Multi-view Stereo Dense Reconstruction using SfM-based PMVS for Image-based Modeling (IBM)

Cho Cherry Aung
University of Computer Studies, Yangon (UCSY)
Yangon, Myanmar
chocherryaung@ucsy.edu.mm

Thin Lai Lai Thein
University of Computer Studies, Yangon (UCSY)
Yangon, Myanmar
tllthein@ucsy.edu.mm

***Abstract*—This paper intends to represent the effective Multiview stereo algorithm based on the image-based modeling (IBM). The system consists of three steps: patch initialization, expansion and filtering. In patch initialization, this approach takes the corresponding camera parameters together with sparse 3D points as inputs. The purpose is to introduce a framework that employs camera calibrating and 3D points from the results of structure-from-motion (SfM) method instead of Harris corner detector and Difference-of-Gaussians (DoG) in feature detection and matching step to initialize the patch. The patch expansion reconstructs at least one patch in every cell of the image. The patch expansion stage may contain outliers. Consequently, they remove the outlier patches in the filtering stage. These patch expansion and filtering are then iteratively implemented until getting the respectable and complete output. The experiments of our proposed framework on various datasets and comparisons between the other method are presented in this paper.**

Keywords— Multiview Stereo (MVS), Patch-based Multiview Stereo (PMVS), Image-based Modeling (IBM), Structure-from-Motion (SfM)

I. INTRODUCTION

Currently, the three-dimensional modeling is a vital component, automatically obtains the intuitive 3D appearance of objects and scenes from the multiple photos or the successive video frames. Accurate 3D modeling from calibrated cameras is indispensable in computer vision but challenging and long studied topic in 3D reconstruction. Archaeology and architecture are beneficial applications for developing and comparing of 3D reconstruction methods. Their applications range from design and prototyping to digital museum and virtual museums, restoration and preservation of cultural heritage, and computer video games. Nowadays, image-based modelling (IBM) emerges as a separate branch in computer vision and provide a fast way of capturing accurate 3D content with low cost. In this paper, we exploit the detailed geometry of 3D objects using IBM in which a scene or object can be observed from an arbitrary view using multiple images.

Multiview stereo (MVS) is one of the most well-known and accurate methods for accurate three-dimensional reconstruction when having enough number of viewpoints. MVS generate a dense, homogeneous, complete point clouds, and need to be robust so that the images with varying content, resolution, and contrast can be processed into one unified model. In addition, MVS should have the following properties. It should be fast and allow the reasonable processing time for large image series; and geometric accuracy is also vital according to the resolution and quality of the input images allow.

According to the previous studies and surveys, they can be classified into four ways: (1) Voxel-based approaches [1,2,3] need the knowledge of a bounding box containing the scene, and their accuracy may vary depending on the resolution of the voxel grid; (2) Deformable polygonal meshes algorithms such as a visual hull model [4, 5], which need a good starting point to initialize their corresponding optimization process, are more suitable for small objects and not great for large scenes. (3) Depth map -based approaches [6, 7] first extract the depth map of each stereo image pairs and combine these individual results into single model. Their key challenge is how to combine them into a single 3D model and they depend on the type of noise present in the depth map. Consequently, (4) patch-based approaches [8] are simple and effective, and produce the scene surfaces by groups of small patches.

In [12], C. Leung et. al. introduced a new 3D reconstruction approach via the design of a 3D voxel volume and the system must ensure the image details and the camera geometry are contained in one feature space. Their system was dependent on the accuracy of the 3D and 4D voxel volume construction. C.Strecha et. al. [13] presents the depth-map reconstruction algorithm which combined sequence of the depth map, recorded the z-depth based on their cameras for 3D coordinates refinements and outliers removing. The patch-based multi-view approach [8] outperforms all other approaches in terms of both correctness and completeness, and the surfaces are characterized by a collection of patches.

This system offered a multi-view dense point cloud reconstruction based on the IBM. In MVS, feature extraction and matching are the key factors for accurate and efficiency of automatically 3D reconstruction and remains the key issues. In this paper, we make an enhanced method using the patch-based approaches. Patch-base Multiview Stereo (PMVS) reconstructs a 3D point cloud of a scene using a set of points, i.e., patches, and consists of the 3 steps: initial reconstruction, patch expansion, and filtering. The first step reconstructs sparse 3D points from the different view images. They find the set of features on each image using Harris corner detector and Difference-of-Gaussians (DoG). The Harris Corner Detector attempt to find the corner regions in the image with large variation in intensity in all the directions. The corresponding feature points between the images are matched by using epipolar constraint and normalized cross-correlation (NCC). They reconstruct the sparse 3D point cloud from these corresponding feature points. For dense reconstruction, the second step increases the number of patches. The image is divided into $N \times N$ – pixel cells and then a patch is projected onto this image. The patch is expanded to the neighboring cells if no other patches occur in the neighboring cells. The patch expansion does not operate on this assumption: the depth discontinuity for the patch even when the other patches are not occurred in the neighboring cells. The third step removes the outliers from the patches using the visibility consistency.

The purpose of this paper is to introduce a framework that employ a set of calibrated cameras C and sparse 3D points x from the outcome of SfM method instead of Harris corner detector and Difference-of-Gaussians (DoG) in feature detection and matching to initialize the patches p. That generates a much denser point cloud and also saves the processing time.

This approach takes a set of photographs with those camera parameters and the sparse 3D points as inputs. In this approach, three stages of work are performed. The first stage is the initialization a small patch p for each of the Structure-from-Motion 3D x points. The second stage is the patch expansion in order to rebuild at least one patch in every cell of the image. The patch expansion stage may contain outliers. Consequently, the third stage is the patch filtering in order to remove these outliers. These patch expansion and filtering are then iteratively implemented until a satisfied percentage of image coverage is completed (3-time iterations are used). Our system does not need any type of initialization such as a bounding box, visual hull or depth ranges.

The remaining parts of the paper are ordered as follows: Section 2 explains about the patch model; our SfM-based PMVS reconstruction pipeline and the proposed patch initialization from Structure-from-Motion (SfM) are introduced in section 3; To prove the effectiveness of the proposed system, a few experiments with sampled images of the scene are reported in section 4 and the conclusion is finally presented in section 5.

II. PATCH MODEL

PMVS provides the main benefit of flexibility and reconstructs a set of patches for the whole scene that directly estimates both the depth and the surface normal. A patch is also called an oriented point that means a 3D point with a surface normal estimation or a local region support as in Fig. 1. The model is characterized by a collection of individual patches P, where each patch p is an approximation of local tangent plane of the surface and a rectangle has assigned center $c(p)$ and normal $n(p)$. $V(P)$ is the set of images where p should be truly visible and they can select one among of these visible images and set it as the reference image $R(p)$ ($R(p) \in V(P)$). In this approach, I_i is the image with the index i and O_i is the optical axis of the assigned camera respectively.

III. SFM -BASED PMVS RECONSTRUCTION PIPELINE

This section describes our Structure-from-Motion based PMVS reconstruction pipeline. The whole pipeline contains 3 steps: (1) initial patch generating (2) expansion of patch and (3) filtering of patch (outlier removing). The pipeline of our reconstructed system is illustrated in Fig. 2.

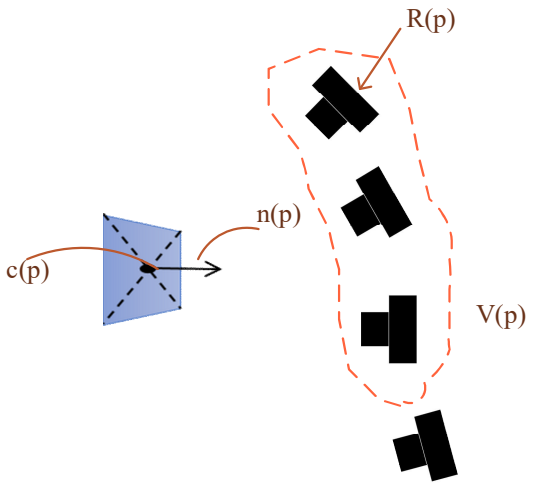


Fig. 1. A Patch and its Associated Set of Images

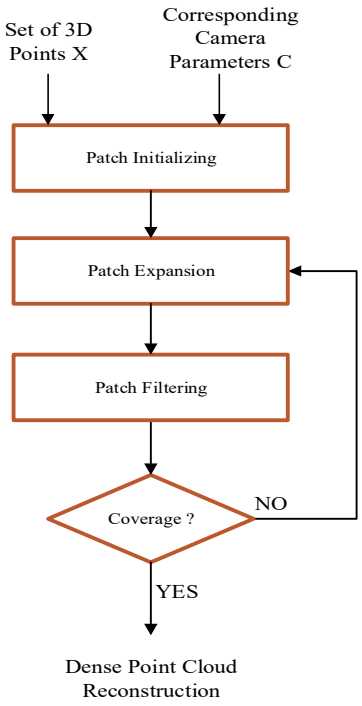


Fig. 2. Pipeline of the SfM-based Dense Multiview Stereo Reconsturction

A. Patch Initialization from Structure-from-Motion (SfM)

In the original PMVS, features are extracted by HARRIS and Difference-of-Gaussian (DoG) operators. The corresponding feature points among images are obtained using epipolar constraint and Normalized Cross Correlation (NCC). 3D point cloud is reconstructed by triangulation from the correspondence among images. In our system, we replace the feature extraction and matching to produce 3D point cloud by a set of calibrated cameras C and sparse 3D points x from the result of SfM as the input. The center of patch $c(p)$ is the 3D point and the normal $n(p)$ is the direction of optical ray from $c(p)$ to O (I_i).

$$c(p) \leftarrow x \quad (1)$$

$$n(p) \leftarrow \overrightarrow{c(p)O(I_i)} / |\overrightarrow{c(p)O(I_i)}| \quad (2)$$

$$R(p) \leftarrow I_i \mid I \in V(p) \quad (3)$$

If the angle between the patch normal $n(p)$ and the direction from the patch to the optical center $\overrightarrow{c(p)O(I_i)}$ is less than a certain threshold t, the patch is assumed to be visible in an image I_i.

$$V(p) \leftarrow \{I \mid n(p) \cdot \overrightarrow{c(p)O(I_i)} / |\overrightarrow{c(p)O(I_i)}| \leq t\} \quad (4)$$

B. Patch Expansion

In this stage, the patches are projected into the images I_i with the purpose of increasing the number of patches and the density of the point cloud. When projecting a patch onto an image, it is divided into $N \times N$ -pixel cells and then recreate at least one patch in every image cell $C_i(x,y)$, as in Fig. 3. If there is a patch already exists in a neighboring cell, or if there is a depth discontinuity when viewed from the camera, the expansion is not performed.

C. Patch Filtering

In this stage, we filter the false positive patches caused by the expansion procedure. The normal filter and local neighborhood filter in [11] are used in this paper. In the first filter, the normal $n(p)$ of the patch is used and check the normal of each point with its 20 neighbors again. If the distance between the current point's normal $n(p)$ and an adjacent point's normal $n(p')$ is less than $\pi/12$, then the current point is assumed as outliers and should be removed. In second filter, the whole point cloud is back projected to its reference image. Then, the point is considered as an outlier if the number of neighborhood points of $c(p)$ is lower than a fixed threshold ε (ε is the half of the average number of the neighbor). Fig. 4 shows the patch filtering stage of the system.

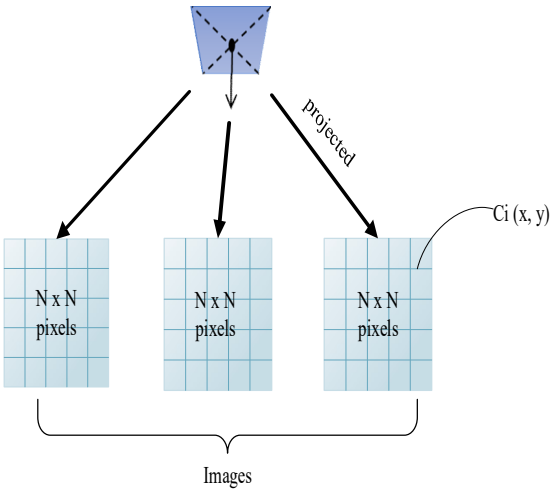


Fig. 3. Patch Expansion

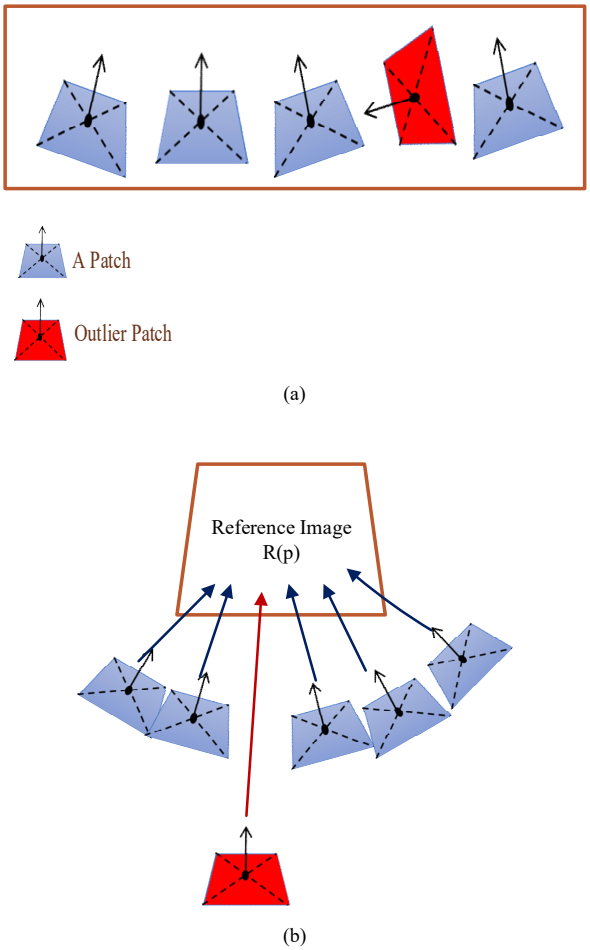


Fig. 4. Patch Filtering using (a) Normal Filter (b) Local Neighbor Filter

IV. EXPERIMENTS AND RESULTS

We experimented our system based on the Core i5 2.50 GHz processor, 8.00 RAM, and openMVG [9] for the structure-from-motion (SfM). We evaluated our approach on two different datasets to compare the performance. The first of these is the Der Hass provided by the official website of Multiview Environment [10] and the second dragon sculpture (12 images) are taken by Huawei nova 2s mobile camera. The comprehensive assessment between the final outcome of PMVS and the proposed system is shown in Fig. 5 and Fig. 6.

In these experiments, Fig. 5 (a) and 6 (a) are the sample image of the datasets. The input for our system is displayed in Fig. 5 (b) and Fig. 6 (b) and these are designed by openMVG [9]. In Fig. 5, the reconstructed mesh model using our proposed method provide more accurate and getting better and it is not rough as Fig. 5 (c). For Dragon Statue, in PMVS, the Dragon's head is not properly reconstructed, but, in the proposed system, the Dragon's head can see clearly as in Fig. 6. In comparison to PMVS, the proposed method offers very high-resolution patches and reconstructs the whole object accurately and completely. Table 1 shows the number of patches comparison between the original PMVS and the our proposed SfM-based PMVS. As shown by Table 1, our improved SfM-based PMVS generates more patches and denser than the PMVS. The runtimes and the number of patches among the algorithms and two different datasets are

shown in Table 2. In Table 2, our system does not significantly reduce the run time, but it does reduce it slightly.

V. CONCLUSION

This paper proposed the dense multi-view stereo reconstruction based on the SfM point cloud. The system focuses on the first stage of the PMVS – patch initialization and modified these patch initialization stage. The proposed system takes the camera parameter and 3D sparse point cloud from SfM instead of HARRIS and DoG feature detection in patch initialization. We tested our system on two different datasets. The experiments described our improved SfM-based PMVS system increases the number of patches and denser than the PMVS. According to the increasing number of patches like this, this will produce in a better, coverage and more accurate 3D model. This system capable of delivering the accurate a dense mesh model in compare to PMVS. In the future, the expansion and filtering stages will be modified to get better results and the implementation will be performed on video frames.

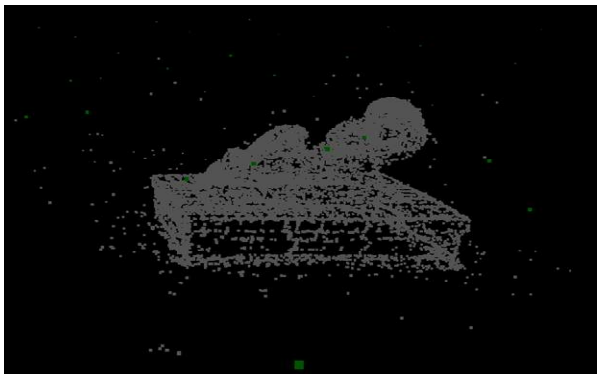
TABLE I. THE NUMBER OF PATCHES COMPARISON

Data	Number of Input Point Cloud	Number of patches by PMVS	Number of Patches by Proposed Method
Der Hass	36, 056	1, 251, 197	1, 259, 997
Dragon	9, 483	528, 445	563, 092

TABLE II. TIME COMPARISON BETWEEN THE PROPOSED SFM-BASED PMVS AND PMVS

Algorithm	Der Hass		Dragon	
	#Patches	#Time (min)	#Patches	#Time (min)
Proposed	1, 259, 997	15	563, 092	3.43
PMVS	1, 251, 197	18. 785	528, 445	4.62

(a)



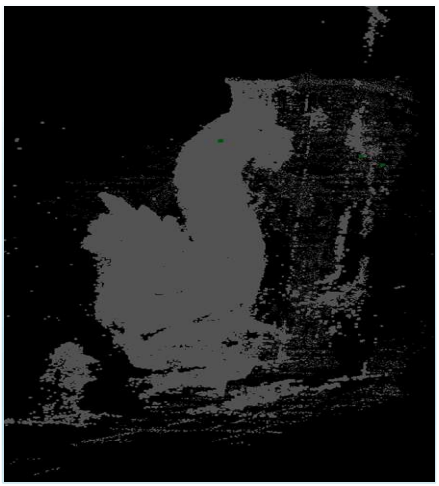
(b)

(c)

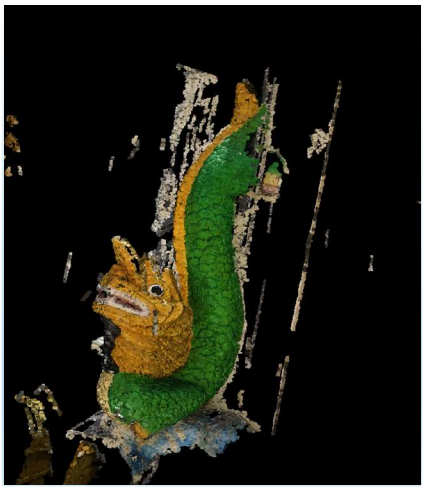
(d)

Fig. 5. Dense Multi-view Stereo Patch Reconstruction and Final Mesh Model (a) Image (b) Input Point Cloud (c) Point Cloud by PMVS (d) Point Cloud by the Proposed Method on Der Hass

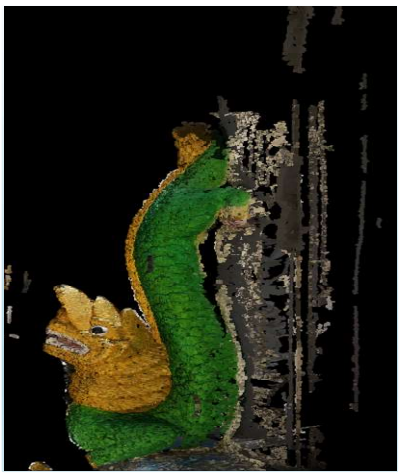
(a)



(b)



(c)



(d)

Fig. 6. Dense Multi-view Stereo Patch Reconstruction and Final Mesh Model (a) Image (b) Input Point Cloud (c) Point Cloud by PMVS (d) Point Cloud by the Proposed Method on Dragon Statue

REFERENCES

[1] A. Hornung and L. Kobbelt, "Hierarchical Volumetric Multi-View Stereo Reconstruction of Manifold Surfaces Based on Dual Graph Embedding", Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2006.

[2] S. Sinha, P. Mordohai, and M. Pollefeys, "Multi-View Stereo via Graph Cuts on the Dual of an Adaptive Tetrahedral Mesh," Proc. Int'l Conf. Computer Vision, 2007.

[3] S. Tran and L. Davis, "3D Surface Reconstruction Using Graph Cuts with Surface Constraints," Proc. European Conf. Computer Vision, 2006.

[4] C. Hern´ andez Esteban and F. Schmitt, "Silhouette and stereo fusion for 3D object modeling," *CVIU*, vol. 96, no. 3, 2004.

[5] Y. Furukawa and J. Ponce, "Carved visual hulls for image-based modeling," *IJCV*, March 2008.

[6] C. Strecha, R. Fransens, and L. V. Gool, "Combined depth and outlier estimation in multi-view stereo," in *CVPR*, 2006, pp. 2394–2401.

[7] D. Bradley, T. Boubekeur, and W. Heidrich, "Accurate multi-view reconstruction using robust binocular stereo and surface meshing," in Fi*CVPR*, 2008.

[8] Y. Furukawa and J. Ponce, "Accurate, Dense, and Robust Multiview Stereopsis," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 8, pp. 1362-1376, Aug. 2010, doi: 10.1109/TPAMI.2009.161.

[9] P. Moulon, P. Monasse, and R. Marlet, Global Fusion of Relative Motions for Robust, Accurate and Scalable Structure from Motion, ICCV, pp. 3248-3255, 2013.

[10] MVE Homepage,https://www.gcc.tu-darmstadt.de/home/proj/mve/.

[11] S. Yao, H. AliAkboarpour, G. Seetharaman and K. Palaniappan, “3D Patch-based Multi-view Stereo for high-resolution Imagery”, Proceedings Volume 10645, Geospatial Informatics, Motion Imagery, and Network Analytics VIII; 106450K (2018) https://doi.org/10.1117/12.2309806.

[12] Leung and B. C. Lovell, “3D Reconstruction through Segmentation of Multi View Image Sequences”, 2003.

[13] C.Strecha, R.Fransens, and L.Y.Gooi. "Combined depth and outlier estimation in multi- view stereo". in CVPR, 2006, pp, 2394-2401.