www.ijsrnsc.org

Available online at www.ijsrnsc.org

IJSRNSC

Volume-5, Issue-3, June 2017
Research Paper

Int. J. Sc. Res. in Network Security and Communication

ISSN: 2321-3256

LIOP based Feature Detection and Matching in SfM for 3D Object Reconstruction

Amit Banda[1*], Rajesh Patil[2]

[1*]Dept. of Electrical Engineering, VJTI, Mumbai, India
[2]Dept. of Electrical Engineering, VJTI, Mumbai, India

Received 06th Mar 2017, Revised 15th Mar 2017, Accepted 28th Apr 2017, Online 30th Jun 2017

Abstract—3D object reconstruction is growing popular due to its various applications such as movie industry and research simulations. Multi-View Stereo (MVS) based reconstruction using camera without extra hardware helps in reducing the cost of the system. The process consists of solving correspondence problem between images cause by camera motion using feature detector and descriptor or the optical flow technique. We propose to use LIOP proposed recentlyas a feature descriptor algorithmin the SFM based method for 3D object reconstruction.Based on the feature descriptor, it is more robust to noise, rotation, translation and monotonic intensity changes. Use of such rich feature descriptors increases the accuracy of reconstruction.

Index Terms—local feature descriptor, local feature descriptor, optical flow, SURF, SIFT, LIOP, multi-view stereo

I. INTRODUCTION

Various sectors like movie and gaming industry, street-view, geology and lots more have an increasing need for 3D models. The active techniques [1] such as LAZER, Structured-Light 3D scanners have high cost for the system. The cost of the system can be brought down by using passive techniques like camera which use available light from the environment. There are various passive techniques in which the one that use single vantage point include shape from texture, shape from occlusion, time to contact and shape from focus. In our experiment, we have used Multi-View Stereo approach where we capture images from various view-points using single camera.

The modern 3D scanners have calibrated rigs, where camera position and orientation can be easily obtained using sensors like encoders with high accuracy which increases the cost of the system. Structure from Motion technique [14] enables to reconstruct the scene without knowing in advance the camera position and orientation. Using SfM algorithm, the object structure is estimated using point cloud representation. First camera motion is determined by solving the correspondence problem between images captured. Using the camera motion, we can then determine points matched in 3D by triangulation technique.

The correspondence problem between two images can be solved either using feature detection and descriptors or the optical flow techniques.

II. LITERATURE REVIEW

In [2], there are 4 classes of MVS reconstruction algorithm. The first class uses volumetric reconstruction method [3]. It makes a single sweep through volume to reconstruct voxels by estimating thecost and comparing it with the threshold. Another class iteratively evolves the surface by using space carving methods [4] while in the third class image-space methods [5] are usedwhere reconstruction is done by computing depth maps and combining those maps. In the fourth class,reconstruction is done by matching a set of feature points [6-9]between images and fit a surface to the features.

In [13], the process of obtaining 3D reconstruction using MVS is explained mathematically.Two-view geometry is practically explained in which correspondence geometry between the two images from camera matrix is given by epipolar geometry of two views.

One method is optical flow algorithm used between the pair of images. By using intensity values of neighboring pixels, an optical flow algorithm OF [15] calculates the displacement of brightness patterns between two images. There are two types of OF algorithms dense OF and sparse OF algorithms. In dense OF algorithms [16], displacement is calculated at each pixel using global constraints but are less robust to noise. While sparse OF algorithm like Lucas-Kanade [17], calculate displacement across selective pixels and is more robust to noise and efficient. Thus, sparse OF is good for practical application.

Another approach would be to use feature detector and descriptor and matching those between the two images. [20] shows comparison of various descriptors such as SIFT [19], differential invariants, complex filters and moment invariants. Wang, Fan and Wu propose LIOP [22] based feature descriptor technique that outperforms other descriptors in accuracy [20].

Chapter III explains the block diagram of proposed system in detail, while Chapter IV explains the performance criteria of the algorithm. Experiments and results obtained are explained in Chapter V while Chapter VI concludes the paper.

III. PROPOSED SYSTEM

The block diagram for typical SFM-MVS based reconstruction isshown in Figure 1 with major implementation for SfM given in [11]

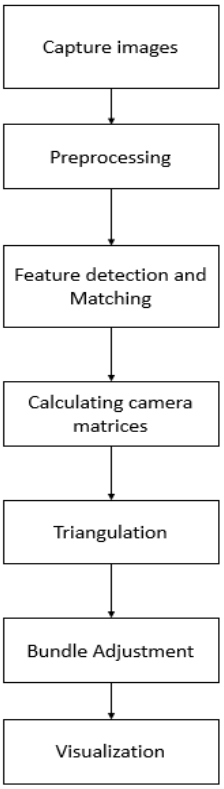

Figure 1: Block Diagram of SFM

A. *Capturing images*

Image acquisition is done by capturing images of the objects from various view-points. The difference between the two view-points must not be larger as it won't lead to affine transformation and the resultant feature matching would be minimum. So care has to be taken to capture images at small incremental distance.

B. *Pre-processing*

In order to decrease the sensitivity due to noise, the images are smoothened with Gaussian filter with σ_p. Harris affine feature detector was used to localize feature position and estimate affine shape of its neighborhood. These detected regions are then normalized to circular regions of fixed diameter. Finally, a Gaussian smoothening of σ_n was applied to the resulting local patch. Figure 2 show the normalized patch for one detected region

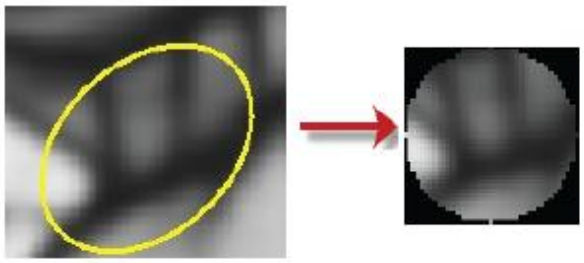

Figure 2: Pre-processing the captured image

C. *Describing features and matching*

Initial steps for describing the feature involves division patch into several sub regions based on intensity order. All pixels in the local patch were sorted in non-descending order and patch divided into B ordinal bins. LIOP for one pixel was computed asin (1).

$$LIOP(x) = V_{N!}^{Ind(YP(x))} = (0,..,0,\ 1,\ 0,..0) \quad (1)$$

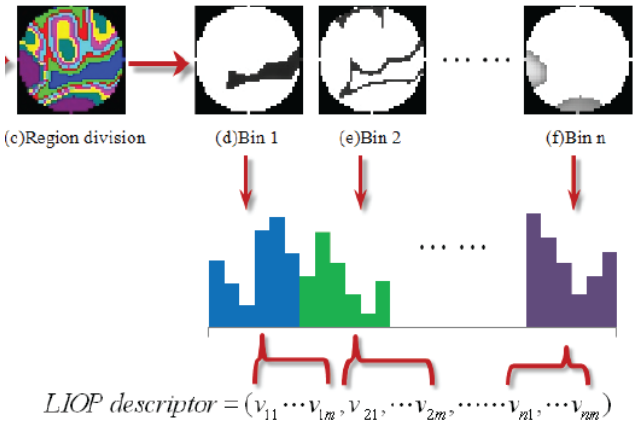

Figure 3: Compting the LIOP descriptor

Finally, complete descriptor was computed as in (2) by concatenating.

$$LIOP\ descriptor=(des_1,des_2,..des_B)$$
$$where\ des_i = \sum_{x\in bini} w(x)LIOP(x) \quad (2)$$

Matching of descriptors between two images was done by FLANN [10] base matching as it provides quick and efficient matching compared to brute force matching.

D. Calculating camera matrices

After finding out matched key points between pair of images, matching points are first aligned in two arrays. On aligned matrices, fundamental matrix F is computed.

Using (3), camera matrices P can be estimated where Rdenotes the rotation and t denotesthe translation.

$$P = [R\ /t\ ] = \begin{bmatrix} r1 & r2 & r3 & t1 \\ r4 & r5 & r6 & t2 \\ r7 & r8 & r9 & t3 \end{bmatrix} \quad (3)$$

The relation between 2D point (image point) denoted by x and 3D point (scene point) denoted by X is given in (4)

$$x = PX \quad (4)$$

Fundamental matrix is the matrix used to represent uncalibrated cameras while essential matrix is used when cameras are calibrated. So, essential matrix can be computed using (5) where K is the intrinsic camera calibration matrix.

$$E = K^T F\ K \quad (5)$$

Essential matrix E is then decomposed into R and t (5) by using singular vector decomposition. However, this decomposition consists of four possible solutions, out of which only one is correct[13]. This can be solved further in triangulation.

$$E=[t]_x\ R \quad (6)$$

E. Triangulation

Linear triangulation [12] was used to compute X which can be estimated by solving equation (7) where we have x and x' as matching key points between two images

$$AX = B \quad (7)$$

The issue with this kind of reconstruction has motion between two images is using given unit of scale which will be variable across multiple cameras.

Re-projection error can be calculated by the difference in the distance between the actual image point and the point obtained by re-projecting a 3D point on the same camera. If this distance is large, this means the point has large error and can be filtered away. Thus, reobtaining camera matrix with four camera matrix found by SVD, correct camera matrix can be found out.

Further reconstructing from multiple views can be done by Perspective-N-Point PNP using the scene points we have already found and other method is Iterative Closest Point (ICP) where more points are triangulated and how they fit into existing geometry. PNP approach was used.

F. Bundle Adjustment

It is the process of optimizing the reconstructed scene. Optimization is done so that the re-projection error is minimized. We used Simple Sparse Bundle Adjustment SSBA library [18] for this process that assumes images were taken from same hardware.

G. Visualization

In the final step, the point cloud was visualized using Point Cloud Library PCL.

IV. PERFORMANCE CRITERIA

The performance criteria for feature detection and matching are explained as in [20-21]. The repeatability score, precision-recall, speedup factor form performance metrics for feature detection algorithms.

Two points a and b are similar if the distance between their descriptors is below an arbitrary threshold $(D_a - D_b) \leq t$. The value of t is varied to obtain the ROC (Receiver operating curves). The pcorrect matches are given as in (8)

$$pcorrect = \frac{\#\ correct\ matches}{\#\ possible\ matches} \quad (8)$$

The number of points detected and tracked depends on the algorithm used and it can become a measure for density of the point cloud constructed. The speed of algorithm can be found by the amount of time required to match two images.

V. EXPERIMENTAL RESULTS

Experiments were carried out on images of various objects. Few samples of bag and dome are shown in Figure.Library used was OpenCV in C++ on Linux platform with i3 dual core CPU.

OpenCV library provides with massive functions for image processing and is highly efficient compared to other software like MATLAB, SciLab.

Few of the sample images of bag and dome used in our experiments is shown in Figure 4-5.

Figure 4: Image acquisition of bag

Figure 5: Image acquisition of dome structure

Figure 6-7 shows feature matching of bag and dome using Harris-affine detector and LIOP based descriptor.

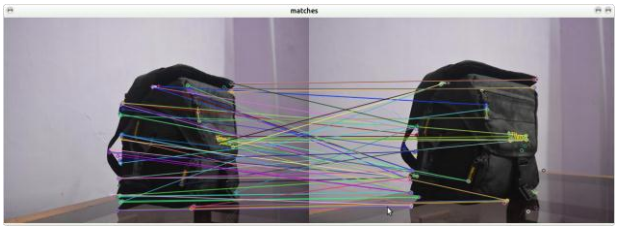

Figure 6: Feature matching of bag using 2 samples

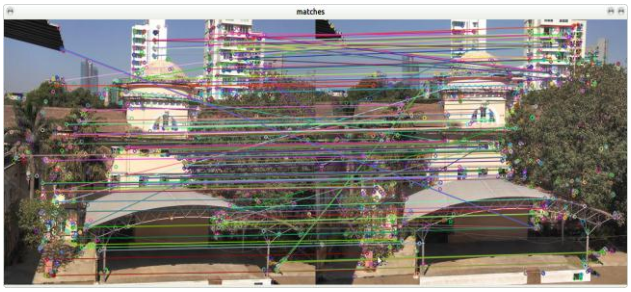

Figure 7: Feature matching of dome using 2 samples

As we can see in Figure 6-7, there were several false matches detected. Those were eliminated by comparing with the threshold distance.

The reconstruction of the bag and dome is visualized as shown in Figure 7-8 using PCL. The detected camera pose for all the cameras are shown in red. The reconstruction was done using set of 20 images. The density of reconstruction with accuracy can be increased by using

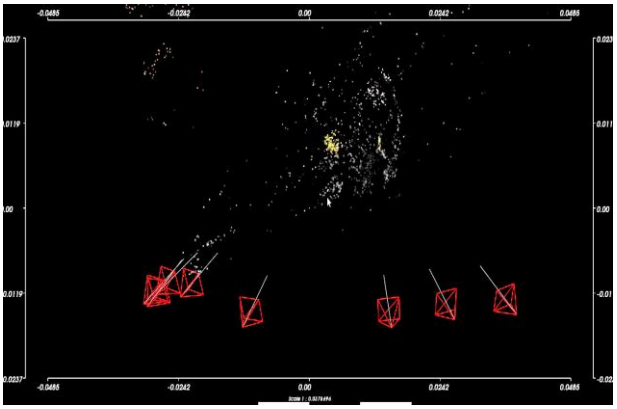

Figure 8: Reconstruction of bag

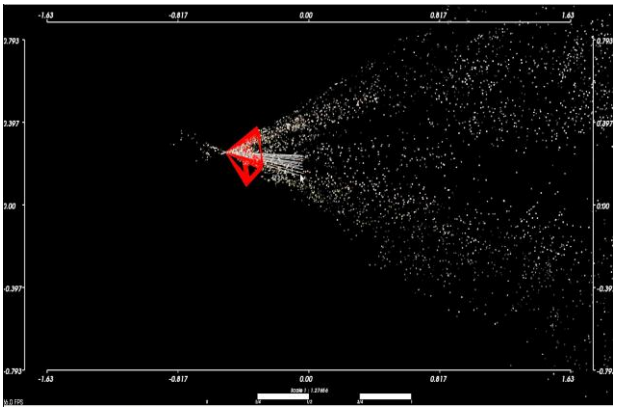

Figure 9: Reconstruction of dome

For calculating metrics, the dataset used was the Grafiti image sequences [20]. Table 1 shows comparison of three techniques based on average number of features matched and average computation time required. The parameters were calculated after finding homography using RANSAC algorithm. There are several feature detectors and descriptor, out of them popular SIFT-SURF was used. In optical flow detection Fast feature detector was used as detector and Lucas Kanade based optical flow technique.

Technique	**Avg. Number of points matched**	**Average Computation time (ms)**
SIFT-SURF	450	2200
Harris affine-LIOP	800	12000
FFD-Optical Flow	4000	7000

Table 1: Comparison of different techniques

The number of detected points and matched points still would depend on the type of detector used and the parameter settings. Based on the computation times, LIOP methods requires lot of time for descriptor computation. The advantage of it is robust to noise like lighting conditions. Although optical flow technique has lot points matched, limitation is that it requires images to be taken from same hardware and is sensitive to noise

VI. CONCLUSION

As in the proposed system, it performs good in monotonic lightning variation conditions. In optical flow, the number of points tracked are higher but it is affected if there are changes in lightning conditions from various viewpoints. Compared to optical flow technique, in feature detector approach, points are matched better if there are lighting changes. The limitation of the feature descriptor approach is that the number of points tracked between images are less as a particular feature may appear and disappear in sequence of images. As a result, density of cloud shall be low.

For good reconstruction, the number of images required are large typically in multiples of 1000 so that maximum features are tracked and can be triangulated. Additionally, the process requires lot of time depending on the number images and features detected. Future work may include the connecting the sparse point cloud to form mesh models.

REFERENCES

[1] Amit Banda, Rajesh A Patil, "*Review on Feature Detection and Matching Algorithms for 3D Object Reconstruction*", IJRTER, Vol. 3, Jan 2017, pp. 219-225

[2] Seitz, Steven M., "*A comparison and evaluation of multi-view stereo reconstruction algorithms*" 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR06), USA, pp. 519-528 ,2006

[3] Soo Mi Choi, "*Volumetric object reconstruction using the 3D-MRF model-based segmentation*", IEEE Transactions on Medical Imaging, Vol.16, Issue.6, pp.887-892, 1997.

[4] K. Kutulakos, S. Seitz., "*A theory of shape by space carving*", IJCV, Vol.38, Issue.3, Issue.3, pp.199-218, 2000

[5] R. Szeliski, "*A multi-view approach to motion and stereo*", In CVPR, Vol.1, Issue.4, pp.157-163, 1999.

[6] O. Faugeras, E. Bras-Mehlman, J.-D. Boissonnat, "*Representingstereo data with the Delaunay triangulation*", Artificial Intelligence, Vol.44, Issue.1-2, pp.41-87, 1990.

[7] A. Manessis, A. Hilton, P. Palmer, P. McLauchlan, X. Shen, "*Reconstruction of scene models from sparse 3D structure*", In CVPR, vol.1, Issue.3, pp.666-673, 2000.

[8] D. Morris, T. Kanade , "*Image-consistent surface triangulation*", In CVPR, Vol.1, Issue.3, pp.332-338, 2000.

[9] C. J. Taylor, "*Surface reconstruction from feature based stereo*", In ICCV, USA, pp. 184-190, 2003.

[10] Marius Muja, David G. Lowe, "*Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration*", VISAPP, Vol.1, Issue.2, pp.331-340, 2009.

[11] DL Baggio, "*Mastering OpenCV with practical computer vision projects*", Packt Publishing Ltd, UK, pp.1-134, 2012.

[12] RI. Hartley, "*Triangulation*", Computer vision and image understanding, Vol.68, Issue.2, pp.146-157, 1997.

[13] R. Hartley, Andrew Zisserman, "*Multiple view geometry in computer vision*", Cambridge university press, UK, pp.34-57, 2003.

[14] E. Hildreth, "*Recovering three-dimensional structure from motion with surface reconstruction*", Vision research, Vol.35, Issue.1, pp.117-137, 1995.

[15] JL. Barron, "*Performance of optical flow techniques*", International journal of computer vision, Vol.12, Issue.1, pp.43-77, 1994.

[16] Berthold KP, Brian G. Schunck, *"Determining optical flow*", Artificial intelligence, Vol.17, Issue.1-3, pp.185-203, 1981.

[17] Kapila Sharma, "*An Effective Approach of Thinning for Morphological Features An Effective Approach of Thinning for Morphological Features*", International Journal of Computer Sciences and Engineering, Vol.3, Issue.10, pp.58-60, 2015.

[18] Zach Christopher, "*Robust bundle adjustment revisited*", European Conference on Computer Vision. Springer International Publishing, Europe, pp.12-19, 2014.

[19] Lindeberg Tony, "*Scale invariant feature transform*", Scholarpedia, Vol.7, Issue.5, pp.10491-10498, 2012.

[20] K. Mikolajczyk, C. Schmid, "*A performance evaluation of local descriptors*", IEEE T-PAMI, Vol.27, Issue.10,pp.12-16, 2005

[21] Miksik Ondrej, Krystian Mikolajczyk, *"Evaluation of local detectors and descriptors for fast feature matching"*, International Conference on Pattern Recognition (ICPR), Japan, pp.11-15, 2012.

[22] Wang Zhenhua, Bin Fan, Fuchao Wu, *"Local intensity order pattern for feature description*", 2011 IEEE International Conference on Computer Vision (ICCV), Spain, pp.603-610, 2011.

