Aerial and Space Images, Birkhäuser Verlag, Basel 1997.

Grün A. (1996) *Generierung und Visualisierung von 3D Stadtmodellen*, Proc. IAPR TC 7 Workshop, Graz.

Grün A., O. Kübler, P. Agouris (1995 and 1997) *Extraction of Man-Made Objects from Aerial and Space Images*. Proc. of two meetings held at Ascona, Switzerland. Published by Birkhäuser-Verlag, Basel.

Haala N. (1994) *Building Detection by Fusion of Range and Image Data*, ZPF, 5/1994, Karlsruhe.

Hartley, Richard I. (1997) *'In Defense of the Eight-Point Algorithm'*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 6, June.

Henricsson O., Streilein A., Grün A. (1996) *Automated 3D Reconstruction of Buildings and Visualization of City Models*, Proceedings of the OEEPE Workshop on 3D-City Models, Inst. für Photogrammetrie, Univ. Bonn.

Kofler M., H. Rehatschek, M. Gruber (1996) *A Database for a 3D GIS for Urban Environments Supporting Photo-Realistic Visualization*. Intl. Archives of ISPRS, Vol. XXXI, Part B2, Comm. III, pp. 198-202.

Kraus, K. (1993) *Photogrammetry, Vol I*, Dümmler, Bonn, Chapter: "Common orientation of the two photographs", pp. 121.

Lang F., Förstner W. (1996) *Surface Reconstruction of Man-Made Objects Using Polymorphic Mid Level Features and Generic Scene Knowledge*, ZPF 6/96, Karlsruhe.

Loffet A. (1996) *3D-Models for Telecommunication - Methods and Experiences*, Proceedings of the OEEPE Workshop on 3D-City Models, Inst. für Photogram-metrie, Univ. Bonn.

Leberl F., R. Kalliany, M. Gruber (eds.) (1998) *Man-Made Objects from Aerial and Space Imagery*, Special Issue, ISPRS J. on Photogrammetry and Remote Sensing, Elsevier, Vol 53, No.2.

Lukes G. (ed.) (1999) *Image Understanding Workshop*, 20-23 November 1998 in Monterey, California. Proc. ISBN 1-55860-583-5, Morgan Kaufmann Publishers Inc., San Francisco. 1318 p.

Maitre H. (1996) *Fusion and Optimization*, Proceedings of the 20th Workshop of the ÖAGM/AAPR, Leibnitz 1996.

Maresch M. (1998a) *'The FotoG matcher'*. Internal publication of Vexcel Corp. Boulder, CO.

Maresch M. (1998b) *'The FotoG Feature Extraction'*. Memorandum, Vexcel Corp. Boulder, Colorado

Quan L. (1996) *Conic Reconstruction and Correspondence from two Views*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 18. no. 2, February.

Schmid C., A. Zisserman (1997) *Automatic Line Matching across Views*. Technical Report, University of Oxford.

Siebe E. (1996) *Requirements of 3D-City Stucture Data from the View Point of a Radio Network Service*, Proceedings of the OEEPE Workshop on 3D-City Models, Institut für Photogrammetrie, Univ. Bonn

Smith, S.M. and Brady, J.M. (1997) *'SUSAN - A New Approach to Low Level Image Processing'*, International Journal of Computer Vision, pg. 45-78, vol. 23, no. 1, May.

Spradley H. and R. Welch (1999*) PC-Based Construction and Visualization of 3D Urban Databases*. ASPRS Annual Convention, May 17-21 1999, Portland, Oregon.

Strat T. (ed.) (1997) *Image Understanding Workshop*, 11-14 May 1997 in New Orleans, LA. Proc. ISBN 1-55860-490-1. Morgan Kaufmann Publishers Inc., San Francisco. 1531 p.