An Effective Multiview Stereo Method for Uncalibrated Images

Peng Cui, Yiguang Liu(✉), Pengfei Wu, Jie Li, and Shoulin Yi

Vision and Image Processing Lab(VIPL), College of Computer Science, SiChuan University, Chengdu 610065, People's Republic of China
liuyg@scu.edu.cn, lygpapers@aliyun.com

Abstract. For most dense multi-view stereo methods, the process of finding correspondences is the basis and is independent of acquiring 3D information, and this often brings about erroneous correspondences followed by erroneous 3D information. To tackle this problem, by expanding matched points and by expanding 3D patches, this paper proposes an effective approach to acquire dense and accurate point clouds from multi-view uncalibrated images. In the approach, two novel algorithms are newly designed and are placed before and after the Bundler: 1) the *match expansion* algorithm, which generates evenly distributed correspondences with geometric consistency; after using Bundler to produce geometry estimation and quasi-dense point clouds which are not dense and accurate, 2) the *point-cloud expansion* algorithm, which is proposed to improve the density and accuracy of point clouds by optimizing the geometry of each 3D patch and expanding each good patch to its neighborhood. A large number of experimental results demonstrate the proposed approach get more accurate and denser point clouds than the state-of-the-art methods. A quantitative evaluation shows the accuracy of the proposed method favorable to PMVS.

Keywords: Multiview stereo · Match expansion · Point-cloud expansion

1 Introduction

Mutli-view stereo (MVS) reconstruction from a set of images, which have been collected from Internet, has achieved great development in the last decade. The construction of realistic object models can be applied to the film, television, and video game industries, etc. According to [10], the state-of-the-art MVS algorithms can be categorized into four classes: The *Voxel* based approaches [11,18] compute a cost function on a three-dimensional (3D) volume by first, and then reconstruct a surface from this volume, they are suitable for small compact objects. The *surface evolution* based methods [2,3] work by iteratively evolving a surface to decrease or minimize a cost function, the algorithms rely on a reliable initial guess which limits their applicability. The *depth maps* based methods [12,16] compute a set of depth maps by first, and then merge the set of depth maps into a 3D scene. However more computations and memory are required.

H. Zha et al. (Eds.): CCCV 2015, Part I, CCIS 546, pp. 124–133, 2015.
DOI: 10.1007/978-3-662-48558-3_13

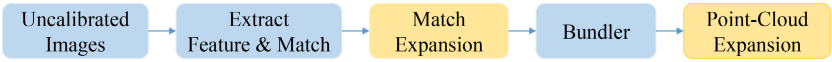

Fig. 1. The pipeline of the proposed approach.

The *feature point* based methods [4] extract and match a set of feature points firstly, and then fit a surface to the reconstructed scenes. They are simple and effective, but rely on the accuracy of the correspondences.

This paper addresses the problem of acquiring dense and accurate point clouds from multiple uncalibrated images. For uncalibrated images, the common methods are based on sparse feature points which are tracked across sequences, and then automate acquisition of a sparse point cloud together with the camera motion, such as [7,8,14]. Although this approach can estimate the camera parameters, it's not always sufficient as limited by the correspondences without geometric constraint, meanwhile it cannot acquire dense point clouds. So we propose *match expansion* technique in this paper, which is similar to [6] but our method replaces ZNCC by DASIY [16] as DASIY is more robust and efficient for wide-baseline stereo, to process initial correspondences and make them more suitable for Bundler.

For calibrated images, many dense stereo methods have been proposed [4,5, 17]. Our *point-cloud expansion* algorithm is similar with the expansion step of PMVS [4], but our algotithm takes these 3D points reconstructed by Bundler as seeds rather than re-extracting feature points, and we only expand once instead of three times, as PMVS does, that can efficiently improve the process. The following point-cloud expansion technique is mainly used to refine and expand 3D patches, and produces accurate and dense point clouds. Fig. 1 shows the pipeline of the proposed approach.

The rest of this paper is organized as follows: Section 2 introduces the overview of the proposed approach. Section 3 presents the detail of the proposed match expansion algorithm. Section 4 describes the point-cloud expansion algorithm. Experimental results and discussion are given in Section 5, and Section 6 concludes this paper.

2 Overview

Our MVS method attempts to reconstruct dense and accurate point clouds from uncalibrated images, and it can be divided into four steps (Fig. 1): 1) *Extract Feature & Match*: Features extracted by *VLFeat* [1] operator, which is a fast dense version of SIFT, are firstly matched across multiple images by *kd-tree*, then that will yield a sparse set of correspondences associated with salient image regions; 2) *Match Expansion*: Based on *best-first* strategy, this step expands the initial matches to their neighborhoods, and generates further dense correspondences which are suitable for geometric computation. The detail will describe in Section 3; 3) *Bundler*: The Bundler [13,14] procedure can estimate the camera postures and simultaneously construct a sparse scene structure by taking these correspondences as input; and 4) *Point-Cloud Expansion*: This step optimizes

and expands these reconstructed 3D points, and then filters erroneous points. The detail will describe in Section 4.

3 Match Expansion

Since the performance of Bundler depends on the correspondences, we propose the *match expansion* algorithm to produce enough good correspondences suitable for geometric estimation. Before discussing the match expansion algorithm, we define the correlation score between two points from different images as:

$$\mathcal{C}_{ij}(\mathbf{x}, \mathbf{x}') = ||\mathcal{D}_i(\mathbf{x}) - \mathcal{D}_j(\mathbf{x}')||. \tag{1}$$

Where $\mathcal{D}_i(\mathbf{x})$ denotes a DAISY descriptor in the $\mathbf{x}$ coordinate of the ith image. The reason why is DAISY used for correlation is that DAISY is rarely affected by perspective distortion and occlusion in wide-baseline situation, and it can be computed much fast.

After the feature extraction and matching, we obtain initial matches of image salient regions, but which contain inevitable errors. In order to effectively avoid these errors, we estimate a fundamental matrix for each pair of images using the Random Sample And Consensus (RANSAC) framework, and remove the outliers to the recovered F-matrix. Then we expand these remaining matches which meet the epipolar constraint.

We divide each image into regular grids of $\beta_1 \times \beta_1$ pixels as in Fig. 2 ($\beta_1 = 2$ in all our experiments), that effectively guarantees the uniqueness of correspondence. Then, we sort the remaining matches for each pair of images by increasing correlation score as seeds. At each step, the match $(\mathbf{x}, \mathbf{x}')$ with the best correlation score is used for current expansion, and simultaneously removed from the list of seeds. Next we collect the neighboring image points $\mathcal{N}(\mathbf{x})$ defined as:

$$\mathcal{N}(\mathbf{x}) = \{\mathbf{q} | \mathbf{q} - \mathbf{x} \in \{(\beta_1, 0), (-\beta_1, 0), (0, \beta_1), (0, -\beta_1)\}\}. \tag{2}$$

For each collected point $\mathbf{q}$ in $\mathcal{N}(\mathbf{x})$, the following expansion procedure is performed to generate new match $(\mathbf{q}, \mathbf{q}')$: We calculate the epipolar line $\mathbf{l}' = F\hat{\mathbf{q}}$, where F is the fundamental matrix between the pair of images, and $\hat{\mathbf{q}}$ denotes the homogeneous coordinate of the point $\mathbf{q}$ [9]. Next we search for the candidate points along the epipolar line $\mathbf{l}'$, and also within the neighborhood of location $\mathbf{x}'$, that can be formalized as:

$$\mathcal{N}'(\mathbf{x}') = \{\mathbf{q}' | {\mathbf{l}'}^T \hat{\mathbf{q}}' = 0, ||\mathbf{q}' - \mathbf{x}'|| < r\} \tag{3}$$

($r = 6.0$ in all our experiments). We also add a *random perturbation* for each $\mathbf{q}'$ in our experiments, that sufficiently improves the robustness of correspondences. Then the candidate matches problem, which satisfied the epipolar constraint and limited by the neighbor region, can be formalized as:

$$\mathcal{N}(\mathbf{x}, \mathbf{x}') = \{(\mathbf{q}, \mathbf{q}') | \mathbf{q} \in \mathcal{N}(\mathbf{x}), \mathbf{q}' \in \mathcal{N}'(\mathbf{x}')\}. \tag{4}$$

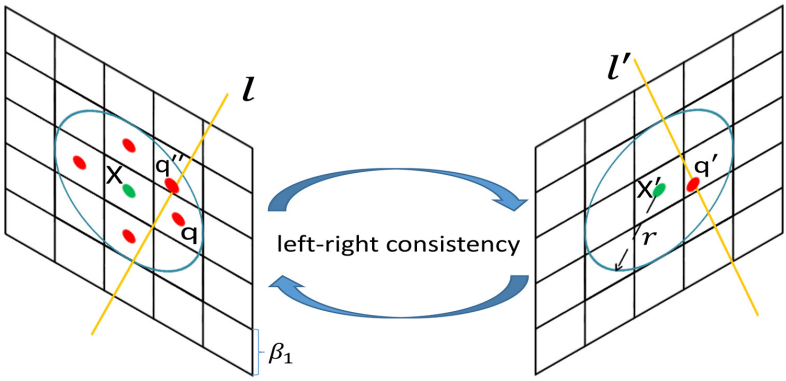

Fig. 2. Possible match $(\mathbf{q}, \mathbf{q}')$ indicated by the red dots around a seed match $(\mathbf{x}, \mathbf{x}')$ indicated by green dots. The point $\mathbf{q}'$ satisfies the epipolar constraint and within the neighborhood of $\mathbf{x}'$. if the match $(\mathbf{q}, \mathbf{q}')$ has the best correlation score and satisfies the *left-right consistency*, we add it to the list of seed matches. See text for more details.

Candidate matches are sorted by increasing correlation score using (1), then we select the best correlation for each $\mathbf{q}$ in $\mathcal{N}(\mathbf{x})$ by *left-right consistency* testing. The left-right consistency testing is just used to find the best correspondence $\mathbf{q}''$ in the first image of $\mathbf{q}'$ by the above procedure conversely. If the formula $||\mathbf{q} - \mathbf{q}''|| < \alpha$ is satisfied and the correspondence has the best correlation score related to the current grid, we add the match $(\mathbf{q}, \mathbf{q}')$ to the current list of seeds($\alpha = 4.0$ in all our experiments). Fig. 2 illustrates the entire procedure. In the match expansion procedure, we choose only the best match that has not been selected, and expands only the reliable matches whose correlation score below a certain threshold γ ($\gamma = 0.4$ in all our experiments). This drastically limits the bad matches for expansion and guarantees the ending of the process.

The expansion procedure produces dense but irregular distribution correspondences. Since these correspondences are not suitable for geometric computation, resampling procedure is proposed to refine these correspondences. We regularize these correspondences by locally selecting the best correspondence. Concretely, we redivide the first image into new regular square grids of $\beta_2 \times \beta_2$ pixels ($\beta_2 = 8$ in all our experiments). For each new grid, we select the correspondence with best correlation score and only accept it when its score below the threshold λ ($\lambda = 0.08$ in all our experiments).

The resampling procedure can effectively filter the erroneous correspondences and generate evenly distributed correspondences with geometric consistency, which are suitable for Bundler. In practice, we keep the original correspondences of feature points as they have better property of tracking than other points obtained by expansion. One example for a real pair of images is illustrated in Fig. 3. Next, we estimate the camera postures by Bundler procedure taking these correspondences as input and simultaneously acquire quasi-dense scene point clouds. One real example is illustrated in Fig. 4a.

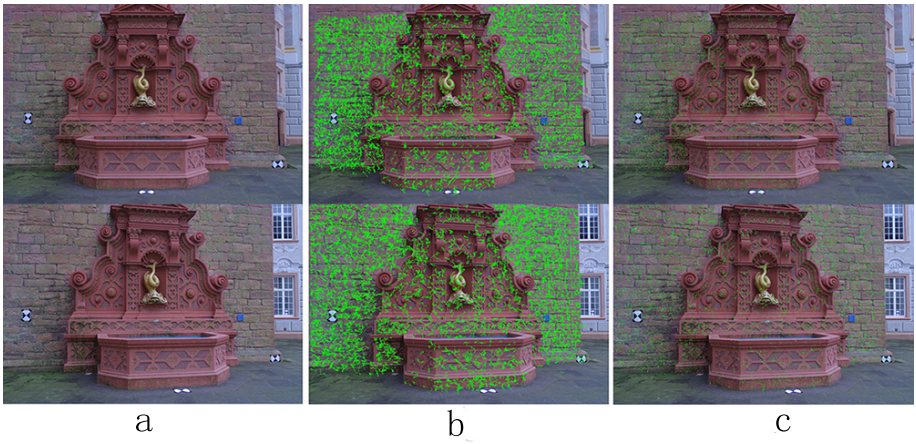

Fig. 3. The examples of *match expansion* algorithm. (a) The initial sparse matches from feature points. (b) The dense correspondences after expansion procedure. (c) The resampled correspondences.

4 Point-Cloud Expansion

As the quasi-dense scene point clouds reconstructed by Bundler are not dense and accurate, we propose the *point-cloud expansion* algorithm to further generate dense and accurate point clouds, which includes optimization and expansion of two steps. The optimization step aims to improve accuracy of point clouds under geometric constraint, and the other expansion step is used to produce dense point clouds combined with the optimization.

4.1 Optimization

Since the quasi-dense point clouds are not accurate, the optimization procedure is proposed to improve the accuracy of the point clouds. In order to achieve the goal and improve the robustness of experiments, we add orientation to each 3D point and draw a local tangent plane at that point, called *patch*. The patch's geometry is fully determined by its center $c(p)$ and unit normal vector $n(p)$. Then we define the photometric discrepancy function $\mathcal{G}(p)$ for patch p as:

$$\mathcal{G}(p) = \frac{1}{|V(p) \setminus R(p)|} \sum_{i=V(p)\setminus R(p)} \mathcal{C}_{ri}(q_r, q_i), \tag{5}$$

where $V(p)$ is a set of images in which p is visible and $R(p)$ denotes the reference image. The symbol backslash represents removing $R(p)$ from $V(p)$. The symbol q_r indicates the intersection of projecting the $c(p)$ into the reference view $R(p)$. The formula $\mathcal{C}_{rj}(q_r, q_i)$ indicates the correlation score between q_r and q_i by using (1). The map from q_r to q_i is the homography induced by the plane where the

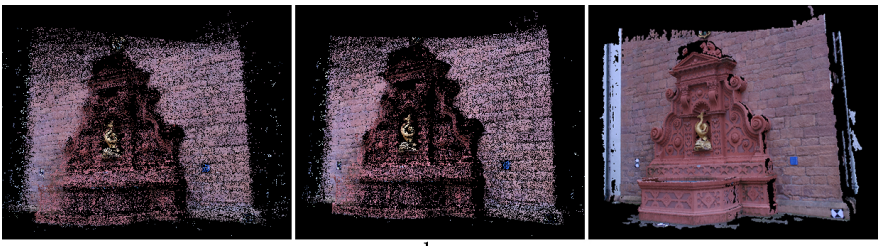

Fig. 4. (a) The quasi-dense point cloud reconstructed by Bundler. (b) The point cloud after optimization. (c) The final point cloud after expansion.

patch p is located. For a pair of images, the camera parameters can be represented by $\{K_i, R_i, C_i\}$ and $\{K_j, R_j, C_j\}$, thus the projection matrixes can denoted as: $P_i = K_i[R_i|-R_iC_i] = [M_i|-M_iC_i]$, $P_j = K_j[R_j|-R_jC_j] = [M_j|-M_jC_j]$, where $M_i = K_iR_i, M_j = K_jR_j$. The homography induced by the plane $f = \{c(p), n(p)\}$ for the cameras P_i and P_j is:

$$H_{ij} = M_jM_i^{-1} + \frac{M_j(C_i - C_j)n^T(p)M_i^{-1}}{n^T(p)(c(p) - C_i)}. \tag{6}$$

Then, the map can be represented by the formula, $q_i = H_{ri}q_r$.

Having defined the photometric discrepancy function $\mathcal{G}(p)$, our optimization strategy is to minimize discrepancy score of the patch. The corresponding parameters of the patch p can be simplified as:

$$c(p) = C_r + \lambda R_{ay}(p), \tag{7}$$

$$n(p) = [sin\theta cos\phi, sin\phi, -cos\theta cos\phi]^T. \tag{8}$$

We constrain $c(p)$ to lie on the viewing ray of p, $R_{ay}(p)$, from the reference camera, such that its image projection in the reference image does not change, reducing its three degrees of freedom to one and solving only for a depth λ. And the normal $n(p)$ can be parameterized by two angles θ and ϕ in spherical coordinate ($|\theta|, |\phi| < \pi/2$ in our experiments). So the optimization problem is reduced to three degrees of freedom and is solved by a conjugate gradient method. After the optimization procedure, we accept p only when its photometric discrepancy satisfies $\mathcal{G}(p') < \eta$ ($\eta = 0.2$ in all our experiments), this drastically limits erroneous patches and effectively improves accuracy of point clouds.

4.2 Expansion

To obtain dense point clouds, we further expand these patches under geometric constraint. For each patch p, the following expansion step is performed to generate new patches: We first find the candidate patches in the neighborhood of patch p, and these candidate patches also on the plane containing p (see an example in Fig. 5). For a candidate patch p', $c(p')$ is initialized as the 3D point

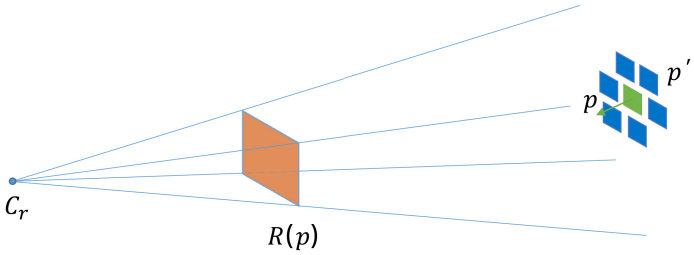

Fig. 5. Expansion three-dimensional(3D) patch p. We find the neighboring patches p' and refine them by the optimization procedure. See the text for more details.

near $c(p)$, and $n(p')$ is initialized by the formula $\frac{C_r - c(p')}{||C_r - c(p')||}$, where C_r is the camera center of the reference image of the patch p. And then, we refine $c(p')$ and $n(p')$ by the optimization procedure described above. After the optimization, we add the patch p' to the queue of expansion, when its photometric discrepancy is small enough.

After the point-cloud expansion procedure, we acquire dense point clouds, but still exist errors. We use some heuristic rules to remove these erroneous patches. Finally, the reconstructed point clouds of scenes or objects are dense and accurate, one real example shows in Fig. 4.

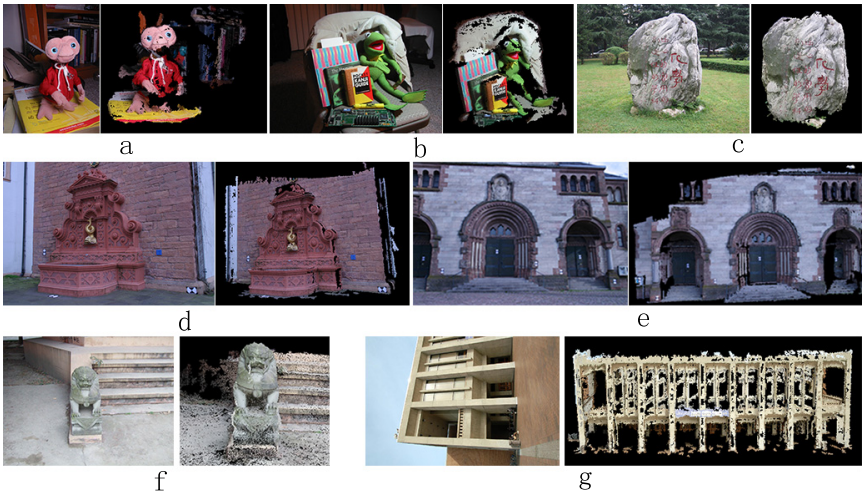

Fig. 6. Sample input images of data sets used in our experiments and corresponding results reconstructed by our method. From left ro right and top to bottom: (a) *ET*, (b) *kermit*, (c) *stone*, (d) *fountain*, (e) *herz-Jesu*, (f) *lion*, (g) *hall*. In each case, one of the images is shown, along with the reconstructed point clouds.

5 Experimental Results and Comparisons

We show our experimental results on some different data sets in Fig. 6. The *stone* and *lion* data sets have been acquired in our lab, while other data sets have been provided by S.Seitz (*ET*, *kermit* [14]), Y.Furukawa (*hall* [4]), C.Strecha (*fountain*, *herz-Jesu* [15]). Fig. 6 shows sample input images of all of the data sets used in our experiments and corresponding results reconstructed by our method. Table 1 lists the number of input images, their approximate size, and the number of 3D points reconstructed. All the experiments are implemented on an Intel 4GHz CPU with 32GB RAM. As illustrated by the figure, the object and scene point clouds reconstructed by our method are quite dense and accurate.

Table 1. Characteristics of the Data Sets

Data set	*ET*	*kermit*	*stone*	*lion*	*fountain*	*herz-Jesu*	*hall*
Num of Images	9	11	9	145	11	8	61
Size of Image	480 × 640	640 × 480	2848×2136	640×480	3072×2048	3072×2048	1200×797
Num of points finally	162650	132110	2182491	1506239	6752176	6087278	1918973

Fig. 7. Compared with PMVS using the benchmark data sets, *fountain-P11* and *herz-Jesu-P8*. The regions framed by the green squares have been enlarged, and displayed in the top left corner.

Before comparing with PMVS [4], we set parameters for the proposed approach and PMVS, and all parameters for the proposed approach have been discussed above. For PMVS, we set its parameters as: *level = 0, csize = 1, threshold = 0.7, wsize = 7*. Fig. 6d, e respectively show sample images of the two benchmark *fountain* and *herz-Jesu* data sets and results reconstructed by the proposed approach. In order to further demonstrate the accuracy and density of our results, we compare our method with PMVS, see Fig. 7. Our method can get denser results than PMVS visually, as the regions denoted by green squares. To quantitatively evaluate the proposed approach, together with PMVS, the quantitatively measure provided by [10] is used in the evaluation. Fig. 8 shows the *accuracy* of PMVS and the proposed approach, and the results demonstrate that our method is more accurate than PMVS.

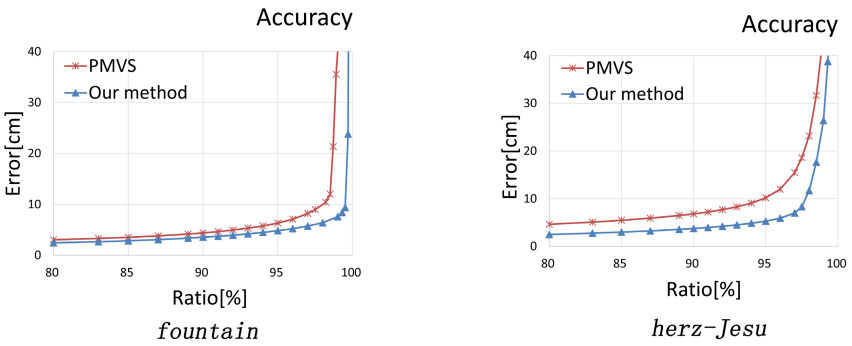

Fig. 8. The quantitative evaluation between the two benchmark data set *fountain* and *herz-Jesu*. For each data set, we evaluate the *accuracy* between ground truth and the reconstructed results by PMVS and the proposed approach.

6 Conclusion

An effective multi-view stereo approach is developed that obtains dense and accurate point clouds from multiple uncalibrated images. The approach mainly contains two algorithms: 1) the *match expansion* algorithm, which is used to expand initial matches to the geometric consistent correspondences, based on which the Bundler procedure can estimate camera parameters and simultaneously reconstruct quasi-dense point clouds; 2) the *point-cloud expansion* algorithm, which is used to further improve the density and accuracy of point clouds under geometric constraint. The experimental results demonstrate our approach is effective and can reconstruct quite dense and accurate point clouds from uncalibrated images. Quantitative evaluation shows that the proposed approach is favorable to the state-of-the-art method PMVS in terms of *accuracy* for the two benchmark data sets.

Acknowledgments. The authors thank the editors and anonymous reviewers for their insights. This work is supported by NSFC under grants 61173182 and 613111154, funding from Sichuan Province (2014HH0048, 2014HH0025) and the Science and Technology Innovation seedling project of Sichuan (2014-033, 2014-034).

References

1. Bosch, A., Zisserman, A., Muoz, X.: Image classification using random forests and ferns. In: IEEE 11th International Conference on Computer Vision (ICCV 2007), pp. 1–8 (2007)
2. Cremers, D., Kolev, K.: Multiview stereo and silhouette consistency via convex functionals over convex domains. IEEE Transactions on Pattern Analysis and Machine Intelligence **33**(6), 1161–1174 (2011)

3. Furukawa, Y., Ponce, J.: Carved visual hulls for image-based modeling. International Journal of Computer Vision **81**(1), 53–67 (2009)
4. Furukawa, Y., Ponce, J.: Accurate, dense, and robust multiview stereopsis. IEEE Transactions on Pattern Analysis and Machine Intelligence **32**(8), 1362–1376 (2010)
5. Goesele, M., Snavely, N., Curless, B., Hoppe, H., Seitz, S.M.: Multi-view stereo for community photo collections. In: IEEE 11th International Conference on Computer Vision (ICCV 2007), pp. 1–8 (2007)
6. Lhuillier, M., Quan, L.: A quasi-dense approach to surface reconstruction from uncalibrated images. IEEE Transactions on Pattern Analysis and Machine Intelligence **27**(3), 418–433 (2005)
7. Liu, Y., Cao, L., Liu, C., Pu, Y., Cheng, H.: Recovering shape and motion by a dynamic system for low-rank matrix approximation in l1 norm. The Visual Computer **29**(5), 421–431 (2013)
8. Liu, Y., Liu, B., Pu, Y., Chen, X., Cheng, H.: Low-rank matrix decomposition in L1-norm by dynamic systems. Image and Vision Computing **30**(11), 915–921 (2012)
9. Liu, Y., Sun, B., Shi, Y., Huang, Z., Zhao, C.: Stereo Image Rectification Suitable for 3D Reconstruction 用于3维重建的图像校正. Journal of Sichuan University (Engineering Science Edition) **45**(03), 79–84 (2013)
10. Seitz, S.M., Curless, B., Diebel, J., Scharstein, D., Szeliski, R.: A comparison and evaluation of multi-view stereo reconstruction algorithms. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 519–528. IEEE (2006)
11. Seitz, S.M., Dyer, C.R.: Photorealistic scene reconstruction by voxel coloring. International Journal of Computer Vision **35**(2), 151–173 (1999)
12. Shen, S.: Accurate multiple view 3D reconstruction using patch-based stereo for large-scale scenes. IEEE Transactions on Image Processing **22**(5), 1901–1914 (2013)
13. Snavely, N., Seitz, S.M., Szeliski, R.: Photo tourism: Exploring photo collections in 3D. ACM Trans. Graph. **25**(3), 835–846 (2006)
14. Snavely, N., Seitz, S.M., Szeliski, R.: Modeling the World from Internet Photo Collections. International Journal of Computer Vision **80**(2), 189–210 (2007)
15. Strecha, C., von Hansen, W., Van Gool, L., Fua, P., Thoennessen, U.: On benchmarking camera calibration and multi-view stereo for high resolution imagery. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008), pp. 1–8, June 2008
16. Tola, E., Strecha, C., Fua, P.: Efficient large-scale multi-view stereo for ultra high-resolution image sets. Machine Vision and Applications **23**(5), 903–920 (2012)
17. Uh, Y., Matsushita, Y., Byun, H.: Efficient multiview stereo by random-search and propagation. In: 2014 2nd International Conference on 3D Vision (3DV), vol. 1, pp. 393–400, December 2014
18. Vogiatzis, G., Torr, P.H.S., Cipolla, R.: Multi-view stereo via volumetric graph-cuts. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), vol. 2, pp. 391–398, June 2005