International Journal of Multimedia and Ubiquitous Engineering
Vol.11, No.8 (2016), pp.29-36
http://dx.doi.org/10.14257/ijmue.2016.11.8.03

3D Reconstruction of Remote Sensing Image Using Region Growing Combining with CMVS-PMVS

Aili Wang[1], Na An[1], Yangyang Zhao[1], Yuji Iwahori[2] and Rui Kang[3]

1. Higher Education Key Lab for Measuring & Control Technology and Instrumentations of Heilongjiang, Harbin University of Science and Technology, Harbin, China
[2] Dept. of Computer Science, Chubu University, Japan
3. State Grid Harbin Electric Power Supply Company, Harbin, China
aili925@hrbust.edu.cn

Abstract

The characteristics of remote sensing image is not obvious and can not reflect the be reconstructed the detail characters of the objects. For the sparse points in multi images fusion results, texture selection and it is relatively difficult problems, for accurate reconstruction of remote sensing image details, this paper presents CMVS-PMVS(cluster multi view stereo- patch based multi view stereo) combining with region growing for computing dense matching points. First, select some seed points by region growing algorithm, and find the matching relationship between the seed points, followed by the matching relation from the seed point to spread until to the entire image. Then the image set are clustered by CMVS in order to reduce the amount of data in the process of reconstruction, and the operation rate and reconstruction accuracy can be improved. PMVS reconstruction method is used to complete the reconstruction task by matching, expanding and filtering three steps. The experimental results showed that the point cloud is dense enough which are reconstructed by the 3d reconstruction algorithm based on regional growth combining CMVS-PMVS and well expressed the practical model of object reconstruction, the reconstruction of objects in remote sensing images has very strong practicability.

Keywords*: 3D reconstruction; Regional growth; CMVS; PMVS*

1. Introduction

With the rapid development of computer technologies, people have more and more demand for 3D models using two-dimensional projective images to achieve the perception. 3D reconstruction based on image sequences has become one of the hot research topics in the fields of image processing because it has the advantages of being low cost, simple and realistic [1]. This should be a motivation to research, design, develop and validate novel easy-to-use, ease-to-learn and a low-cost framework for 3D modeling and further understanding of virtual environments using multiple data sources, so that the whole 3D modeling community has access to an affordable, transferable, functional and usable framework of methods and tools [2].

In the category of 3D reconstruction, the existing multi view stereo vision method is based on the research of the matching and reconstruction of the points. The local methods such as gradient descent, level set, or global method such as image segmentation, have high reconstruction precision. But the process of rebuilding demands extra information, such as depth maps, bounding box, visual shell and so on, which are not suitable for the reconstruction of chaotic image sets and outdoor scenes [3-5].

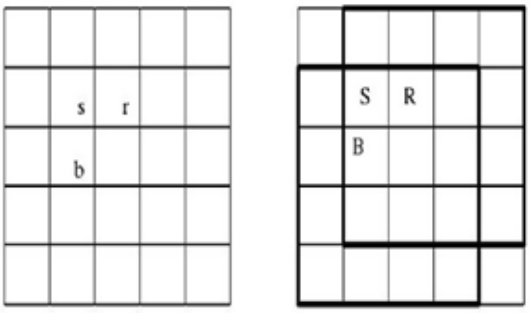
ISSN: 1975-0080 IJMUE

In this paper, the CMVS-PMVS dense matching algorithm based on region growing can automatically detect and remove the obstacles in the scene and the output is small and directional patch set covering the entire surface of the target.

2. Regional Growth Algorithm Description

The first step of the dense matching algorithm based on region growing is to select some seed points, and to find matching pairs between the seed points. Regional growth is the second step of the algorithm; the matching relation began to spread from the seed point until to spread to the entire image. Compared to the traditional point by point method, this method can greatly improve the matching efficiency by the continuity constraint [6].

The basic principle is: if it is known that a point s on the left and right image and the point S on the left image are a correct match point, then the matching points of r and b which are adjacent to s should be R the B which must be located around the S point in the right image. So searching for the R and B points can only be done near the S point. In general, size of searching window is 4×4 or 3×3 pixels, the target function is used to calculate the matching value, the maximum matching value is considered to be the correct matching to point S. The accuracy and efficiency of the algorithm can be greatly improved through this technique. Once a pair of the accurate matching points is obtained, through the relationship of relative position between two points will expanse rapidly matching to the whole image. Figure 1 shows a region growing strategy for the use of seed points to the bottom right of the image, where the search area of R and point B are limited to Figue1 (b), as shown in two bold line box, where the size of the search window is 4×4 pixels[7].

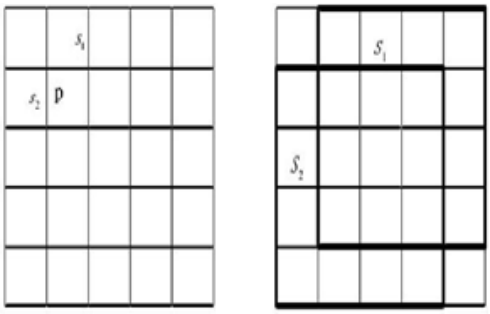


(a) Image on the Left (b) Image on the Right

Figure 1. Growing Strategy from Seed Point to Region

When the matching relation is extended from seed point to the other regions of the image, the search area is relatively easy to determine given a known pair of seed points. Otherwise if there are two seed points adjacent to the current point in the left image, the search area in the right image is determined by two pairs of seeds [8].

As shown in Figure 2, there are two points s_1 and s_2 are adjacent top in the left image, respectively, but their corresponding points S_1 and S_2 are not adjacent in the right image, then the search area of the P points is determined by the two rough line box shown in Figure 2 (b). The two ways are regional growth strategies. The first strategy is basic mode of matching transmission. By second strategy, searching region of current point is not totally dependent on a certain seed point, and is determined by the matching status of the points around the seed points, which is more reliable and accurate for matching transmission [9-11].



(a) Image on the Left (b) Image on the Right

Figure 2. Searching Region

3. Cluster Multi View Stereo Algorithm

Step 1. SFM filtering. The visual information of some SFM feature point is searched in a local neighborhood, then the information of the location is taken as the average value of the neighborhood. This can effectively reduce the number of input points set. This step is repeated, and the final output point set is formed as $\{P_j^2\}$.

Step 2. Image selection. The image is removed according to the coverage constraint mentioned above. Note that the search order is in accordance with from low to high image resolution, so that the low resolution image can be removed firstly.

Step 3. Clustering classification. The image size is the constraint condition and the coverage of the range constraint is not considered for image segmentation. Clusters that are not satisfied with the size of the image are classified into smaller clusters. $e_1 m$ is edge value of the image(I_1, I_m) pairs representing reconstruction contribution to characteristic point MVS. It means $\frac{\sum P_j \in \Theta^{1m} f\left(P_j,(I_1,I_m)\right)}{f(P_j,V_j)}$, Θ^{1m} are SFM point set that can be seen to the imageI_1, I_m. As a result, the greater the contribution to the MVS reconstruction, the greater the edge value and the more likely to be removed.

Step 4. Image increasement. For each SFM feature P_j that is not joined, compute $C_k = argmax_{c_1} f(P_j, C_j)$ to determining a image cluster C_k. For feature points P_j, construct $\{(\mathrm{I} \rightarrow C_k)\}$ to join $\mathrm{I}(\in V_j \notin C_k)$ into cluster C_k. $\mathrm{g} = \mathrm{f}(P_j, C_k \cup I_j)$ is used to measure the efficiency and each SFM feature point has only one g value, and the max g is selected to join the cluster.

Step 3 does not consider the coverage of the constraints, step 4 does not consider the constraints of image size, so step 3 and 4 should repeat until the final results meet coverage and image size constraints.

4. Patch Based Multi View Stereo Algorithm

Image clusters are obtained through the CMVS step, then PMVS is used to reconstruct the image [12].

4.1. Matching

Firstly, the algorithm is based on Harris and DoG operator to detect corner points and feature points in all images. In order to meet the uniform convergence, in each image, a rectangular grid of pixels size of $\beta_2 \times \beta_2$ is taken, and the two operators in each grid are selected as the point of the maximum response value and the feature points are detected(in

experiments $\beta_{2=32}$). When all the images of the feature points are detected, image feature points will be matched to reconstruct sparse patches which will be stored in the grid$C(i,j)$.

A given imageI_i, the corresponding camera optical center is expressed as OI_1. Any feature point f which is detected in I_i , a set of levy points f' (Harris or DOG) called as (f,f') in the other images should be located in the corresponding to the polar line with the distance of two pixels and the use of triangulation method to reconstruct 3D points. According to the order of increasing distance to consider these points and these points are regarded as a potential center of the patch, reconstructed these points by the following method.

For given feature point (f,f'), reconstruct candidate patches p and its center, the normal and reference image are initialized as follows:

$$n(p) \leftarrow \overrightarrow{c(p)\,O(I_i)} / \overrightarrow{c(p)\,O(I_i)} \tag{1}$$

$$R(p) \leftarrow I_i \tag{2}$$

So far only to reconstruct sparse patches and in the initial matching stage there may be a lot of mistakes. Assume the angle between the patch normal and the line which starts from the corresponding patches pointing to the camera optical center $O(I_i)$ is less than a certain threshold τ , image I_i is considered visible to the patch. V(p) can be determined by formula (3):

$$V(p) \leftarrow \{I \,|n(p)\cdot c(p)\,O(I_i)/|c(p)\,O(I_i)| > \cos(\tau)\} \tag{3}$$

After initialization for all parameters of the candidate patches, the patch optimization method is used to optimize the c(p) and n(p). Then update the visual image set V(p) and $V^*(p)$. In the optimization process, c(p) is always located in the projection ray of R(p). If $|V^*(p)| \geq \gamma$ at least γ images with low illumination difference exist, thus patch generation is successful and p is stored in the corresponding visible image cell to update $Q_i(i,j), Q_i^*(i,j)$. In order to accelerate the computing speed, once a patch has been rebuilt and stored in a cell, other characteristics of this cell are removed and not to participate in the calculation.

4.2. Expansion

This step reconstructs at least one patch in each image cell $C_i(i,j)$ and repeats on existing patches to generate a new patch in the empty neighbor region. Specifically, given patch p, first determine the a certain rules neighbors in the image cell set C(p), then proceed expansion strategy for each cell patch.

(a) Determination the extension of cell: for patch p, collect cell image of neighbors in each visible image to initialize C(p):

$$C(p) = \{C_i(x',y') | p \in Q_i(x,y), |x-x'|+|y-y'|=1\} \tag{4}$$

If a patch has been rebuilt, it is not necessary to expand. Specifically, if an image cell $C_i(x',y') \in C(p)$ has a neighbor patch p' to p, remove $C_i(x',y')$from C(p). When satisfies (5) patch p and p' are believed to be neighbor facets:

$$|(C(p)-C(p'))*n(p)| + |(C(p-C(p'))*n(p')| < 2\rho_1 \tag{5}$$

ρ is the displacement depth distance between the midpoint of c(p) and $c(p')$ to ρ_1 pixels in a image belongs to R(p)。

(b) Expansion. For each image cell $C_i(x,y)$ belongs to C(p), implement the following process to produce new facets p'. Firstly, initialize $n(p')$, $r(p')$ and $V(p')$ using the corresponding to the p value. $C(p')$ is initialized as the intersection of visible ray trough

the center of $C_i(x,y)$ and the plane including patch p. Use patch optimization method to optimize$\mathbf{c}(p')$ and $\mathbf{n}(p')$. In the optimization process, in order to ensure that the patch can always be in image cell $C_i(x,y)$, make $\mathbf{c}(p')$ lie in ray of I_i remaining the projection position unchanged. After the optimization, add a series of visible images to $\mathbf{V}(p')$ which are obtained by depth measurement which is calculated for each image cell instead of a single pixel) and then update $V^*(p')$. Finally, if meet $|\mathbf{V}^*(\mathbf{p}')| \geq \lambda$, p' will be accepted and update$Q_i(\mathrm{x},\mathrm{y})$ and $Q_i^*(\mathrm{x},\mathrm{y})$ for its visible images. Note that in the initial matching stage, α is set to 0.6 and 0.3 before and after the optimization, respectively. However, after each execution of an extension process, each value is increased 0.2. The purpose is to facilitate the processing region with fewer textures in the subsequent iterative process.

4.3. Filtering

Use visual consistency constraints to remove the patches which fall outside the correct surface. Let U(p) be the set of patch p' that is not consistent with the current visual information. If the following inequality is satisfied, p is treated as an external point:

$$|V^*(\mathrm{p})|\big(1-g^*(\mathrm{p})\big) < \sum_{p_i \in U(p)} 1-g^*(\mathrm{p}) \quad (6)$$

If p is an external point,$\big(1-g^*(\mathrm{p})\big)$ and $|V^*(\mathrm{p})|$ should be small and p is easily to be deleted.

5. Experimental Results and Analysis

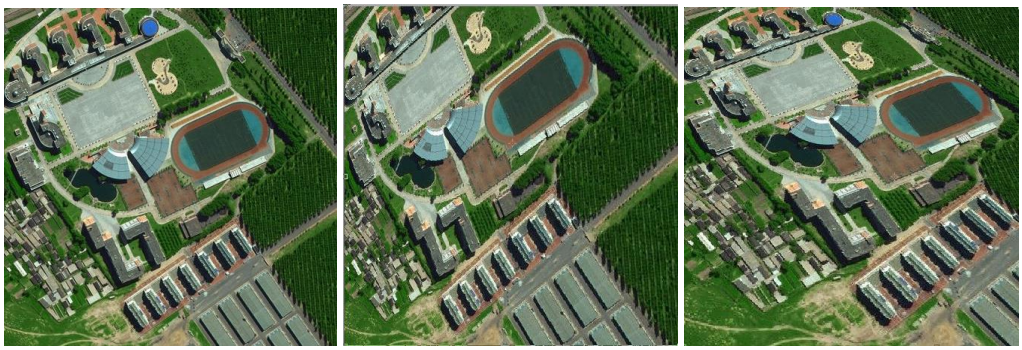
In this paper, we use 42 remote sensing images size of 1600×1200 to test the reconstruction performance of proposed algorithm. The experimental environment is for CPU Intel I5 2.6GHz, 4Gbyte memory, memory is 2Gbyte, the operating system for windows 8.1. This algorithm uses region growing method to extract 167824 feature points, and then 42 images are reduced to 27 clusters, which effectively reduce the amount of data and runs for 3.5 hours. Comparing to the reconstruction algorithm without the region growing algorithm that algorithm extracts a total of 158961 feature points and 42 image data volume reduces to 32 clusters, and running time is close to four hours. Table 1 is the location information for 27 image clusters.

Figure 3 (a), (b) and (c) are three images selected from 42 images sequence, (d) is the distribution of feature points marked by red color, (e) and (f) are 3D reconstructed results by CMVS-PMVS and our proposed algorithm respectively. It can be clearly seen that the use of regional growth algorithm is better than the results of the reconstruction method without region growing algorithm. The reconstruction results obtained by our proposed method are more complete and real. It can be clearly observed from the image (e), the reconstruction results have been distorted and shifted, and can observe the image surface ambiguity slightly higher than the image (f), because the number of three-dimensional reconstruction points found by CMVS-PMVS is far lower than the use of CMVS-PMVS combining with regional growth algorithm.

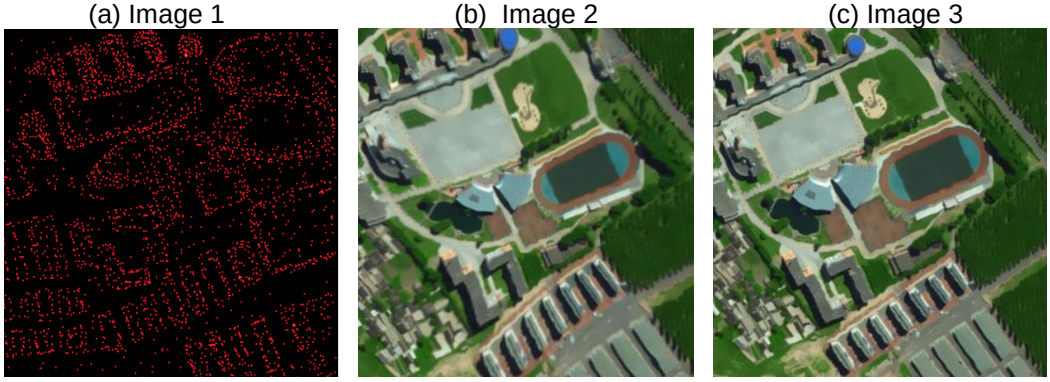
Table 1. Position Information after Cluster (m)

No.	x	y	Z	No.	x	y	z
1	-3.567 7	-4.985 5	-4.056 4	15	-3.400 6	0.721 4	-7.823 1
2	-5.642 1	-8.635 7	-7.021 0	16	-3.752 2	-2.365 8	-8.065 4
3	-6.368 9	-9.235 5	-9.536 6	17	-3.960 0	-2.095 4	-7.645 0
4	-6.536 3	-9.634 2	-12.635 0	18	-1.489 9	-0.806 9	-7.037 1

5	-7.201 2	-7.396 3	-15.032 2	19	-2.019 4	8.695 2	-7.646 2
6	-6.702 4	4.489 6	-11.604 7	20	-3.965 7	-5.879 2	-10.081 3
7	-7.396 5	3.697 4	-13.075 9	21	-5.036 5	-5.673 9	-12.631 2
8	-6.412 7	3.746 0	-11.090 2	22	-4.936 2	-5.312 0	-14.093 1
9	-5.496 3	2.806 9	-10.198 6	23	-6.151 7	-4.406 1	-16.503 7
10	-5.702 6	2.765 7	-8.862 5	24	-6.602 8	-6.482 4	-13.754 0
11	-5.294 0	0.968 9	-9.680 4	25	-7.076 2	-7.177 9	-13.965 1
12	-5.983 9	-2.681 0	-7.067 5	26	-4.961 4	-5.316 2	-17.109 3
13	-3.658 3	-2.923 4	-6.929 5	27	-2.718 7	-3.367 3	-13.706 1
14	-3.408 6	-3.319 2	-7.093 3				



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



(d) Feature Points Distribution (e) CMVS-PMVS (f) Our Proposed Method

Figure 3. 3D Reconstruction Results for Remote Sensing Images

6. Conclusion

The generation of 3D model for buildings and crossroads presents a challenge due to the complexity of manmade objects and lack of image understanding algorithms. In this paper, a, CMVS-PMVS combining with region growing algorithm is proposed for computing dense matching points and 3D reconstruction. By using remote sensing sequence images, the experimental results indicate that the proposed framework can always get higher correctness of image matching in automatic way, compared to the standard CMVS-PMVS matching technology

References

[1] W. Zhang and Q. Liu, "Three- Dimensional Scattering and Inverse Scattering from Objects with Simultaneous Permittivity and Permeability Contrasts", vol. 53, no. 1, **(2015)**, pp. 429-439.
[2] J. P. Pons, R. Keriven and O. D. Faugeras, "Multi View Stereo Reconstruction and Scene Flow Estimation with a Global Image-Based Matching Score", International Journal of Computer Vision, vol. 72, no. 2, **(2007)**, pp. 179-193.
[3] M. Lhuillier and L. Quan, "Match Propagation for Image-Based Modeling and Rendering", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 8, **(2002)**, pp. 1140-1146.
[4] L. Falkenhagen, "Depth Estimation from Stereoscopic Image Pairs Assuming Piecewise Continuous Surfaces". In European Workshop on Combined Real and Synthetic Image Processing for Broadcast and Video Productions. Hamburg, **(1994)**, pp. 115-127.
[5] H. Yuhang and Y. Jun, "Research and Implementation Based on Multi-View Dense Matching by Applying CMVS/PMVS", Journal of Geomatics, vol.38, no. 3, **(2013)**, pp. 20-23.
[6] M. Lhuillier and L. Quan, "Edge-Constrained Joint View Triangulation for Image Interpolation", In Proceedings of the Conference on Computer Vision and Pattern Recognition. Hilton Head, South Carolina, **(2000)**, pp. 218-224.
[7] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms", International Journal of Computer Vision, vol. 47, no. 1, **(2002)**, pp. 7-42.
[8] P. Keju, "3D Reconstruction Using Image Sequences", National University of Defense Technology of China, **(2012)**, pp. 58-62.
[9] S. Limin, G. Fusheng and H. Zhanyi, "An Improved PMVS through Scene Geometric Information", Acta Automatica Sinica, vol. 37, no. 5, **(2011)**, pp. 560-568.
[10] R. Bishop, "A Survey of Intelligent Vehicle Applications Worldwide", Proceedings of the IEEE Intelligent Vehicles Symposium. Dearbom, MI, USA, **(2000)**, pp. 25-30.
[11] J. C. Rosito and K. C. Roberto, "A Lane Departure Warning System Using Lateral Offset with Uncalibrated Camera", IEEE Conference on Intelligent Transportation Systems. Vienna, Austria, **(2005)**, pp. 102-107.