

Occlusion-Model Guided Anti-Occlusion Depth Estimation in Light Field

Hao Zhu, Qing Wang
 School of Computer Science
 Northwestern Polytechnical University
 qwang@nwpu.edu.cn

Jingyi Yu
 University of Delaware
 jingyiyu@udel.edu

Abstract

Occlusion is one of the most challenging problems in depth estimation. Previous work has modeled the single-occluder occlusion in light field and get good results, however it is still difficult to obtain accurate depth for multi-occluder occlusion. In this paper, we explore the multi-occluder occlusion model in light field, and derive the occluder-consistency between the spatial and angular space which is used as a guidance to select the un-occluded views for each candidate occlusion point. Then an anti-occlusion energy function is built to regularize depth map. The experimental results on public light field datasets have demonstrated the advantages of the proposed algorithm compared with other state-of-the-art light field depth estimation algorithms, especially in multi-occluder areas.

1 Introduction

Depth estimation has been researched for decades. A common and important assumption is the photo-consistency assumption, *i.e.*, the colors of a point observed from different views ought to be similar. This assumption holds for non-occlusion points. However, it fails for occlusion points as these points can not be observed from all views.

Many works have been done to handle occlusions. Kolmogorov *et al.* [5] encodes the visibility constraint and introduces an occlusion term to smooth it. Woodford *et al.* [18] adds an additional second order smoothness terms and use Quadratic Pseudo-Boolean Optimization to solve it. Then, Bleyer *et al.* [11] applies the asymmetric occlusion model to improve depth estimation. However, due to the wide baseline and the lack of views, these

works only notice the image of the occlusion, *i.e.*, the occlusion point is visible in reference view and invisible in other views, and heavy occlusion can not be well handled.

Light-field cameras from Lytro [7] and Raytrix [10] obtain a 4D light field by inserting a micro-lens array into the traditional camera [9], which provides hundreds of views in a single shot. Apart from this, the baseline between these views is very small, which means no aliasing occurs and the consistent correspondences still holds in this case [19]. Combining the multi-view and the micro-baseline, a more detailed and complete representation of occlusion appears, which renews the method to handle occlusion.

Wanner *et al.* [15, 16] applies structure tensor to analyze the Epipolar Plane Image (EPI). This method only takes advantage of angular samples in one dimension, and the tensor field becomes too random to analyze in heavy occlusion. Yu *et al.* [20] encodes the constraints of 3D lines and introduces Lind Assisted Graph Cuts (LAGC) to improve depth estimation. However, the 3D lines are partitioned into small and incoherent segments in heavy occlusion which leads to wrong estimation. Chen *et al.* [4] proposes to select the un-occluded views by using a bilateral metric in angular space. However, this selection lacks the guidance of the physical model and the number of un-occluded views is also a predefined parameter. Wang *et al.* [13, 14] analyzes the formation of occlusion and finds the consistency between the spatial patch and the angular patch in occlusion boundaries. They select the un-occluded views according to the edges in spatial patch. However, their method failed in multi-occluder areas where the local patch can not be divided into two regions by a straight line, and it leads to over smooth results in these areas (Fig. 7, 8). Although Wang *et al.* have a more recent work [12] for depth estimation, it aims at solving shadings which is not applicable in our case.

In this paper, we explore the light field occlusion theory for multi-occluder occlusion, and propose an algorithm to regularize depth map. Our main contributions are:

- 1) The light field occlusion theory for multi-occluder occlusions.
- 2) An algorithm to accurately select the un-occluded views in angular space with the guidance of the occlusion theory which does not need any predefined parameters.
- 3) A depth estimation algorithm which is robust to multi-occluder occlusion.

In Section 2, we model the multi-occluder occlusion in light field, and derive the occlusion-consistency between the spatial and angular space, *i.e.*,

the occluded views in angular space are projections of the occluder in spatial space. With the guidance of occluder-consistency, we select the un-occluded views for each candidate occlusion point using a partition in spatial patch in Section 3.1, and obtain an initial depth map using the un-occluded views in Section 3.2. Then, we detect the occlusion points from the initial depth map according to the visible constraint in Section 4.1. Finally the occlusion map and the un-occluded views are used to build an anti-occlusion energy function to refine depth in Section 4.3. In Section 5, we provide complete comparisons with other state-of-the-art algorithms, both in quantitative and in qualitative, showing great advantages compared with previous works [15, 16, 20, 4, 13].

2 The Light Field Occlusion Model

In this section, we will analyze the formation of the occlusion based on the physical model, and explore the occlusion theory for multi-occluder occlusion.

2.1 Definitions and Notations

Before building the light field occlusion model, we first give some definitions and notations of the light field model.

Definition 1. Single-occluder occlusion refers to the occlusion that occluded views and un-occluded views can be divided half-and-half.

Definition 2. Multi-occluder occlusion refers to the occlusion that there are more occluded views than un-occluded views.

Point (x, y, u, v) in 4D light field describes a ray emitting from the world point (X, Y, Z) , and intersecting with two parallel planes, namely, the uv camera plane and the xy image plane, which we refer to as angular and spatial space. Tab. 1 presents a list of terms used throughout this section.

2.2 Occlusion Model

Previous work [13] has proved the occluder-consistency for single-occluder occlusion, *i.e.*, when refocused to the correct depth, the edge which separates the un-occluded and occluded views in the angular patch has the same orientation as the occlusion edge in spatial patch. This property is useful for single-occluder occlusions, however it fails in multi-occluder situation as

Table 1: Notations of the light field occlusion model.

Term	Definition
(u, v)	angular/camera plane coordinates
(x, y)	spatial/image plane coordinates
(X, Y, Z)	world coordinates
\vec{e}_A	the vector in angular coordinate system
\vec{e}_I	the vector in spatial coordinate system
\vec{e}_W	the vector in world coordinate system

the un-occluded and occluded pixels in angular patch can not be divided into two regions by a straight line.

We first consider a simple multi-occluder occlusion (Fig. 1). Considering a pixel (X_0, Y_0, F) on the focal plane (the left image in Fig. 1(a)), and an occluder intersects at (X_0, Y_0, Z_0) ($0 < Z_0 < F$). Note that the occluder has two edges, and the directional vectors of these two edges in the plane $Z = Z_0$ are,

$$\begin{aligned}\vec{e}_W^1 &= (k_{X_1}, k_{Y_1}) = (X_1 - X_0, Y_1 - Y_0) \\ \vec{e}_W^2 &= (k_{X_2}, k_{Y_2}) = (X_2 - X_0, Y_2 - Y_0).\end{aligned}\tag{1}$$

The larger angle between these two vectors denotes occluded areas (the golden areas in Fig. 1). Without loss of generality, we assume $k_{Y_j} > 0, j = 1, 2$.

For any other pixel (X_i, Y_i, F) on the focal plane, it will be observed by the view (u_0, v_0) iff it meets the following inequalities,

$$\begin{aligned}k_{Y_1}(X_i - X_0) - k_{X_1}(Y_i - Y_0) &> 0 \\ k_{Y_2}(X_i - X_0) - k_{X_2}(Y_i - Y_0) &< 0.\end{aligned}\tag{2}$$

We then project these inequalities from world coordinate system to the image system (the right image in the Fig. 1(a)). The corresponding directional vectors of the \vec{e}_W^1 and \vec{e}_W^2 are $\vec{e}_I^1 = (k_{x_1}, k_{y_1})$ and $\vec{e}_I^2 = (k_{x_2}, k_{y_2})$ respectively. $\vec{e}_I^1 = \lambda_1 \vec{e}_W^1$ and $\vec{e}_I^2 = \lambda_1 \vec{e}_W^2$, λ_1 is a scale factor to denote the scaling relationship between the world coordinate system and the image coordinate system. For any other point (x_i, y_i) on the image, it is a background point iff,

$$\begin{aligned}k_{y_1}(x_i - x_0) - k_{x_1}(y_i - y_0) &> 0 \\ k_{y_2}(x_i - x_0) - k_{x_2}(y_i - y_0) &< 0.\end{aligned}\tag{3}$$

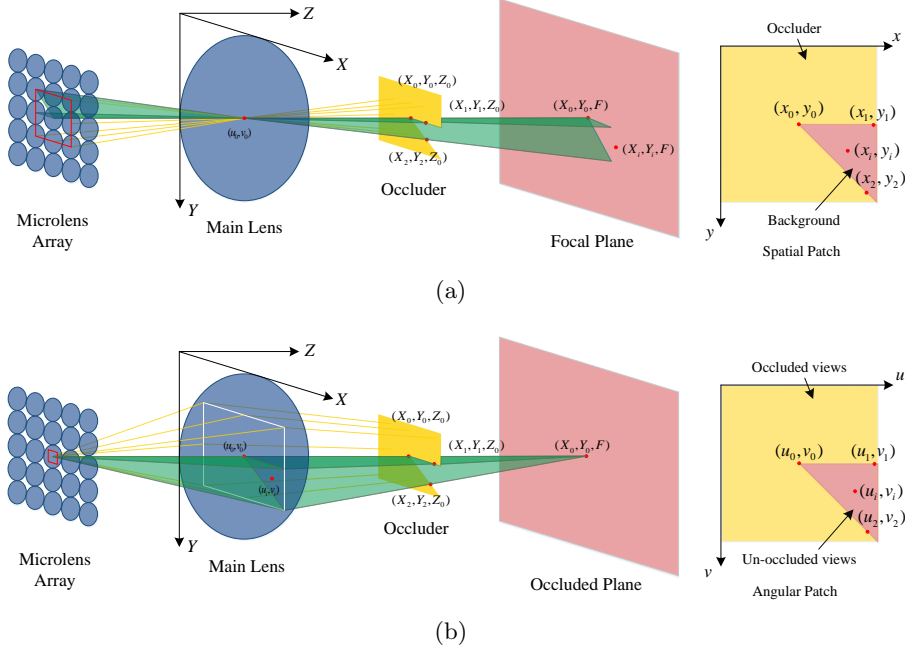


Figure 1: The light field camera model with occlusion. The left image in (a) denotes the the image captured from the view (u_0, v_0) , and the right image in (a) is a local patch centered at (x_0, y_0) from the view (u_0, v_0) . The left image in (b) denotes the light field is refocused in depth F , and only views constrained by two green planes can see the point (X_0, Y_0, F) , the images formed from other views describe the occluder. The right image in (b) is the angular patch of the point (x_0, y_0) .

Then considering the main lens plane (the left image in Fig. 1(b)). The light field is refocused to the depth F . For any other view (u_i, v_i) on the main lens plane, it can capture the pixel (X_0, Y_0, F) iff

$$\begin{aligned} k_{v_1}(u_i - u_0) - k_{u_1}(v_i - v_0) &> 0 \\ k_{v_2}(u_i - u_0) - k_{u_2}(v_i - v_0) &< 0. \end{aligned} \quad (4)$$

where $\vec{e}_A^j = (k_{u_j}, k_{v_j})$, $\vec{e}_A^j = \lambda_2 \vec{e}_I^j$, $j = 1, 2$, λ_2 is a scale factor to denote the scaling relationship between the image coordinate system and the angular coordinate system.

Revisiting the Eqn. 3 and 4, it is noticed Eqn. 3, 4 have the same inequalities and (u, v) , (x, y) are one-one correspondence, as (u_0, v_0) and

x_0, y_0 are on a same line, the following proposition can be obtained,

Proposition 1. *The occluded views in angular space are projection of the occluder in spatial space.*

In other words, in a local spatial patch, the corresponding views of the occluder are the occluded views, and the corresponding views of the background are the un-occluded views. This proposition is called occluder-consistency later.

Note that, the proposition mentioned above is obtained under a simple multi-occluder assumption. For a more complex multi-occluder, the boundaries of the occluder can be divided into more small straight lines by following the idea of the Calculus, and the Eqn. 3, 4 will contain more inequalities. No matter how many inequalities, the numbers of inequalities in Eqn. 3, 4 are equal and inequalities are one-one correspondence, and the Prop. 1 always holds.

2.3 Projection Radius

For occluded points, the occluded views in angular space are projection of the occluder in spatial space in a local patch. We derive the radius of the patch in a 2D light field. In Fig. 2, the purple lines denote the background at depth α_0 , the orange lines denote the occluder at depth α_1 , blue lines denote the camera plane, x denotes a pixel in the background, and $u_0 - u_4$ denote different views in light field.

Firstly, the light field is refocused to the background at depth α_0 (Fig. 2(a))[8],

$$LF_{\alpha_0}(x, u) = LF_0(x + (u - u_2)(1 - \frac{1}{\alpha_0}), u), \quad (5)$$

where LF_0 is the input light field, LF_{α_0} is the refocused light field at depth α_0 , u_2 is the central view of the light field. It is noticed the light from views u_0, u_1, u_2 converge to the point x_2 , and the light from u_3, u_4 are blocked by the occluder. u_3, u_4 are the occluded views in angular space, and the images of these two views come from points x_3, x_4 . In other words, the horizontal distance d between x_4 and x is the projection radius.

Then, the light field is refocused to the foreground at depth α_1 (Fig. 2(b))

$$LF_{\alpha_1}(x, u) = LF_0(x + (u - u_2)(1 - \frac{1}{\alpha_1}), u), \quad (6)$$

where LF_{α_1} is the refocused light field at depth α_1 . It can be seen the light from all views converge to the point x' . As the horizontal distance between

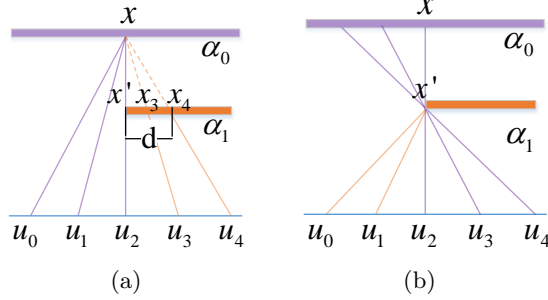


Figure 2: Projection radius. In (a), the light field is refocused to the background in depth α_0 , where the light from u_0, u_1, u_2 converge to point x and the light from u_3, u_4 are blocked by the occluder. In (b), the light field is refocused to the foreground in depth α_1 , and light from all views converge to point x' .

x' and x_2 is 0, the projection radius d is obtained

$$\begin{aligned}
 d &= |x + (u_4 - u_2)(1 - \frac{1}{\alpha_1}) - x - (u_4 - u_2)(1 - \frac{1}{\alpha_0})| \\
 &= |(u_4 - u_2)(\frac{1}{\alpha_0} - \frac{1}{\alpha_1})|.
 \end{aligned} \tag{7}$$

3 Depth Initialization

In this section, we will show how to use the occluder-consistency between the spatial and angular space to select the un-occluded views, and how to obtain an initial depth estimation using these un-occluded views.

3.1 Un-Occluded Views Selection

We first give an important assumption about occlusion of the proposed algorithm. The occluder has a different color of the occluded point. For the situation that the occluder is similar to occluded point, as far as we know, no work can handle it. Based on this assumption, the following proposition holds,

Proposition 2. *An occlusion point is an edge point but an edge point may not be an occlusion point.*

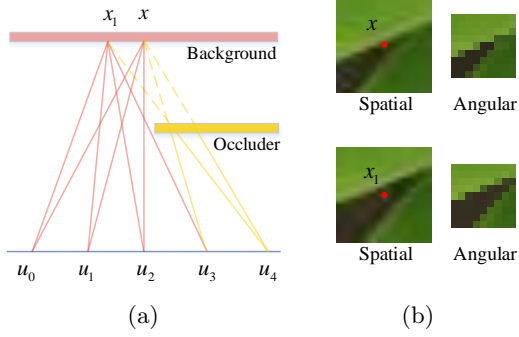


Figure 3: The occlusion in different views. (a) demonstrates the model of occlusion in central view and in other views respectively. Point x is occluded in central view and point x_1 is occluded in other views. (b) demonstrates the image of occlusion in central view and in other views respectively. It is noticed x is an edge point, and there are many occluded views in its angular patch when refocused to its true depth. The point x_1 is not an edge point, however as it near the edges, there are many occluded views in its angular patch too.

With Prop. 2, the canny edge detector is firstly applied to find the candidate occlusion points S_{occ} . Then the K-means clustering [1] is applied for the local image patch¹ centered at each occlusion point p from S_{occ} (the feature is the RGB color, and the number of labels is 2). For each patch, the pixels which share the same label with the center pixel are labeled as background or un-occluded points. According to the occluder-consistency mentioned in Prop. 1, the corresponding views in angular patch of the center pixel p are labeled as un-occluded views Ω_p^{uocc} . These pixels are candidate occlusion pixels in the central view (Fig. 3). For pixels occluded in other views, the un-occluded views are voted from its neighborhood system (Fig. 4).

3.2 Depth Estimation

With the un-occluded views selection, a robust initial depth estimation is obtained based on the classical photo-consistency in un-occluded views.

¹The patch size is set as half of the angular resolution of light field initially since we do not have depth map here.

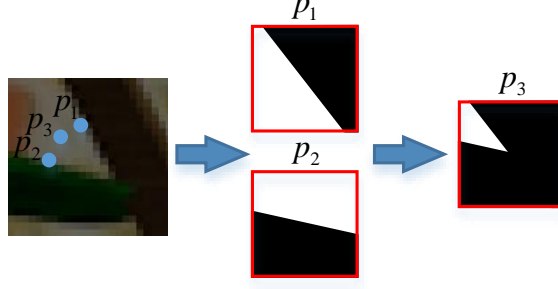


Figure 4: The voting strategy of obtaining un-occluded views for pixels occluded in other views. The first col shows that points p_1, p_2 are occluded in central view and p_3 is occluded in other views. The second row shows the un-occluded views of p_1, p_2 (the areas labeled with white are un-occluded views). The third col is the un-occluded views of p_3 after voting.

We refocus the light field to different depth α ,

$$LF_\alpha(x, y, u, v) = LF_0(x + u(1 - \frac{1}{\alpha}), y + v(1 - \frac{1}{\alpha}), u, v), \quad (8)$$

where LF_0 is the input 4D light field, LF_α is the refocused light field in depth α , (x, y) are the spatial coordinates, and (u, v) are the angular coordinates. Then, the matching cost of each pixel p is defined as,

$$C_\alpha^{uocc}(p) = \frac{1}{|\Omega_p^{uocc}|} \sum_{(u,v) \in \Omega_p^{uocc}} |A_{p_\alpha}(u, v) - A_{p_\alpha}(0, 0)|, \quad (9)$$

where Ω_p^{uocc} is the un-occluded views set of the point p (Sec. 3.1), and $|\cdot|$ denotes the size of the set \cdot , and $A_{p_\alpha}(u, v)$ denotes the angular image of pixel p at depth α (In other words, $A_{p_\alpha}(u, v) = LF_\alpha(x_p, y_p, u, v)$, (x_p, y_p) are the spatial coordinates of pixel p), and $A_{p_\alpha}(0, 0)$ is the color of pixel p in central view.

Then, the initial depth estimation of each pixel p is obtained,

$$\alpha_{ini}(p) = \arg \min_{\alpha} C_\alpha^{uocc}(p). \quad (10)$$

4 Depth Regularization

In this section, we will show how to find the occlusion and regularize it with a global energy function.

4.1 Occlusion Detection

We find the occlusion point using the visible constraint, *i.e.*, the occlusion point is visible in reference view and invisible in other views. In other words, if the difference of disparities of two neighboring pixels is larger than 1 pixel, there is an occlusion point here. In light field, this constraint can only find the occlusion point in the central view due to the multiple views, and the threshold value ϵ_{occ} of the difference of the disparities ought to be relaxed to find the occlusion in other views,

$$\epsilon_{occ} = \frac{1}{\lfloor N_{uv}/2 \rfloor}, \quad (11)$$

where N_{uv} is the angular resolution ².

As the initial depth estimation in occluded points is unreliable, and sometimes it is too smooth and random to distinguish the occlusion point accurately from only two neighboring pixels. We select a disparity patch centered at each candidate occlusion point (the detected points by the Canny operator), and use the K-means clustering to divide the patch into 2 classes, then the difference of disparities of each class Δdis is determined by the subtraction of two centers,

$$\Delta dis = |dis_1 - dis_2|, \quad (12)$$

where dis_i is the center of the i -th class.

Finally the candidate occlusion point p is determined by comparing the ϵ_{occ} and $\Delta dis(p)$,

$$Occ(p) = \begin{cases} 1 & \Delta dis(p) \geq \epsilon_{occ} \\ 0 & \Delta dis(p) < \epsilon_{occ}. \end{cases} \quad (13)$$

4.2 Un-occluded Views Re-Selection

For each occlusion point p , we apply the K-means clustering for its local depth patch to find the background depth α_0 and occluder depth α_1 , and the projection radius d is determined using Eqn. 7. Then, each patch is resized to the angular resolution of light field. The following procedure is the same as Sec. 3.1.

Since the occlusion points can not be fully detected in Sec. 4.1, we retain the un-occluded views selection for other candidate points obtained in Sec. 3.1.

²The angular resolution of the light field that we use in the experiment is 9×9 .

4.3 Final Depth Regularization

Finally, given the occlusion cues, we regularize with Markov Random Field (MRF) for a final depth map.

$$E = \sum_p E_{uocc}(p, \alpha_p) + \lambda_s \sum_{p,q} E_{smooth}(p, q, \alpha_p, \alpha_q) \quad (14)$$

where α_p is the depth of pixel p , and p, q are neighboring pixels, and λ_s ($=0.35$ in our experiment) is to control the smooth term.

The data term E_{uocc} measures the photo-consistency in un-occluded views,

$$E_{uocc}(p, \alpha_p) = 1 - e^{-\frac{C_{\alpha_p}^{uocc}(p)^2}{2\sigma_{uocc}^2}}, \quad (15)$$

where σ_{uocc} ($=3$ in our experiment) controls the sensitivity of the function to large distance, and the definition of $C_{\alpha_p}^{uocc}(p)$ can be found in Eqn. 9.

The smooth term E_{smooth} encodes the smoothness constraint between two neighboring pixels,

$$E_{smooth}(p, q, \alpha_p, \alpha_q) = \omega_{pq} |\alpha_p - \alpha_q|, \quad (16)$$

$$\omega_{pq} = e^{-\frac{(Occ(p)-Occ(q))^2}{2\gamma_{occ}^2} - \frac{(I_e(p)-I_e(q))^2}{2\gamma_e^2} - \frac{(I(p)-I(q))^2}{2\gamma_c^2}},$$

where I_e is the edge map of the central view image I , and $(\gamma_{occ}, \gamma_e, \gamma_c)$ are three weighting factors. Comparing with previous works, we introduce the occlusion term and the edge term into the weighting function ω_{pq} to preserve occlusion boundaries and keep the depth of occlusion boundaries similar.

The final depth map is obtained by minimizing the Eqn. 14 [2, 3, 6],

$$\alpha^* = \arg \min_{\alpha} E. \quad (17)$$

The full description of the proposed algorithm is given in Algo. 1. First, edge detection is applied on central view image I to find all possible occlusion point I_e . Then, the un-occluded views Ω^{uocc} of each candidate occlusion point are selected based on a K-means clustering. After that, an initial depth map α_{ini} is estimated by using the un-occluded views. Moreover, the occlusion map Occ is detected by using the initial estimation. Finally, based on the un-occluded views and the occlusion map, the depth map α_{final} is regularized with a MRF energy function.

Algorithm 1 The proposed algorithm

Input:4D light field LF **Output:**Final depth map α_{final} **Process:**

I_e	$= \text{canny}(I)$	▷ Sec. 3.1
Ω^{uocc}	$= \text{kmeans}(I_e, I)$	▷ Sec. 3.1
α_{ini}	$= \text{InitialDepthEst}(LF, I_e, \Omega^{uocc})$	▷ Sec. 3.2
Occ	$= \text{OccDetection}(\alpha_{ini}, I_e)$	▷ Sec. 4.1
E	$= E_{uocc}(LF, \Omega^{uocc}) + E_{smooth}(I, I_e, Occ)$	▷ Sec. 4.3
α^*	$= \arg \min_{\alpha} E$	▷ Sec. 4.3

5 Experimental Results

We compare our results with the globally consistent depth labeling (GCDL) by Wanner *et al.* [15], the line-assigned graph-cut(LAGC) by Yu *et al.* [20], the Bilateral Consistency Metric (BCM) by Chen *et al.* [4] and the occlusion-aware depth estimation (OADE) by Wang *et al.* [13]. Note that, the results of GCDL come from their published papers [17], the results of LAGC and OADE are obtained by running their published codes or executable files, and the results of BCM are provided by the authors.

The performance of the proposed algorithm is evaluated by using the most popular light field datasets [17]. This datasets are synthesized by the Blender, and each data includes a 9×9 light field and its ground truth depth. For runtime, on a 3.4 GHz Intel i7 machine with 16 GB RAM, our MATLAB implementation takes about 1 hour on a $9 \times 9 \times 768 \times 768$ color light field. Considering the precise results and the low-speed of MATLAB, this time cost is acceptable.

5.1 Un-occluded views selection

A consensus on depth estimation in computer vision is that more effective views lead to more accurate depth estimation. So, the precision and the recall of the selected un-occluded views are important. We count the F-measure (the harmonic mean of precision and recall compared with the ground truth) of the un-occluded views in occlusion using our algorithm, and compare it with previous work [13]. The quantitative comparisons are listed in Tab. 2, and the qualitative comparisons are shown in Fig. 5. It

Table 2: The F-measure of un-occluded views selection

	Buddha	Buddha2	Horses	Medieval	Mona	Papillon	StillLife
OADE[13]	0.71	0.73	0.66	0.58	0.69	0.59	0.59
Our method	0.71	0.74	0.62	0.60	0.79	0.79	0.72

can be seen that our selection method outperforms previous work in the un-occluded views selection. And this advantage is more obvious especially in multi-occluder areas. In Fig. 5, our method can always select accurate un-occluded views, however, Wang’s *et al.* [13] method always selects more occluded views and these selections will lead to over smooth results in occlusion areas (Fig. 7, 8). It is noticed our method performs not good in Horses. The reason is that there are many textures near the occlusion boundaries in background, and the K-means clustering based on color can not divide the background and occluder accurately in the complex texture areas.

5.2 Occlusion Boundaries

For each data, we detect its occlusion boundaries using the depth map, and compute its F-measure. Then, we compare it with other state-of-the-art algorithms. The quantitative comparisons are listed in Tab. 3, and the qualitative comparisons are shown in Fig. 6. Our algorithm outperforms the previous works. Note that the results of GCDL [15] and BCM [4] are not contained as it is difficult to run their codes in our experimental environment. However, as previous works [20, 13] have demonstrated their advantages to [15] and [4], these comparisons are convincing. Our method performs not good in StillLife (the third row in Fig. 6). That is because there are many weak occlusions (the difference of disparities is small) in StillLife. The difference of disparities in the boundaries of the bee is small, the occlusion detection method in Sec. 4.1 can not handle these occlusions well.

5.3 Depth maps

The quantitative comparisons of the RMS errors of recovered disparity maps are listed in Tab. 4. Note that all results are obtained in a same parameters setting. Our algorithm outperforms previous state-of-the-art algorithms in almost all datasets.



Figure 5: The comparisons of the un-occluded views selection. The areas labeled with white are the selected un-occluded views.

Table 3: The F-measure of occlusion boundaries.

	Buddha	Buddha2	Horses	Medieval	Mona	Papillon	StillLife
LAGC[20]	0.54	0.41	0.55	0.32	0.64	0.53	0.49
OADE[13]	0.71	0.70	0.75	0.47	0.75	0.65	0.82
Our method	0.75	0.85	0.80	0.56	0.81	0.76	0.71

The qualitative comparisons of the recovered disparity map are shown

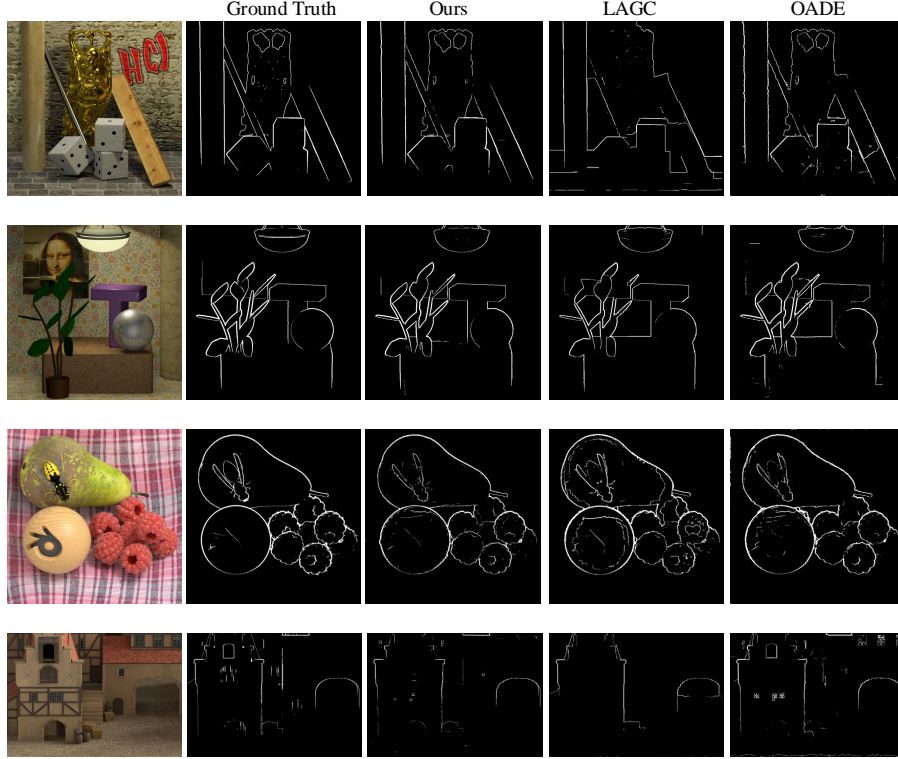


Figure 6: The detected occlusion boundaries.

in Fig. 7, 8. It can be seen that, our algorithm yields sharper occlusion boundaries. As our selection method for un-occluded views can always find them accurately (Fig. 5, Tab. 2), and do not select any other occluded

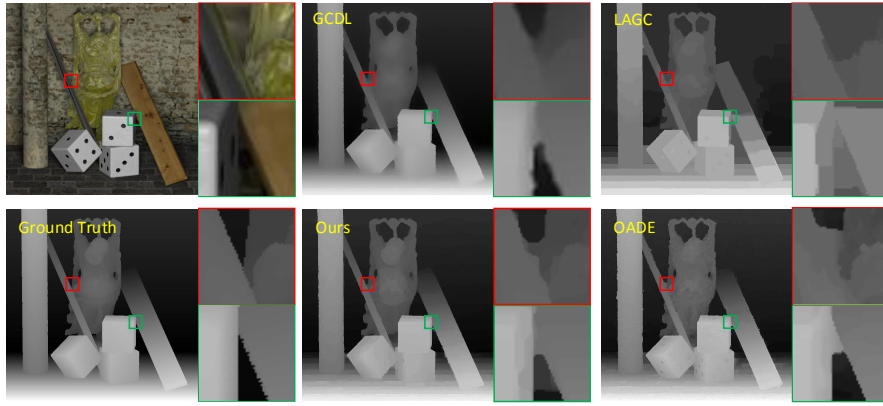
Table 4: RMS errors of recovered disparity for all pixels.

	Buddha	Buddha2	Horses	Medieval	Mona	Papillon	StillLife
GCDL[15]	0.079	0.094	0.163	0.111	0.096	0.158	0.184
LAGC[20]	0.134	0.179	0.188	0.144	0.119	0.406	0.150
BCM[4]	0.057	0.139	0.122	0.129	0.077	0.108	0.113
OADE[13]	0.095	0.107	0.140	0.115	0.089	0.125	0.212
Our method	0.069	0.051	0.074	0.101	0.071	0.148	0.110

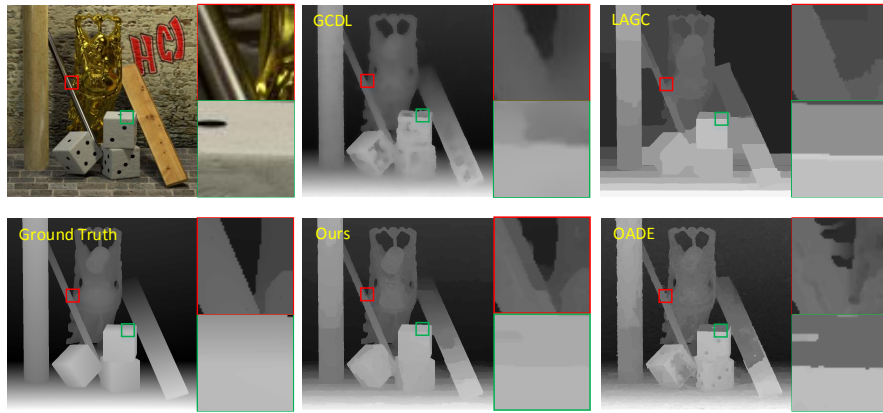
views, our results are sharp in the multi-occluder areas.

It is noticed the proposed algorithm performs not as good as OADE in the un-occluded views selection for Horses, however we get better results in depth map. This is because our energy function can preserve occlusion boundaries better. Moreover, we get the best results in the depth estimation for StillLife, however the F-measure of detected boundaries is not the best. The reason is that our algorithm performs much better than OADE in multi-occluder boundaries which have a larger difference of disparities.

Apart from the heavy occlusion, the proposed algorithm also performs



(a) Buddha



(b) Buddha2

Figure 7: The disparity maps on the Buddha and Buddha2.

well for shadings (Fig. 7(b)), although it is not taken into account in our model. Comparing with the Buddha (Fig. 7(a)), there are more shadings in the buddha2 (Fig. 7(b)). Although all algorithms perform good in the Buddha, only our algorithm maintains the same level in the Buddha2. The reason for this phenomenon worth further study.

However, our algorithm can not handle the situation where the background has a similar texture or color to the occlusion. In the Fig. 8(b) (the green box), as the color of cloth in background is similar to the foreground, it is difficult to recover the true depth. Apart from this, our algorithm can not handle textureless areas just like all previous methods.

6 Conclusion and Future works

In this paper, we propose a new anti-occlusion depth estimation algorithm by modeling the formulation of the occlusion. The model reveals an important property, the occluders are consistent between the spatial and angular space. Utilizing this information, we improve the depth estimation in occlusion areas in two ways. Firstly, the un-occluded views are accurately selected by a clustering in spatial space, and the classical photo-consistency is enforced in these views. Secondly, the occlusion map is detected using the edges and the initial depth map, and then combined into the smooth term in the MRF function to keep the occlusion boundaries sharp. We have demonstrated the advantages of the proposed algorithm compared with other state-of-the-art algorithms in synthetic datasets.

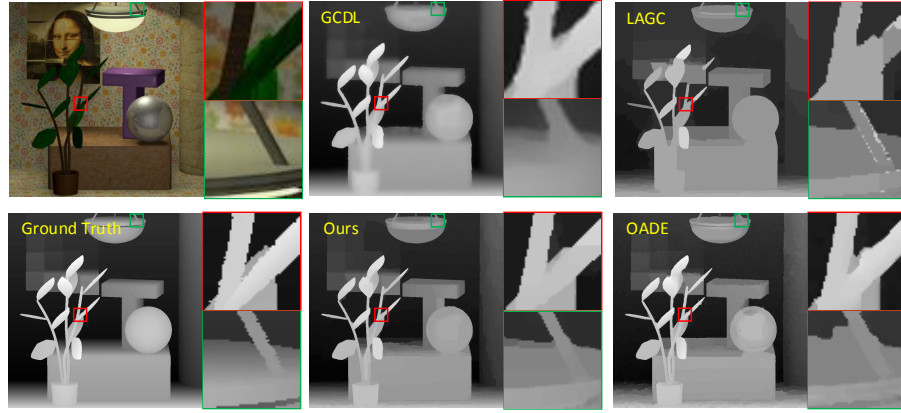
Just as we mentioned in Sec. 5.3, our algorithm produces unexpected good results in shading situations although the shading is not considered in our model. It is worthy for us to investigate the phenomenon. Furthermore, as the light fields captured by real light field cameras have more noise compared with the synthetic datasets, it is essential to do more experiments on real data to better evaluate the performance of the proposed algorithm.

References

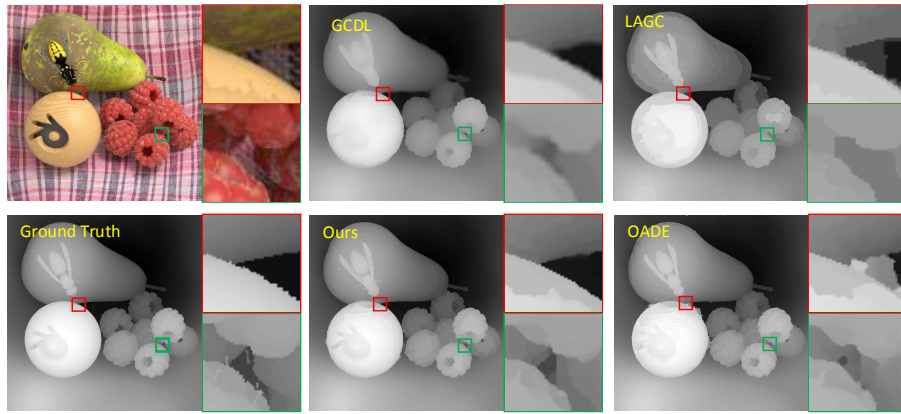
- [1] Yuichiro Anzai. *Pattern Recognition & Machine Learning*. Elsevier, 2012.
- [2] Yuri Boykov and Vladimir Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *Pattern*

- Analysis and Machine Intelligence, IEEE Transactions on*, 26(9):1124–1137, 2004.
- [3] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(11):1222–1239, 2001.
 - [4] Can Chen, Haiting Lin, Zhan Yu, Sing Kang, and Jingyi Yu. Light field stereo matching using bilateral statistics of surface cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1518–1525, 2014.
 - [5] Vladimir Kolmogorov and Ramin Zabih. Multi-camera scene reconstruction via graph cuts. In *Computer Vision – ECCV 2002*, pages 82–96. Springer, 2002.
 - [6] Vladimir Kolmogorov and Ramin Zabih. What energy functions can be minimized via graph cuts? *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(2):147–159, 2004.
 - [7] Lytro. Lytro redefines photography with light field cameras. <http://www.lytro.com>, 2011.
 - [8] Ren Ng. *Digital light field photography*. PhD thesis, stanford university, 2006.
 - [9] Ren Ng, Marc Levoy, Mathieu Brédif, Gene Duval, Mark Horowitz, and Pat Hanrahan. Light field photography with a hand-held plenoptic camera. *Computer Science Technical Report CSTR*, 2(11), 2005.
 - [10] raytrix. Raytrix lightfield camera. <http://www.raytrix.de>, 2012.
 - [11] Carsten Rother, Vladimir Kolmogorov, Victor Lempitsky, and Martin Szummer. Optimizing binary mrfs via extended roof duality. In *Computer Vision and Pattern Recognition, 2007. CVPR’07. IEEE Conference on*, pages 1–8. IEEE, 2007.
 - [12] Ting-Chun Wang, Manmohan Chandraker, Alexei Efros, and Ravi Ramamoorthi. Svbrdf-invariant shape and reflectance estimation from light-field cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
 - [13] Ting-Chun Wang, Alexei A Efros, and Ravi Ramamoorthi. Occlusion-aware depth estimation using light-field cameras. In *Proceedings of the*

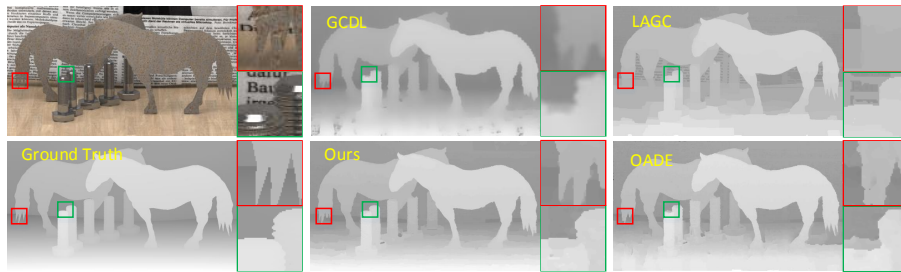
- IEEE International Conference on Computer Vision*, pages 3487–3495, 2015.
- [14] Ting-Chun Wang, Alexei Alyosha Efros, and Ravi Ramamoorthi. Depth estimation with occlusion modeling using light-field cameras. *Pattern Analysis and Machine Intelligence, IEEE Transactions on (In press)*, 2016.
 - [15] Sven Wanner and Bastian Goldluecke. Globally consistent depth labeling of 4d light fields. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 41–48. IEEE, 2012.
 - [16] Sven Wanner and Bastian Goldluecke. Variational light field analysis for disparity estimation and super-resolution. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(3):606–619, 2014.
 - [17] Sven Wanner, Stephan Meister, and Bastian Goldluecke. Datasets and benchmarks for densely sampled 4d light fields. In *VMV*, pages 225–226. Citeseer, 2013.
 - [18] Oliver Woodford, Philip Torr, Ian Reid, and Andrew Fitzgibbon. Global stereo reconstruction under second-order smoothness priors. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(12):2115–2128, 2009.
 - [19] Zhaolin Xiao, Qing Wang, Guoqing Zhou, and Jingyi Yu. Aliasing detection and reduction in plenoptic imaging. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3326–3333, 2014.
 - [20] Zhan Yu, Xinqing Guo, Haibing Lin, Andrew Lumsdaine, and Jingyi Yu. Line assisted light field triangulation and stereo matching. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2792–2799, 2013.



(a) Mona



(b) StillLife



(c) Horses

Figure 8: The disparity maps on the Mona, StillLife and Horses.