The 12th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST (2011) *Short and Project Papers*
M. Dellepiane, F. Niccolucci, S. Pena Serna, H. Rushmeier, and L. Van Gool (Editors)

The V-City Project

J. Himmelstein[1] O. Balet[1] F. Ganovelli[2] E. Gobbetti[3],

M. Specht[5] P. Mueller[5] C. Engels[4] L. van Gool[4] J.B. de la Rivière[6] A. Cavazzini[7]

[1] Diginext, France [2] Visual Computing, ISTI-CNR, Italy [3] Visual Computing, CRS4, Italy
[4] KU Leuven, Belgium [5] Procedural, Inc., Switzerland [6] Immersion SAS, France [7] BlomCGR, Italy

Abstract

3D geoinformatics have entered the digital age, hesitantly in some areas, and rampantly in others. Google Earth and Microsoft Virtual Earth are household names. However, these projects are limited to textured 3D landscapes, aerial 2D images and a few boxy building envelopes. The V-City project is a European research initiative to surpass these limitations, and create a system for intuitively exploring large urban areas with a high degree of detail. Bringing together technologies from geoinformatics, virtual reality, computer graphics, and computer vision, the system constructs detailed 3D city models from geopositioned aerial images and building footprints. For networked browsing, city models are compressed and streamed for interactive viewing of entire landscapes. A unique tactile table has also been developed to let multiple users visualize the same city model in stereo 3D, and interact with it simultaneously using hand gestures.

Categories and Subject Descriptors (according to ACM CCS): H.5.1 [HCI]: Multimedia Information Systems—

1. Introduction

In the cultural heritage (CH) domain, virtual reality is recognized for its ability to transport users directly within a virtual environment, including 3D reconstructions of modern cities. At the same time, the data available about cities is both more numerous and of higher quality, including building footprints, elevation models, and both satellite and aerial imagery. However, high-quality reconstructions can typically only be achieved manually, and with great effort. Visualizing entire cities at interactive rates, especially from a bird's-eye view, is also a challenge. Finally, both virtual reality and GIS tools are usually designed for use by individuals instead of groups. In contrast, many cultural heritage researchers work collaboratively, and promoting their work involves showing their results to large numbers of people.

The V-City research project[†] has been funded by the European Commission to address these challenges by enabling automatic reconstructions of entire cities, and visualization of these reconstructions in a collaborative manner, either via the web browser or through a tactile and stereoscopic table for multiple users. This short project paper provides an overview the results obtained.

2. Related Work

Our system combines and extends state-of-the-art solutions for reconstruction, visualization and exploration of detailed large-scale urban environments. Covering related work in all of these areas is beyond the scope of this short paper. We refer the reader to recent surveys on modeling urban spaces [VAW*10], massive model visualization [eYGKM08], and individual and group support in tabletop interaction [MTNP*10] for more details. Concerning system integration efforts, there are at least two major systems described in the recent literature that aim to support reconstruction, navigation and analysis of 3D cities. The Autodesk LandXplorer system was used to create a virtual model of modern Berlin [DKL*06]. It contained components for automatic and manual building reconstruction based on aerial photography, visualization on a large-screen projector, and editing 3D landmarks. There was no focus on collaborative or immersive interfaces, however. MapCube, the automatically constructed 3D models of ma-

---

[†] `vcity.diginext.fr`

DOI: 10.2312/PE/VAST/VAST11S/057-060

jor Japanese cities, has a virtual reality program called UrbanViewer [TSSS03]. The authors also developed dedicated touch screens, though it was unclear if this allowed collaborative work.

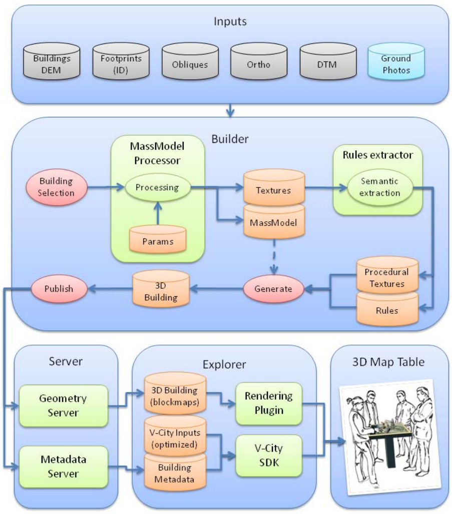

Figure 1: *Typical workflow of the V-City system. The Builder accepts multiple data sources as input and transforms them into textured 3D buildings. These are published on the Server, and downloaded by the Explorer. Multiple users can simultaneously interact and visualize using the Map Table.*

3. Major steps of system workflow

Since the users who reconstruct cities may have different technical backgrounds from those who analyze the results, we chose to segment the solution in two functionality sets: the Builder and the Explorer (see Fig. 1). The major steps of the system workflow are the following:

- Gathering and cataloging geopositioned aerial images
- Automatic reconstruction of 3D buildings from multiple sources
- Visualization of large and detailed urban environments at interactive rates
- Navigation and interaction within the urban environment
- Collaborative exploration techniques in stereoscopic 3D

Gathering and cataloging geopositioned aerial images. The Blom Group has the largest and most versatile aircraft fleets in Europe, which enables them to meet a variety of needs with respect to accuracy, altitude, resolution, medium and area. Their resources include pressurized cabins enabling projects to be undertaken from both high and low altitude. All aircraft are equipped with survey cameras (digital and analogue), laser-scanners (LiDAR) and hyperspectral scanners, providing the ability to acquire imagery in both the visible and non-visible parts of the light spectrum. The LiDAR can capture elevation as accurate as 1.5cm in a variety of environments, including urban areas with many obstacles. The large format digital and oblique cameras produce superior quality images, with a ground pixel resolution below 5 cm. The cameras record five different radiometric channels simultaneously; B/W, R, G, B and IR. All digital cameras are mounted in gyro stabilized camera mounts, equipped with Integrated Position Systems and inertial measurements units (GPS / INS).

Figure 2: *Fleet of BlomCGR in Parma, Italy.*

Automatic reconstruction of 3D buildings from multiple sources. The user first imports georeferenced data from various sources into one scene and then selects the footprint of the building she wishes to reconstruct. The position, shape, and metadata (e.g., building height) associated with the building are used in conjunction with spatially overlapping aerial images and LIDAR points to automatically extrude the footprint, add a roof mesh, and project textures to obtain a textured mass model.

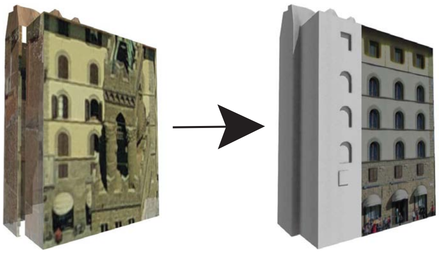

Figure 3: *Procedural refinement. The textured mass model's front facade is largely occluded. With the aid of a ground-based image, shape grammar rules can be extracted and used to generate a detailed 3D model.*

The textured mass model is a first representation of the building. Depending on the quality of the data and occluding objects (e.g. buildings and vegetation), the textured facades can already be quite accurate. However, there is a lack of 3D detail and often the image resolution is limited or only the upper parts of the facades are visible. The quality of the building models can be improved by adding ground-based images and replacing the flat facades with procedurally generated 3D models (s. Fig. 3). These manual steps are facilitated by the semi-automatic *Facade Wizard*. With its aid, the user defines a subdivision scheme for the facade.

The output is a set of CGA (*Computer Generated Architecture* [MWH*06]) rules which are used to generate a textured 3D model of the facade. The rules encode the exact positions of floors, windows, and other semantic data about the facade. Finally, the 3D building model is exported to a Collada file (plus textures). Its metadata and position is written to an XML file, which is included in a compressed archive that is sent to the Server via HTTP.

Visualization of large and detailed urban environments at interactive rates. The end result of automatic reconstruction and procedural enhancement is a large dataset, with lots of geometric and texture detail. Transmitting over the network and rendering such models is a very hard problem. In V-City, we choose to render landmark buildings using classic LODs only at closeup distances, and to revert to a compact, unified, and approximate LOD structure for the overall city. At the core of our approach is the BlockMap data structure [DCG*09], originally introduced for encoding coarse representations of blocks of buildings to be used as direction-independent impostors when rendering far-away city blocks [CDBG*07]. We have extended the BlockMap representation to support sloped surfaces and input-sensitive sampling of color, introduced a novel sampling strategy for building accurate BlockMaps, and improved the expressiveness of urban model rendering by finding an efficient method for calculating and integrating an ambient occlusion term. At run-time, we support networked browsing of very large urban environments through a client-server architecture enabling multiple clients to explore city models stored on a remote server. Since the BlockMap rendering algorithm is GPU bound [DCG*09], we were able to also implement it effectively using scripting languages on OpenGL-enabled web browsers [DBPGS10].

Navigation and interaction within the urban environment. From talking with end-users, we found that two different viewpoints were necessary: an aerial (or bird's-eye view) and ground (or "walking" view). We offer different ways to control the navigation on each device (e.g. mouse, 3D mouse, and tactile interfaces). In the "walker" mode of navigation, the user explores the city in much the same way as they might control a first-person video game. This mode is also ideal to visit indoor environments. Users can import their own custom data (rasters, vectors, elevation, and 3D models) as well as connect to web services (WFS, WMS, etc.). As well as visualizing the result with high-quality rendering including ambient occlusion and shadow effects, we provide tools for geolocation, measurement, altitude profiles, annotations, and exporting image and video.

Collaborative exploration techniques in stereoscopic 3D. Compared to a physical mockup, a digital 3D model of a city is highly flexible and permits numerous operations, but may lack certain properties that facilitate collaboration. Multitouch tables offer an innovative form factor that favors group collaboration as well as direct and intuitive tactile interaction with the digital content. However, visualizing 3D models on a tabletop requires specific hardware, as a perspective projection displayed on a standard multitouch table would appear upside down for the users standing on the opposite side, and would not be able to take advantage of stereo visualization. Likewise, interacting with a 3D city with 2D multiuser tactile input requires specific interaction techniques. The hardware and software solutions we developed are illustrated in Fig. 5. This table displays two different stereo viewpoints of the same 3D mockup simultaneously, one for each of the two users standing on the sides. Head tracking is supported for both users, to bring the experience as close as possible to the observation of a physical mockup. Specific interaction techniques have been developed, from the bird's eye view navigation above the 3D city, which takes advantage of the typical rotation / translation / scale multitouch gestures, to the various measurements operations, which take advantage of bimanual input to allow combining navigation and measurement.

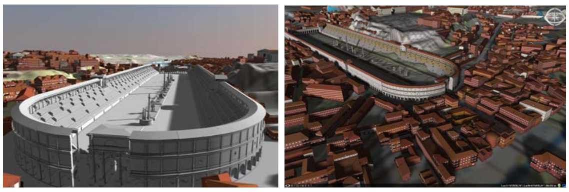

Figure 4: *Two perspectives on the Circus Maximus 3D reconstruction (courtesy of Archeovision).*

Figure 5: *The concept (left) and the reality (right) of two users viewing a single 3D mockup and manipulating the 3D environment using dedicated interaction techniques.*

4. User Evaluation

We tested the proposed system with 32 users, coming from a variety of CH, GIS, 3D, archaeological, and engineering backgrounds. Using a short questionnaire, we evaluated their expertise level regarding GIS, 3D, and programming. We then guided them through typical tasks for which the system was designed for, evaluated their satisfaction with these tasks, and monitored their comments. We chose several data sets for users to test: Toulouse, Florence, and ancient Rome. Toulouse was constructed with aerial photography and building footprints augmented by elevation data assigned to several roof sections, planes of which were then estimated using computer vision techniques. Florence was also constructed with aerial photography, but used LIDAR data directly for elevation information. Ancient Rome, meanwhile, was constructed using procedural methods, based on information provided by archaeologists for the RomeReborn project [DMU*09]. For the Toulouse dataset, we concentrated on the Canal du Midi region, for which we used 146 images at $4008x2672$ resolution. There were 1722 building footprints divided into 7805 sections. The entire city consists of over $117K$ building sections. The Florence dataset consists of $18.5K$ building footprints, 671 oblique and ortho images at $4872x3248$ resolution, a digital terrain model of $663K$ points, and a digital elevation model of $11M$ points collected from a LIDAR scanner.

Builder The main feature of the Builder, the automated reconstruction, was seen as a major advancement beyond the state-of-the-art, both in quality and in ease of use. Users also found the procedural reconstruction (performed with the Rome dataset) pleasantly fast, but remarked that it required more specialized knowledge than the automated reconstruction, especially for generating new rules. The visual quality indicators were also appreciated, and users requested that even more detail be displayed, such as texture quality. Users also wished to be able to incorporate multiple images, including ground photos, more easily.

Explorer The visual quality (aided by the ambient occlusion and shadow effects) were the most appreciated features of the Explorer. Users also enjoyed the ability to navigate from city to city, a typical feature of a globe-viewer, but also within a city and even within a building (such as the reconstructed Circus Maximus in Rome). Users wished for better Internet connectivity for accessing their data from the web as well as to use the Explorer within a web browser.

Map Table The Map Table was by far the most remarked component of the V-City project. Thanks to the democratization of tactile devices, users quickly found how to manipulate it. The novel inclusion of multi-user stereoscopic view was also highly appreciated. Users found the head-tracking feature brought the virtual model closer to its physical counterpart, although at extreme angles they wished for an additional tilt control. There are of course limitations of using the Table beyond two users, and in future work we plan on adding a vertical screen to allow groups to oversee what the users are manipulating, even if this view is not in 3D.

5. Conclusion

This paper has presented the V-City project and its intermediate results. We proposed an integrated approach that reconstructs detailed city models from a variety of data sources, including aerial photography, and publishes them so that they can be viewed over the web or on a unique tactile table that encourages collaborative work. We have discussed the solutions to technical challenges in visualization, interaction, and computer vision that we have devised. We also presented our user testing procedure and its preliminary results, which has given us a number of possibilities for future work. In general, we are convinced that a tight integration is valued by users who are normally obligated to apply multiple complex steps in order to get their data from a raw format to a viewable and workable form. We are currently evaluating the system with a larger variety of users.

Acknowledgments. This research is partially supported by the European project V-City (reference ICT-231199-V-CITY).

References

[CDBG*07] Cignoni P., Di Benedetto M., Ganovelli F., Gobbetti E., Marton F., Scopigno R.: Ray-casted blockmaps for large urban models visualization. *Computer Graphics Forum 26*, 3 (2007), 405–413. 3

[DBPGS10] Di Benedetto M., Ponchio F., Ganovelli F., Scopigno R.: SpiderGL: a javascript 3d graphics library for next-generation www. In *Proc. Web3D* (2010), ACM, pp. 165–174. 3

[DCG*09] Di Benedetto M., Cignoni P., Ganovelli F., Gobbetti E., Marton F., Scopigno R.: Interactive remote exploration of massive cityscapes. In *Proc. VAST* (October 2009), pp. 9–16. 3

[DKL*06] Döllner J., Kolbe T. H., Liecke F., Sgouros T., Teichmann K.: The virtual 3D city model of Berlin: Managing, integrating, and communicating complex urban information. In *Proc. UDMS* (2006), pp. 15–17. 1

[DMU*09] Dylla K., Müller P., Ulmer A., Haegler S., Frischer B.: Rome reborn 2.0: A framework for virtual city reconstruction using procedural modeling techniques. In *Proc. CAA* (2009). 4

[eYGKM08] eui Yoon S., Gobbetti E., Kasik D., Manocha D.: *Real-time Massive Model Rendering*. Synthesis Lectures on Computer Graphics and Animation. Morgan and Claypool, August 2008. 1

[MTNP*10] Müller-Tomfelde C., Nacenta M. A., Pinelle D., Gutwin C., Mandryk R.: Individual and group support in tabletop interaction techniques. In *Tabletops - Horizontal Interactive Displays*, Mı̈ller-Tomfelde C., (Ed.), Human-Computer Interaction Series. Springer London, 2010, pp. 303–333. 1

[MWH*06] Müller P., Wonka P., Haegler S., Ulmer A., Van Gool L.: Procedural modeling of buildings. *ACM Transactions on Graphics (TOG) 25*, 3 (2006), 614–623. 2

[TSSS03] Takase Y., Sho N., Sone A., Shimiya K.: Automatic generation of 3d city models and related applications. *Archives XXXIV*, Table 1 (2003), 113–120. 1

[VAW*10] Vanegas C. a., Aliaga D. G., Wonka P., Müller P., Waddell P., Watson B.: Modelling the Appearance and Behaviour of Urban Spaces. *Computer Graphics Forum 29*, 1 (Mar. 2010), 25–42. 1