Incremental 3D Reconstruction Using Bayesian Learning

Ze-Huan Yuan[1], Lu Tong[1,2,*], Hao-Yi Zhou[1], Chen Bin[1], and Jia-Ning Li[1]

[1] State Key Laboratory of Software Novel Technology, Nanjing University, China 210093
[2] Jiangyin Institute of Information Technology of Nanjing University, China
zhyuan001@gmail.com, lutong@nju.edu.cn

Abstract. We present a novel algorithm for 3D reconstruction in this paper, converting incremental 3D reconstruction to an optimization problem by combining two feature-enhancing geometric priors and one photometric consistency constraint under the Bayesian learning framework. Our method first reconstructs an initial 3D model by selecting uniformly distributed key images using a view sphere. Then once a new image is added, we search its correlated reconstructed patches and incrementally update the result model by optimizing the geometric and photometric energy terms. The experimental results illustrate our method is effective for incremental 3D reconstruction and can be further applied for large-scale datasets or to real-time reconstruction.

Keywords: Incremental reconstruction, Bayesian model, PMVS.

1 Introduction

In computer vision, 3D reconstruction has been one of the widely researched areas in the recent decades, and automatic geometric reconstruction plays a key role in automated intelligent systems. With the decreasing costs of video equipments, we now have the opportunity and an urgent need to run automated and accurate 3D reconstruction algorithms directly on multiple photographs or video clips. Indeed, the most important technological ingredients towards this goal are already in place. We have known that feature matching algorithms [6] can provide accurate correspondences, structure-from-motion (SFM) algorithms use these correspondences to evaluate accurate camera pose, and multi-view-stereo (MVS) methods finally reconstruct dense and accurate surface models of complex objects from a moderate number of calibrated images. Actually, the existing MVS algorithms has nearly achieved surface coverage of about 95% and depth accuracy of about 0.5 mm from a set of low resolution (640x480) images as reported [1, 18].

MVS plays an important role in automatic acquisition of geometric objects. Existing state-of-the-art MVS algorithms can be roughly categorized into four classes: *voxel*, *mesh*, *depth maps* and *patch* based methods. *Voxel-based* MVS methods (VMVS) [2], [3], [4], [5] represent geometry on a regularly sampled 3D grid (volume), either as a discrete occupancy function or a function encoding distance to the closest surface.

* Corresponding author.

H. Jiang et al. (Eds.): IEA/AIE 2012, LNAI 7345, pp. 754–763, 2012.

Algorithms based on *deformable polygonal meshes* [7] [8] represent a surface as a set of connected planar facets and operate by iteratively evolving a surface to decrease or minimize a cost function. Approaches based on *multiple depth maps* [9], [10] model a scene as a set of depth maps and fuse individual depth maps into a single 3D model. Finally, *patch-based* MVS (PMVS) [1] algorithms output a dense collection of small oriented rectangular patches covering the observed surface obtained from pixel-level correspondences. Recently, CMVS [17] is approved effective in reconstructing from images of crowed scenes without any initialization process.

However, the mentioned methods still face the following difficulty. They cannot well handle incremental reconstruction tasks. In another word, the input images should be well sequenced before reconstruction. Once a geometric object is obtained, it cannot be incrementally updated when facing a new input view image.

Thus in this paper, we propose a novel algorithm aiming at incrementally reconstructing a 3D model using the Bayesian framework. We first select a group of key views uniformly distributed on our view sphere to create an initial 3D surface modeled by PMVS as stated above. Then when a new calibrated image is input, we 1) map it into a triangle on our view sphere, 2) search the correlated patches with the new input view, and 3) automatically update the initial 3D model using the photometric consistency constraint and geometric smoothness priors under the Bayesian inference framework. Note that once a new image is added, more geometric details can be extracted and integrated to incrementally optimize the final 3D model.

Our method has two main contributions. First, we propose a novel incremental 3D reconstruction framework, which makes full use of new views to incrementally update an existing 3D model. As a result, the reconstruction process is more efficient and convenient, especially useful for automatic 3D reconstruction from a large number of real-life images or videos and real-time reconstruction. Second, to our knowledge, no previous work has attempted to reconstruct 3D models using the Bayesian learning framework, where pixel-level information and geometric level constraints are well integrated to optimize the final model. As a result, the reconstruction accuracy can be effectively improved.

2 Our Method

In this section, we give our incremental reconstruction algorithm in details. Our method can be briefly summarized as the following three steps:

1. Map the given multi-view images set I_{source} to a view sphere $S_{initial}$ and select uniformly distributed key views to initialize a 3D model;
2. For each new input image i_{new}, map it to $S_{initial}$ and search its related patches set P_{update} on the 3D model;
3. Re-calculate the patches of P_{update} using the Bayesian learning framework to incrementally refine the 3D model.

Step 2 and 3 are repeated until there are no new input images. Note that in Step 2, only a subset P_{update} (named *seed patches set*) on the previous 3D model is chosen to be updated for any new input image rather than all the patches on the model. It is based on the following fact that in each incremental recursion step, the existing patches on the previous 3D model may have different *correlations* to i_{new} and we need not update those patches having low *correlations*. For example, there is no (or too low) *correlation* between i_{new} and another patch that is completely invisible to it. This helps reduce the computational cost, simultaneously without losing accuracy in our incremental reconstruction.

2.1 Initialize a 3D Model

Given a calibrated image set I_{source}, we need firstly to select an image subset uniformly distributed in different viewpoints to reconstruct an initial 3D model. The initial key views are selected as follows: 1) map each view image in I_{source} to a view sphere $S_{initial}$ (see Fig. 1(a)), with its coordinate determined by the corresponding image plane, namely the normalized principal axis vector obtained from its projection matrix, and 2) sample the key views uniformly across the sphere.

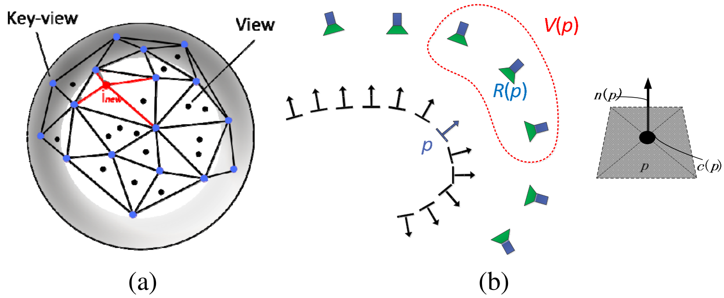


Fig. 1. (a) The view sphere. (b) The patch model.

Next, we triangulate $S_{initial}$ by grouping the neighboring key views on it into triangles using the Delaunay Triangulation algorithm [11]. 3D initial geometric model $\overline{S}$ can be simultaneously reconstructed using [12] from key views. Note that the geometric contour is reconstructed using the patch-based approach [1], where a 3D surface is covered by a plenty of patches, and a patch p is essentially a local tangent plane approximation of the surface. A patch p here has three geometric attributes (see Fig. 1(b)): $c(p)$, $n(p)$ and $R(p)$, where $c(p)$ denotes the geometric center, $n(p)$ is the unit normal vector oriented toward the camera observing it, while a reference image $R(p)$ is an image chosen from $V(p)$ where p is *truly* visible on the condition that the retinal plane of $R(p)$ is nearly parallel to p within a tiny distortion.

As a result, a triangulated view sphere and a 3D patch model are obtained as the initializations of our incremental updating system.

2.2 Search Related Patches for a New Input Image

In our incremental reconstruction step, we first search a corresponding patch subset from the previous 3D model for any new input calibrated image, and then extend the subset to make the model more uniform and well-sampled.

2.2.1 Search Seed Patches for Any Input Image

To search the seed patches P_{update} for any incrementally input image i_{new}, we first search a proper triangle T on $S_{initial}$, where i_{new} can be mapped into using SIFT [6] as follows:

$$T \leftarrow \arg\max_{T} \sum\nolimits_{v\in T} | x_{i_{new}}^{v} | \tag{1}$$

where $x_{i_{new}}^{v}$ is a set of matches between i_{new} and the key view v corresponding to a vertex in triangle T. Then we search the correlated patch subset P_{update} from the reconstructed 3D model by

$$P_{update} = \bigcup_{v\in T} \{ p \mid p \in \overline{S}, v \; is \; R(p) \} \tag{2}$$

Obviously, i_{new} provides more useful reconstruction details for the patches in P_{update} than those outside it. Then we update $S_{initial}$ as follows: 1) add a new vertex representing the new image; 2) add a pyramid of triangles by connecting the new image to the three vertices of T, and 3) delete T with i_{new} located in. As a result, we can simultaneously obtain an updated view sphere (see i_{new} in Fig. 1(a)).

2.2.2 Extend the Seed Patches

Next, we extend the patch model to obtain a relatively uniform patch density along different viewpoints over the surface. The extension is associated with the orientation of the new view and the average density of the existing global surface. Note that during this process, we may create new patches under the local geometric constraints to improve patch density where patches are too sparse. Our extension has the following steps:

- Estimate local density D_p for every patch p in 3D model. We count its neighbors $N(p)$ to evaluate the local density equivalently as follows:

$$N(p) = \{ p' \mid p' \in \overline{S}, |(c(p) - c(p')) \cdot n(p)| + |(c(p) - c(p')) \cdot n(p')| < \rho \} \tag{3}$$

$$D_p = | N(p) | \tag{4}$$

  where ρ can be computed relating to the distance at the depth of the center of $c(p)$ and $c(p')$ corresponding to an image displacement of u pixels in $R(p)$(u=2 in our experiment);
- Compute the global average density D_g by averaging all estimated local densities;
- For every seed patch in P_{update} with its local density less than 0.5* D_g, use the SMOTE [13] to oversample new ones whose initialization can be seen in Table 1 between the seed patch and its neighbors (see Fig. 2). As a result, the original geometric constraints can be well maintained;
- Add the new patches into P_{update}.

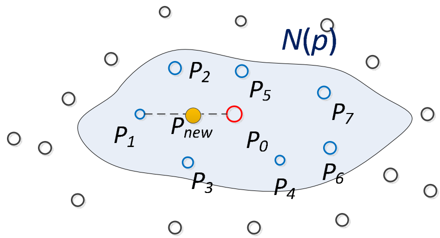


Fig. 2. Seed patches extension, where P_{new} is generated along the line combining a seed patch P_0 and one of its neighbors P_1

2.3 Incremental Surface Reconstruction Using Bayesian Learning

This section introduces the Bayesian model used in our incremental reconstruction. We aim at discovering the photometric consistency and geometric smoothness constraints to obtain high-quality incremental reconstruction results.

Suppose i_{new} is a *measurement* to our camera from the real scene modeled by PMVS in our method. Let S be the real scene to be modeled, we need reconstruct the most likely surface S_{MAP} given the *measurement* i_{new}. This can be achieved by maximizing the Bayesian posterior probability $p(S|i_{new})$ in the solution space Ω

$$p(S \mid i_{new}) = \frac{1}{Z} p(i_{new} \mid S) p(S), S \in \Omega \tag{5}$$

$$S_{MAP} = \arg\min\ (-\log p(i_{new} \mid S) - \log p(S)) \tag{6}$$

in order to reduce the parameter dimensions, we constraint Ω to the expanded patches subset P_{update} as mentioned in Section 2. Note that the constant related to Z is ignored in (6). $p(i_{new} \mid S)$ specifies the likelihood of the *measurement* i_{new} agreeing with S. In other words, it measures how well the normal and coordinate of a patch match the real surface according to the information hidden in i_{new} and the other correlated images. It can be defined by the use of photometric discrepancy function [1], which we choose to express the photometric consistency:

$$p(i_{new} \mid S) \propto \exp(-\eta E_p) \tag{7}$$

$$E_p = \frac{1}{|S|} \sum_{p \in S} \frac{1}{|V(p)|-1} \sum_{i \in V(p)/i_{new}} h(p, i_{new}, i) \tag{8}$$

whereη is a control coefficient, and $h(p, i_{new}, i)$ is equal to one minus the pair-wise normalized cross correlation concerning to the patch projection into images i_{new} and i.

We use two constraints to define the prior $p(S)$:

$$p(S) \propto \exp(-\{\lambda E_1 + \zeta E_2\}) \tag{9}$$

where E_1 and E_2 are two geometric smoothness energy terms, and λ, ζ are weighted coefficients. E_1 is used to assure the smoothness of the reconstructed surface. For a natural 3D object, we can model its surface smoothness by accumulating sub-linear potentials of surface curvature similar to [14]. Concretely, we define E_1 as follows:

$$E_1 = \frac{1}{|S|}\sum_{p\in S}\frac{1}{|N(p)|}\sum_{v\in N(p)} f(p,v) \tag{10}$$

$$f(p,v) = \sqrt{(n(p)-n(v))^T(n(p)-n(v))} \tag{11}$$

where $N(p)$ is the neighboring patches set of p defined in (3), $f(p,v)$ is the square-root potential with $f(p,v)=0$ if $n(p)=n(v)$ and positive otherwise.

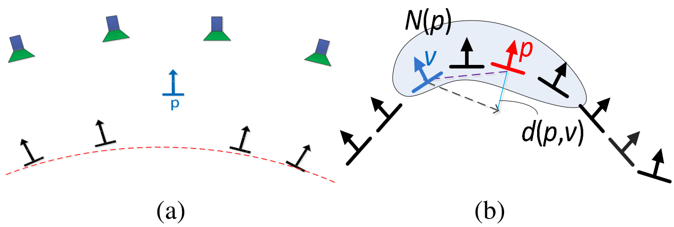


Fig. 3. Geometric smoothness terms. (a) The blue patch p is an outlier; however it has a continuous normal with its neighboring patches. (b) $d(p,v)$ is the absolute distance between two patches p and v along $n(p)$.

However, there still may exist exceptions even (10) is met. For example, in Fig. 3(a), the patch p is an outlier while having well sub-linear continuous relations with normals of its neighbors in $N(p)$. Considering although such a patch has a continuous normal, its geometric location is far away from the real surface, we use another geometric smoothness energy term E_2 to minimize such errors as follows:

$$E_2 = \frac{1}{|S|}\sum_{p\in S}\frac{1}{|N(p)|}\sum_{v\in N(p)} d(p,v) \tag{12}$$

$$d(p,v) = |n(p)\cdot(c(v)-c(p))| \tag{13}$$

where $d(p,v)$ is the distance between two patches p and v along $n(p)$(see Fig. 3(b)).

This minimization problem requires us to adjust $c(p)$ and $n(p)$ for any patch in S from the initial value to the final convergent solution. It is actually a sparse energy minimization optimization problem. To simplify the complexity and reduce the dimension of variables, we constrain $c(p)$ lie on a ray to assure the projection into $R(p)$ is not changed. Simultaneously, we model $n(p)$ with Euler angles. Thus for every patch, only three parameters participate in the optimization problem, greatly reducing the dimension of the solution space and improve stability in the search process. We use the conjugate gradient descent to solve the global optimization. In this process, the derivatives for geometric smoothness prior can be directly computed and those for the photometric consistency term are currently estimated numerically.

As a summary, our incremental updating algorithm is shown in Table 1.

Table 1. The incremental algorithm

Input : $S_{initial}$ and 3D patch model $\overline{S}$ reconstructed by PMVS

Output : an improved well-sample, high-resolution and more accurate patch model

While Input an image i_{new}

Locate i_{new} in $S_{initial}$ and find a corresponding triangle T using SIFT

For any p in the 3D patch model

$$N_p \leftarrow \{p' \mid p' \in \overline{S}, |(c(p)-c(p'))\cdot n(p)| + |(c(p)-c(p'))\cdot n(p')| < \rho\}$$

$$D_p \leftarrow |N(p)|$$

$$P_{update} \leftarrow \bigcup_{v \in T} \{p \mid v \ is \ R(p)\}$$

Update $S_{initial}$

Compute D_g by averaging all local density

For any p in P_{update}

If D_p<0.5* D_g

Generate a new patch k

$c(k)$,$n(k) \leftarrow$ oversampling method *smote*(N_p, *sample-rate,p*).

$R(k) \leftarrow R(p)$

$V(k) \leftarrow V(p)$

Add k into P_{update}

For any patch p in P_{update}

$$c(p), n(p) \leftarrow \arg\min(\lambda E_1 + \zeta E_2 + \eta E_p), p \in P_{update}$$

end while

3 Experiments and Discussions

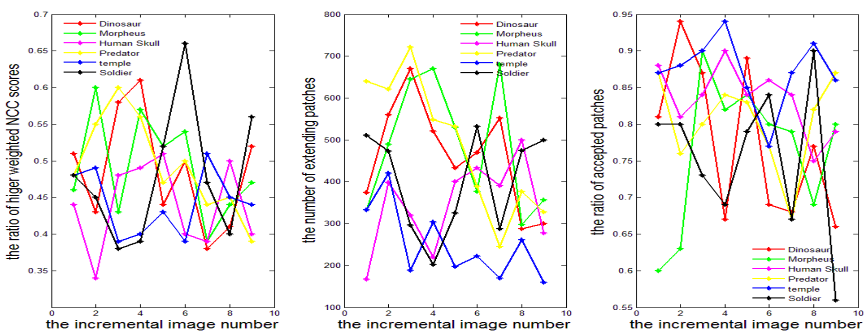
We have implemented our incremental reconstruction algorithm on C++ platform. The datasets [15][16] used in our experiments are shown in Table 2, with the number of the input images, their approximate sizes, the number of the key views we choose and the patch number of the reconstructed initial model using PMVS [12]. In our incremental process, we set λ, ζ and η 0.3, 0.2 and 0.7, respectively.

Fig. 5 gives the incremental reconstruction results of different models, where Column (a) and Column (b) correspond to example 2D images and their initial result models reconstructed from key views, respectively. After gradually adding new images, the result models are incrementally updated, as shown in the rest three columns (c)-(e). It can be seen that the result models can be dynamically optimized and enriched with more details during these processes.

Table 2. The datasets used in our experiments

Name	Images	Image size	Key views	Initial patches
Toy Dinosaur	24	2000*1500	15	27267
Morpheus	24	1400*1200	15	18433
predator	24	1800*1800	15	29620
Human Skull	24	2000*1800	15	45223
temple	312	640 * 480	209	32317

To evaluate our method quantitatively, we adopt the weighted sum of normalized cross correlation (NCC) [1] to model the accuracy of a patch. During each incremental step, we calculate the ratios of those patches with larger weighted NCC scores in P_{update} (see Fig. 4). Fig. 4 is a discrete figure where different points on curves have no relations and can be replaced by tables if enough space available. It can be seen that after adding a new image, the NCC accuracy of nearly 50% of its related patches are improved averagely, illustrating the effectiveness of our method.



(a) (b) (c)

Fig. 4. The overall statistic analysis. (a) the ratio of patches having higher photometric consistency scores, (b) the number of extended patches, and (c) the ratio of accepted extending patches for different incremental images.

We also find that in Fig. 4(a), the ratio changes with image quality and position on our view sphere during the incremental reconstruction steps. It is due to that for poor-quality images, geometric smoothness term plays an important role in the optimization, and thus the accuracy may be reduced simultaneously because of over-smoothing.

Fig. 4(b) illustrates the number of the extended patches in each incremental reconstruction step with the *sample-rate* as 200% in our experiments. Obviously, the number greatly depends on the viewpoint of 2D images and more patches need to be generated in sparse regions. Note that not all the extended patches are finally added to the result model due to the global geometric constraints and the pixel-level information. Fig. 4(c) gives the accepted patch ratios in our experiments.

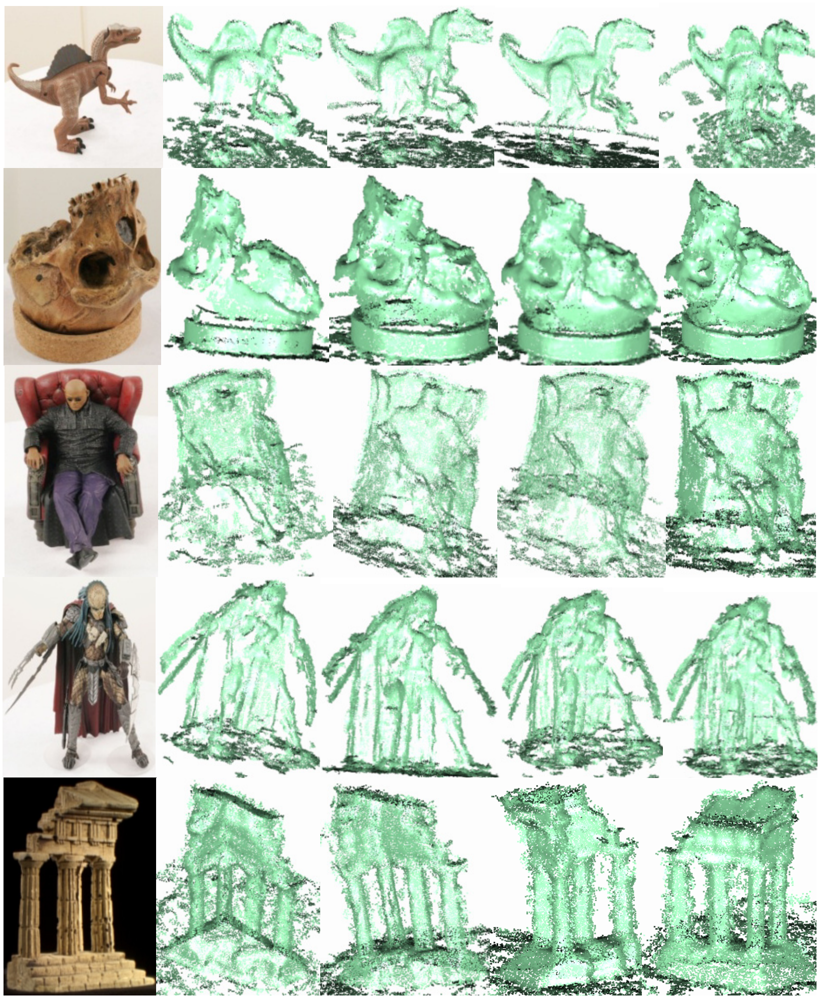


Fig. 5. Our incremental reconstruction results. (a) 2D sample images, (b) the initial 3D model, (c)-(e) the incremental reconstruction results. From top to bottom, the datasets are *dinosaur*, *human skull cast*, *Morpheus*, *predator* and *temple*.

4 Conclusions

We have developed a novel incremental reconstruction algorithm for calibrated multi-view stereo. Our method first initializes a 3D patch model using the selected key views, and then when inputting a new image interactively, seed patches for which the new image provides useful reconstruction details are searched and then extended to make surface of the 3D target uniform. We end up the incremental learning under Bayesian framework. We focus our future work on directly reconstructing crowed scene models from real-life videos and online real-time reconstruction. Another improvement may lie on better evaluating 3D model reconstruction methods, especially for incremental reconstruction applications.

Acknowledgements. The work described in this paper was supported by the New Century Excellent Talent of Ministry of Education Project under NCET-11-0232, National Science Foundation of China under Grant No. 61021062, and the 973 Program of China under Grant No. 2010CB327903.

Reference

1. Furukawa, Y., Ponce, J.: Accurate, dense, and robust multi view stereopsis. PAMI 32(8), 1362–1376 (2010)
2. Pons, J.-P., Keriven, R., Faugeras, O.D.: Multi-view stereo reconstruction and scene flow estimation with a global image-based matching score. IJCV 72(2), 179–193 (2007)
3. Tran, S., Davis, L.: 3D Surface Reconstruction Using Graph Cuts with Surface Constraints. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3952, pp. 219–231. Springer, Heidelberg (2006)
4. Vogiatzis, G., Torr, P.H., Cipolla, R.: Multi-view stereo via volumetric graph-cuts. In: CVPR, pp. 391–398 (2005)
5. Hornung, A., Kobbelt, L.: Hierarchical volumetric multi-view stereo reconstruction of manifold surfaces based on dual graph embedding. In: CVPR, pp. 503–510 (2006)
6. Lowe, D.: Distinctive image features from scale-invariant keypoints. IJCV 60(2), 91–110 (2004)
7. Hernández Esteban, C., Schmitt, F.: Silhouette and stereo fusion for 3D object modeling. CVIU 96(3), 367–392 (2004)
8. Furukawa, Y., Ponce, J.: Carved visual hulls for image-based modeling. IJCV 81(1), 53–67 (2009)
9. Goesele, M., Curless, B., Seitz, S.M.: Multi-view stereo revisited. In: CVPR, pp. 2402–2409 (2006)
10. Strecha, C., Fransens, R., Gool, L.V.: Combined depth and outlier estimation in multi-view stereo. In: CVPR, pp. 2394–2401 (2006)
11. `http://www.cse.unsw.edu.au/~lambert/java/3d/delaunay.html`
12. `http://grail.cs.washington.edu/software/pmvs`
13. Chawla, N.V., Bowyer, K.W., Hall, L.O., Philip Kegelmeyer, W.: SMOTE: Synthetic Minority Over-sampling Technique. JAIR 16, 321–357 (2002)
14. Diebel, J.R., Thrun, S.: A Bayesian Method for Probable Surface Reconstruction and Decimation. ACM Transactions on Graphics - TOG 25(1), 39–59 (2006)
15. `http://www.cs.washington.edu/homes/furukawa/research/mview/index.html`
16. `http://vision.middlebury.edu/mview/data/`
17. Furukawa, Y., Curless, B., Seitz, S.M., Szeliski, R.: Towards Internet-scale Multi-view Stereo. In: CVPR, pp. 1434–1441 (2010)
18. Furukawa, Y., Ponce, J.: Accurate camera calibration from multi-view stereo and bundle adjustment. IJCV 84(3), 257–268 (2009)