Improving PMVS Algorithm for 3D Scene Reconstruction from Sparse Stereo Pairs

Bo Li[1], Y.V. Venkatesh[2], Ashraf Kassim[3], and Yijuan Lu[1]

[1] Department of Computer Science, Texas State University, San Marcos, USA 78666
[2] Interactive and Digital Media Institute, NUS, Singapore 119613
[3] Department of Electrical and Computer Engineering, NUS, Singapore 119613
{b_l58,lu}@txstate.edu, yv.venkatesh@gmail.com, ashraf@nus.edu.sg

Abstract. 3D scene reconstruction resulting from a limited number of stereo pairs captured by a 3D camera is a nontrivial and challenging task even for current state-of-the-art multi-view stereo (MVS) reconstruction algorithms. It also has many application potentials in related techniques, such as robotics, virtual reality, video games, and 3D animation. In this paper, we analyze the performance of the **PMVS** (Patch-based Multi-View Stereo software) for scene reconstruction from stereo pairs of scenes captured by a simple 3D camera. We demonstrate that when applied to a limited number of stereo pairs, PMVS is inadequate for 3D scene reconstruction and discuss new strategies to overcome these limitations to improve 3D reconstruction. The proposed Canny edge feature-based PMVS algorithm is shown to produce better reconstruction results. We also discuss further enhancements using dense feature matching and disparity map-based stereo reconstruction.

1 Introduction

3D reconstruction of a scene from multiple stereo views is called image-based modeling (**IBM**) technique, which is important for diverse applications ranging from robotics vision, electronic earth maps, and virtual reality to 3D film production, computer games and animation. However, it is known that depth information is lost when 2D images of a 3D world scene are captured by a single camera or multiple cameras. In a single camera image, the depth information is implicit (in the form of shading, shadows, texture and others). In contrast, in two (i.e., binocular) or more (i.e., multi-view) camera images, the same information is explicitly available in what is called *disparity* between pixels in a binocular stereo pair and also in multi-view stereo images, which can be organized in a sequence as stereo pairs. One example demonstrating the disparity information existing in stereos, captured using our 3D camera, is shown in Fig. 1. Therefore, based on the geometry of projection of a 3D scene as a pair of 2D images using a 3D camera, which can be modeled as a pair of pinholes, we can relate the 3D scene point to (i) the disparity between the corresponding points that are projections of the same 3D scene point on the 3D camera images; and (ii) focal lengths (expressed in terms of pixel sizes) of the two lenses and other *intrinsic*

B. Huet et al. (Eds.): PCM 2013, LNCS 8294, pp. 221–232, 2013.

(a) left stereo (b) right stereo (c) overlaid stereo

Fig. 1. An example of disparity information as the foundation of multi-view stereo techniques. Left stereo (a) is coded in red and right stereo (b) is coded into cyan color in (c). 3D depth information can be easily and clearly perceived from the feature stereo pairs by utilizing a cyan-red filter (goggles).

and *extrinsic* parameters of the 3D camera. When multiple 3D cameras are employed (or the same 3D camera is used at multiple locations), this relationship between disparity and 3D scene information can be extended to other binocular pairs. For diagrammatic and theoretical details, please see [6], for instance.

In this paper, our primary goal is reconstruction of a 3D scene from sparsely sampled stereo pairs captured by a single 3D camera in different locations. We present our experimental findings related to the application of the PMVS (Patch-based Multi-View Stereo software) [4] [5] package and its modified version, meant explicitly for *self-calibration* of cameras and scene reconstruction from multiple camera images, to stereo pairs of a laboratory scene having finite depth, as captured by a 3D camera in a limited number of positions. This is an interesting but challenging problem and promising for related applications, i.e., 3D robot vision and real-time video surveillance, which require an efficient and effective 3D scene reconstruction based on only several stereo pairs. Our major findings and contributions are as follows:

- PMVS is inadequate for a complete, accurate and robust 3D scene reconstruction of typical laboratory scenes based on sparse stereo pairs, due to the nature of the extracted Difference-of-Gaussian (DoG) [9] and Harris [7] features used for stereo correspondence.
- For an indirect verification of the relevance of the DoG+Harris features (currently employed in PMVS) to disparity (and hence scene-depth) computation, we have stereo-perceptually analyzed the feature stereo pairs by using the cyan-red filter (goggles) and found that a significant amount of depth perception, and hence of depth information existing in the original stereo pair of images, have been lost.
- In contrast, our experimental results demonstrate that feature stereo pairs obtained from an application of the Canny-operator to the original stereo pairs do retain a significant amount of perceptual depth information of the original stereo pairs. Moreover, these features lead to denser and more accurate feature matches, in contrast with those based on Harris+DoG operator-based features. Better 3D scene reconstruction results also have been obtained and demonstrated on three experimental scenes.

The rest of the paper is organized as follows: Section 2 reviews the related work. An overview of PMVS is presented in Section 3. In Section 4, we present and evaluate the results of PMVS-based scene depth estimation and reconstruction. By way of overcoming the limitations of PMVS, in Section 5 we demonstrate better results arising from a choice of features different from those used in PMVS, that is Canny detector-based features, as well as a proposal of dense matching and disparity-based stereo reconstruction algorithm which is also based on the Canny features for further improvement. We conclude the paper and list the future work in Section 6.

2 Related Work

The literature on 3D scene reconstruction is too vast to be reviewed here satisfactorily. For a taxonomy of algorithms for 3D scene reconstruction from multi-view stereo pairs, please see [12], which builds a benchmark composed of high-quality calibrated multi-view stereo images, ground truth 3D models as well as evaluation methodology. In contrast, without any prior knowledge of the camera or the scene, recent literature deals with the estimation of camera parameters simultaneously with the 3D depth map of the scene for 3D scene reconstruction from one or more image pairs, with or without stereo. This is called the *self-calibration* approach of which bundle adjustment (**BA**) [13] is an example. While this approach has been in use in photogrammetry, its application in computer vision is very recent. It is used to refine scene reconstruction by simultaneously estimating both the 3D structure of the scene and the projection matrices that match its corresponding images in a chosen optimal (i.e., least squares) sense.

Alternative techniques for multi-view 3D scene reconstruction are: (1) the patch-based multi-view (**PMVS**) software [4] [5] which requires camera parameters and outputs a set of oriented points, where both the 3D coordinate and the surface normal are estimated at each oriented point; (2) surface normal estimation and best viewpoint selection as found in [14]. For establishing correspondence among the images, other strategies include photo-consistency [8] as a similarity measure, and correlation-based metrics, which are more robust due to their invariance to global scaling of intensity.

3 Overview of PMVS

PMVS (Patch-based Multi-View Stereo software) has camera parameters for a set of views and possibly view visibility information (about the images that can see common 3D points; used to speed-up computations) as input and reconstruct the 3D structure information based on patch-based multi-view stereopsis. The framework of the PMVS is described in Fig. 2.

PMVS extracts two features using Difference-of-Gaussian (DoG) [9] and Harris [7] operators. The DoG operator performs edge detection by subtracting the results of two Gaussian blurs with different radii. As an approximation of Laplacian of Gaussian, it can filter out high frequency signals such as random noise,

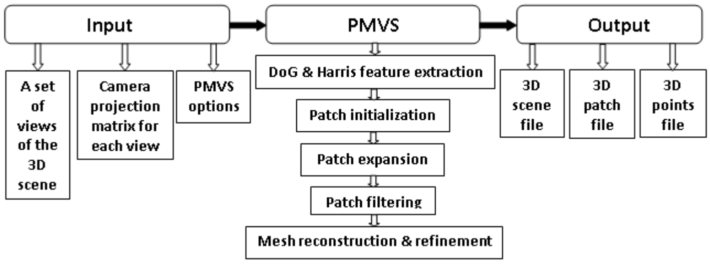


Fig. 2. Frameworks of PMVS

thus making it suitable for processing a highly noisy image. The Harris operator employs local auto-correlation operation to improve the edge consistency by extracting both the edge and corner features of an image. The Harris response is positive in corner regions, negative in edge regions, and small in flat regions. Further, in the PMVS, with image sample points as seeds, epipolar lines are used to decide the corresponding region (within 2x2 pixel) in another image, thereby generating patches (each defined by its center, normal, and visibility) to meet constraints on visibility, and leading to patch-based correspondence between images. Multi-view matching in the PMVS is patch-based and depends on mean photo-consistency of all *visible* pairs. A patch is reconstructed by maximizing the mean value of photo-consistency, and then accepted only if the number of visible images is greater than or equal to three.

PMVS has several input parameters that have different options of choice, such as decomposition level (*level*), cell size (*csize*), threshold value (*threshold*) for photometric consistency measure, window size (*wsize*), and minimum number of images (*minImageNum*) that a 3D point must be visible in. PMVS output comprises the followings: (1) reconstructed 3D scene geometry/model, i.e., 3D coordinates, normal and color data of a set of reconstructed 3D points; (2) patch information (center, normal, and visibility image set); and (3) a 3D point set listing the coordinates and normals of the 3D points, and it is generated as the input of two candidate post-processing algorithms after PMVS: Poisson-based 3D surface reconstruction (PSR) algorithm [5] and Visual Hull (VH) reconstruction [5].

4 PMVS-Based 3D Scene Reconstruction: Evaluations and Limitations

In this section, we target identifying the limitations of PMVS when applied to perform 3D scene reconstruction on a sparse set of stereo pairs captured in a controlled environment.

(a) Scene 1 (b) Scene 2 (c) Scene 3

Fig. 3. Sample images of three 3D scenes

(a) Left stereo (b) Right stereo (c) Left stereo (d) Right stereo

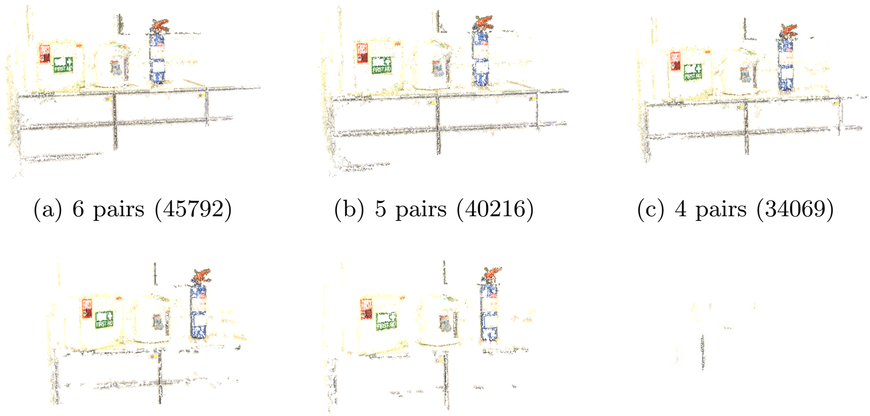
Fig. 4. Sample stereo pairs of Scene 1 from two locations: (a) and (b) are for one pair; (c) and (d) are for the other

4.1 3D Data

We used a 3D camera ("FUJIFILM REAL 3D W1") which has two Fujinon optical zoom lenses (left and right) to capture stereo pairs in a controlled (i.e., measurable) laboratory environment. We captured typical 3D scenes in a laboratory with objects with/without textural information at varying depths in both 2D JPEG image and 3D MPO (Multi Picture Object) file formats. The MPO files are then processed to generate the 3D stereo pairs (left and right). The resulting images obtained for various scenes are shown in Fig. 3. Fig. 4 shows two stereo pairs of Scene 1 from two different locations. Data from Scene 1, 2 and 3 contain 6, 13 and 15 3D stereo pairs sampled roughly in $\frac{1}{4}$ circle, respectively.

4.2 Evaluations

We evaluate PMVS with our captured 3D stereo images with respect to its following two aspects: (1) sparsity degree of 3D stereo pairs; (2) comparison with 2D image-based reconstruction results. Unless stated otherwise, we adopt the PMVS default parameter options in [4] [5], such as *level*, *csize*, *threshold*, *wsize*, and *minImageNum*. We also need to mention that in order to have a comparison based on a common reconstruction process, we only compare the main PMVS reconstruction results till the step of patch filtering (see Fig. 2) and do not consider the final step of mesh reconstruction which still has several options for different types of scenes. We also have found that sometimes the initial scene reconstruction based on the Poisson surface software is unsatisfactory, especially when the reconstructed 3D point set is sparse. It is not easy to tune the parameters as well. It seems to perform well when the surface is a flat plane but not when dealing with curved surfaces.



(a) 6 pairs (45792) (b) 5 pairs (40216) (c) 4 pairs (34069)

(d) 3 pairs (23496) (e) 2 pairs (14283) (f) 1 pair (577)

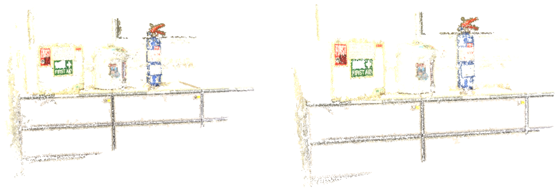
Fig. 5. PMVS reconstruction results comparison on Scene 1 with respect to the sparsity degree of 3D stereo pairs. The digit in a parenthesis is the number of vertices of the 3D reconstruction result. (f) shows the reconstruction result when setting $minImageNum$ to be 2 since there is only two images.

Evaluation of PMVS with Respect to the Sparsity Degree of 3D Stereo Pairs. Take Scene 1 as an example, starting from 6 pairs, we sequentially reduce the number of pairs from the end of the view list (Section 4.1) to only 1 pair, that is, testing 6, 5, 4, 3, 2 and 1 pairs. Fig. 5 compares their PMVS reconstruction results. From the above example on Scene 1 as well as on the other two scenes we have tested, we have found that PMVS has apparent limitations in dealing with sparse stereo pairs. For example, it cannot reconstruct a meaningful 3D scene geometry for Scene 1 if only using the first captured pair and for Scene 2, it cannot do that even with first two pairs. At least 4$\sim$5 pairs are needed to reconstruct a relatively complete 3D scene for these two cases.

More pairs and denser view sampling will improve the reconstruction quality. However, the computational complexity will also increase rapidly. For example, Table 1 lists the computational time of Scene 1 with respect to different numbers of input pairs using a computer with an AMD Opteron processor 6174, with 2.19 GHz clock and 16.0 GB memory on a Windows 7 64-bit operating system. It is also found that the patch expansion process takes up a large part (about 90%) of the total computational time. We also note that a more efficient algorithm named Clustering Views for Multi-View Stereo (CMVS) [3] based on PMVS and view clustering is available. However, it mainly concentrates on a large number

Table 1. Computational time required for 3D scene reconstruction of Scene 1

Number of stereo pairs	1	2	3	4	5	6
Computational time (sec)	15	155	309	342	350	416



(a) six 3D stereo pairs (b) six 2D images

Fig. 6. Comparison of reconstruction results using of Scene 1

of views and still does not solve the issue of sparse view sampling, which in fact is a general limitation of patch-based reconstruction techniques. Similar situation is with [2] which is also based on PMVS but especially deals with reconstruction of a scene which comprises highly structured objects with little textural information. Therefore, the inadequacy of PMVS to reconstruct a 3D scene using a limited number of 3D camera stereo pairs seems to be *largely* due to the fact that, in the course of patch-based optimization, a patch is accepted only when the number of *visible* views is greater than or equal to three.

Evaluation of PMVS with Respect to the 3D Stereo Pairs versus 2D Images. As mentioned in Section 4.1, at one position of a camera we also obtained one 2D JPEG image. Based on the 2D images of a scene, we perform PMVS-based 3D reconstruction and compare the obtained result with the one based on the 3D pairs. Fig. 6 shows the comparison for Scene 1. It seems that using 2D images directly will obtain at least comparable results as using 3D pairs. However, the number of images employed is reduced by half, thus the reconstruction speed is much faster. We think the main reason for this is because of the small difference in the two baselines of a pair, that is *narrow-baseline*. The 2D correspondence is liable to errors because of the small differences in the coordinates of corresponding feature points in two stereo pair images. In fact, the authors of PMVS [5] also mentioned that one limitation of the algorithm is that the reconstruction is unsatisfactory for narrow-baseline stereo pairs. Therefore, rather than estimating the depth information, we recommend disparity-based reconstruction using a dense stereo matching framework [11].

4.3 Limitations

Though PMVS has demonstrated good performance and ability in their paper [5], it shows several shortcomings when applied to scene reconstruction from a limited number of 3D camera stereo pairs.

(1) PMVS is dependent on the accuracy of camera calibration results. We have found that SIFT feature-based Bundler camera calibration algorithm has a limitation on sparse view sampling and usually the camera estimation results are inaccurate when the input views are sparse or lack texture features. Therefore, a

more robust and accurate camera estimation technique should be considered to provide the important camera parameters input for the PMVS reconstruction.

(2) PMVS 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. The selection of two Gaussian blur radii for the DoG operator and the response threshold for the Harris operator are subject to the influences of manual settings. A feature extractor (like the Canny edge detection based on automatic thresholding) which avoids manual settings seems to be better suited for implementation in the PMVS package.

(3) PMVS has apparent limitations in dealing with sparse stereo pairs, especially for those captured by a *narrow-baseline* 3D camera. Our finding is that at least four to five pairs are needed to reconstruct a relatively complete 3D scene for these two cases. Compared to 2D images-based reconstruction, 3D stereo pairs-based reconstruction does not achieve an apparently better quality. Both of the above two disadvantages limit the scope of PMVS applications.

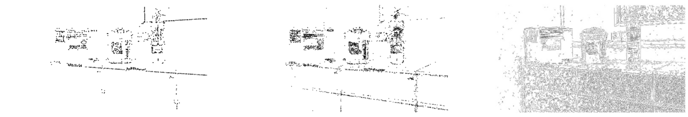
5 PMVS Improvement: Canny-Based PMVS

Motivated by the above limitations found during the results analysis of applying PMVS (together with Bundler) to our sparse stereo pairs-based 3D scene reconstruction scenarios, we propose to use Canny features to replace the DoG and Harris features for the feature extraction part of PMVS and have achieved better quality for the reconstruction results for the three sample scenes. Based on this, we further propose a preliminary disparity-based stereo reconstruction algorithm which also utilizes Canny features, and it serves as our next work.

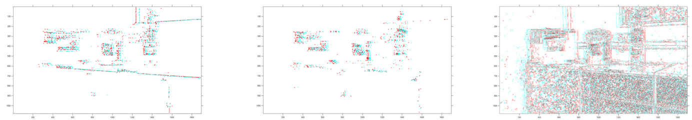
5.1 Disparity Analysis of Different Features

Feature extraction is a very important module of a multi-view stereo reconstruction algorithm. PMVS employs Harris and DoG features. However, it lacks a justification of their feature selection based on a sound analysis of their contributions to the quality of reconstruction results, that is to say, the corresponding relationship between the 2D features selected and the 3D features reconstructed. For example, (1) What 3D features will be highlighted and which 3D features will be missing due to the limitations of the 2D features selected? (2) Whether the 2D features selected are representative enough for the 3D reconstruction? In other words, do the various features contain enough information to facilitate computation of a depth map for the scene and hence the 3D scene reconstruction? (3) Since the 2D feature point sets are extracted and selected as the seeds for the patch definition, then what impact they will have on the accuracy of the patch initialization and expansion till to the final results?

Motivated by the above questions, we compare different candidate features, such as the above mentioned Harris+DoG features, as well as Canny (using automatic thresholding) and SIFT features, in terms of the completeness of the



(a) Harris+DoG features (b) SIFT features (c) Canny features



(d) Harris+DoG feature stereo (e) SIFT feature stereo (f) Canny feature stereo

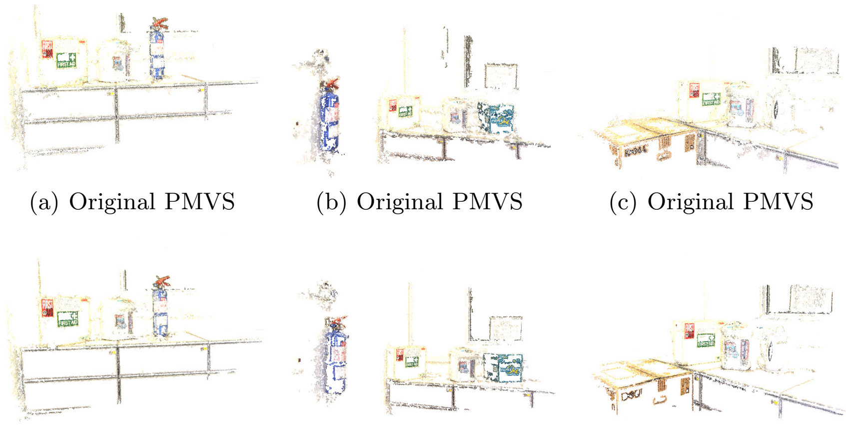
Fig. 7. (a)$\sim$(c): Harris+DoG, SIFT, and Canny features of an example stereo image of Scene 1. (d)$\sim$(f): SIFT-, Harris+DoG- and Canny operator-based overlaid corresponding feature stereo pairs of Scene 1.

disparity information they contain. Please note that *disparity information is the foundation for depth map estimation and hence 3D scene reconstruction*. Our intention to perform the comparison is based on the following assumption: a 2D feature which shows more complete and discernible disparity information for diverse objects in a 3D scene will improve the accuracy and robustness of the 2D feature correspondence, then the 3D triangulation which is directly based on the correspondence results, thus the final reconstructed 3D scene.

Fig. 7 (a)$\sim$(c) compare the three sets of features for the left stereo image of an example stereo pair of Scene 1 and Fig. 7 (d)$\sim$(f) are for a comparison of the overlaid corresponding feature stereo pairs. In fact, we can even view *stereoscopically* the various feature stereo pairs of Fig. 7 (d)$\sim$(f) using the cyan-red filter, and discover that Canny operator-based feature stereo pair contains significantly more depth information of the 3D scene than either the SIFT or Harris+DoG feature stereo pair. More specifically, for the scene under consideration, the number of DoG+Harris feature points in the left image is 1194; and in the right, 1195. For the SIFT operator, the corresponding numbers are 3715 and 3308, respectively; and for the Canny operator, they are 210922 and 194928. Thus, a much more denser feature pair will be obtained if based on Canny operator-based feature extraction. Then, a better disparity map and more accurate reconstruction result can be also expected.

5.2 Canny-Based PMVS

The stereo correspondence algorithm computes disparity at the feature points from which the disparity map can be generated. It has been found that (a) the disparity map computed from matching the Canny operator-based stereo pair is much *denser* than the one obtained from a similar matching operation on either SIFT or DoG+Harris operator-based stereo pair; and (b) the Canny



(a) Original PMVS (b) Original PMVS (c) Original PMVS

(d) Canny-based PMVS (e) Canny-based PMVS (f) Canny-based PMVS

Fig. 8. Comparison of reconstruction results, for Scene 1∼3 respectively, using (a) Original PMVS (first row); and (b) Canny features as input to PMVS (second row)

feature-based PMVS algorithm, which we named **Canny-based PMVS**, gives a *denser* depth map and more complete boundaries of objects in the scene, in contrast with the sparse depth maps generated from SIFT- and DoG+Harris operator-based PMVS algorithms.

It should be noted here that, unlike the SIFT or the Harris+DoG operator, the Canny edge detector is based on an automatic threshold selection according to the content of the images, thus it does not involve manual assignment for the parameter selection. We also need to mention that since our Canny edge detection is directly performed on the original images, we need to set the parameter *level* to zero, which means we do not consider the multi-level texture features as in PMVS. However, we will demonstrate that even on a single level for the feature extraction, we can achieve better reconstruction results. Finally, Fig. 8 compares the reconstruction results of the original PMVS and the proposed Canny features-based PMVS for the three scenes.

As can be seen, we have the following findings: (1) Canny-based PMVS algorithm achieves cleaner and more salient edge reconstruction results; (2) More 3D points can be reconstructed if using Canny edge features compared to the SIFT or Harris+DoG features, therefore we have obtained a denser depth map. Please find out the details about the above findings by zooming in on the above figures. For example, stronger and apparent edge information can be found, such as the "***FIRST***" label in Fig. 8 (d)∼(f), the "***DELL***" label and the box object in (f), and the fire extinguisher in (d) and (e). It is likely that a combination of Harris+DOG and Canny features as inputs to PMVS will further improve the results of the Canny-based PMVS algorithm.

Canny Feature and Disparity-Based Stereo Reconstruction Algorithm. As demonstrated above, Canny-based PMVS has achieved better 3D reconstruction quality after utilizing Canny features to replace the Harris+DoG feature. However, as a limitation demonstrated and analyzed before (see Sections 4.2∼4.3), PMVS does not perform comparably well for *narrow-baseline* stereo pairs. Thus, rather than using a patch-based technique PMVS, a dense feature matching and disparity-based technique is more promising to further apparently improve the 3D reconstruction quality and robustness. Motivated by this, we propose an initial Canny feature and disparity-based stereo reconstruction algorithm.

(1) Canny edge feature detection. This is to obtain a set of feature points existing in the left and right images of a pair.

(2) Dense feature matching-based disparity map estimation. We consider all the edge points and utilize a dense feature matching algorithm specially designed for narrow-baseline pair correspondence, such as [10]. Then, we construct a dense disparity map based on the feature correspondence results.

(3) Fundamental matrix computation. Utilizing the matching points and disparity map, we compute the Fundamental matrix **F** based on the RANSAC [1] method (in a loop manner). For implementation, we can refer to the OpenCV function of *cvFindFundamentalMat()*.

(4) Projection matrix computation. We compute the left and right projection matrices **P1** and **P2** from the fundamental matrix **F** [6].

(5) 3D point reconstruction based on triangulation. For this, we can refer to the OpenCV function of *cvTriangulatePoints()*.

6 Conclusions and Future Work

We have first presented results when applying the PMVS package to a limited number of stereo image pairs of typical laboratory scenes captured by a 3D camera. Then, the limitations of PMVS have been identified with respect to the application scenario of sparse stereo pairs-based 3D scene reconstruction. It is demonstrated that the features (Harris+DoG for PMVS, SIFT for Bundler), which are presently extracted as an integral part of the package, lead to not only unsatisfactory camera calibration parameters but also loss of many details in depth map estimates. In contrast, it is shown that Canny operator-based edge features extracted from the stereo pairs retain more depth information (than other features extracted in the PMVS package), and generate denser depth maps, which is important to generate better scene reconstruction. Thus, a Canny feature-based PMVS algorithm has been proposed and better 3D reconstruction results have been achieved on the same example scenes. To further overcome the limitations of the patch-based technique PMVS and SIFT features-based Bundler, we further propose a preliminary disparity map-based stereo reconstruction algorithm based on Canny features and dense feature mapping. Both proposed approaches are promising for related applications which require effective 3D scene reconstruction from a set of sparsely sampled pairs. As the main future work, we plan to further improve the Canny-based PMVS algorithm, as well as its descendent algorithm presented in Section 5.2, for the challenging problem of constrained sparse stereo pairs-based 3D scene reconstruction.

Acknowledgements. This work is partially supported by the Texas State University Research Enhancement Program (REP) and Army Research Office grant W911NF-12-1-0057 to Dr. Yijuan Lu, and the project Grant NRF2007IDM-IDM002-069 on "Life Spaces" from the IDM Project Office, Media Development Authority of Singapore. The authors wish to express grateful thanks to Professor Lawrence Wong (Deputy Director, IDMI) for his keen interest and encouragement.

References

1. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM 24(6), 381–395 (1981)
2. Furukawa, Y., Curless, B., Seitz, S.M., Szeliski, R.: Manhattan-world stereo. In: CVPR, pp. 1422–1429. IEEE (2009)
3. Furukawa, Y., Curless, B., Seitz, S.M., Szeliski, R.: Towards internet-scale multi-view stereo. In: CVPR, pp. 1434–1441. IEEE (2010)
4. Furukawa, Y., Ponce, J.: Accurate camera calibration from multi-view stereo and bundle adjustment. International Journal of Computer Vision 84(3), 257–268 (2009), `http://homes.cs.washington.edu/~furukawa/research/pba/`
5. Furukawa, Y., Ponce, J.: Accurate, dense, and robust multiview stereopsis. IEEE Trans. Pattern Anal. Mach. Intell. 32(8), 1362–1376 (2010)
6. Harltey, A., Zisserman, A.: Multiple view geometry in computer vision, 2nd edn. Cambridge University Press (2006)
7. Harris, C., Stephens, M.: A combined corner and edge detector. In: Proc. of Fourth Alvey Vision Conference, pp. 147–151 (1988)
8. Kutulakos, K.N., Seitz, S.M.: A theory of shape by space carving. International Journal of Computer Vision 38(3), 199–218 (2000)
9. Marr, D., Hildreth, E.: Theory of edge detection 207(1167), 215–217 (1980)
10. Pollefeys, M., Gool, L.J.V.: Visual modelling: from images to images. Journal of Visualization and Computer Animation 13(4), 199–209 (2002)
11. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal of Computer Vision 47(1-3), 7–42 (2002)
12. Seitz, S.M., Curless, B., Diebel, J., Scharstein, D., Szeliski, R.: A comparison and evaluation of multi-view stereo reconstruction algorithms. In: CVPR (1), pp. 519–528. IEEE Computer Society (2006)
13. Triggs, B., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A.W.: Bundle adjustment – A modern synthesis. In: Triggs, B., Zisserman, A., Szeliski, R. (eds.) ICCV-WS 1999. LNCS, vol. 1883, pp. 298–372. Springer, Heidelberg (2000)
14. Zabulis, X., Daniilidis, K.: Multi-camera reconstruction based on surface normal estimation and best viewpoint selection. In: 3DPVT, pp. 733–740. IEEE Computer Society (2004)