Layer-Constraint-Based Visibility for Volumetric Multi-view Reconstruction

Yumo Yang, Liuxin Zhang, and Yunde Jia

Beijing Laboratory of Intelligent Information Technology,
School of Computer Science, Beijing Institute of Technology,
Beijing 100081 P.R. China
{yangyumo,zhangliuxin,jiayunde}@bit.edu.cn

Abstract. Visibility estimation is one of the most difficult problems in multi-view reconstruction using volumetric approaches. In this paper, we present a novel approach called layer-constraint-based visibility (LCBV) to estimating visibility. Based on the layered state of a scene and photo-consistency constraint, this method can determine the more accurate visibility for every point in a scene. We use LCBV in multi-view reconstruction using volumetric graph cuts and obtain satisfactory results on both synthetic and real datasets. We also discuss quantitative error analysis to judge visibility techniques.

1 Introduction

The basic principle in volumetric multi-view reconstruction is to find a classification for all the voxels within a discrete volume whether they belong to the surface of the 3D object or not. There are many volumetric techniques like Space Carving [1], level set [2] and volumetric graph cuts [3,4,5,6]. These methods obtain a uniform 3D representation; in particular, graph cuts are able to find a global optimization [7].

Volumetric approaches usually use the photo-consistency measure of the voxel within the volume to evaluate how consistent would be the reconstructed surface at the voxel. The only requirement to compute the photo-consistency is that visibility is available. However, accurate visibility estimation is very difficult for reconstruction. To estimate the true visibility of some surface point, one needs to know the true scene geometry and vice versa. To solve this chicken-and-egg problem, previous methods compute approximation of visibility. Some papers [8,9] develop probabilistic formulations with the visibility reasoning to evaluate visibility. Hernandez et al. [8] present probabilistic visibility in which the visibility of a point depends only on the probabilistic depth measurements of sensors along optical rays that go through the point. This approach is both computationally and memory efficient, but it is not computed independently. There are also some simple and independent visibility techniques, such as state-based visibility [10] and oriented visibility [11] as shown in Fig. 1a and 1b respectively. State-based visibility uses the current estimate of the geometry to predict visibility for every point on that surface globally. In detail, a point is considered visible

S. Boll et al. (Eds.): MMM 2010, LNCS 5916, pp. 15–24, 2010.

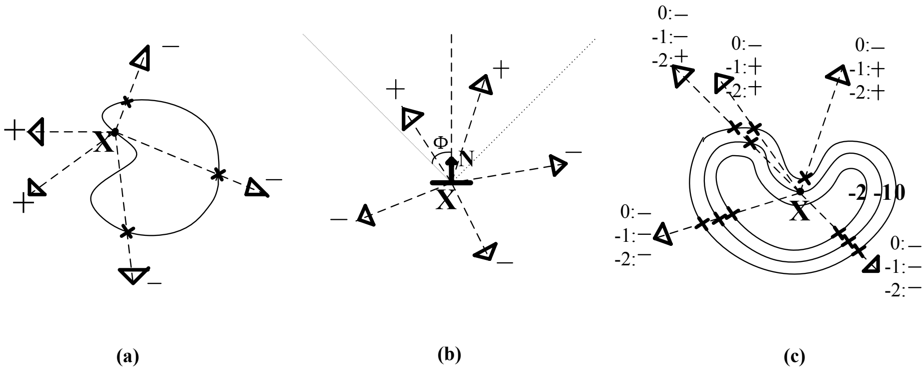


Fig. 1. Three approaches to visibility reasoning. (a) state-based visibility: the visibility for X on its surface is estimated globally. (b) oriented visibility: a local patch (X, N) is considered visible from the viewpoints within a predefined angle from the normal direction N. (c) layer-constraint-based visibility: the layered state and photo-consistency constraint are used to estimate the visibility information for all the voxels within the volume globally. Signs 0, -1, -2 denote layer 0, layer -1, and layer -2 respectively.

to a viewpoint if it is not occluded with current scene configuration. This global technique is often used in the approaches based on iterative process [1,2], which is not guaranteed to converge to the global minimum. Oriented visibility is a local algorithm which infers the visibility of a patch by using the patch position and the normal direction. Concretely, the patch is considered visible to a viewpoint if and only if it locates in front of this viewpoint according to a predefined angle from the normal direction. This method is simple and independent from initialization but, because of unknown the normal directions of the inner points, the accurate visibility for every point inside the volume is not determined. In many papers [4,5,6,12] with this technique, the visibility for the inner point is approximated with the visibility of the closest surface point to it.

In this paper, we propose a novel visibility approach called layer-constraint-based visibility (LCBV) as shown in Fig. 1c. Based on the layered state of the scene and photo-consistency constraint, LCBV can determine the more accurate visibility information for every voxel in the volume in volumetric multi-view reconstruction. Moreover, a quantitative error analysis is presented to compare the visibility methods in detail.

Our solution is inspired by the work of Zhang et al. [10] and Jonathan et al. [13]. Zhang et al. [10] used state-based visibility on the iterated surface which is regarded as the zero of level set. Jonathan et al. [13] divided the scene space into a set of depth layers and the visibility is derived by testing for occluders both on the depth layer and base surface, however, they did not reason visibility in detail. We combine both methods and divide visual hull [14] into several virtual layers like zero level sets. Then state-based visibility is extended to the virtual layers with photo-consistency constraint.

Contrast to the traditional visibility methods such as state-based visibility and oriented visibility, the main benefit of layer-constraint-based visibility is that the visibility information of the points within the volume can be determined accurately. Theoretically, we propose a quantitatively analysis on computing the visibility information for the points within the base surface, while the traditional methods only give a qualitative analysis on the visibility for the points within the base surface. For example, in the traditional visibility techniques, the visibility information for the inner points is approximated with the visibility of the closest surface point to it. In terms of application, the key advantage of LCBV is its ability to produce the better intermediate result to be used in multi-view reconstruction.

Intuitively, the accurate computation of the photo-consistency of the right point (the point is on the true surface) is more important than the computation of the photo-consistency of the wrong point (the point is not on the true surface) in reconstruction. LCBV can infer the more accurate visibility information for these right points to benefit producing a better intermediate result in multi-view reconstruction. According to LCBV, we also give an energy functional which is minimized by graph cuts.

The rest of the paper is organized as follows. Section 2 introduces our layer-constraint-based visibility approach. In Section 3, we describe an energy function which is minimized by graph cuts in multi-view reconstruction. Section 4 presents a significant error analysis and experimental results on synthetic dataset and produces several significant reconstruction results on both synthetic and real datasets. We discuss the paper's main contributions and the future work in Section 5.

2 Layer-Constraint-Based Visibility

In this section, we describe the layer-constraint-based visibility (LCBV) for evaluating the more accurate visibility information for the points within the volume in volumetric multi-view reconstruction. This algorithm is divided into two steps: one step is to develop a method called layered visibility (LV) for computing visibility roughly. The next is to use photo-consistency constraint to obtain the accurate visibility.

2.1 Layered Visibility for Volumetric Reconstruction

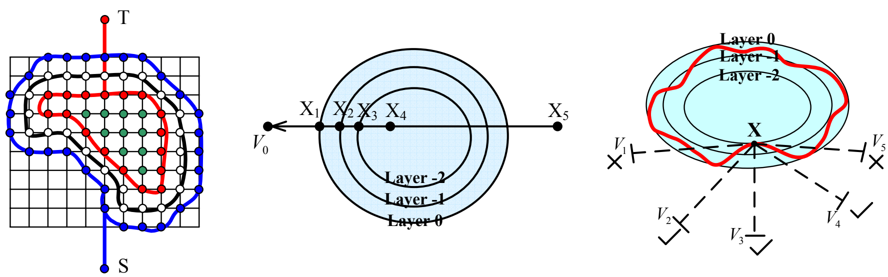
We assume that the whole scene is located within a bounding volume $\Omega \in \Re^3$. Let $I_1, \ldots, I_N$ denote the sequence of input calibrated images and $S_1, \ldots, S_N$ denote the foreground silhouettes obtained from these input images. The visual hull VH, the maximal shape consistent with $S_1, \ldots, S_N$, is considered as the initial space. Similar to the work [13], we divide visual hull VH space into a set of virtual layers $V = \{l_0, \ldots, l_D\}$. Each voxel in the virtual layers forms a node in an undirected graph and edges are defined in the graph to link adjacent voxels with a capacity corresponding to the energy function for minimization.

The first layer l_0 corresponds to an external set of voxels and the final layer l_D an internal set, each connected to a source s and sink t node respectively with infinite edge capacity as illustrated in Fig. 2a [13]. We compute LV on multiple layers and modify the function ϕ in [10] to define a new function $\varphi : \Re^3 \rightarrow \Re$ in Ω, which represents the scene layers as $[-D, 0]$ ($-D$ represents the D-th layer in visual hull VH space)of φ, with the following properties:

$$\begin{cases} \varphi(\mathbf{X}) > 0 & \text{for } \mathbf{X} \text{ outside } V \\ \varphi(\mathbf{X}) < -D & \text{for } \mathbf{X} \text{ inside } V \\ \varphi(\mathbf{X}) = 0 & \text{for } \mathbf{X} \in l_0 \subset V \\ \vdots & \quad \vdots \quad \vdots \\ \varphi(\mathbf{X}) = -D & \text{for } \mathbf{X} \in l_D \subset V \end{cases}$$

where φ is similar to a level set function; the layers in V are similar to zeros of level set function; $\mathbf{X}$ denotes a voxel in the bounding volume. The values of φ are easy to derive from layered state.

We use ray tracing [10] to multiple layers and determine in which layer the point in the volume is visible. Let $\mathbf{V_0}$ denote a viewpoint in Ω, then a point $\mathbf{X_1}(\mathbf{X_1} \neq \mathbf{V_0}, \mathbf{X_1} \in \Omega)$ is considered invisible to the viewpoint $\mathbf{V_0}$ for layer $l_i (i = 0, 1, \ldots, D)$ if it lies inside the layer l_i or if it is occluded by another point $\mathbf{X_2}(\mathbf{X_2} \in l_j, j \in [i, \ldots, D])$ in the radial direction $\overrightarrow{\mathbf{X_1 V_0}}$ (Fig. 2b). Under this condition, we shoot a ray towards the viewpoint, and check whether there is a point being inside the layer in this ray. We can determine visibility for a point simply by looking at the value of function φ. That is, $\mathbf{X}$ is inside the layers



(a) Discrete graph for volumetric reconstruction via graph cuts

(b) Ray Tracing for multiple layers

(c) Layer visibility with photo-consistency constraint

Fig. 2. 3D reconstruction via volumetric graph cuts, ray tracing and photo-consistency constraint. (a) A discrete graph is constructed for volumetric reconstruction via graph cuts, and outer (blue line) and inner (red line) layers are connected to source S and sink T respectively. (b) A point is considered visible for some layer when it is not occluded by the same or inner layer, e.g. $\mathbf{X_1}$ is visible for layer 0; $\mathbf{X_2}$ is invisible for layer 0 but visible for layer -1; $\mathbf{X_3}$ is invisible for layer 0, -1 but visible for layer -2; $\mathbf{X_4}$ and $\mathbf{X_5}$ are invisible for three layers. (c) Red line denotes the true surface. According to layered visibility, point $\mathbf{X}$ is considered visible for v_1, v_2, v_3, v_4 and v_5, but in fact, $\mathbf{X}$ is invisible for v_1 and v_5. After imposing photo-consistency constraint, the last result is that $\mathbf{X}$ is visible for v_2, v_3 and v_4 (v_1 and v_5 are excluded).

when $\varphi(\mathbf{X}) < -D$, outside the layers when $\varphi(\mathbf{X}) > 0$, and on some layer when $\varphi(\mathbf{X}) = -i, i \in \{0, \ldots, D\}$.

2.2 Photo-Consistency Constraint

As discussed in Section 2.1, we approximately evaluate from which views the voxel on some layer is visible. In this part, we obtain the final accurate visibility by imposing photo-consistency constraint as shown in Fig. 2c. The photo-consistency criteria is not used in the traditional visibility techniques, but usually used in the optimization in multi-view reconstruction after computing the visibility. In this paper, we impose photo-consistency constraint on computing the visibility to improve the results. Basically, if a point is on the true surface, it has high photo-consistency. In contrast, if a point has high photo-consistency, it is not always on the true surface. For example, under weak texture, even points very far from the surface may have a strong photo-consistency. However, we mainly consider that accurately computing the visibility information of the right points (see Section 1), since the visibility information of the wrong points (see Section 1) has little influence on multi-view reconstruction.

In this paper, the normalized cross-correlation (NCC) is used as a photo-consistency metric. Let VC denotes the set of all views, VB denotes the set of the views for which the voxel is visible after computing layered visibility and VE denotes the set of the views for which the voxel is visible after using photo-consistency constraint. Obviously, we have $VB \subset VC, VE \subset VC$. Assume, to begin with, after computing layered visibility (see Section 2.1) we obtain the rough visibility information for the voxels and we have $VB = \{v_1, v_2, v_3, v_4, v_5\}$, v_1,v_2,v_3,v_4,$v_5 \in VC$,$VE = \Phi$. Views v_1, v_2, v_3, v_4, v_5 array in order, and the middle view is chosen as reference view so that its retinal plane is close to parallel to the tangent surface through the voxel with little distortion. Then the closer to the reference the other views are, the more accurate they are.

We first choose the middle value v_3 in the set VB as reference view and turn left or right from v_3 to search a value, e.g. the first value v_2, each time alternately. If such the NCC computed by views v_2 and v_3 is above some threshold W, that means the voxel is invisible for view v_2 and the program is end ($VE = \Phi$). Otherwise, we put v_2 into the set $VE(VE = \{v_3, v_2\})$. Next, v_4 is searched out and the mean NCC computed by views v_2, v_3, v_4 is obtained again to determine whether v_4 is put into the set VE or not. Until all the values in the set VB are searched, we acquire a final accurate visibility set VE. The algorithm excludes inexact views in VB through imposing photo-consistency constraint and produces the more accurate set of views. The results of the synthetic experiment prove that layered visibility with photo-consistency constraint, namely layer-constraint-based visibility (LCBV), is more accurate than oriented visibility.

3 Volumetric Multi-view Reconstruction

Most of multi-view reconstruction methods relying on the traditional visibility techniques need a good initial scene surface to evaluate the visibility information for the points on or in the initial surface [4,5,6]. If the current scene state is far from the true state, e.g. concavity, the traditional visibility methods will approximate the true visibility with significant errors. Therefore, the traditional visibility algorithms are not thought to be good choice in the situation when no good initialization is given. In this section, we deduce an energy functional for multi-view reconstruction based on layer-constraint-based visibility. Compared with other methods, the main benefit of our reconstruction is that concavity can be better recovered.

3.1 Volumetric Graph Cuts

The energy functional is generally defined in terms of a data term E_{data} that imposes photo-consistency and a regularization term E_{smooth} introducing spatial smoothness [13]. Following Section 2.1, we obtain a layered configuration (see Section 2.1). In contrast with the photo-consistency constraint used in visibility (see Section 2.2), here a new photo-consistency metric $\rho(u)$ used in optimization is computed for each node u and edges are constructed between nodes (u, v) using capacities corresponding either to the data term or smoothness term of the cost function.

$$E = E_{data} + E_{smooth}. \tag{1}$$

Data edges are introduced between nodes in adjacent layers (l_i, l_{i+1}) using photo-consistency score at layer $-i$. Smoothing edges are introduced between adjacent nodes within each layer l_i using an average of the photo-consistency. Edge capacities are normalized by edge length and a relative weight $o < k < 1$ controls the degree of smoothing in reconstruction (similar to [13]),

$$E_{data}(u, v) = 1 - k\frac{\rho(u)}{\|x(u) - x(v)\|}, u \in l_i, v \in l_{i+1}, 0 < k < 1, \tag{2}$$

$$E_{smooth}(u, v) = 1 - k\frac{\rho(u) + \rho(v)}{2\|x(u) - x(v)\|}, u, v \in l_i, 0 < k < 1, \tag{3}$$

where $x(u), x(v)$ are the coordinates of node u and v respectively; the photo-consistency score for node u, $\rho(u)$ is the normalized cross correlation (NCC) between the pairs of local image patches that node u projects to in the different views:

$$\rho(u) = \sum_{C_i, C_j \in Vis(u)} NCC(p(C_i, u), p(C_j, u)), \tag{4}$$

where $Vis(u)$ notes the visibility for node u, and $Vis(u)$ is evaluated with layer-constraint-based visibility; C_i, C_j note two camera centers; $p(C_i, u)$ is the local image patch around the images of u in the i -th image I_i.

The global minimization of the energy functional corresponding to an optimal surface is derived from computing the max flow/min cut method [15].

4 Experimental Results

Because visibility results are just intermediate results in multi-view reconstruction and are difficult to be represented by graphic or image, to judge whether a visibility method is better or not is rather difficult. Most of the papers judge the visibility in the indirect way that the final reconstruction results determine whether the visibility technique is better or not. We propose a direct quantitative analysis on the visibility methods by the data form. We demonstrate the performance of our approaches on synthetic dataset, and give a significant error analysis (Table 1). We define two types of errors:

(1)E_{r_1} which denotes the wrong rate of set VE (Section 2) relative to set VT (real visibility set). If there exists a camera $X(X \in VE, X \notin VT)$, the set VE is considered wrong, namely the visibility for the point is wrong.

$$E_{r_1} = \frac{M_{VE}}{N_{VE}} \times 100\%, \tag{5}$$

where M_{VE} denotes the number of the points which have the wrong visibility in V, and N_{VE} denotes the number of all the points in V. Note that there is a problem about E_{r_1}, that is, if $VE \subset VT, VE \neq VT$, then VE is considered right, because of little influence on the photo-consistency, but not exact. Therefore another type of error is presented.

Table 1. Error analysis for the visibility. E_{r_1} denotes wrong rate of set VE(section 2) relative to VT . E_{r_2} denotes the similar degree of VE and VT. Theshold W is 0.3.

Error	E_{r_1}	E_{r_2}(viewpoint)
Oriented visibility	36.56%	4.177
Our method	26.76%	1.237

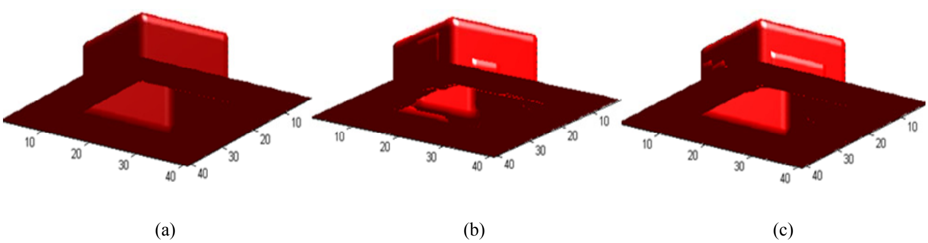


Fig. 3. A synthetic scene contains a non-closed surface with a large cubic concavity. (a) the true surface. (b) the reconstruction result using oriented visibility. (c) the reconstruction result using layer-constraint based visibility.

Table 2. Quantitative analysis on the accuracies for reconstructions using oriented visibility and our method. Accuracy and completeness are similar to the work in [16].

Reconstruction	Accuracy (mm)	Completeness
Oriented visibility	0.209996	70.29%
Our method	0.169690	78.84%

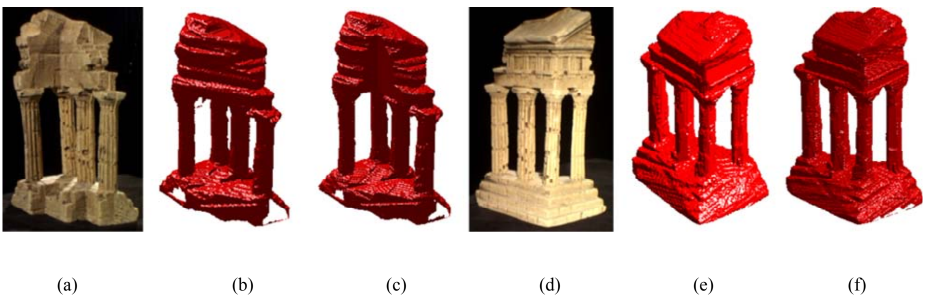


Fig. 4. Reconstruction for temple in Middlebury datasets. (a,d) the true temple images (16 views). (b,e) visual hull. (c,f) the reconstruction result using our method.

(2)E_{r_2} which denotes the similar degree of VE and VT, then

$$E_{r_2} = \frac{\sum |VT - VE|}{N}, \text{when} VE \subset VT, \tag{6}$$

where N denotes the number of investigated voxels. Viewpoint is used as unit of measure, e.g. E_{r_2}=1.900viewpoint, that means when $VE \subset VT$, the average number of the viewpoints which can see the point but not be computed right is 1.9.

Note that the synthetic scene is a non-closed surface with a large cubic concavity which makes it a challenging test (Fig. 3a). 21 viewpoints in line face to the concave surface. We compare the results of our method and oriented visibility to the true visibility (we can get it) respectively on the synthetic image. Table 1 quantitatively analyses the two types of errors on the visibility methods. Fig. 3b,3c show the 3D reconstruction results on the synthetic images using both the methods. Table 2 quantitatively analyses the accuracy and completeness of reconstructions on the synthetic image. The results present that our method is better than oriented visibility in multi-view reconstruction. Similar to the work in [16], we use an accuracy threshold of 80% and a completeness threshold of 0.1mm.

In Fig. 4, we apply our visibility algorithm to reconstruct to the temple (16 images) in Middlebury datasets [17] and acquire a promising result.

5 Conclusion and Future Work

In this paper, a novel visibility method called layer-constraint-based visibility (LCBV) for volumetric multi-view reconstruction has been proposed to determine the accurate visibility for every voxel within the volume. Using this method, an energy functional which can be minimized by graph cuts is presented. Moreover, quantitative error analysis between visibility approaches is first developed. Above are our main contributions. In the future work, we will modify the photo-consistency constraint and improve the results on real datasets. Our results in real datasets will be improved and submitted to Middlebury. Because we put emphasis on the study of the visibility, the reconstruction results via volumetric graph cuts are not perfect. We will also improve graph cuts optimization.

Acknowledgements

This work was partially supported by the Natural Science Foundation of China (90920009) and the Chinese High-Tech Program (2009AA01Z323).

References

1. Kutulakos, K.N., Seitz, S.M.: A theory of shape by space carving. International Journal of Computer Vision 38(3) (2000)
2. Faugeras, O., Keriven, R.: Variational principles, surface evolution, PDEs, level set methods, and the stereo problem. IEEE Transactions on Image Processing 7 (1998)
3. Zeng, G., Paris, S., Quan, L., Sillion, F.: Progressive surface reconstruction from images using a local prior. In: ICCV (2005)
4. Vogiatzis, G., Torr, P., Cipolla, R.: Mult-view stereo via volumetric graph-cuts. In: CVPR (2005)
5. 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)
6. Ladikos, A., Benhimane, S., Navab, N.: Multi-view reconstruction using narrow-band graph-cuts and surface normal optimization. In: BMVC (2008)
7. Kolmogorov, V., Zabih, R.: What energy functions can be minimized via graph cuts? In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2352, pp. 65–81. Springer, Heidelberg (2002)
8. Hernandez, C., Vogiatzis, G., Cipolla, R.: Probabilistic visibility for multi-view stereo. In: CVPR (2007)
9. Boykov, Y., Lempitsky, V.: From photohulls to photoflux optimization. In: BMVC (2006)
10. Liuxin, Z., Yumo, Y., Yunde, J.: State-based visibility for 3D reconstruction from multiple views. In: ECCV Workshop (2008)
11. Lempitsky, V., Boykov, Y., Ivanov, D.: Oriented visibility for multiview reconstruction. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3953, pp. 226–238. Springer, Heidelberg (2006)

12. Esteban, C., Schmitt, F.: Silhouette and stereo fusion for 3D object modeling. Computer Vision and Image Understanding 96(3) (2004)
13. Starck, J., Hilton, A., Miller, G.: Volumetric stereo with silhouette and feature constraints. In: BMVC (2006)
14. Laurentini, A.: The visual hull concept for silhouette-based image understanding. TPAMI 16(2) (1994)
15. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. PAMI 26(9) (2004)
16. Seitz, S., Curless, B., Diebel, J., Scharstein, D., Szeliski, R.: A comparison and evaluation of multi-view stereo reconstruction algorithms. In: CVPR (2006)
17. The middlebury multi-view stereo page, `http://vision.middlebury.edu/mview/`