

3D-MODELS OF THE HUMAN HABITAT FOR THE INTERNET

Franz Leberl

Institute of Computer Graphics and Vision, Graz University of Technology, Graz, Austria

Michael Gruber

Microsoft Photogrammetry, Graz, Austria

Keywords: Internet, Geo-data, Photogrammetry, 3-dimensional objects.

Abstract: The Internet has inspired an enormous appetite for 3-dimensional geo-data of the urban environment to support location-aware applications. This has in fact become the surprising „killer application“ of such 3-dimensional data. In March 2005, at the occasion of his 50th birthday, Bill Gates went public with his “Virtual Earth Vision” for local search in the Internet and stated: “You’ll be walking around in downtown London and be able to see the shops, the stores, see what the traffic is like. Walk in a shop and navigate the merchandise. Not in the flat, 2D interface that we have on the web today, but in a virtual reality walkthrough.”

The key words are „walk in a shop“. This implies the need for an enormous detail, and the associated computing power, communications bandwidth, miniaturization of computing, increase of storage capacity and in the ability to model the human habitat (the Earth) in great detail in 3 dimensions.

This paper seeks to evangelize the current capabilities of the Virtual Earth system, focuses on the creation of 3D data, but also points to some pieces of new science in the 3D-analysis of overlapping imagery of the human habitat at sub-pixel accuracies

1 FROM 2-DIMENSIONAL NAVIGATION TO 3-DIMENSIONAL SEARCH

The Internet-use of geographic information dates back to the mid-1990s in the form of navigational support for cars and trucks. Digital street maps have made it into car navigation systems. Microsoft offered Streets and Trips, later MapPoint, as stand-alone PC-based solutions for travel planning, and when augmented by GPS, also supporting the navigation application. Subsequently, digital street maps have also appeared on the Web by various vendors to support travel plans and online routing as well as navigation via the Internet. Some of the more prominent global offerings are by MapQuest, in the form of an Internet-enabled MapPoint by Microsoft, as mapping service by Google, Yahoo, Ask and other search services (maps.google.com, maps.yahoo.com, maps.ask.com, maps.live.com). In nearly every industrialized country, one or even multiple regional systems have come into existence,

oftentimes on the basis of an existing telephone directory business. Under <http://www.klicktel.de/routenplaner/> one finds one sample solution for Germany.

These 2-D navigation and route planning systems were soon augmented by an aerial and satellite photography backdrop in the form of so-called ortho-photos. Google may have been the pioneering provider of imagery via its acquisition of Keyhole Inc in 2004 and the subsequent release of Google Earth in 2005. Microsoft released its Virtual Earth website in June 2005, also augmenting MapPoint by aerial imagery, to include bird’s eye aerial coverage as collected by Pictometry Inc. under an exclusive contract with Microsoft.

3-D building models were first introduced into Virtual Earth by Microsoft in November 2006 and grew to a coverage of all major US cities and some cities outside North America. The media response was enormous. Google followed suit by initially providing its own data sets consisting of Lego-type building blocks without photo texture.

Table 1: UltraCam-evolution via an exploitation of advances in CCD-technology. „Level 1“ are raw data from a total of 13 area array CCD-sensors @ 2 byte per pixel; „Level 2“ are the 9 panchromatic image tiles merged into one large format panchromatic image, plus the 4 separate color bands in red-green-blue and near-infrared, geometrically matched to the panchromatic image, but at a geometric resolution that is 3 times reduced. „Level-3“ consists of the 4 high resolution color bands resulting from fusing the panchromatic and color bands („pansharpened“). From the 4 Level-3 component images one can create a true color (red-green-blue) as well as a false color infrared image (red-green-infrared).

ULTRACAM MODEL	CCD Pitch μm	PIXEL ACROSS	PIXEL ALONG	SINGLE-CCD ACROSS	SINGLE-CCD ALONG	RAW MB LEVEL1	NET MB LEVEL 2	NET MB LEVEL 3
D	9	11,500	7,500	3,994	2,662	276	258	690
X	7,2	14,430	9,420	4,992	3,328	432	405	1,087
XP	6	17,310	11,310	5,990	3,994	622	583	1,566



Figure 1: Image formats of the UltraCam-versions D, X und XP. Obviously, a larger format will increase the efficiency of image collection resulting in fewer aircraft flight miles and fewer individual images and data files.

The third dimension reflects the idea of a human experience of our environment – we see a 3D world, not the flat mapping version of the world. As we navigate urban spaces we interact with buildings, building floors, vegetation, billboards, interior spaces and other items defining the world around us. We search in 3D. From a Virtual Earth we move into a digital model of the spaces we live in -- a Virtual Habitat.

2 LOCATIONAL AWARENESS IN THE INTERNET?

One may want to denote the internet-based mapping and search services as “locationally aware”. We see a marriage between geo-data and the Internet with its vast locationally unorganized data repository. “Geodata” have come a long way in the last 2

decennia with a transition from 2-D paper maps to, initially, the 2-D geographical information system GIS. We now see a slow emergence of the third dimension in geo-data production. As a result the GIS transits from 2D data with a third dimension encoded as an attribute to the GIS-elements, to a true 3-D model of the world. The main source of such 3D geo-data is the field of photogrammetry. One may be able to track the evolution via the quadrennial scientific-technical Congresses of the International Society for Photogrammetry and Remote Sensing ISPRS as follows (ISPRS, 2008):

1992 (Baltimore) -- first film scanning systems get introduced, thereby supporting the digital processing of images in photogrammetric geo-data processing, but based on film sources;

1996 (Vienna) – first digital stereo (also called softcopy stereo) systems come online and start the demise of the 100-year optical stereo-technology relying on the human stereo viewing ability;

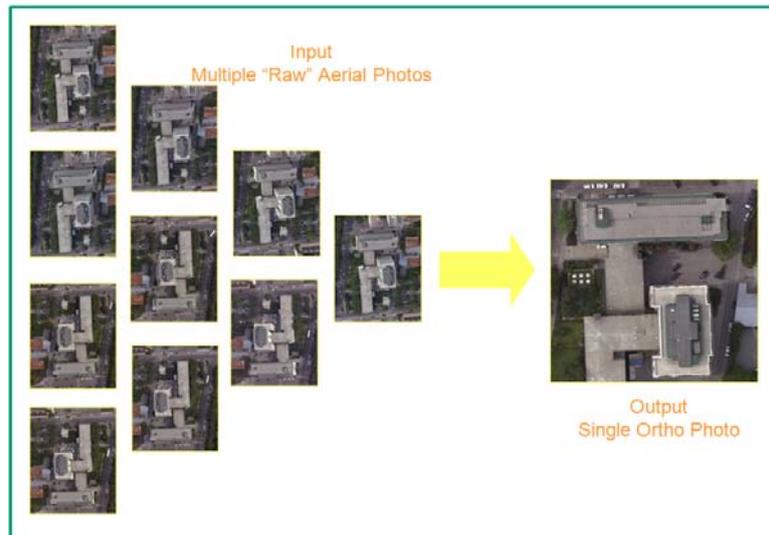


Figure 2: Illustrating an image coverage with 80% forward and 60% side-lap, resulting in a building being imaged 10 times. From these 10 inputs, a single output is being generated in the form of a 2-D ortho-photo.

2000 (Amsterdam) -- the first digital aerial cameras get announced but have yet to get introduced into the application in photogrammetry;

2004 (Istanbul) – first reports of the systematic application of digitally collected imagery and its advantages in reducing cost and increasing quality; the end of the film-based geo-data creation is near;

2008 (Beijing) – this is the Congress celebrating the marriage of the Internet with geo-data in 2 and 3 dimensions.

The Internet changes the geo-data user's paradigm. It moves from the realm of the expert users into everyday live. *At all places* one can interact *with all places* and do so *at all times*. The idea of interacting "*with all places*" may include the galaxy, as demonstrated by Microsoft's „Worldwide Telescope“ and <http://www.google.com/sky/>.

These Internet data systems thereby evolve into a societal force by providing the global geographic knowledge to all of humanity in an easy and intuitive manner. The business model initially relies on advertisement. But we can expect that this model will broaden into computer games, e-commerce, the Internet-of-things, the so-called „Ambient Living“ and others.

3 SIX CHALLENGES TO MEET FOR 3D GEODATA

3.1 Superior Quality in Aerial Imagery

To develop a 3D model of our habitat economically and thus automatically we need to be able to rely on

excellence in aerial image coverage. In the case of Microsoft's Virtual Earth, the source imagery for 3D urban data is being created by the UltraCam digital aerial camera. Table 1 summarizes the numbers of pixels being collected per image by three different camera models. In its most recent implementation, a single image covers 17K by 11K pixels. Figure 1 augments that parameter by a view at the image formats.

“Quality” is being defined by the geometric and radiometric performance of the sensor. Photogrammetric standards require that an image be *geometrically* stable and accurate to within $\pm 2 \mu\text{m}$ across its format. Without such accuracy the sensor would not be considered photogrammetric.

The greater challenge is radiometry. Because automated procedures must rely on image details in both very bright and very dark objects, one develops an insatiable appetite for the number of grey values such a sensor can separately produce. In the UltraCam's case, one achieves a radiometric range in excess of 7,000 grey values or nearly 13 bits per pixel and color channel.

3.2 Intelligent Image Coverage from the Air

If one were to add an image within an image creation mission, one would not add any variable costs. The zero-variable-cost paradigm of digital sensing, without any film, any chemicals to develop the film, any film scanning reduces the costs to just the cost of flying a mission. This is an invitation to increase the redundancy in the number of images

and thereby achieve a greater degree of automation, reduced errors, less manual work. While with film one needed to minimize the number of images collected, to control costs with digital sensing, one can increase the forward overlap within a flight line from the basic stereo case of 60% to perhaps 80% or even 90%, and the side-lap between flight lines from the basic 20% to perhaps 60%, sometimes to 80%. These overlaps eliminate any concerns one might otherwise have had about occlusions due to surfaces and objects hidden behind tall buildings. Figure 2 illustrates how this leads to each building being imaged in 10 separate images, and how these 10 images need to be combined into one single ortho-photo data product.

3.3 Automated 3d Image Analysis

Automation is the key to making a global 3D model of urban spaces feasible; the process was described by Leberl (2007). This begins with an automated *aerial triangulation* of all images of one single project or urban space to be modeled, encompassing perhaps thousands of aerial photographs. The result is a precise reconstruction of all camera positions and attitudes, a geometric marriage of all individual images into a coherent image block, and a non-dense collection of 3D terrain points.

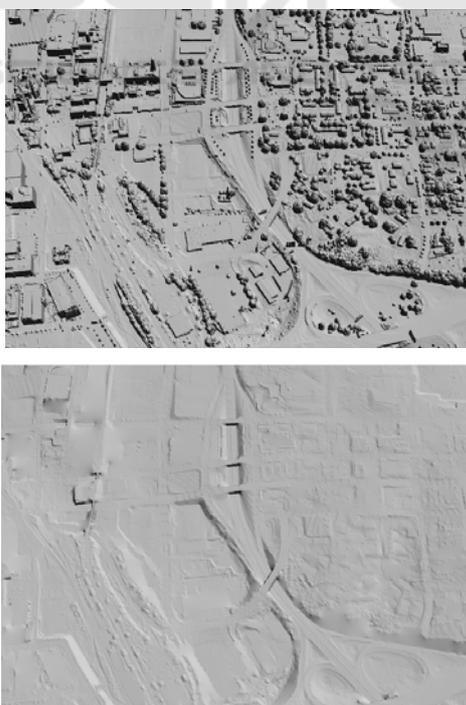


Figure 3: A dense surface model (DSM, above) from a project in Winston-Salem (USA). Below is the “bald Earth” without the buildings and vegetation.

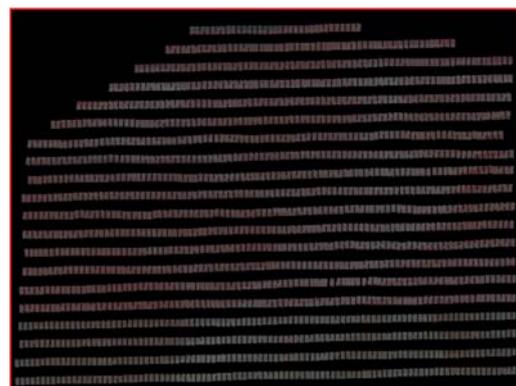


Figure 4: „Dragonfly“ derives from Seadragon (Livelabs, 2008) and is used as part of an automated aerial triangulation system UltraMap. It supports the interaction with thousands of images, each with perhaps 0.5 Gbytes. Changes vis-à-vis Seadragon are the 16 bits per color channel, 4 colors to include infrared, the ability to deal with overlaps in images and interactive point measurements (Reitinger, 2008).

This is being followed by a so-called *dense matching* process to develop a surface model with a posting of terrain points every 2 or 3 pixels, using the 10 or so input images per terrain point and a process developed by Klaus (2007). This often also denoted as *multi-view matching* (as opposed to the traditional stereo matching with just 2 images). The dense surface model is the basis to create the ortho-photo, using the triangulated input photographs, as shown in Figure 2, and using it as the 2D backdrop for the Internet-data system. That ortho-photo is in turn the basis for an *image classification* to separate the terrain objects into buildings, trees, shrubs, circulation spaces, water and grass surfaces.

In the transition to the 3D data products, the dense surface model DSM needs to be separated into the *bald Earth* and the set of vertical objects on top of that bald Earth (see Figure 3). Those vertical objects are now being labeled using the result of the classification.

In spite of automation there is a need to interact with the imagery to review, control quality, make measurements. For a mid-size city like Graz (Austria) with its 150 km², the aerial coverage at 8 cm per pixel results in 3000 images. Such large image numbers and data quantities require clever interaction approaches. Based on SeaDragon, as described by Livelabs (2008), the Dragonfly-system was developed to use 4 instead of only 3 colors, to cope with 16 bits per color channel instead of 8, and to deal with overlaps in imagery. This is integrated into a fully automated aerial triangulation system UltraMap (Gruber and Reitinger, 2008). Figure 4

illustrates the user interface of Dragonfly with the aerial image coverage of an entire city (Reitinger, 2008).

3.4 Taking Advantage of Redundancies

Collecting more images than the minimally needed 2 images per object point, for stereo support, provides many advantages:

- better geometric accuracy (see Figure 5);
- improved automation with fewer catastrophic failures and fewer manual interventions;
- reduced occlusions of urban surfaces;
- fewer gross errors in the automated computations.

This all points in the direction of decreased costs in the image analysis and improved quality of the 3D geo data.

3.5 High Geometric Accuracy

It may surprise that a 3D geo-data system for Internet mapping and search needs to be very accurate. Figure 6 illustrates the effect of a poor dense surface model on an ortho-photo, along the edges of a building, thus along the roof line. Unless the roof line is well defined in 3D, one will suffer from visually disturbing errors with ground texture on the roof and vice-versa. The geometric errors in a photogrammetric 3D system can be characterized in image coordinates by the following component errors:

- Laboratory calibration of the aerial camera..... $\pm 0.5 \mu\text{m}$
- Merging the 9 image tiles into a single image..... $\pm 0.6 \mu\text{m}$
- Field calibration by an actual aerial triangulation, σ_0 $\pm 1.0 \mu\text{m}$

These numbers need to be seen with respect to the physical size of the CCD-pixels in the range between $6 \mu\text{m}$ and $7.2 \mu\text{m}$. And they show that one operates well within a sub-pixel domain.

3.6 Limitless Detail

(a) Human Scale Geo-Data: From “Earth to “Habitat”

Humans do not experience urban spaces from a bird’s eye perspective. We all walk a city, or drive in it, we visit buildings and move in interior spaces. That is exactly the essence of Bill Gates’ March-2005 address that kicked-off Microsoft’s Virtual Earth initiative.

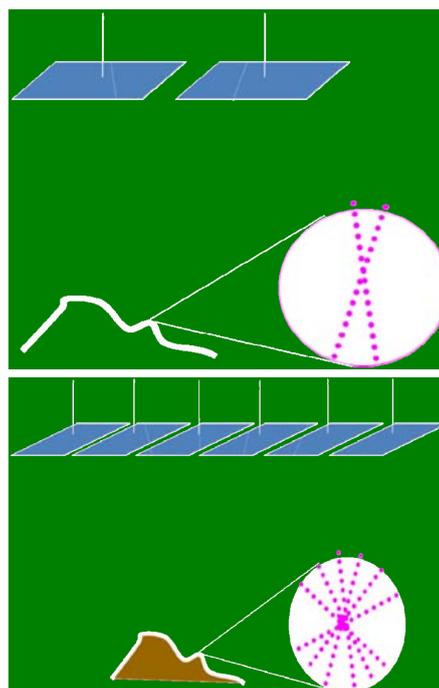


Figure 5: Comparison between the computation of a 3D point from two images (above) and 6 images (below). At 10 cm pixels, the 2-image approach may result in a vertical coordinate error of $\pm 30 \text{ cm}$. With 8 images, this error reduces to $\pm 5 \text{ cm}$ (Gruber and Ladstätter, 2006). The intersection geometry improves as the images are farther apart (compare images 1 and 6 to the right).



Figure 6: Segments of an ortho-photo. Above is the effect of a less well defined DSM, produced with an older DSM-algorithm. Below is the ortho-photo achieved with the DSM from the current state-of-the-art.

We argue that a 3D urban model needs the building facades with their signage, the shop windows and building details such as balconies, the street views with their fire hydrants and bicycle racks, the parking meters, drive ways etc. But we also want public building interiors, courtyards, shopping malls, musea, churches etc.

(b) Streets

Right from the 2005-start of the Virtual Earth project, Microsoft embarked on ambitious vehicle-based image collection campaigns. A technology preview website illustrated how the data were to be used, separately from the aerial imagery. Most recently, this data collection has evolved into a fairly complex enterprise with multiple cameras, lasers, positioning and attitude sensing (Figure 7). The use of the collected data is envisioned in a framework defined by the aerial data so that the street-side imagery is used to refine the building details originally collected from the air.



Figure 7: A Microsoft-Virtual-Earth car with a setup using 12 cameras and 4 laser scanners.

(c) Interior Spaces

We are concerned with public spaces and buildings as defined in item (a) above. This must be based on procedures and workflows for interior 3D modeling. Early examples were shown in the Vienna National Library (Gruber and Sammer, 1995) or the Graz Congress (Gruber, 1997).

4 INVOLVING THE USERS IN SUPPLYING DATA

The premise of urban building details at a level of perhaps 2 cm per pixel, and of all the World's cities, calls for an approach that involves the users of such geo-data.

While we, as citizen-users, may not be experts in geo-data, we certainly are experts of our

neighborhood. Goodchild (2008) and many others are advocating the idea of the neo-geographer as a person contributing local mapping data to an Internet-based geo-data system in analogy to Wikipedia. An interesting development is Photosynth by Microsoft Research (Livelabs, 2008). The capability is based on the ability to perform an (aerial) triangulation with uncalibrated photographs and without any knowledge of a focal length.

Photosynth has been released as a global service in August 2008 (<http://photosynth.net/>). Figures 8 and 9 illustrate their basic idea with an example of the Tummelplatz (square) in Graz. Note how the sequential viewing of 2D images create a 3-D "illusion".

Photosynth is now part of the Virtual Earth organization within Microsoft. A related development exists at the University of Washington (2008) under the name Phototourism (Snavely, Seitz et al., 2008; Snavely, Garg et al., 2008)

5 AN EXABYTE OF LOCATIONAL DATA

We have suggested that the aerial imagery for Virtual Earth is at a resolution of 10 to 15 cm. The higher the resolution, the easier it will be to obtain the required delineation of roof lines and human scale objects from the air.

Street-level imagery is expected to be at a resolution of perhaps 2 cm, to ensure that all street signs can be read and interpreted. And indoor data would by necessity be at an even higher resolution, perhaps as high as 0.5 cm. How much data will a complete 3D model of the World consist of?

Given the Earth's land masses with about 150 million km², a basic 15 cm image backdrop in 2D would consist of more than 20 PB (peta-bytes), in color pixels and as an output. But if we operate with a 10-times redundancy, these 20 PB will really require one to deal with 200 PB at the input side. Currently, a majority of uses are 2-dimensional. Figure 10 is a representative application to queries about real estate values in the USA.

Street-side data also need to get collected with redundancy. One may want to make assumptions about the mileage of streets along which the images get collected, per city, and the number of cities. Finally, one will want to add the indoor data for a subset of the buildings in a city and the 10-times redundancy in indoor data.

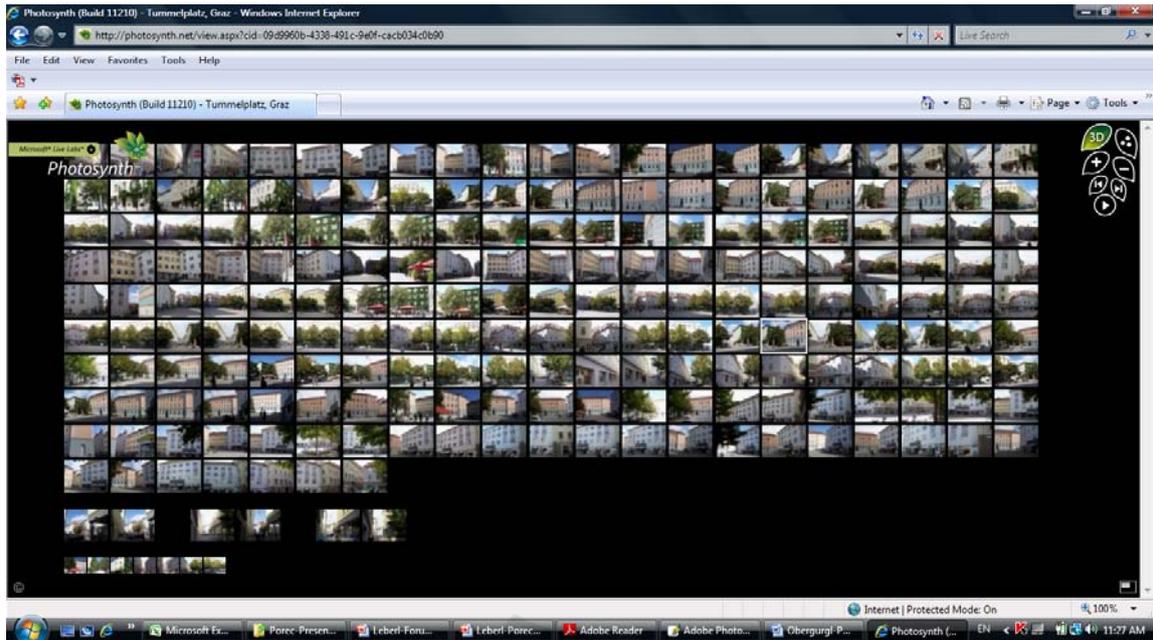


Figure 8: Screenshot. User-images taken in a city square. These images are being loaded into Microsoft's system „Photosynth“ and get oriented automatically without a known focal length or image calibration. Such imagery is then available to Virtual Earth. This will augment the detail in the basic Virtual Earth system. The example is from Tummelplatz in Graz, Austria, with ~ 200 photographs, a vast excess beyond what is minimally required.

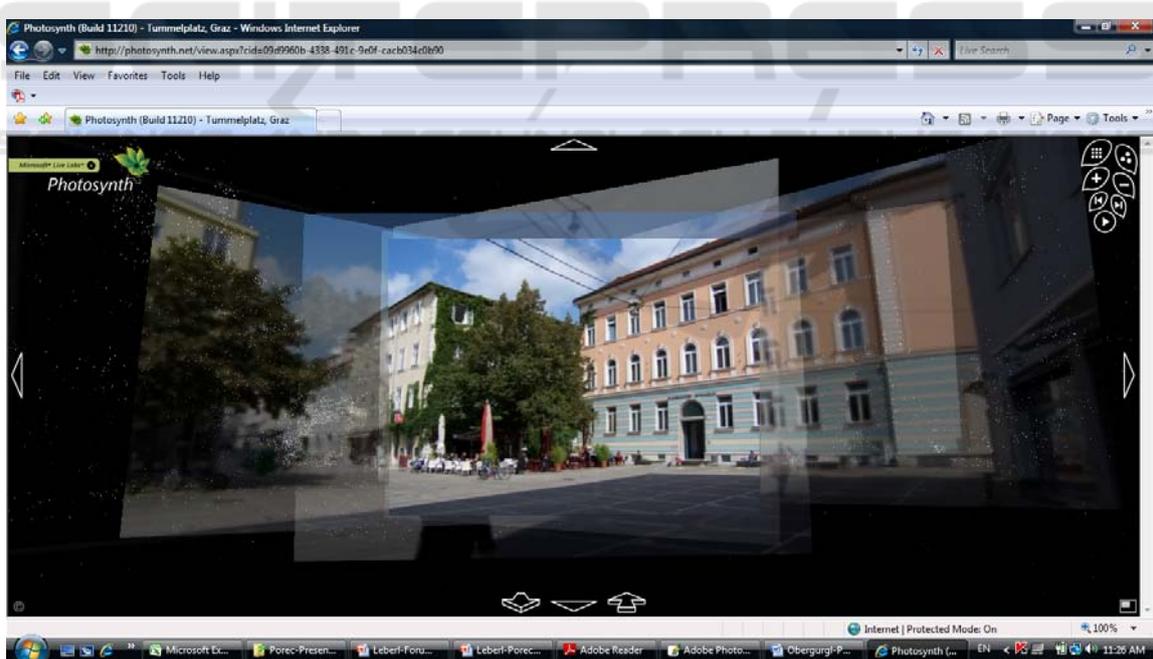


Figure 9: Screenshot. The ~ 200 images from Figure 8 are now available in a 3D coordinate system and one can navigate the image data set as if we were moving in 3D space. The 3D visual impression results from viewing individual 2D images, but oriented in 3D.

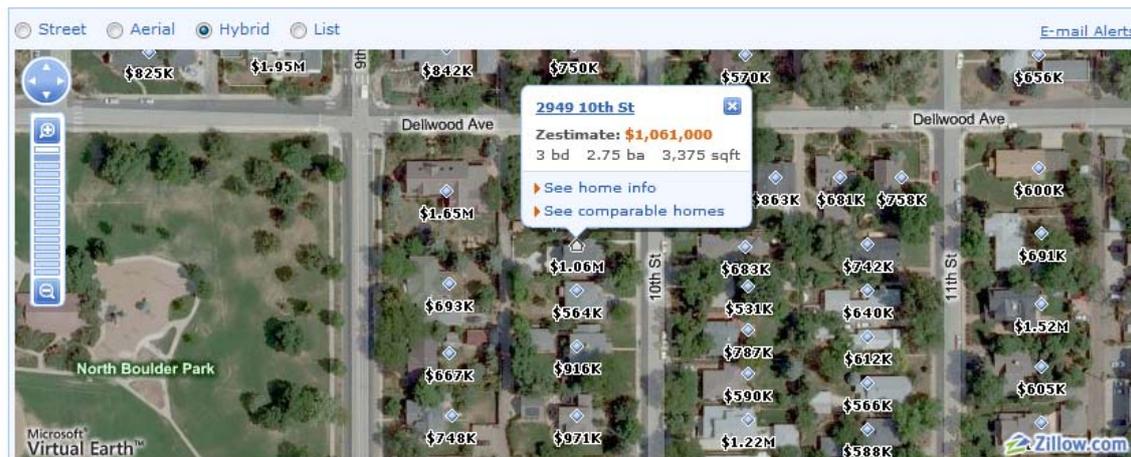


Figure 10: Example of a 2D application of the Virtual Earth system by www.zillow.com. It presents the values of real estate properties, based on property tax records in the USA. One enters with an address and the system returns the value of that address and of its surroundings.

These data sets, whether contributed partly by users, say 200 images of a single square in Graz, or systematically collected by the provider of the 3D system, will exceed 1 EB (exa-bytes).

6 ISSUES AND RESEARCH FOCUS

The creation of 3D geo-data has certainly been a significant focus of the developmental work within the Virtual Earth initiative. Today, an end-to-end workflow exists to produce complete 3D models of every building of a city at a rate of perhaps 300 cities per year.

Of course this has been built on innovations in the automated analysis of blocks of highly overlapping aerial imagery, as described earlier. However, to meet the goal of a high resolution 3D model of the human habitat at the level of human scale objects, considerable additional innovation is needed addressing street level source data, even indoor data. Once a 3D model exists, its presentation on a computer monitor over the Internet is a matter of “visualization”. While today the 3D models are based on relatively crude geometric models of buildings with photographic texture, one should expect that this will evolve into much finer detail in geometry and an interpretation of the image content. Current research work along those lines in our own teams in Graz is being documented by Kluckner et al. (2009); Ischara et al. (2009), Schall and Schmalstieg (2009) and Ladstätter (2009).

Interaction with true 3D models of urban spaces requires some computer power in the user’s hands. This is currently still a significant limitation. Replacing the 3D-models by a clever approach presenting just 2D imagery, as in Photosynth, is a work-around until such time that users do have sufficient power to deal with 3D. This 2D approach as a “3D-look-alike” versus true 3D is a topic of current relevance.

Keeping the geo-data current, and finding smart ways of defining areas of change, is another topic of interest in this new world of locationally aware Internet applications.

ACKNOWLEDGEMENTS

The material for this paper was developed through the period 2006 to 2008 by various people at the Institute for Computer Graphics and Vision at Graz University of Technology, especially the 12 members of the Virtual Habitat Group, and by the team at Microsoft-Graz. We thank the members of those two teams.

REFERENCES

- Goodchild M., 2008. *Assertion and authority: the science of user-generated geographic content*. Proceedings of the Colloquium for Andrew U. Frank's 60th Birthday. GeoInfo 39. Department of Geoinformation and Cartography, Vienna University of Technology.
- Gruber M. and R. Ladstätter, 2006. *Geometric issues of the digital large format aerial camera UltraCamD*.

- International Calibration and Orientation Workshop EuroCOW 2006, Proceedings, 25-27 Jan. 2006, Castelldefels, Spain.
- Gruber M., Reitinger B., 2008. *UltraCamX and a new way of photogrammetric processing*. Proceedings of the ASPRS Annual Conference 2008, Portland 2008.
- Gruber M., Sammer P., 1995. *Modeling the Great Hall of the Austrian National Library*. International Journal of Geomatics 9/95, Lemmer 1995.
- Gruber M., 1997. „*Ein System zur umfassenden Erstellung und Nutzung dreidimensionaler Stadtmodelle*“, Dissertation, Graz University of Technology, 1997.
- Irschara A. H. Bischof, F. Leberl, 2009. *Kollaborative 3D Rekonstruktion von urbanen Gebieten*, in 15. Intern. Geodätische Woche Obergurgl 2009, Wichmann – Heidelberg.
- ISPRS, 2008. <http://www.isprs.org/congresses/beijing2008/Default.aspx>
- Klaus A., 2007. *Object Reconstruction from Image Sequences*, Dissertation, Graz University of Technology, 2007.
- Gluckner S., Georg Pacher, H. Bischof, F. Leberl, 2009. *Objekterkennung in Luftbildern mit Methoden der Computer Vision durch kombinierte Verwendung von Redundanz, Farb- und Höheninformation*, in 15. Internationale Geodätische Woche Obergurgl 2009, Wichmann – Heidelberg.
- Ladstädter R., 2009. *Untersuchungen zur geometrischen Genauigkeit der UltraCamD/X*, in 15. Internationale Geodätische Woche Obergurgl 2009, Wichmann – Heidelberg.
- Leberl F., 2007. *Die automatische Photogrammetrie für das Microsoft Virtual EarthSystem*. in 14. Internationale Geodätische Woche Obergurgl 2007, Wichmann – Heidelberg, S. 200 – 208.
- LiveLabs, 2008. *Seadragon*, Microsoft Live Labs, 2008. <http://livelabs.com/seadragon>.
- Reitinger B., 2008. *Interactive Visualization of Huge Aerial Image Datasets*; International Archive for Photogrammetry and Remote Sensing, Volume XXXVII, Beijing, 2008.
- Schall G., D. Schmalstieg, 2009. *Einsatz von mixed reality in der Mobilen Leitungsauskunft*, in 15. Intern. Geodätische Woche Obergurgl 2009, Wichmann – Heidelberg.
- Snively N., S. M. Seitz, and R. Szeliski, 2008. *odeling the world from Internet photo collections*. International Journal of Computer Vision, 80(2):189-210, November 2008.
- Snively N., Rahul Garg, Steven M. Seitz, and Richard Szeliski, 2008. *Finding Paths through the World's Photos*. ACM Transactions on Graphics (SIGGRAPH 2008).
- University of Washington, 2008. <http://phototour.cs.washington.edu/>

BRIEF BIOGRAPHY

Franz W. Leberl received his degrees from Vienna University of Technology (Dipl.-Ing., 1967; Dr.techn., 1972). Worked in the Netherlands, California, Minnesota, Colorado and Austria. Today he is a chaired professor of Computer Science at Graz University of Technology.

As a business man, formed Vexcel Corporation in Boulder (Colorado, 1985) and Vexcel Imaging GmbH (Austria, 1993, manufacturer of the UltraCam Digital Large Format Aerial Camera, www.microsoft.com/ultracam). As a research manager, he was CEO of the Austrian Research Centers (1996-1998) with 1000 employees. In 1980 founded the “Institute for Digital Image Processing” at Joanneum Research in Graz, Austria.

His current outlook on life is defined by the sale of Vexcel Corp. and Vexcel Imaging GmbH to Microsoft Corp. (USA) in mid-2006. This resulted in a position as a Director of Microsoft Virtual Earth. Since completion of that assignment in November 2007 returned full-time to academia and now serve as Dean of Computer Science at Graz University of Technology (2008-2011).