Virtual Building Model Construction from Real Images

Jorge Luis Reyes Ortiz – s0783280@inf.ed.ac.uk
School of Informatics, University of Edinburgh

In the last years an increasing interest in the virtual reconstruction of architectural scenes has produced new and useful computer graphics and computer vision applications. In this document the state of the art of this field is presented and the most relevant techniques are depicted.

These techniques consist basically in the reconstruction of virtual 3D models based on real data acquisition from buildings (i.e. photographs, range images and video). Afterwards, the sensor input is processed by using a reconstruction algorithm that estimates depths of the elements in the scenes and incorporates techniques such as epipolar geometry [1], triangulation [11], structure from motion [2], etc. Once the model is obtained a rendering step creates a virtual 3D representation of the environment including both geometry and texture of the objects. Refer to Figure 1 for an example of the process pipeline.

Virtual modelling methodologies have been implemented at different scales, starting from indoors to entire cities. For instance, in the accurate reconstruction of archaeological and cultural heritage sites in order to provide interactive visualization, preserve documentation and potential studies of their evolution through time. Another interesting application is the use of advanced GPS navigation assistants which exploit 3D models of urban areas to guide the drivers along the roads. (See section 2 for other examples).

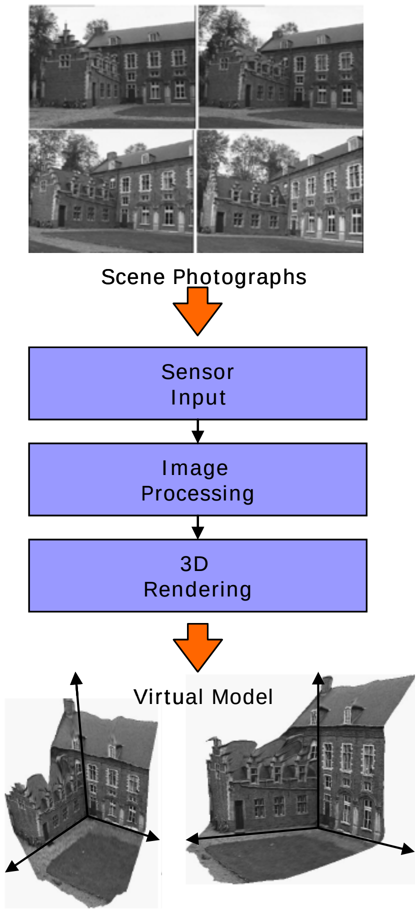
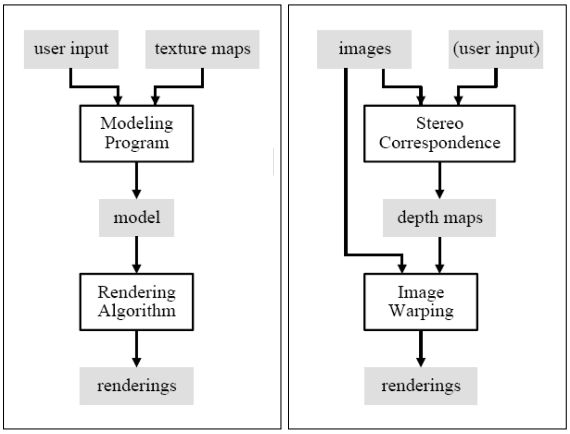


Figure 1. 3D scene virtual Modelling. Basic process pipeline. Pictures from [3].

Contents

1 Methodologies

1.1 Photographs



(a) (b)

Figure 2. Methods for modelling scenes from a sparse set of still photographs. a) Geometry-Based. b) Image Based. [5].

These techniques solve the problem of the building virtual 3D modelling from a sparse set of uncalibrated images. Two different approaches for modelling have been employed for this purpose:

a. Geometry-Based: Exploits information about the building to reconstruct it. Compelling models can be reconstructed with this technique but they are generally based on simple geometric primitives (buildings as basic structural forms such as cubes, spheres, pyramids, etc). As a result, the details of the scenes are not really modelled but treated as texture. This also requires the user participation in the initialization stages to manually position the elements in the scene, so this step can be labour-intensive. Refer to Fig. 2a.

b. Image-Based: Real images are used as inputs. The depth map of the building can be found semi-automatically using computer vision algorithms such as epipolar geometry [1] stereo or trinocular vision among images. An example of this is proposed by [3]. Refer to Fig. 2b.

Hybrid approaches combine geometry and image based techniques to produce more accurate, and photorealistic results by using strengths of both methods [5],[7]. Constraints of parallelism and orthogonality of the architectural scenes are commonly present in the models [4]. This assumptions help to simplify the camera calibration and the model recovery methods.

The modelling systems start with the camera calibration [6] (estimation of intrinsic and extrinsic parameters) for each viewpoint if they are not known. This procedure generally requires human intervention to help in the buildings' localization such as giving objects edges or generating a basic volumetric model of the scene. Consequently, the associated projection matrices (refer to eq. 1) of each camera are computed and pairs of images are matched to find correspondence maps among them. These maps are then linked together with all the correspondences of different viewpoint pairs to obtain an accurate depth map of the scene. These maps are used to generate triangular surfaces meshes and finally render the 3D model in a virtual environment including additional texture data from the photographs.

$$\lambda i = \begin{bmatrix} ui \\ vi \\ 1 \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} Xi \\ Yi \\ Zi \\ 1 \end{bmatrix} \quad (1)$$

Equation 1. The projection from a three dimensional space to an image can be represented in homogeneous coordinates by a 3x4 camera projection matrix P.

$$\lambda m = PM \quad (2)$$

Equation 2. Simplified form of eq. 1. M are the homogeneous coordinates of a 3D point and m are the homogeneous coordinates of the image point.

Further techniques have been used to realistically render the 3D model virtual viewpoints. An example is called view-dependent texture mapping [5] and basically what it does is to interpolate the textures from the available images from the scene depending on each user's point of view. Results of this approach are depicted in Figure 3.

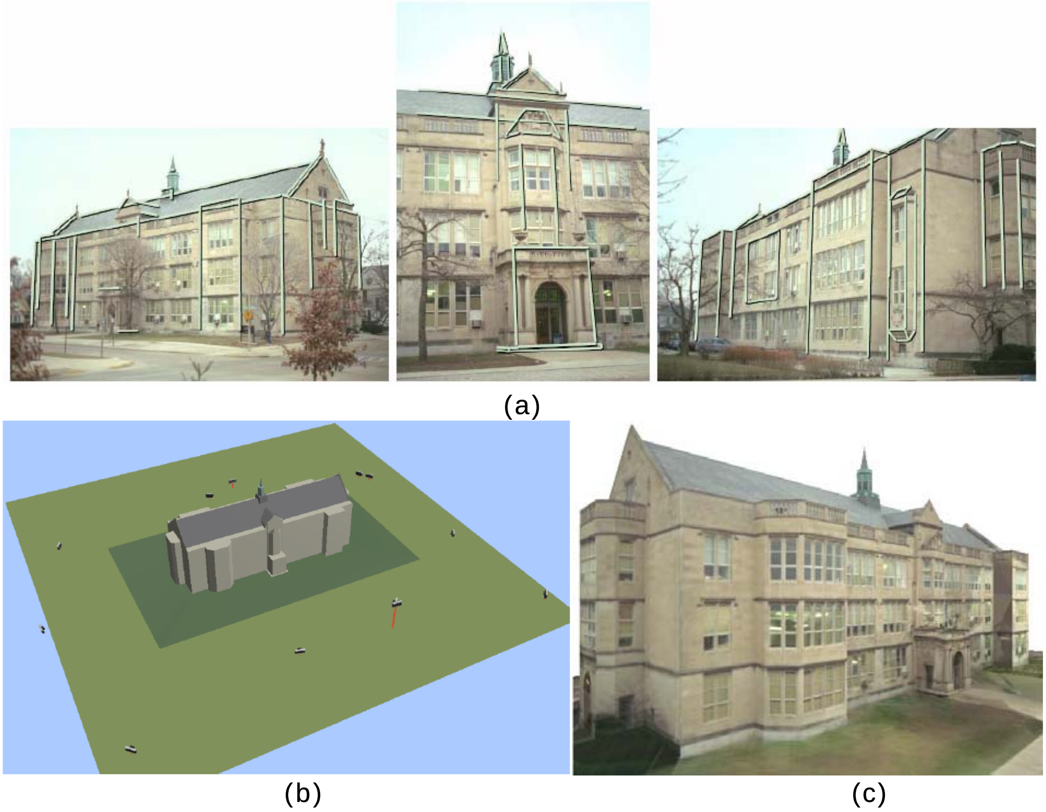


Figure 3. a) Examples of photographs used in the building reconstruction. b) Simulation showing the 12 recovered camera positions. c. synthetic view of the building.

1.2 Range Images

In this approach 3D models of urban scenes are reconstructed using range images [8] which are obtained from laser scanners (see triangulation [11]). The scanning methodology can be classified in two different types which are stationary and vehicle-borne systems.

a. In stationary systems a laser scanning system is mounted on fixed platforms (i.e. [10]). The scenes are measured from multiple viewpoints and then integrated by the extraction of similar geometrical features between them. These systems have some disadvantages, for instance in data acquisition, range images from consecutive viewpoints need to keep a degree of overlay, so the viewpoints' localization can be traced by registering range data. The planning for the selection of the position and number of viewpoints in large environments becomes a hard task.

b. In Vehicle-borne systems the laser scanning equipment is mounted on a mobile platform and continuously registers range data from the scenes. Each registration is integrated in the 3D virtual model according to the vehicle trajectory[9]. One of the biggest issues of these systems is the accurate estimation of that trajectory. Some alternatives to solve this have been proposed such as the integration of GPS (Global Positioning System), INS (Inertial Navigation Systems), and the use of the laser scanners also as positioning sensors.

Recently the laser technology has developed eye-safe lasers which prevent damage to the eyes' retina and permits the use of this equipment in populated areas. In this methodology, photographs of the environment are also obtained as part of the sensor input and then incorporated in the model to add textured and more realistic simulations. Finally there is a method of reconstructing 3D models of architectural environments by fusing data from the inputs (range, localization and images).

In the next figure an example of a vehicle-borne approach used in terrestrial urban reconstruction[9]. This system has three sensory inputs: a Laser scanning system for depth map estimation, linear video camera for texture and GPS/INS for trajectory calculation, all these are mounted in a vehicle. The results of this research show a very accurate 3D reconstruction of a quadrangle.

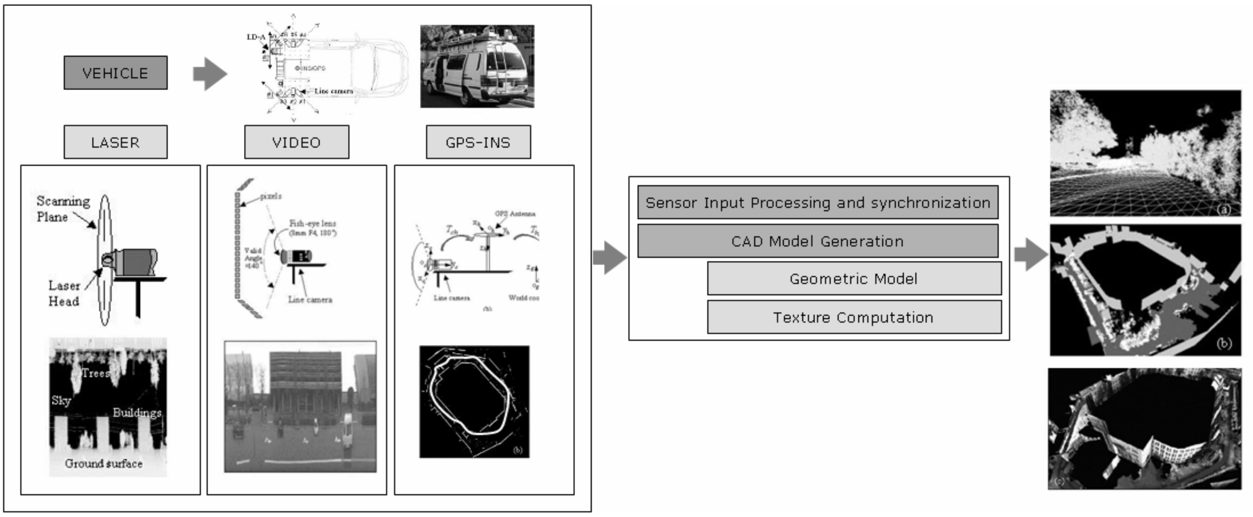


Figure 4. Vehicle-borne laser system approach (Pictures from [9]).

1.3 Video

3D models of architectural scenes are reconstructed with this technique by using image sequences. This consists of an arrangement of video cameras (1 or more) located in a mobile platform (vehicle) and translated around the scene working in continuous mode to obtain the video data which is processed to obtain the virtual model. The camera poses for each frame can be estimated for example, in a geo-spatial coordinate system by using correspondence between each frame and GPS/INS data or from the image sequence itself using structure from motion techniques [2] (or combined as proposed in [12]). Then depth maps can be obtained performing stereo matching between the images and fused together to obtain the three-dimensional model.


Figure 5. Dense reconstruction of a urban scene using video data. [12]

The Structure from Motion techniques are computationally costly so an on-line operation for 3D reconstruction is not possible in the current state of research. The common procedure is to store the video data and the localization information in hardware (i.e. hard disks), and then after the capture sessions are finished, the data can be processed and rendered.

Texture information can also be extracted from the images, so this means that with only one passive device (video cameras), depth maps, localization and texture can be extracted simultaneously. This technique at the current stages is not as precise as laser data acquisition, but it is less expensive than the laser technology. Consequently, the challenges for this system are the improvements of the accuracy in the 3D reconstruction from the video data. A reconstruction example using this technique is depicted in figure 5.

2 Architectural Scenes

All the three previously mentioned methodologies for building reconstruction are used at different scales and environments. The selection of any of them depends on the purpose, requirements and resources of each particular application. Some applications also include combinations of these techniques. In the following section a collection of existing research in architectural virtual modelling is classified according to the size and type of the structures modelled.

2.1 Indoors

- Projection-based Registration Using a Multi-view Camera for Indoor Scene Reconstruction. [13] (Photographs, Image-based)
- First Experiences with Terrestrial Laser Scanning for Indoor Cultural Heritage Applications Using Two Different Scanning Systems. [15]. (Laser scanning, Stationary system)
- 3D Reconstruction of Indoor and Outdoor Scenes Using a Mobile Range Scanner. [16] (Range data, Stationary system).

2.2 Outdoor

2.2.1 Façades

- Automatic 3D Model Acquisition from Uncalibrated Image Sequences [3]. (Image based)
- Modelling and Rendering Architecture from Photographs: A hybrid geometry- and image-based approach [5].
- Solutions to 3D Building Reconstruction from Photographs [7]. (Hybrid approach)
- Reconstructing Textured CAD Model of Urban Environment Using Vehicle-Borne Laser Range Scanners and Line Cameras [9].
- Extracting Windows from Terrestrial Laser Scanning [10]. (Stationary system)
- 3D laser measurement system for large scale architectures using multiple mobile robots [14].

2.2.2 Aerial Image Analysis

- The utilisation of airborne laser scanning for mapping [19].
- Reconstruction of Urban Scenes from Aerial Stereo Imagery: A Focusing Strategy [20]. (Photographs)

2.2.3 Urban modelling

- Towards Urban 3D Reconstruction From Video [12].
- Constructing 3D city models by merging ground-based and airborne views [17] (Laser scanning, vehicle-borne)
- 3D model generation for cities using aerial photographs and ground level laser scans [18]

3 References

[1] A, Fusiello. Elements of Geometric Computer Vision. Avaliable: http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/FUSIELLO4/tutorial.html in CVOnline: On-Line Compendium of Computer Vision [Online]. R. Fisher (ed). Available: http://homepages.inf.ed.ac.uk/rbf/CVonline/. 20 Jan 2008.

[2] T, Jebara, A. Azarbayejani and A. Pentland. The Structure from Motion Task. Avaliable: http://www1.cs.columbia.edu/~jebara/htmlpapers/SFM/node2.html in CVOnline: On-Line Compendium of Computer Vision [Online]. R. Fisher (ed). Available: http://homepages.inf.ed.ac.uk/rbf/CVonline/. 30 Jan 2008.

[3] R. Koch, M. Pollefeys and L. Van Gool. Automatic 3D Model Acquisition from Uncalibrated Image Sequences [PDF]. K.U.Leuven, Dept. Elektrotechniek, ESAT-PSI. Kardinaal Mercierlaan 94, B-3001 Leuven, Belgium.

[4] H. Cantzler, R. Fisher and M. Devy. Improving architectural 3D reconstruction by plane and edge constraining [PDF]. Informatics Research Report EDI-INF-RR-0147. Division of Informatics. Institute of Perception, Action and Behaviour. August 2002.

[5] P. Debevec, C. Taylor and J. Malik. Modeling and Rendering Architecture from Photographs: A hybrid geometry- and image-based approach [PDF]. SIGGRAPH 96 conference proceedings. University of California at Berkeley.

[6] R. Owens. Camera calibration. Avaliable: http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/OWENS/LECT9/node2.html in CVOnline: On-Line Compendium of Computer Vision [Online]. R. Fisher (ed). Available: http://homepages.inf.ed.ac.uk/rbf/CVonline/. 28 Jan 2008.

[7] G. Farinella and G. Mattiolo. Solutions to 3D Building Reconstruction from Photographs [PDF]. Dipartimento di Matematica e Informatica, University of Catania, Italy. Image Processing Laboratory. Eurographics Italian Chapter Conference (2006).

[8] Helmut Cantzler. An overview of range images. Avaliable: http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/CANTZLER2/range.html. in CVOnline: On-Line Compendium of Computer Vision [Online]. R. Fisher (ed). Available: http://homepages.inf.ed.ac.uk/rbf/CVonline/. 28 Jan 2008.

[9] H. Zhao and R. Shibasaki. Reconstructing Textured CAD Model of Urban Environment Using Vehicle-Borne Laser Range Scanners and Line Cameras [PDF]. Center for Spatial Information Science, Univ. of Tokyo.

[10] S. Pu and G. Vosselman. Extracting Windows from Terrestrial Laser Scanning [PDF]. International Institute for Geo-information Science and Earth Observation (ITC), Hengelosestraat 99, P.O. Box 6, 7500 AA Enschede, The Netherlands ISPRS Workshop on Laser Scanning 2007 and SilviLaser 2007, Espoo, September 12-14, 2007, Finland.

[11] S. Price. Fundamentals: Laser ranging using triangulation. Avaliable: http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/MARBLE/low/fundamentals/triang.htm. in CVOnline: On-Line Compendium of Computer Vision [Online]. R. Fisher (ed). Available: http://homepages.inf.ed.ac.uk/rbf/CVonline/. 30 Jan 2008.

[12] A. Akbarzadeh, J.-M. Frahm, P. Mordohai, B. Clipp, C. Engels, D. Gallup, P. Merrell, M. Phelps, S. Sinha, B. Talton, L. Wang, Q. Yang, H. Stewénius, R. Yang, G. Welch, H. Towles, D. Nistér and M. Pollefeys, "Towards Urban 3D Reconstruction From Video" [PDF], Third International Symposium on 3-D Data Processing, Visualization and Transmission, Chapel Hill, North Carolina, USA, June 2006.

[13] S. Kim and W. Woo. Projection-based Registration Using a Multi-view Camera for Indoor Scene Reconstruction [PDF]. GIST U-VR Lab. Gwangju 500-712, S.Korea

[14] R. Kurazume, Y. Tobata, Y. Iwashita and T. Hasegawa. 3D laser measurement system for large scale architectures using multiple mobile robots. Kyushu University [PDF]. 744, Motooka, Nishi-ku, Fukuoka, 819-0395, JAPAN

[15] Th. Kersten, H. Sternberg and E. Stiemer. First Experiences with Terrestrial Laser Scanning for Indoor Cultural Heritage Applications Using Two Different Scanning Systems. Hamburg University of applied sciences, Department of geomatics [PDF]. Hamburg, Germany. 2nd photogrammetry workshop 2005.

[16] Y. Sun, J. K. Paik, A. Koschan, and M. A. Abidi, 3D reconstruction of indoor and outdoor scenes using a mobile range scanner [PDF]. Proc. Int. Conf. Pattern Recognition, Canada, Vol. III, pp. 653-656, Quebec, Canada, August 2002.

[17] Frueh, C. Zakhor, A. Constructing 3D city models by merging ground-based and airborne views [PDF]. Video & Image Process. Lab, Univ. of California, Berkeley, CA, USA; Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference Publication Date: 18-20 June 2003.

[18] C. Fruh and A. Zakhor. 3D model generation for cities using aerial photographs and ground level laser scans [PDF]. Video & Image Process. Lab., California Univ., Berkeley, CA, USA; Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference. Publication Date: 2001

[19] G. Vosselman, P. Kessels and B. Gorte. The utilisation of airborne laser scanning for mapping [PDF]. Delft University of Technology, The Netherlands, Faculty of Aerospace Engineering, Kluyverweg 1, NL-2629 HS Delft, The Netherlands. 15 October 2004.

[20] C. Baillarda, and H. Maîtreb. 3-D Reconstruction of Urban Scenes from Aerial Stereo Imagery: A Focusing Strategy [PDF]. a Institut Géographique National, Laboratoire MATIS, 2, av. Pasteur, Saint Mandé, 94160, France. ENST, Département TSI, 46, rue Barrault, Paris, 75013, France. 21 June 1999.