AUTOMATED 3D MODELING OF URBAN ENVIRONMENTS

Ioannis Stamos

Department of Computer Science
Hunter College, City University of New York
695 Park Avenue, New York NY 10065
istamos@hunter.cuny.edu
http://www.cs.hunter.cuny.edu/~ioannis

KEY WORDS: LIDAR, 3D Modeling, Urban Scenes

ABSTRACT:

The photorealistic modeling of large-scale scenes, such as urban structures, requires a fusion of range sensing technology and traditional digital photography. This paper summarizes the contributions of our group in that area. We present a system that integrates automated 3D-to-3D and 2D-to-3D registration techniques, with multiview geometry for the photorealistic modeling of urban scenes. The 3D range scans are registered using our automated 3D-to-3D registration method that matches 3D features (linear or circular) in the range images. A subset of the 2D photographs are then aligned with the 3D model using our automated 2D-to-3D registration algorithm that matches linear features between the range scans and the photographs. Finally, the 2D photographs are used to generate a second 3D model of the scene that consists of a sparse 3D point cloud, produced by applying a multiview geometry (structure-from-motion) algorithm directly on a sequence of 2D photographs. A novel algorithm for automatically recovering the rotation, scale, and translation that best aligns the dense and sparse models has been developed. This alignment is necessary to enable the photographs to be optimally texture mapped onto the dense model. Finally, we present a segmentation and modeling algorithm for urban scenes. The contribution of this work is that it merges the benefits of multiview geometry with automated registration of 3D range scans to produce photorealistic models with minimal human interaction. We present results from experiments in large-scale urban scenes.

1 INTRODUCTION

The photorealistic modeling of large-scale scenes, such as urban structures, can be achieved by a combination of range sensing technology with traditional digital photography. Laser range scanners can produce highly-detailed geometry whereas color digital cameras can produce highly-detailed photometric images of objects. Our main focus is the geometric and photorealistic reconstruction of individual buildings or large urban areas using a variety of acquisition methods and interpretation techniques, such as ground-base laser sensing, air-borne laser sensing, and ground and air-borne image sensing. The ultimate goal is the reconstruction of detailed models of urban sites, i.e. digital cities, by the efficient combination of all possible sources of information. The creation of digital cities drives other areas of research as well: visualization of very large data sets, creation of model databases for GIS (Geographical Information Systems) and combination of reconstructed areas with existing digital maps. Recently, intense commercial interest for photorealistic reconstruction of city models is eminent in systems such as Google Earth, or Microsoft Virtual Earth.

3D models of cities can be acquired by various techniques such as aerial imagery, ground-based laser range-scanning, existing architectural CAD modeling, and traditional photogrammetry. Aerial-based methods produce crude box-like models, whereas ground-based laser range-scanning methods produce highly accurate models. The latter models though consist of irregular and heavy geometry.On the other hand purely image-based approaches have presented significant progress, and are now able to produce impressive 3D models (Pollefeys et al., 2008, Seitz et al., 2006), that are still inferior to laser-based models. Finally, web-based platforms (such as Google Earth or Microsoft Virtual Earth), are able to receive and display light-weight 3D models of urban objects, whereas rapid-prototyping machines are able to build such models. Therefore, the generation of photorealistic 3D content of urban sites at various resolutions and from various sensors is a very important current problem. Some of the systems that combine 3D range and 2D image sensing for 3D urban modeling include the following: (Früh and Zakhor, 2003, Sequeira and Concalves, 2002, NRC, 2008, Zhao and Shibasaki, 2003, Stamos and Allen, 2002, Zhao et al., 2005).

The framework of our system is shown in Fig. 1. Each of the framework elements listed below, is a distinct system module in Fig. 1.

- A set of 3D range scans of the scene is acquired and co-registered to produce a dense 3D point cloud in a common reference frame.
- An independent sequence of 2D images is gathered, taken from various viewpoints that do not necessarily coincide with those of the range scanner. A sparse 3D point cloud is reconstructed from these images by using a structure-from-motion (SfM) algorithm.
- A *subset* of the 2D images are automatically registered with the dense 3D point cloud acquired from the range scanner.
- The *complete* set of 2D images is automatically aligned with the dense 3D point cloud. This last step provides an integration of all the 2D and 3D data in the same frame of reference. It also provides the transformation that aligns the models gathered via range sensing and computed via structure from motion.
- Finally, segmentation and modeling of the 3D point clouds follows.

2 3D MODELING PIPELINE

In this section we present the status of our 3D modeling system: 3D-to-3D Registration (Sec. 2.1). 2D-to-3D registration

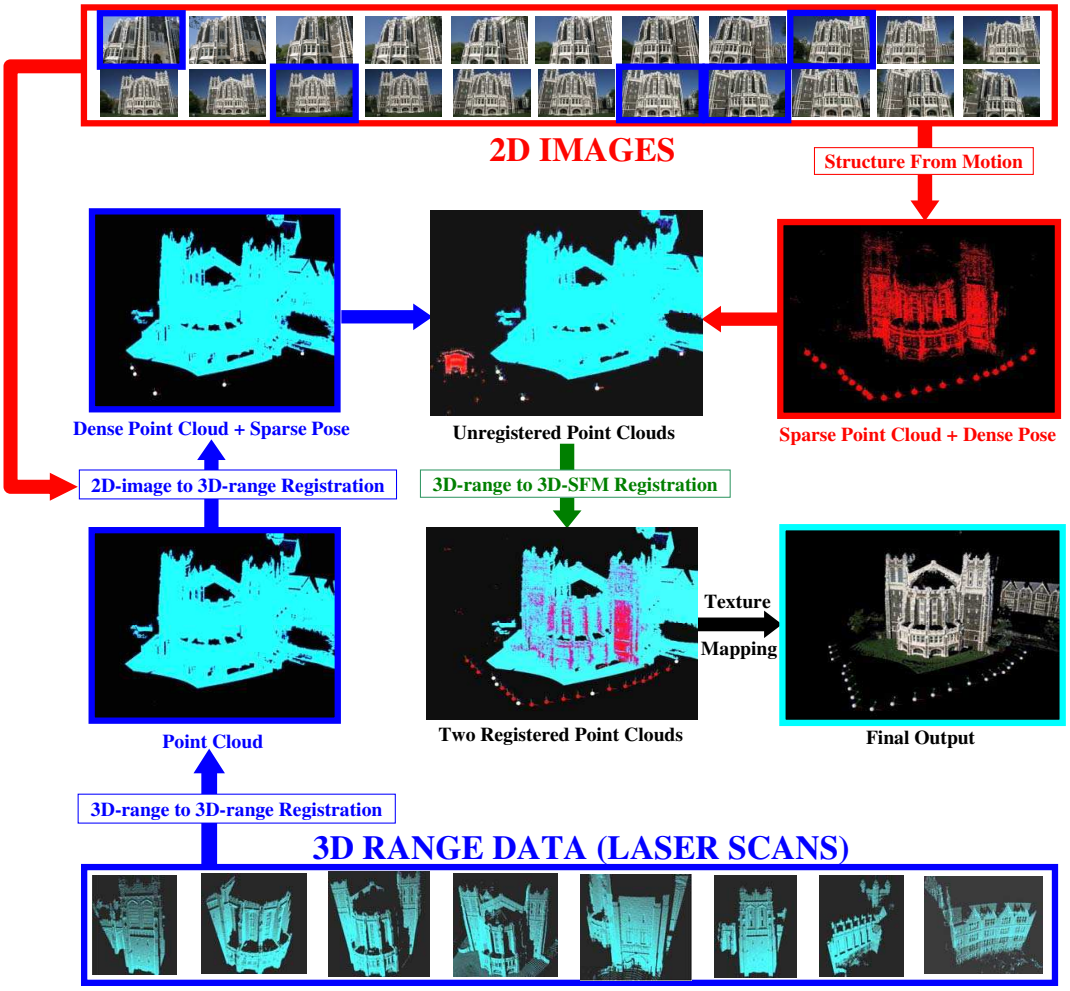


Figure 1: System framework (Stamos et al., 2008). Several registered range scans of Shepard Hall (CCNY) constitute a dense 3D point cloud model M_{range} shown in the leftmost column. The five white dots correspond to the locations of five of the 26 color images (shown as thumbnails on top row) that are independently registered with the model M_{range} via a 2D-to-3D image-to-range registration algorithm. The rightmost image of the second row depicts the 3D model M_{sfm} produced by SFM. The points of M_{sfm} as well as **all** the recovered camera positions for the sequence of 2D images that produced M_{sfm} are shown as red dots in the figure. Since SFM does not recover scale, M_{range} and M_{sfm} are not registered when brought to the same coordinate system, as shown in the second row. The 3D range model M_{range} overlaid with the 3D model M_{sfm} is shown in the third row of the figure after a 3D-range to 3D-SFM registration module aligns them together. The recovered camera positions from SFM can now be used to project the26 color images onto M_{range}, which now properly sits in the M_{sfm} coordinate system, to produce the richly textured 3D model (Final Output) shown in the right column.

(Sec. 2.2),and 3D modeling (Sec. 2.3). More details can be found on some of our papers: (Stamos et al., 2008, Liu and Stamos, 2007, Chao and Stamos, 2007, Liu et al., 2006, Yu et al., 2008).

2.1 3D-to-3D Range Registration

Our 3D registration techniques are based on automated matching of features (lines, planes, and circles) that are extracted from range images. We have applied our automated methods for registration of scans of landmark buildings. In particular we have acquired and registered: interior scans of Grand Central Terminal in NYC, Great Hall at City College of New York (CCNY), as well as exterior scans of St. Pierre Cathedral in Beauvais (France), Shepard Hall at CCNY, Thomas Hunter building at Hunter College, and Cooper Union building (NYC). As a result, all range scans of each building are registered with respect to one selected pivot scan. The set of registered 3D points from the K scans is called M_{range} (Fig. 1).

2.2 2D-to-3D Image-to-Range Registration

We present our automated 2D-to-3D image-to-range registration method used for the automated calibration and registration of a single 2D image I_n with the 3D range model M_{range}. The computation of the rotational transformation between I_n and M_{range} is achieved by matching at least two vanishing points computed from I_n with major scene directions computed from clustering the linear features extracted from M_{range}. The method is based on the assumption that the 3D scene contains a cluster of vertical and horizontal lines. This is a valid assumption in urban scene settings.

With this method, a few 2D images can be independently registered with the model M_{range}. The algorithm will fail to produce satisfactory results in parts of the scene where there is a lack of 2D and 3D features for matching. Also, since each 2D image is independently registered with the 3D model, valuable information that can be extracted from relationships between the 2D images (SfM) is not utilized. In order to solve the aforementioned problems, an SfM module final alignment module (Stamos et al., 2008, Liu et al., 2006) has been added into the system. These two modules increase the robustness of the reconstructed model, and improve the accuracy of the final texture mapping results. Therefore, the 2D-to-3D image-to-range registration algorithm is used in order to register a few 2D images (five shown in Fig. 1) that produce results of high quality. The final registration of the 2D image sequence with the range model M_{range} is performed after SfM is utilized.

Our recent contributions (Stamos et al., 2008, Liu and Stamos, 2007, Liu, 2007) with respect to 2D-to-3D registration can be summarized as follows:

- We have developed a working system that is able to independently register 2D images to 3D models at interactive rates. This system requires minimal user interaction. Note that after a few 2D images are registered to the 3D model the multiview geometry approach (SfM) is utilized for registering all images with the 3D range model.
- The whole space of possible matches between 3D and 2D linear features is explored efficiently. That improves the possibility of convergence of our algorithm.
- Our method utilizes 3D and 2D linear features for matching without significant grouping. This increases the generality of our algorithm since we make fewer assumptions about the 3D scene. Scenes with various layers of planar facades, or without clear major facades can thus be handled.

2.3 Modeling

We have developed novel algorithms (Yu et al., 2008, Chao and Stamos, 2007, Chen, 2007) for extracting planar, smooth non-planar, and non-smooth connected segments, and then merging all these extracted segments from a set of overlapping range images. Our input is a collection of registered range images. Our output is a number of segments that describe urban entities (e.g. facades, windows, ceilings, architectural details). In this work we detect different segments, but we do not yet identify (or recognize) them. A flowchart of our current technique can be seen in Fig. 2.

In addition to segmenting each individual scan, our methods also merge registered segmented images. The merging results in coherent segments that correspond to urban objects (e.g. facades, windows, ceilings) of a complete large scale urban scene. Based on this, we generate a different mesh for each object. In a modeling framework, higher order processes can thus manipulate, alter, or replace individual segments. In an object recognition framework, these segments can be invaluable for detecting and recognizing different elements of urban scenes. Results of our segmentation and modeling algorithms can be seen at Fig. 3.

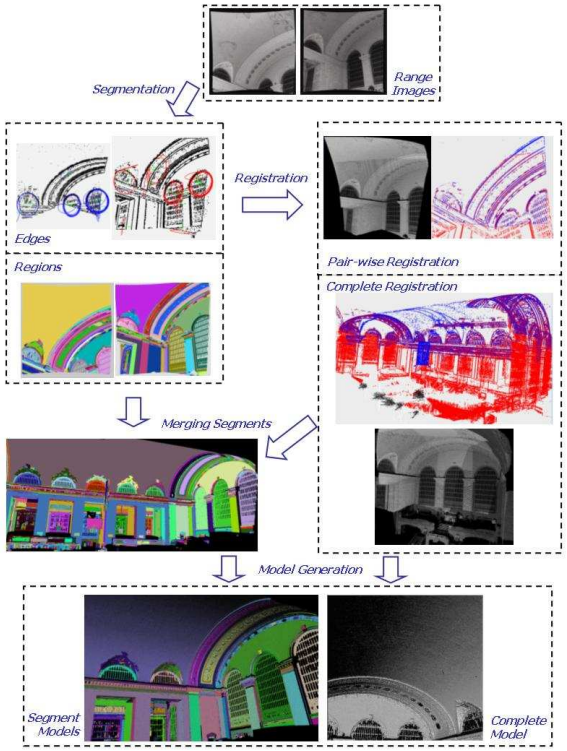


Figure 2: Our segmentation and modeling framework (Chen and Stamos, 2005, Chao and Stamos, 2007, Chen, 2007).

3 FUTURE WORK

The generated 3D models are complex triangular meshes. Mesh simplification is thus important. Unfortunately, simplification approaches suffer from the fact that their input is a complicated mesh. A mesh is a low-level heavy collection of triangles that does not take into account the high-level abstraction of urban structures. A high-level model should identify facades, doors, windows, and other urban entities. An important avenue of exploration is an automated high-level representation of the final 3D urban model.

ACKNOWLEDGEMENTS

This work was supported in part by the following grants: NSF CAREER IIS-02-37878, NSF MRI/RUI EIA-0215962, and NSF CNS-0821384.

REFERENCES

Bernardini, F. and Rushmeier, H., 2002. The 3D model acquisition pipeline. Computer Graphics Forum 21(2), pp. 149–172.

Chao, C. and Stamos, I., 2007. Range image segmentation for modeling and object detection in urban scenes. In: The 6th International Conference on 3-D Digital Imaging and Modeling, Montreal, Canada.

Chen, C., 2007. Range Segmentation and Registration for 3D Modeling of Large-Scale Urban Scenes. PhD thesis, City University of New York.

Chen, C. and Stamos, I., 2005. Semi-automatic range to range registration: A feature-based method. In: The 5th International Conference on 3-D Digital Imaging and Modeling, Ottawa, pp. 254–261.

Früh, C. and Zakhor, A., 2003. Constructing 3D city models by merging aerial and ground views. Computer Graphics and Applications 23(6), pp. 52–11.

Liu, L., 2007. Automated Registration of 2D Images with 3D Range Data in a Photorealistic Modeling System of Urban Scenes. PhD thesis, City University of New York.

Liu, L. and Stamos, I., 2007. A systematic approach for 2D-image to 3D-range registration in urban environments. In: VRML Workshop, 11th International Conference on Computer Vision, Rio de Janeiro, Brasil.

Liu, L., Stamos, I., Yu, G., Wolberg, G. and Zokai, S., 2006. Multiview geometry for texture mapping 2D images onto 3D range data. In: IEEE Conf. Computer Vision and Pattern Recognition, Vol. II, New York City, pp. 2293–2300.

NRC, 2008. Visual Information Technology Group, National Research Council, Canada.
http://iit-iti.nrc-cnrc.gc.ca/about-sujet/vit-tiv_e.html.

Pollefeys, M., Nistr, D. and etal., 2008. Detailed real-time urban 3d reconstruction from video. International Journal of Computer Vision 78(2–3), pp. 143–167.

Seitz, S. M., Curless, B., Diebel, J., Scharstein, D. and Szeliski, R., 2006. A comparison and evaluation of multi-view stereo reconstruction algorithms. In: IEEE Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. 519–526.

Sequeira, V. and Concalves, J., 2002. 3D reality modeling: Photo-realistic 3D models of real world scenes. In: Intl. Symposium on 3D Data Processing, Visualization and Transmission, pp. 776–783.

Stamos, I. and Allen, P. K., 2002. Geometry and texture recovery of scenes of large scale. Journal of Computer Vision and Image Understanding 88(2), pp. 94–118.

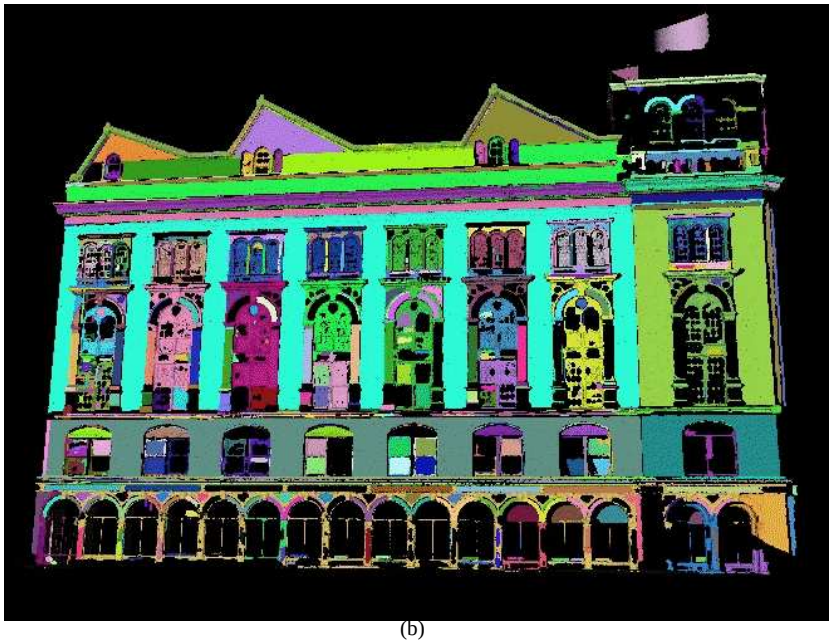
Stamos, I., Liu, L., Chao, C., Wolberg, G., Yu, G. and Zokai, S., 2008. Integrating automated range registration with multiview geometry for the photorealistic modeling of large-scale scenes. International Journal of Computer Vision 78(2–3), pp. 237–260.

Yu, G., Grossberg, M., Wolberg, G., Stamos, I. and Zokai, S., 2008. Think globally, cluster locally: A unified framework for range segmentation. In: Fourth International Symposium on 3D Data Processing, Visualization and Transmission, Georgia Institute of Technology, Atlanta, GA.

Zhao, H. and Shibasaki, R., 2003. Reconstructing a textured CAD model of an urban environment using vehicle-borne laser range scanners and line cameras. Machine Vision and Applications 14(1), pp. 35–41.

Zhao, W., Nister, D. and Hsu, S., 2005. Alignment of continuous video onto 3D point clouds. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(8), pp. 1305–1318.

(a)



(b)

Figure 3: (a) Segmentation and modeling result: 15 range images of Grand Central Terminal dataset. Different colors correspond to different segments that have been automatically extracted and modeled via the Ball Pivoting algorithm (Bernardini and Rushmeier, 2002). Cylindrical ceiling, planar facades, as well as other more complex areas (windows, etc.) have been correctly segmented. (b) Segmentation and modeling result of Cooper Union dataset: 10 range images (one facade is shown). Planar facades, and complex window and arch elements have been correctly segmented. Note, that in both (a) and (b) each segment is represented as a dense triangular mesh.