Overview of 3D Reconstruction Methods Based on Multi-view

Mengxin Li [1], Dai Zheng [2], Rui Zhang[3], Jiadi Yin[4], Xiangqian Tian[5]
School of Information and Control Engineering
Shenyang Jianzhu University
Shenyang, China
e-mail: limengxin@sjzu.edu.cn, zhengdai2014@163.com, zhangrui@sjzu.edu.cn, 1304007591@qq.com, 528321894@qq.com

***Abstract*—3D reconstruction based on multi-view is a hot research topic in computer vision due to more information and wide scope of the view of multiple views. Feature detection and matching, fundamental matrix estimation, camera self-calibration, 3D reconstruction and dense surface reconstruction are key proportions of 3D reconstruction. This paper summarizes the main algorithms in these parts, analyses and compares merits and drawbacks of the methods. The aim is to provide a concise, complete understanding of the subject so that the researchers in this field can have a holistic view of the problem and provide more efficient and unique solutions in the future.**

Keywords- Multi-view 3D Reconstruction; Feature Detection and Matching; Fundamental Matrix Estimation; Camera Self-calibration; Dense Surface Reconstruction

I. INTRODUCTION

Multi-view 3D reconstruction is to obtain 3D model of the scene and scenery objects from a series of images taken by a camera, more commonly known as recovering structure of 3D from the camera motion (SFM), is one of the means to solve the 3D modeling in the field of computer graphics and computer vision. It has very important practical significance to improve the 3D realistic modeling and real time large-scale and complex scene.

At present, people mainly through three ways to obtain the 3D model: the first way directly according to the traditional geometric modeling techniques; the second way is to scan the target by 3D scanning device, and then rebuild the 3D model of it; the third way based on multiple images taken at different angles, using the computer vision theory to reconstruct 3D model of target objects eventually.

In three methods described above, the traditional geometric modeling technology is the most mature. 3DMAX, AUTOCAD and CREATOR are most commonly used software. Using this method can get very precise 3D model, let people better control light and texture, and it has been widely used in machinery and construction these engineering fields, the entertainment industry, such as film and television animation. But using this kind of modeling technology needs a long cycle, a complex operation and skilled operators, for many irregular natural or man-made objects, differences are still exiting with the real scene. 3D scanning equipment such as the laser scanner is widely applied to obtain object geometric model of high precision in reverse engineering, the protection of cultural relics and other fields. However, even if ordinary equipment also cost hundreds of thousands, it's expensive, and normally, the recovery of large-scale 3D scene is often powerless. In addition, the use of 3D scanner for modeling usually can only obtain the 3D point cloud model of the target object and not the color on the surface of it. Compared with other modeling methods, modeling technology based on multi-view can reconstruct large complicated outdoor scene, because photos is easy to get, it doesn't need to spend a lot of effort and skill. What's more, no shape and size requirements for modeling object, only take photos from different angles by a camera, the 3D reconstruction of target is completed by these pictures.

3D reconstruction based on multi-view is an important issue in various applications, scene planning and navigation of autonomous mobile robot. Although, some progresses have been made during the last few decades, still there are no methods satisfying the requirement of high accurate as well as robust results.

II. MAIN RESEARCH CONTENTS

Generally speaking, 3D reconstruction based on multi-view includes several main proportions such as feature detection and matching, fundamental matrix estimation, camera self-calibration, 3D reconstruction, dense surface reconstruction and so on. The basic diagram is shown in Figure1.

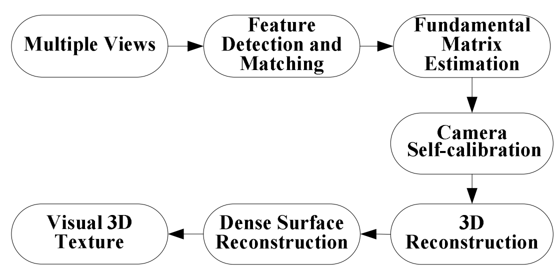

Figure 1. Basic diagram of 3D reconstruction based on multi-view.

A. *Feature Detection and Matching*

Local features of the image usually contain feature edge and feature point. Extracting the feature points as the local characteristics is very common, because the boundary element can't obtain stable performance in the problem of description and matching etc.

978-1-4799-8646-0/15 $31.00 © 2015 IEEE
DOI 10.1109/IHMSC.2015.117

Moravec put forward the feature point extraction algorithm. Then Harris et al. improved it, proposed Harris corner detection algorithm [1] based on image gray value. But this method can only work on an image according to a fixed proportion, two images can't be matched when ratio is changed much, it doesn't have the affine deformation and scale invariance. Difference of Gaussian (DoG) is representative of the scale invariant, affine invariant [2]. Theoretically, Harris operator tends to extract the sharp points of gradient change, and the DOG operator generally extracts the center of the homogeneous regions, we call it spots. Columbia University Lowe proposed Scale-Invariant Feature Transform algorithm (SIFT) in 1999 [3], which is invariant to scale, rotation, brightness change, can maintain a degree of stability to the affine transformation, visual angle variation, noise and so on, has been applied to the image matching field successfully. But in the process of feature extraction, SIFT algorithm needs to construct image pyramid, the dimension of descriptor is high, thus the calculation process is complicated and the matching speed is slow. Later Ke et al. [4] proposed PCA-SIFT detection algorithm, using principal component analysis, the original 128 dimensional vector was reduced to 36 dimension, it reduced the complexity of the SIFT operator. Bay [5] proposed a faster Speeded Up Robust Features (SURF) to replace the SIFT operator, using a similar scale and rotation invariant method combined with the efficient approximation to speed up the calculation, in order to speed up the pace, using the box filtering to replace Gauss filter. Independent Component Analysis (ICA) [6], as the extension of the PCA, it overcame the PCA orthogonal constraint limits, more suitable as a non-Gaussian random signal statistical model, is available for wide baseline matching problem of multi-view images, made a complete 3D model reconstruction with fewer images, can reduce the complexity of reconstruction in time and space.

B. Fundamental Matrix Estimation

Luong's doctoral dissertation proposed fundamental matrix is used to solve the geometric relationship between two images, described the basic properties of epipolar geometry in projective space [7]. Epipolar geometry is also called projective geometry. It describes inner relations between two images, only decided by the intrinsic parameters of camera and relative information of position. The epipolar constraint represented by a 3×3 matrix is called the fundamental matrix. Through epipolar geometry, the problem of 3D reconstruction of images is transformed into robust parameters estimation of fundamental matrix.

The traditional algorithms [8, 9] are seven-point and eight-point method, they are only useful in no noise images. Due to clear geometric significance of epipolar geometric relationship, scholars began to study nonlinear algorithm to solve fundamental matrix, such as minimizing back-projection error or Sampson error. Nonlinear iterative algorithm can handle Gaussian noise well, but for wrong matching points it could do nothing. Therefore, many robust parameter estimation methods have been put forward. For example, Random Sample Consensus (RANSAC) algorithm [10], by reason of its simple implementation, it is widely applied to eliminate the outlier. However, to restore the structure, it is suitable only for two or three images. It doesn't have superiority in the process of tracking in the image sequences, and will miss some outliers. In addition, it depends on the selection of the threshold T and sees interior points' contribution as equal, but in fact interior points have different error, they're not entirely consistent. In recent years, in order to improve the efficiency and performance of the RANSAC algorithm, researchers have proposed many improvement calculations. Maximum Likelihood Estimation by Sample and Consensus (MLESAC) [11, 12], modeling interior points and outliers respectively is its biggest characteristic. It has two shortcomings: one is the outlier modeling's weakness; the other is the accuracy of estimation needs to be improved. In order to improve the precision of the fundamental matrix estimation, Chojnacki et al. [13] introduced covariance information to describe position of the matching point in the fundamental matrix estimation. Chen Ze zhi [14] put forward the weighted normalized linear algorithm and the introduction of new double pole constraint's nonlinear algorithm. The fundamental matrix estimation method based on a genetic algorithm or swarm intelligence has appeared in recent years, but the introduction also increases the complexity of the algorithm.

C. Camera Self-Calibration

Over the years, a large number of self-calibration techniques have been developed for converting a projective reconstruction into a metric one, which is equivalent to recovering the unknown calibration matrices associated with each image.

It is a direct calibration progress depends solely on the relationship of corresponding points between the multiple images. This method is more flexible than the Traditional calibration method (such as DLT, Tsai method) and Calibration method based on active vision except complex calculation and low precision. In recent years, camera self-calibration technique has achieved a great progress.

Faugeras [15] from projective geometry perspective proved there are quadratic nonlinear constraints in each of two images in the form of Kruppa equation, and parameters can be solved by directly solving the Kruppa equation. In the consideration of it's hard to solve, stratified gradually calibration methods were proposed, such as QR Decomposition method [16], Absolute Quadric method of Triggs [17], Silhouette Constraints method of Pollefeys [18]. Pollefeys also gave a new more practical camera self-calibration method under variable intrinsic parameters. Heinrich et al. presented camera self-calibration method [19] for the target of maximum likelihood estimation, the experiment showed that this method is robust and effective, can greatly improve the precision, it is especially suitable for less image and noisy place. Gherardi proposed a practical camera self-calibration method [20], it is fast, easy to realize, has good convergence performance etc. Nowadays, foreign countries have been studying in camera self-calibration with the time-varying intrinsic parameters.

As we know, there are less successful calibration techniques on multi- view. Moreover, existing technologies are limited to linear calibration under specific camera motion. Accordingly, the research on self-calibration based on multi-view without motion constrains is of considerable value in theory and application.

D. 3D Reconstruction

The most accurate way to recover structure and motion is to perform robust nonlinear minimization of the measurement (re-projection) errors, which is commonly known in the computer vision as bundle adjustment (BA).

BA method appeared in 2000. Firstly, it defines cost function about a set of parameters (for example, 3D spot coordinate). The cost function is usually the distance between image points and minimum re-projection points. BA requires a fairly accurate initial result of first image pair constrains to perform incremental reconstruction. Unfortunately, it usually gets initial value by a linear method that doesn't have a geometric meaning and is vulnerable to the outlier disturbance. Another drawback of BA is it's easy to fall into local optimum, because of multiple minima of optimizing the cost function.

For two views, employing corresponding paired points and internal, external parameters calculated with triangulation to produced 3D points. More than three views for 3D reconstruction, the most common method is Gold Standard Algorithm (GSA) based on Levenberg Marquardt (LM) iteration. The objective function of LM method is L_2-norm. It is easy to figure out that solving the L_2-norm for more than two cameras is a hard nonconvex problem, and it is unwarranted to obtain the global optimal value. Some authors proposed methods using L_∞-norm or L_1-norm instead of L_2-norm in minimizing the residual error of measured feature and back-projection of 3D points. Using L_∞-norm method can simplify the cost function and obtain the global optimal solution, it also does not depend on the initial value, the defect is it is sensitive to outlier, and the amount of calculation is large. In addition, Lourakis [21] used a trust domain concept, proposed Dog-Leg algorithm, its performance is better than LM algorithm. Ni [22] used image segmentation to optimize BA for large scale scene reconstruction. Wu proposed a new optimization method of triangle with the application of basic cone and Lagrange multiplier method [23], on the premise of considerable accuracy obtained a suboptimal solution, reduced the complexity of the algorithm, the defect is it's applicable only to the two view situation. Graber presented an interactive recovery 3D scene system [24], the whole reconstruction process is online, and the user may observe the current reconstruction results at any time and get the detail information of specific parts by moving the camera.

E. Dense Surface Reconstruction

In general, matching by extracting feature points can only reconstruct some sparse 3D points cloud and roughly identify the shape of the object. However, in many applications, the visualization is needed. Therefore, dense surface reconstruction is necessary.

Dense stereo methods include traditional direct stereo matching method and the body method proposed recently, but both of them need to compute the dense matching points, furthermore, there are obvious shortcomings, they only reconstruct smooth layer and need large computation, high memory. Lhuillier and Quan et al. [25, 26] proposed a quasi-dense method by adding sparse feature point information on the basis of 2D image information, which overcame the disadvantages of dense matching algorithm, can descript 3D scene accurately and still maintain a high computational efficiency in the same time, it is a preferable solution for surface reconstruction of 3D reconstruction problem. Goesde et al. [27] presented a very time-consuming algorithm based on confidence value. Firstly, get the scene boundary, then calculate normalized cross-correlation values of each depth on line of each pixel and camera optical center reversely, confidence value as well. By adopting the method of volumetric fusion reliable point depth maps to generate the scene surface, in addition to the sparse texture exists loopholes, other parts can be recovered accurately.

In 2009, Furukawa submitted a Patch-based Multi-view Stereo (PMVS) algorithm in literature [28]. It mainly involves three steps:

1) *Matching:* detecting feature points through Harris and DoG operators, then matching these feature points in order to get a series of sparse patches and the corresponding image regions. Giving these initial matches, repeating the next two step N times.

2) *Diffusion:* extending the initial matching to nearby pixels to obtain a series of dense patches.

3) *Filtering:* removing the error matching located before or after the object.

The advantages of PMVS involve that it doesn't need any initialization, and can automatically detect and ignore external point, which is also considered as an accurate, simple and efficient algorithm. Though PMVS has been demonstrated of good performance and ability, it shows several shortcomings when applied to scene reconstruction from a limited number of 3D camera stereo pairs, such as PMVS is dependent on the accuracy of camera calibration results, it employs Harris and DoG operators to extract features which act as seeds for patch definition, however, limited feature representativeness has been found for certain scenes even based on a combinational utilization of these two features, and PMVS has apparent limitations in dealing with sparse stereo pairs, especially for those captured by a narrow-baseline 3D camera. Recently, many scholars applied self-adaption and geometric constraints to improve PMVS algorithm, some people also proposed to use Canny features to replace the DoG and Harris features for the feature extraction part of PMVS.

III. PROBLEMS AND FUTURE WORKS

3D reconstruction based on multi-view has achieved a plenty of results, but there are still some questions need to be solved.

1) *Fundamental matrix estimation:* although there have been some robust algorithms like RANSAC and MLESAC,

which expect to avoid the outlier in the sample to obtain the optimal solution through random sampling way, these approaches don't eliminate outliers completely, and depend on the initial value of fundamental matrix, which may contain outlier data, so the accuracy of initial estimation still needs to be improved.

2) Camera self-calibration: in recent years, the camera self-calibration has made many breakthroughs, but its robustness and accuracy remain to be further improved. There are still defects such as redundant parameters, ill-conditioned equation. If the complete 3D model reconstruction is seen as the goal, the camera self-calibration requires higher accuracy.

3) The surface reconstruction problem of uneven and noisy point cloud data: point cloud obtained by 3D reconstruction based on multi-view are sparse and inaccuracy. Therefore, using pre-exiting surface reconstruction algorithms, such as Poisson surface reconstruction, a number of problems will emerge, for example, the surface holes, pits, humps etc. For uneven point cloud data, more robust surface reconstruction algorithms are needed.

As the academic circles has a deep research on feature detection and matching, there have been many mature algorithms available for use. While the dense surface reconstruction technology is one of the research hotspots in computer vision in recent years, a lot of problems are worth studying. Further study on this aspect will be conducted next.

IV. CONCLUSIONS

In this article, we have summarized many algorithms from five main parts, feature detection and matching, fundamental matrix estimation, camera self-calibration, 3D reconstruction and dense surface reconstruction. 3D reconstruction based on multiple views has an extensive research and application background in computer graphics, computer vision and virtual reality. How to solve the questions existing in it, improve the robustness and performance of the algorithm are still the main problems of 3D reconstruction based on multi-view.

REFERENCES

[1] C. Harris, M. Stephens, A combined corner and edge detector[C]. 4th Alvey vision conference, Manchester, UK, 1988, 15-50

[2] K. Mikolajczyk, C. Schmid, A Performance Evaluation of Local Descriptors, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003, 2: 257-26

[3] Lowe, Object recognition from local scale-invariant features [C], Proceedings of International Conference on Computer Vision, Washington, USA, 1999:1150-1157

[4] Y. Ke, R. Sukthankar, PCA-SIFT: A more distinctive representation for local image descriptors[C], Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington DC, USA, 2004, 2,506-513

[5] H. Bay, A. Ess, T. Tuytelaars, L. V. Gool, Speeded-up robust features (SURF) [J], Computer vision and image understanding, 2008, 110(3): 346-359

[6] P. Comon, Independent Component Analysis, a New Concept? Signal processing, 1994,36(3):287-314

[7] J. Zhou, L. T. Chen, Q. Liu, Random sample consensus algorithm based on sequential probability and local optimization, 2012 [J], Chinese Journal of scientific instrument, 33 (9): 2037-2044

[8] J. Zhou, L. T. Chen, Y. M. Li, Optimize Preview Model Parameters Evaluation of RANSAC [J], International Journal of Advancements in Computing Technology, 2012, 4(16): 18-25

[9] H. Longuet-Higgins, A computer algorithm for reconstructing a scene from two projections [J], Nature, 1981, 293(10):133-13

[10] M. A. Fischler, R. C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography [J], Communications of the Association for Computing Machinery, 1981, 24(6): 381-3

[11] P. H. S. Torr, A. Zisserman, MLESAC: a new robust estimator with application to estimating image geometry [J], Computer Vision and Image Understanding, 2000, 78(1): 138-15695

[12] B. J. Tordoff, D. W. Murray, Guided-MLESAC: faster image transform estimation by using matching priors [J], IEEE Transactions on Pattern Analysis and Machin Intelligence, 2005,27(10): 1523-1535

[13] W. Chojnacki, M. J. Brooks, A. V. D. Hengel, D. Gawley, On the fitting of surfaces to data with CO variances [J], IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(11): 1-10

[14] Z. Z. Chen, C. K Wu, A high precision estimation of the fundamental matrix of the linear algorithm M, Journal of software, 2002, 13 (4): 840-84

[15] O. Faugeras, What can be seen in three dimensions with an uncalibrated stereo rig? [C], Proceedings of the Second European Conference on Computer Vision, Santa Margherita Liguere, Italy, 1992, 563-578

[16] R. Hartley, Estimation of relative camera positions of uncalibrated cameras. In: Proceedings of European Conference on Computer Vision, NLCS588, Springer-Verlag, 1992, 579-387

[17] R. Hartley, Eucliden reconstruction and invariants from multiple images, IEEE Trans on Pattern Analysis and Machine Intelligence, 1994, 16(10):1036-1041

[18] M. Pollefeys, L. Van Cool, A. Osterlinck, The modulus constraint: A new constraint for self-calibration, In: Proceedings of International Conference of Pattern Recognition, Vienna, 1996, 349-353

[19] S. B. Heinrich, W. E. Snyder, J-M. Frahm, Maximum likelihood autocalibration [J], Image and Vision Computing, 2011, 29: 653-665

[20] R. Gherardi, A. Fusiello, Practical autocalibration [C], Proceedings of the 11th European Conference on Computer vision, Berlin Heidelberg, Germany, 2010: 790-801

[21] M. Lourakis, A. Argyros, Is levenberg-marquardt the most efficient optimization algorithm for implementing bundle adjustment? [C], 10th IEEE International Conference on Computer Vision, Beijing, China, 2005, 2, 1526-1531

[22] K. Ni, D. Steedly, F. Dellaert, Out-of-core bundle adjustment for large-scale 3d reconstruction [C], IEEE 11th International Conference on Computer Vision, Rio de Janeiro, Brazil, 2007, 1-8

[23] F. C. Wu, Q. Zhang, Z. Y. Hu, Efficient suboptimal solutions to the optimal triangulation [J], International Journal of Computer Vision, 2011, 91(1): 77-106

[24] G. Graber, T. Pock, H. Bischof, Online 3D reconstruction using convex optimization [C], Proceedings of IEEE 13thD. International Conference on Computer Vision Workshops, 2011: 708-711

[25] M. Lhuillier, L. Quan, A quasi-dense approach to surface reconstruction from uncalibrated images [J], IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(3): 418-433

[26] L. Quan, P. Tan, G. Zeng, et al, Image-based plant modeling [J]. ACM Transactions on Graphics, 2006, 25(3): 599-604

[27] M. Goesele, B. Curless, S. M. Seitz, Multi-view stereo revisited [C], Proceedings of IEEE Computer Vision and Pattern Recognition, New York, USA, 2006, 2: 2402-2409

[28] Y. Furukawa, J. Ponce, Accurate, dense, and robust multi-view stereopsis [J], IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(8): 1362-1376