3D Reconstruction of a Scene from Multiple Uncalibrated Images Using Close Range Photogrammetry

Abdoulaye Mahamat Zakaria
Department of Computer and Information Sciences,
Universiti Teknologi PETRONAS, Tronoh, Malaysia
abdoumht@gmail.com

Abas Md Said
Department of Computer and Information Sciences,
Universiti Teknologi PETRONAS, Tronoh, Malaysia
abass@petronas.com.my

***Abstract*--- In order to contribute on the search of solving the 3D reconstruction problem we present in this paper a workflow of 3D reconstruction from multiple view images. Our system for the time being takes a set of uncalibrated images as an input and produces a depth map. We divided the workflow in three major parts which are finding corresponding points between images, depth estimation and 3D model generation. We use the global method and dynamic programming method to produce the depth maps.**

Keywords; 3D reconstruction; feature matching; multiview stereo; model generation

I. INTRODUCTION

Three dimensional (3D) reconstruction of scenes from uncalibrated images is considered as one of the most challenging problem in computer vision and photogrammetry [1]. In order to reconstruct a model from 2D images taken from different views or locations, the intrinsic camera parameters and the relative motion of the images with respect to each other must be known. But in our case the camera parameters and the scene parameters are assumed to be unknown. That makes our task more challenging. To simplify the calibration task we will apply the camera self-calibration method used by (U. Topay in [2]). He mentioned that the cameras are composed of five parameters. Therefore, the aim of camera calibration is to determine these five unknown camera parameters. In order to estimate the camera parameters, relation between the images must be determined. The algebraic relation between two images is determined by using corresponding 2D points. These points are considered as the positions of the projected 3D scene points on these two images. After words, by using the algebraic relations between these images, the camera parameters can be estimated. As a next step, the relative motion of the images is also determined by using the camera intrinsic parameters, algebraic relation between the images and the corresponding points.

Another important part of the 3D structure is the algebraic relation between images. The images of an object or a scene are obtained by taking pictures of the same object from different views and locations by using same or different cameras. By using corresponding points between two images, the algebraic relation between these can be determined. This relation between the two images is represented by the well-known Epipolar geometry [3]. Epipolar geometry relates every point in one image to a corresponding line in the other image. And the epipolar constraint can be represented by what do they called, *Fundamental matrix*. In order to determine the Fundamental matrix, the corresponding points between the two images are required. Always the data points are corrupted by noise and the matches can be spurious or incorrect, the Fundamental matrix may not determine exactly [2].

To achieve the goal of obtaining a 3D model of a scene from uncalibrated images taken from different view with a hand-held camera, we planed to proceed according to the following major processes which are.

- Finding corresponding points between images;
- Depth estimation;

Each one of these processes is composed of sub-problems; we will give deep explanation to each in the coming steps.

II. LITERATURE REVIEW

Obtaining 3D model from multiple images is a difficult task. This comes from the fact that only very little information are available to start with. Both the scene parameters and the camera parameters are assumed to be unknown. Only very general assumptions are made, e.g. rigid scene, piecewise continuous surfaces, mainly diffuse reflectance characteristics for the scene and a pinhole camera model for the camera..

The general idea we used to achieve our task is to separate the problem in a number of sub problems to be more

978-1-4244-6716-7/10/$26.00 ©2010 IEEE

manageable, which can then be solved by separate modules. Mostly interaction or feedback is needed between these modules to debrief the necessary information from the images. Certainly for the first steps when almost no information has been debriefed, feedback to verify the obtained hypotheses is very important. Gradually we will get more information.

Camera Self-Calibration:

The camera self-calibration is an important research topic in computer vision for about 20 years. In this problem [4], there are only images taken from different locations and orientations to estimate camera intrinsic parameters. By only using these images, the relation between pairs of images can be represented algebraically with Fundamental matrices. However, the Fundamental matrix contains both camera intrinsic parameters and relative motion between the two images or cameras. Therefore, a formulation must be defined which does not change by the relative motion between the two images. As expected, while moving the camera, the positions of the 3D points change on the image inversely proportional to their distances from camera. For example, one see the view of the moon while it does not change by looking from different locations.

Due to this contradictory relation, there must be a virtual conic which is quite far away from camera while viewing locations and its projection depends on only the camera intrinsic parameters. By defining the relation between this virtual conic and intrinsic camera parameters, the equations for camera self calibration are defined [3].

There are so many different methods developed for camera self-calibration. The first one was developed in 1992 by Maybank and Faugeras [5]. In this method, the nonlinear quadratic equations, called as Kruppa equations, are constructed by using Fundamental matrices and unknown camera matrices. These equations have been solved in different ways [5], [6], [7], [8]. Other methods, which do not solve Kruppa equations, generally determine the camera intrinsic parameters and the position of the plane of such a virtual conic by using the relation between the virtual conic and the camera intrinsic parameters [9]. Afterwords they update the projective reconstruction to metric reconstruction. In a recent method, which is developed by Pollefeys [10], projective reconst uction is updated to affine reconstruction by using the position of the plane of the virtual conic determined by solving a number of constraints [11]. Then, the affine calibration is updated to a metric one using the estimated camera intrinsic parameters determined by solving the general camera self-calibration equations. This method is called Stratified calibration, since one move between different stratums (i.e. affine to metric) during reconstruction.

These used methods in a general term assuming that the images are taken by the same camera. However, these images can be taken with cameras with different intrinsic parameters or a hand-held camera with varying focal length. Some techniques are also there to solve such cases, such as [12]. However, the camera self-calibration for such conditions will be investigated in the next years.

Feature matching:

Given a feature in an image, what is the corresponding projection of the same 3D feature in the other image? This is an ill-posed problem [2] and therefore in most cases very hard to solve. When some assumptions are satisfied, it is possible sometimes to automatically match points or other features between images. If the images are not too different in term of illumination or similar pose for example, the coordinates of the features and the intensity distribution around the feature are similar in both images. This allows to restrict the search range and to match features through intensity cross-correlation [1].

It is clear that not all possible image features are good for matching. Often points are used since the other modules most easily handle them, but line segments [13] or other features (such as regions) can also be matched. It is clear that not all points are suited for matching. Sometimes a lot of points can be located in symmetrical regions were almost less or no information is available to differentiate between them. It is therefore important to use an interest point detector, which extracts a certain number of points useful for matching. These points should clearly satisfy two criteria. The extraction of the points should be as much as possible independent of camera pose and illumination changes and the neighborhood of the selected points should contain as much information as possible to allow matching. Many interest point detectors exist (e.g. Harris [14]). In [15] Schmid et al. concluded that the Harris corner detector gives the best results according to the two criteria mentioned above.

In fact the feature matching is mostly coupled tightly with the 3D structure from sequence images estimation described in the next paragraph. Hypothetical matches are used to compute the scene and camera geometry. The obtained results are then used to drive the feature matching.

Uncalibrated structure from sequence images:

Researchers have been working for many years on the automatic extraction of 3D structure from image sequences. This is called the structure from motion problem: Given an image sequence of a rigid scene by a camera undergoing unknown motion, reconstruct the 3D geometry of the scene. To achieve this, the camera motion also has to be recovered simultaneously. When in addition the camera calibration is unknown as well, one speaks of uncalibrated structure from motion. In this case the structure of the scene can only be recovered up to an arbitrary projective transformation. In the two-view case the early work was done by Faugeras [7]. They obtained the fundamental matrix as an equivalent for the essential matrix. This matrix completely describes the projective structure of the two-view geometry.

Since then many algorithms have been proposed to compute the fundamental matrix from point matches. Based on these methods, robust approaches were developed to obtain the fundamental matrix from real image data (Torr et al. [16, 17] and Zhang et al. [18]). Relationships between more views have also been studied.

Dense stereo matching:

The structure from motion algorithms only extracts a restricted number of features. Although textured 3D models have been generated from this, the results are in general not very convincing. Often some important scene features are missed during matching resulting in incomplete models. Even when all-important features are obtained the resulting models are often dented.

However once the structure from motion problem has been solved, the pose of the camera is known for all the views. In this case correspondence matching is simpler (since the epipolar geometry is known) and existing stereo matching algorithms can be used. This then allows obtaining a dense 3D surface model of the scene.

Many approaches exist for stereo matching. These approaches can be broadly classified into feature- and correlation-based approaches. Some important feature based approaches were proposed by Pollard, Mayhew and Frisby [19] (all relaxation based methods), and Ohta and Kanade [20] (using dynamic programming). Successful correlation-based approaches were for example Cox et al. [21]. The latter was recently refined by Falkenhagen [22]. Finally [23] proposed a mixed correlation-based and dynamic programming algorithm.

3D reconstruction systems:

It should be clear from the previous paragraphs that obtaining 3D models from an image sequence is not an easy task. It involves solving several complex subproblems.

The resulting models are therefore restricted to a limited number of planar patches. A recent approach developed by Debevec and Malik [25, 24] at Berkeley proved very successful in obtaining realistic 3D models from photographs. A mixed geometric- and imagebased approach is used. The texture mapping is view dependent to enhance photorealism. An important restriction of this method is however the need for an approximate a priori model of the scene. This model is fitted semi-automatically to image features. Shum, Han and Szeliski [26] recently proposed an interactive method for the construction of 3D models from panoramic images. In this case points, lines and planes are indicated in the panoramic image.

By adding constraints on these entities (e.g. parallelism, coplanarity, verticality), 3D models can be obtained. Pollefeys [27] designed a metric 3D reconstruction pipeline, which we will follow to design our system.

Finally some commercial systems exist which allow generating 3D models from photographs e.g. Iwhitness, V. Star and PhotoModeler. These systems require a lot of interaction from the user (e.g. correspondences have to be indicated by hand) and some calibration information. The resulting models can be very realistic. It is however almost impossible to model complex shapes.

III. RESULT AND DISCUSSION

A. Finding corresponding points between images

The reconstruction procedure started with a pair of images as an input “figure 1”, without any prior knowledge about the camera and the scene parameters. The goal of this step is to find the corresponding points between the two images, and the correspondence can be expressed as a disparity value p(x,y) and q(x’,y’). In this step firstly we have used the well known epipolar geometry “figure 2”, to reduce the search space to find corresponding points between the images. The epipolar geometry is used here only to make the correspondence task easier.

Figure 1. The input pair of images

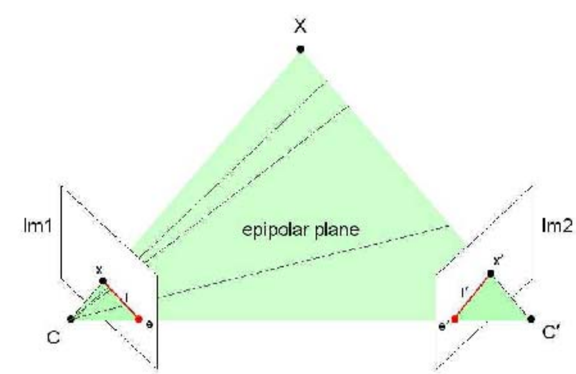

Figure 2. The epipolar geometry concept

The pixels between the pair of images are matched and the result of this process "figure3", allows us to estimate the depth.

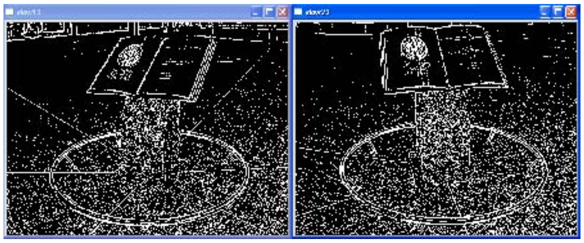

Figure 3. The edge view of our input images

B. Depth estimation

In this step we use the graph cut with dynamic programming approach to extract the correspondence search, first, matching the pixels between the images and linked them, the result of this process "figure 4", lets us estimate the 3D position and the depth of a point

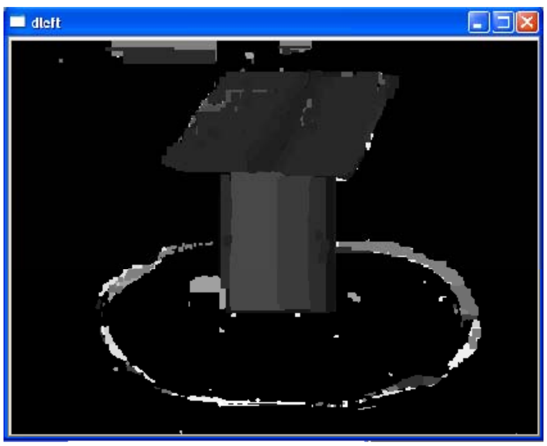

Figure 4. The depth map

IV. CONCLUSION AND FUTURE DIRECTION

The paper has presented an implementation of the 3D reconstruction chain from uncalibrated images. This work shows the possibility to estimate the depth of images taken from different angles. Our work will help in the improvement of the automated 3D reconstruction in its various domains. Even though the result is not perfect, but this work has shown the possibility to obtain a 3D model from uncalibrated images.

The future direction of this study is to develop a complete 3D model from the depth that we have now.

REFERENCES

[1] M. Pollefeys, R. Koch, M. Vergauwen and L. Van. "Automated reconstruction of 3D scenes from sequences of images", ISPRS journal of photogrammetry & remote sensing 55, 251-267, 2000.

[2] U. Topay, "3D reconstruction from uncalibrated images", Master thesis, Dept. of Electrical and Electronics engineering, The graduate school of natural and applied sciences, 2002.

[3] O. Moslah, A. Valles-Such, V. Guitteny, S. Couvet and S. Philipp-Foliguet. "Accelerated Multi-View Stereo using Parallel Processing Capabilities of the GPU's", 3DTV Conference, 2009.

[4] D. Bardsley, "A Practical Framework for 3D reconstruction and its Applications", PhD thesis, School of computer science and IT, University of Nottingham, 2008.

[5] F. Olivier, L. Tuan, S. Maybank, "Camera Self-Calibration: Theory and Experiments, Proceedings of the 2nd European Conference on Computer Vision", 321-334, 1992.

[6] L. Tuan, F. Olivier, "Self-Calibration of a Moving Camera from Point Correspondences and Fundamental Matrices", International Journal of Computer Vision, 22(3), 261-289, 1997.

[7] Z. Cyril, F. Olivier, "Camera Self-Calibration from Video Sequences: the Kruppa Equations Revisited", INRIA, 1996.

[8] A. Fusiello, "Uncalibrated Euclidean Reconstruction: A Review", Image and Vision Computing, 18, 555-563, 2000.

[9] P. Marc, G. Van, "Self-calibration from the absolute conic on the plane at infinity", Proceeding Computer Analysis of Images and Patterns, 175-182, 1997.

[10] P. Marc, G. Van, "A stratified approach to self-calibration", Proceeding 1997 Conference on Computer Vision and Pattern Recognition, 407- 412, 1997.

[11] P. Marc, G. Van, A. Oosterlinck, "The Modulus Constraint: A new Constraint for Self Calibration", Proceeding 13th International Conference on Pattern Recognition, 349-357, 1996.

[12] P. Marc, R. Koch, G. Van, "Self-Calibration and metric reconstruction in spite of varying and Unknown Internal Camera Parameters", Proceeding International Conference on Computer Vision, 90-95, 1998.

[13] C. Schmid and A. Zisserman, "Automatic Line Matching across Views", Proc. IEEE Conference on Computer Vision and Pattern Recognition, IEEE Computer Soc. Press, 666-671, 1997.

[14] C. Harris and M. Stephens, "A combined corner and edge detector", Fourth Alvey Vision Conference, 147-151, 1988.

[15] C. Schmid, R. Mohr and C. Bauckhage, "Comparing and Evaluating Interest Points", Proc. International Conference on Computer Vision, Narosa Publishing House, 230-235, 1998.

[16] T. Moons, "A Guided Tour Through Multiview Relations", Proceeding SMILE Workshop (post- ECCV'98), Lecture

Notes in Computer Science 1506, Springer- Verlag, pp.304-346, 1998.

[17] S. Pollard, J. Mayhew and J. Frisby, "PMF: A Stereo Correspondence Algorithm Using a Disparity Gradient Limit", Perception 14(4), 449-470, 1985.

[18] Y. Ohta and T. Kanade, "Stereo by Intra- and Inter-scanline Search Using Dynamic Programming", IEEE Trans. on Pattern Analysis and Machine Intelligence 7(2), 139-154, 1985.

[19] I. Cox, S. Hingorani and S. Rao, "A Maximum Likelihood Stereo Algorithm", Computer Vision and Image Understanding, Vol. 63, No. 3, May 1996.

[20] L. Falkenhagen, "Depth Estimation from Stereoscopic Image Pairs assuming Piecewise Continuous Surfaces", European Workshop on Combined Real and Synthetic Image Processing for Broadcast and Video Productions, Hamburg, Germany, 1994.

[21] L. Falkenhagen, "Hierarchical Block-Based Disparity Estimation Considering Neighbourhood Constraints". Proceeding International Workshop on SNHC and 3D Imaging, Rhodes, Greece, 1997.

[22] Liang Wang, Miao Liao, Minglun Gong, Ruigang Yang and David Nister, "High Quality Real-time Stereo using Adaptive Cost Aggregation and Dynamic Programming", Third International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT), 2006.

[23] P. Debevec, C. Taylor and J. Malik, "Modeling and Rendering Architecture from Photographs: A Hybrid Geometry- and Image-Based Approach", Siggraph, 1996.

[24] C. Taylor, P. Debevec and J. Malik, "Reconstructing Polyhedral Models of Architectural Scenes from Photographs", Computer Vision - ECCV'96, Lecture Notes in Computer Science, Vol. 1065, vol.II, pp 659-668, 1996.

[25] P. Debevec, Y. Yu and G. Borshukov, "Efficient View-Dependent Image-Based Rendering with Projective Texture Mapping", Proceeding SIGGRAPH '98, ACM Press, New York, 1998.