LOW COST 3D URBAN SCENE RECONSTRUCTION BY FUSING 3D DATA AND 2D OPTICAL IMAGES

Nesrine Grati [a,b,*], Imen Khanfir Kallel [b], Dorra Sellami Masmoudi [b], Frank Boochs [a]

[a] i3mainz, Mainz University of Applied Sciences, Lucy-Hillebrand Str. 2, 55128 Mainz, Germany - (nesrine.grati, boochs)@geoinform.fh-mainz.de
[b] ICOS, Sfax Engineering School, BP 1173, 3038 Sfax, Tunisia- imen.khanfir@batotex.com.tn, dorrasellamimasmoudi@yahoo.com

Commission VI, WG VI/4

KEY WORDS: Low cost 3D scene reconstruction, Urban scene, Laser scanner, Optical scanner, 2D-3D data fusion.

ABSTRACT:

3D objects modelling from 2D images is a major task in many application fields including robotics, virtual and augmented reality, and entertainment. According to the equipment used to acquire and process data, a reconstruction process can be based on laser or optical scanners or on digital images. Laser scanner provides a highly detailed and accurate representation of even complex surface shapes but generally, they are expensive, the data processing consumes much time and requires expert interaction with the machine. On the other hand, image based multi-view stereovision techniques provide nowadays an exciting alternative for 3D scene reconstruction because images can be easily obtained. Given a set of 2D images, we propose a novel approach for 3D urban scene reconstruction with low cost by extending unstructured spatial data normally generated by stereovision approaches by linear structures found in the same images. Our suggested system try to resolve two proposed issues: (1) The generation of dense 3D point clouds from 2D images (2) The integration of the available structured 2D data for the improvement of the detection/reconstruction of urban structures.

1. INTRODUCTION AND MOTIVATIONS

3D object modelling has been a topic of intense effort for many years by computer vision and photogrammetric communities. An increasing interest on 3D virtual model reconstruction has produced new and useful applications such medical research, games, multimedia, cultural heritage documentation, digital archiving, 3D city modelling and underwater environment visualization, etc. Obviously, these applications require an augmented degree of realism, high level of precision and many other supplies.

Reconstruction of 3D models based on real data acquired from urban environment (buildings, cities, road, etc) has long been an active topic of research which led to a number of successful methods. Creating such models can be achieved either with images or with optical scanners. Active sensors are promising and provide a highly detailed and accurate representation of surface shapes and their advantage grows with increasing complexity of surface shapes. But generally, such systems are still expensive; the data processing consumes much time and requires expert assistance for object modelling and extraction. Due to recent advances in multi-view stereovision, new solutions are coming up, which provide an exciting alternative for 3D scene reconstruction based on 2D data. And images can be easily obtained through cheap sources, like digital cameras, or even downloaded from the Internet. Nevertheless, point clouds generated from multi-view methods suffer from some limits like the existence of outliers and gaps in the resulted 3D model. This is due to well known problems of image based reconstructions like effects of perspective, occlusions and extreme lighting conditions (weak contrast, reflections, etc).

That's why many works has been invested to improve 3D scene reconstruction processes based on 2D images. Especially in case of urban environment with its typical linear structures additional information exists, which might help to come to better results. Together with the low cost aspect of images it therefore could be interesting to find a solution for urban scenes. According to the requirements of such an application, the suggested solution tries to answer a number of aspects including:

- adequate geometric accuracy
- morphological correctness
- complete modelling (with more details)
- high degree of automation
- low costs
- flexibility
- portability of the modelling alternative
- etc...

Geometrical and morphological precise 3D urban scene reconstruction techniques based on cheap digital images are still unavailable. By the actual work, we try to make a step forward. From our point of view, we think that we can profit more from the 2D images. In fact, we believe that structured 2D information can be used as an additional data to the previously generated 3D point clouds in the reconstruction process ensuring better 3D models. The goal of our work is to develop an efficient method for urban scene reconstruction based on digital photographs by fusing 2D optical data and 3D point cloud. The presented research makes an attempt to emphasize the possibility to combine 2D and 3D data together since it produce more efficient and accurate results generating a most realistic architectural model while neither 2D nor 3D based technique alone will reach the same efficiency degree while resolving the problem. To facilitate the task, the target problem can be subdivided on two sub-problems where the first one treats the 3D point cloud generation from 2D images. While the second one discusses the fusion of the 2D details with the derived 3D model.

This paper is structured as follows: section two gives an overview of actual existing strategies of 3D scene reconstruction based on multiple images. In the third section, we expose the suggested concept. Finally we conclude and show next planned steps.

2. RELATED WORKS

The reconstruction process of 3D scene has occupied a wide area of computer vision for the past three decades. In this field, many different approaches and algorithms were developed with different automation levels, image source types, terrain complexity and application fields etc. 3D modelling methods can be classified according to automation degree as: manual, semi-automated or automated approaches.

2.1 Manual approaches

In such methods, scene components are manually extracted by means of modelling software. Nevertheless, this procedure is time-consuming and has an expensive cost especially for rich urban scene environment since a variety of object shapes and structures can exist.

2.2 Semi-automatic methods:

Concerning the semi-automatic methods, the user initializes the 3D reconstruction process by some manual measurements, and then specific algorithm extracts scene elements based on 2D images. Such methods are based on user interactions and automatic algorithm processing (Debevec, et al., 1996), (Dick, et al., 2004), (Rother, et al., 2002). A few numbers of images are required although preliminary models with sufficient quality are often not available.

2.3 Automatic methods:

Focus mainly on the full automation of the process but generally produce results which are mainly good for nice-looking, real-time 3D recording or simple visualization.

Two families of methods are proposed in the literature: model-driven approaches (top-down strategy) and the data-driven approaches (bottom-up strategy). The main difference between them is how to detect and reconstruct urban structures.

2.3.1 Data-driven based approaches

Data-driven approaches, can be seen as a bottom-up approaches, aims to extract from the input data geometrics primitives like points, segments, lines, and gathered them to form more complex urban model structure by means of hypothesis process.

In (Taillandier, et al., 2002), (Heuel, et al., 2001), (Schindler, et al., 2006), (Bauer, et al., 2002) and (Mayer, 2007), a set of 2D features like points, segments, lines, planes, are firstly detected then matched. As result, new 3D entities are constructed. Some works take into account the geometric constraint as well as relations such as incidence, equality, parallelism and orthogonality between lines and planes. Problems that arise this class of approaches are especially related to the visibility, where objects cannot be observed in the input data due to total or partial occlusions of architectural structures, or caused by unconstructed area in the final model suffering from errors and incompleteness forms (Mayer, 2007).

Although many concepts try to cope with errors and occlusions (Taillandier, et al., 2002), the problem persist. In fact, specific hypothesis as well as small baseline settings where a high volume of overlapping information is ensured are applied in the reconstruction process.

2.3.2 Model-driven based approaches

In contrast with bottom-up approaches, top-down model-driven approaches incorporate in the 3D modelling process a prior knowledge about the scene. The model-driven approaches look the most appropriate model among primitive models contained in a model library (Maas, et al., 1999). By verifying the hypotheses against the input data, the 3D models can be subsequently refined in order to best match the input data with the predefined models allowing the system to make inferences about the scene. The sample models are particularly used to control and to restrict the generation of the final model. An automatic building reconstruction from aerial images is presented in (Taillandier, et al., 2004) . The system starts by the extraction of planar primitives. The arrangement of this planar primitives allow the creation of acceptable forms of buildings through maximal cliques search in 3D graph. The graph was then simplified in order to obtain polyhedral, generic and non-complex models by the introduction of external knowledge about the structure of encountered forms. The reconstruction performance is laid to the quality of extracted geometrical features, the quality of images, their calibration and the accuracy of detection algorithm.

(Schindler, et al., 2003) proposed a detailed reconstruction of buildings from close range images based on plane detection and principle direction. Then, a specific feature of facades such as primitives are employed to reconstruct the fine details. An hybrid approach known as plane sweeping based method was described by (Werner, et al., 2002) who use sparse point and line correspondences to discover the ground and facade planes from wide baseline images. The presented method reconstructs facades and estimates features, such as windows and balconies, by sweeping polygonal primitives. Two generic models are used to fit some details in the final model.

In contrast with the general suggested methods, these approaches depend on strong architectural cues and are limited to scene reconstructions where their numerous assumptions are verified. However, they still limited to the available models in the library which cannot describe all shapes of the architectural objects.
Further on, the model driven and the data driven approach can be both integrated as alternative achieving promising results.

3. SYSTEM OVERVIEW

Within the literature presented in section 2, it’s clear that tow different approaches for 3D point clouds generation from 2D images are taking place. In the first one, “interest points” are detected from 2D images throw global feature descriptor technique. Assuming that several types of interest points have been identified, the following step is to search the correspondence by matching them across images. This approach achieves as result a dense 3D model. In the second kind of approaches, different techniques extract geometrical primitives like segments, line or points from 2D images, and then 3D structured model is produced by matching these features. The major handicap of this approach is the incompleteness of structures which is especially caused by error detection or occlusion parts in the input data.

Deriving a complete, detailed, dense and realistic model still until now unresolved task in image-based research particularly for outdoor scene reconstruction as the objects structures are various and complex. Our project focus on the automatic urban scene modelling from 2D images sequence. The proposed solution tends to obey to the project requirements as cheap cost, portability, and efficiency.

Our research is based on the assumption that the fusion of structured optical data with spatial data can lead to a more reliable urban model. 2D data is used again as supplement information to enhance the quality of detection or reconstruction of the architectural objects in the urban environment which is usually rich of linear structures. Such assumption involves a lot of questions and challenges such as: (1) which kind of features will be extracted from images for fusion step? (2) How can we combine different types of data without losing time and performance?

In fact, in our work two proposed issues should be resolved:

- The generation of 3D point clouds from 2D images
- The integration of the available 2D data in order to improve the detection or reconstruction of urban structures.

The workflow of our suggested concept is illustrated in figure 1.

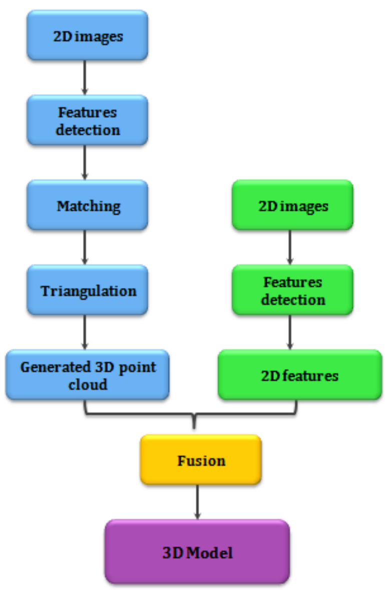

Figure1. Suggested 3D reconstruction solution.

3.1 3D point clouds generation

Generally, image-based approaches include three conjunctive steps: features extraction, matching and 3D coordinate computation by triangulation. In the first time, 2D features are extracted using adapted algorithms. The main challenge in this step is which kind of features will be extracted? And how feature can be characteristic and invariant to image transformations?

In fact, 2D features can be related to geometrical aspect of the architectural elements or it can be related to the global aspect of input data where interest points are extracted with detectors such as Förstner detector (Förstner, et al., 1987), Harris detector (Harris, et al., 1988) or Scale-invariant feature transform (SIFT) operator (Lowe, 2004) which is widely used nowadays. Once these features are extracted from each image, correspondences need to be established. The effectiveness of matching step depends essentially on the quality of the extracted features and the choice of the appropriate matching algorithm. Image noise and occlusions or errors in feature detection basically affects the matching results. The set of correspondences between images allows us to compute 3D point positions as well as camera poses.

In this phase, we can take advantage of some existing solutions for multi image based point cloud. Among the available open source systems, we can note PMVS2+bundler, Photofly (Autodesk) and Arc3D web service tools.

PMVS2, an available binary source code tool, was developed at the University of Illinois by Drs. Furukawa and Ponce (Furukawa, et al., 2010a) (Furukawa, et al., 2010b). PMVS2 is multi-view stereo software that takes a set of images and camera parameters to reconstruct the 3D structure of an object or a scene visible in the images. But giving that PMVS2 does not perform the Structure from Motion (SfM) computation, Bundler tool (Snavely, et al., 2006), proposed by Drs. Snavely and Seitz, is used in conjunction with PMVS2.

Photofly (Autodesk, 2011) technology was acquired on May 2008 from Realviz. After few years of research and development conducted by Autodesk Lab, Photofly was released on July 22, 2010.

Arc3D (Vergauwen, et al., 2006) is a free service web integrating the entire chain of treatment for 3D point cloud reconstruction from images. Users should at first download a small application which sends the selected images to be processed to a distant server. The registration as a user service is needed because in further steps, the resulted model is sent to the user through his e-mail account. In fact, the process of reconstruction is transparent for Arc3D consumers.

Actually, the mentioned 3D modelling systems are based on the structure-from-motion concept and almost of them are based on the web service strategy (user query- distant server reply). We tend to test and evaluate these techniques in order to select the more suitable solution for our project. To guide our choice, we should answer at this question: Which system can be extended by an extra dimension for the fusion step before producing the final 3D urban model? And which technique can support optical knowledge as a supplement data to enrich the reconstruction process?

3.2 2D features detection for fusing with 3D data

As mentioned before, the key step of our suggested solution is principally presented via the introduction of 2D data for the refinement of 3D generated model. The initial generated 3D model by multi-view-stereovision concept is especially at edges in general not precise and complete. Therefore, it can be refined in second phase by introducing 2D data which yields a higher effectiveness of 3D scene modelling. The question here which kind of 2D data can be applied for fusion purpose?

Generally, the dominant geometric forms present in urban scene are either linear forms which describe mainly architectural element´s boundaries or planar surfaces drawing walls, doors, windows, roof, etc. These features can be detected from 2D images and then introduced as supplement details for reconstructing a more accurate model.

Once the 3D point cloud is generated in the first stage of modelling, linear features can be integrated in order to refine geometric structure and details. Through 2D images, the extraction of linear features is easily achieved as a variety of edge operators like Prewitt, Canny, Sobel and also linear edge detector as Hough linear transformation.

4. CONCLUSION

A multi-view passive stereo vision method is promising alternative for 3D urban scene modelling. It could make scene reconstruction cheaper than laser scanner devices by providing some very exciting results. In the attempt to enhance the quality of derived 3D models, an efficient integration of optical data and range data can be made. The two modality data are characterized by their complementary nature aspect to extract and reconstruct 3D features. 3D modelling systems are available for use on the Web or locally in a binary format. However, we focus on future steps to select the suitable technical solutions which can support 2D-3D fusing step and to improve some existing modules and instructions. At the meantime, surveying some existent 3D-2D data algorithms is also concerned.

REFERENCES

Autodesk Project Photofly, Autodesk, 01-12-2011, http://labs.autodesk.com/technologies/photofly/ (accessed 01 dec 2011).

Bauer, J., Klaus, A., Karner, K., Zach, C., Schindler, K., 2002. MetropoGIS: A feature based city modeling system, International archives of photogrammetry remote sensing and spatial information sciences, 34(1), pp. 22-27.

Debevec, P E., Taylor, C J. and Malik, J., 1996. Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach, Proceedings of the 23rd annual conference on Computer graphics and interactive techniques, pp.11-20.

Dick, A R., Torr, PHS., and Cipolla, R., 2004. Modelling and interpretation of architecture from several images. *International Journal of Computer Vision*, 60, pp. 111-134.

Förstner, W. and Gälch, E., 1987. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In: *Proc. ISPRS intercommission conference on fast processing of photogrammetric data*, Interlaken, Switzerland, pp. 281-305.

Furukawa, Y., Curless, B., Seitz, S.M. and Szeliski, R., 2010b. Towards internet-scale multi-view stereo, In Computer Vision and Pattern Recognition (CVPR2010), San Francisco, CA, pp. 1434-1441.

Furukawa, Y., and Ponce, J., 2010a. Accurate, dense, and robust multiview stereopsis. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 32, pp. 1362-1376.

Harris, C., and Stephens. M., 1988. A combined corner and edge detector. In: *4th Alvey Vision Conference*, Manchester, Vol. 15, pp. 50.

Heuel, S., and Forstner, W., 2001. Matching, reconstructing and grouping 3d lines from multiple views using uncertain projective geometry. In: *Proc.IEEE Conference on Computer Vision and Pattern Recognition (CVPR'01)*, Vol. 2, pp. II-517.

Lowe, D.G., 2004. Distinctive image features from scale invariant keypoints. *International journal of computer vision*, 60(2), pp. 91-110.

Maas, H.G., and Vosselman, G., 1999. Two algorithms for extracting building models from raw laser altimetry data. *ISPRS Journal of photogrammetry and remote sensing*, 54, pp. 153-163.

Mayer, H., 2007. 3D reconstruction and visualization of urban scenes from uncalibrated wide-baseline image sequences. *Photogrammetrie fernerkundung Geoinformation*, 2007(3), pp. 167.

Rother, C., and Carlsson, S., 2002. Linear multi view reconstruction and camera recovery using a reference plane. *International Journal of Computer Vision*, 49, pp. 117-141.

Schindler, G., Krishnamurthy, P., and Dellaert, F.,2006. Line-based structure from motion for urban environments. In: *3th International Symposium on 3D Data Processing*, Visualization, and Transmission, pp. 846-853.

Schindler, K., and Bauer, J., 2003. A model-based method for building reconstruction. In: First IEEE International Workshop on Higher-Level Knowledge in 3D Modeling and Motion Analysis (HLK 2003), Nice, France, pp. 74-82.

Snavely, N., Seitz, S.M., and Szeliski, R., 2006. Photo tourism: exploring photo collections in 3D, *ACM Transactions on Graphics (TOG)*, Vol. 25, pp. 835-846.

Taillandier, F., and Deriche, R.,2004. Automatic buildings reconstruction from aerial images: a generic bayesian framework. *Proceedings of ISPRS*, vol. XXXIV/3A, pp. 267-272

Taillandier, F., and Deriche, R.,2002. Reconstruction of 3D linear primitives from multiple views for urban areas modelisation. *International archives of photogrammetry remote sensing and spatial information sciences: Natural ressources*, Canada, Vol. 34 (3/B), pp. 267-272.

Vergauwen, M., and Van Gool, L., 2006. Web-based 3D reconstruction service, *Machine vision and applications*, 17(6), pp. 411-426.

Werner, T., and Zisserman, A., 2002. New techniques for automated architectural reconstruction from photographs, Computer Vision ECCV 2002. pp. 808-809.