Scalable Light Field Disparity Estimation with Occlusion Detection

Yan Li
LISA department
Université Libre de Bruxelles
Brussels, Belgium
yali@ulb.ac.be

Gauthier Lafruit
LISA department
Université Libre de Bruxelles
Brussels, Belgium
gauthier.lafruit@ulb.ac.be

ABSTRACT

An occlusion-aware framework is proposed to robustly estimate the disparities of light field images. It is mainly realized by leveraging multiple edge cues to occlusion detection and then integrate it with local costs into an energy function. To check the performance, the quantitative and/or qualitative evaluations are performed on both synthetic and natural light field datasets. It demonstrates that the proposed framework is robust to the density and disparity range of the light field, advancing the state-of-the-art light field disparity estimation frameworks on aspect of accuracies.

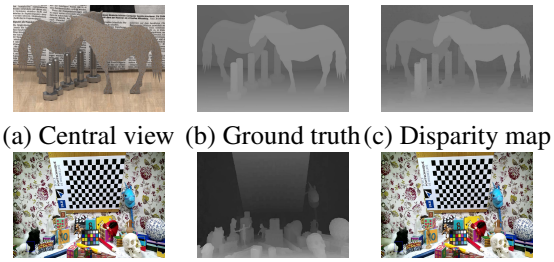
Keywords

Light Field, Disparity/Depth Estimation, Occlusion Detection, Global Optimization.

1 INTRODUCTION

Contrary to a traditional 2D image, the light field records not only the radiance but also the direction of a light ray. This richer information of the light field motivates a large range of computer vision and graphics applications, including disparity/depth estimation [Wanner and Goldluecke, 2012, Kim et al., 2013, Chen et al., 2014, Jeon et al., 2015, Wang et al., 2015, Zhang et al., 2016, Zhu et al., 2017], digital refocusing [Ng et al., 2005] and super-resolution [Wanner and Goldluecke, 2014], etc. In this work, our focus is put on disparity/depth estimation, which is employed as a module of view synthesis for virtual reality (VR) [Huang et al., 2017, MPEG-I, 2017].

Disparity estimation, is a long-term challenging issue in computer vision, which finds correspondences from stereo image pairs. The well-generated disparity maps from this task could bring benefits to various applications, such as view synthesis [Stankiewicz et al., 2013], superpixel segmentation [Stutz et al., 2018], semantic segmentation [Zhang et al., 2010], etc. A common solution for disparity estimation is to employ two views, namely stereo matching [Scharstein and Szeliski, 2002]. Since more viewpoints are available in the light field (Fig. 2), more accurate disparity estimations are possible than in stereo matching.



(a) Central view (b) Ground truth (c) Disparity map

(d) Central view (e) Depth map (f) Synthesized map

Figure 1: Disparity/Depth estimation results on light field images. The top shows the proposed disparity map of the central view in a dense light field; The bottom shows the proposed depth map of the central view and its corresponding synthesized/virtual map in a sparse and large disparity range of the light field, in which the synthesized map is obtained by view synthesis using two neighboring views and their depth maps.

Nowadays, the state-of-art light field disparity estimation references achieve a significantly high accuracy when the disparity range between sub-aperture images is narrow and the light field is densely sampled. However, we observe that the accuracy still remains an issue when the disparity range between sub-aperture images is larger and the light field is sparser. Moreover, this is non-trivial in virtual view rendering of MPEG-I activities [MPEG-I, 2017] in which a sparser and larger disparity range of the light field is being used.

To cope with this issue, a scalable framework for light field disparity estimation is proposed in the paper. More specifically, the kernel density estimation and

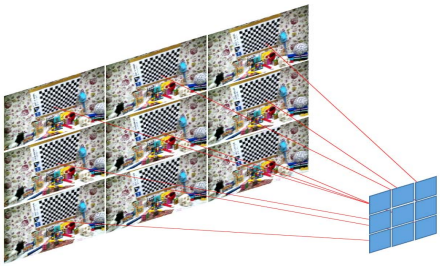


Figure 2: Light field images, also called sub-aperture images, are captured from an equally spaced 2D camera array.

size-adaptive window filter are introduced to locally estimate disparities in which an adaptive size is considered not to be sensitive to the disparity range (Sec 3.2). Since there are more ambiguities at occlusion areas, an occlusion handling method, i.e., occlusion detection and score-volume recomputation, are proposed (Sec 3.3), followed by using an occlusion-aware optimization to improve disparity continuity and enforce global consistency (Sec 3.4). The experimental results show that the proposed framework produces a high accuracy of disparity/depth maps in multiple densities and disparity ranges of light fields, see Fig. 1.

The contributions of our work are summarized below:

1. The proposed framework significantly advances state-of-the-art reference work on both synthetic and natural light field datasets in disparity/depth accuracy.
2. The accuracy of disparity maps from the proposed method is more robust to the density and disparity range of the light field, achieving at least 14.9%, 19.6%, 29.3% Root Mean Square Error (RMSE) gains on average in 9x9, 5x5 and 3x3 synthetic light fields respectively when compared to the state-of-art works.
3. An occlusion handling technique using multiple edge cues is put forward, which always benefits disparity estimations at occlusion regions.

2 RELATED WORK

Since the light field has the redundant information, some works formulate the problem of disparity estimation as the slope calculation of the Epipolar Plane Image (EPI) without explicit occlusion detection. [Wanner and Goldluecke, 2012] introduce a structure tensor technique to light field disparity estimation, and use it to compute the slope of EPI, followed by the integration into a variational-based energy function. However, its accuracy is confined to a narrow disparity range. [Kim et al., 2013] put forward another EPI-based framework using a fine-to-coarse strategy. Since it relies on a pixel to match the corresponding EPI-pixel, it seems robust at occlusion regions but the light field has to be guaranteed densely sampled. [Zhang et al., 2016] propose a spinning parallelogram operator onto both horizontal and vertical EPI-lines, preserving sharp disparity edges. [Jeon et al., 2015] present a disparity estimation framework in sub-pixel accuracy. However, both methods are subject to insufficient disparity quality at textureless regions, though they are not much influenced by the density and disparity range of light field.

Some other works turn to the Surface Camera (SCam)-based strategy to perform disparity estimation with explicit occlusion detection. In fact, the occlusion is a tough issue in matching correspondences. Multi-view stereo matching makes an early effort to this occlusion issue. [Kolmogorov and Zabih, 2002] describe a graph cut framework in which the visibility term is formulated into an energy function. [Wei and Quan, 2005] propose an asymmetric model to overcome occlusions in an efficient way. However, the heavy occlusion still remains an issue, even with a large number of views. To handle (heavy) occlusions, some recent light field-based methods are proposed [Chen et al., 2014, Wang et al., 2015, Zhu et al., 2017]. [Chen et al., 2014] introduce a bilateral consistency metric to predict the occlusion at SCam reliably. [Wang et al., 2015] describe an angular patch occlusion model, i.e., a single-occluder model, in which edge detection is required to obtain the occlusion boundary. When a multiple-occluder appears, it cannot work well because the single-occluder assumption does not hold. To overcome this drawback, [Zhu et al., 2017] describe a multiple-occluder modeling and then adopt an un-occluded view selection and re-selection scheme. Since its accuracy relies on the occlusion boundary, a disparity edge map is combined with an edge map to improve occlusion boundary detections. However, these works are somewhat restricted to dense light fields.

In contrast, the proposed framework, which is modeled by filter-based kernel density estimations with a separate occlusion-aware optimization technique, is not limited to the dense or narrow disparity range of the light field. The experimental results demonstrate that our method achieves a significantly high accuracy in multiple densities and disparity ranges of light fields, advancing the state-of-the-art performance.

3 APPROACH

Fig. 3 shows an overview of our approach. Taking a central view of light fields (Sec 3.1) for instance, the local disparity map (LDM) is initially produced from a winner-take-all strategy onto score-volume computations (Sec 3.2). Then a disparity edge map (DEM), canny edge map (CEM), superpixel edge map (SEM), occluded pixels map (OPM) are put into the occlusion handling site to extract an occlusion boundary map (OBM) and tweak the score of occluded pixels (Sec 3.3). With the aid of these occlusion detection results,

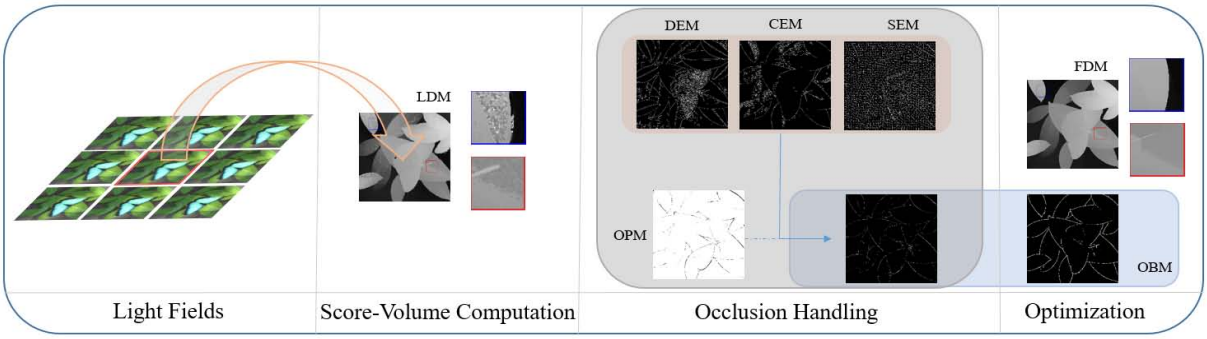


Figure 3: The proposed framework.

the final disparity map (FDM) is better generated under optimizations when compared with the LDM (Sec 3.4).

3.1 Light Fields

The light field, in the paper, is represented by two-plane parametrization (2PP) in which a camera plane is parametrized by the coordinate system (s,t) and the image plane (u,v). Then it could be simply seen as a collection of a plane of views with radiance values r in the RGB color space, described as $r = L(s,t,u,v)$, in which (s,t) represents a camera coordinate and (u,v) indicates a coordinate of a pixel on the image plane. The light field view, which is being estimated, is denoted by R_{s^*,t^*}. Then, according to this view, a radiance set $R_{s,t,u,v}(d)$ is easily built by assigning a hypothetical disparity d to light rays or pixels, as given in Eq. 1:

$$R_{s,t,u,v}(d) = \{L(s,t,u+d*(s^*-s),v+d*(t^*-t)) \\ |s=1,2,...,M;t=1,2,...,N\} \tag{1}$$

where (M, N) denotes the angular resolution of the light field. The subscript (u,v) that corresponds to the pixel or light ray in a view is replaced with p in the following texts for simplicity.

3.2 Score-Volume Computation

Our score-volume computation is composed of two steps: 1) initially computing the score volume, 2) filtering the score volume. The (filtered) score volume indicates a 3D array (u,v,d) that stores the scores/probabilities of candidate disparities d for a pixel p in a light field view.

A kernel density function is employed to the initial score volume calculations, which is formulated as follows:

$$S_p(d) = \frac{1}{|\Omega|}\sum_{s,t\in\Omega} K_h\left(R_{s^*,t^*,p} - R_{s,t,p}(d)\right) \tag{2}$$

where $S_p(d)$ is the score of the pixel p of the being estimated view R_{s^*,t^*} at the candidate disparity d where the maximum value corresponds to the true disparity in volume, and Ω represents a number of valid views for score computations. $K_h(\cdot)$ corresponds to the Epanechnikov kernel that is given in Eq. 3 and h is its bandwidth parameter (= 0.02), which controls the accuracy of the density estimation. Actually, a higher value of h increases the accuracy and robustness to noise. However, it will lose fine details.

$$K_h(x) = \begin{cases} 1 - \left\|\frac{x}{h}\right\|^2 & \left\|\frac{x}{h}\right\| \le 1 \\ 0 & otherwise \end{cases} \tag{3}$$

Rather than the increase of h, a window-based filter, i.e., an edge-aware preserving filter [He et al., 2010], is introduced to filter out some noises, which is computed as follows:

$$\widetilde{S}_p(d) = \sum_q W_{pq} S_q(d) \tag{4}$$

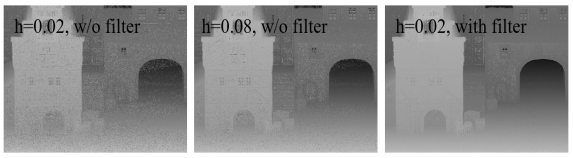
where $\widetilde{S}_p(d)$ is the filtered score of $S_p(d)$ and $S_q(d)$ is the score of the neighboring pixel q in a window. Since the filtering adopts an integral image based technique, it has a low complexity burden O(N). The weight of this filter is computed as below,

$$W_{pq} = \frac{1}{|\omega|^2}\sum_{k:(p,q)\in\omega_k}\left\{1 + \frac{(I_p-\mu_k)(I_q-\mu_k)}{\sigma_k{}^2+\varepsilon}\right\} \tag{5}$$

where $W_{p,q}$ gives a higher weight to the pixel on the same side of the edge and a lower weight to the pixel on opposite sides of the edge in a window ω_k centered at the pixel k. The side length of this window ω_k is adaptive to the spatial resolution (w,h) of the light field, i.e., $max(\lfloor max(w,h)^2/(256*min(w,h))\rfloor, 3)$. I is a guided image, namely the light field view R_{s^*,t^*} that is being estimated; μ_k and σ_k are the mean and variance of the window ω_k in I respectively; ε is set to 0.01; $|\omega|$ is the number of the pixels in ω_k. The more effectiveness of this technique than the only increase of the h is shown in Fig. 4, clearly reducing the speckle noise.

3.3 Occlusion Handling

Assuming that the scene in light fields is lambertian, the scene point that is seen from different viewpoints shares



RMSE 0.4183 *RMSE* 0.2891 *RMSE* 0.1139

Figure 4: Compared with the increase of h, the edge-preserving filter demonstrates its higher ability (a lower RMSE) to remove the noises without losing fine details.

the same color, exhibiting the photo-consistency. However, this is not true for the point that is occluded. Some pixels from such a point in the score-volume computation step might be correctly estimated due to the edge-aware score volume computation. Nevertheless, the disparities of pixels at heavy occlusion regions still remain difficult to be well-estimated due to ambiguities. As a result, a pixel with a wrong disparity may be assigned a highest score. To address this issue, the occluded pixel detection (OPD), occlusion boundary detection (OBD) and score-volume recomputation (SVR) are proposed.

3.3.1 Occluded Pixel Detection

Some pixels disappear in parts of the views due to occlusions, breaking off the photo-consistency. Assuming that the scene is lambertian, a simple thresholding technique could be applied to detect these occluded pixels and obtain the occluded pixel map OPM, as given below,

$$C_p(d) = \frac{1}{|\Omega|}\sum_{\Omega}(1 - exp(-|R_{s^*,t^*,p} - R_{s,t,p}(d)|)) \quad (6)$$

where $C_p(d)$ indicates the occlusion confidence of the pixel p of the view R_{s^*,t^*} at the estimated disparity d. If the confidence of a pixel is larger than a specified threshold τ (= 0.05), it is masked as an occluded pixel ($OP = 1$); otherwise it is unoccluded ($OP = 0$).

3.3.2 Occlusion Boundary Detection

Occlusion boundary detection is a significant step for the occlusion handling as its accuracy makes differences for the following disparity re-estimation and occlusion-aware optimization. To guarantee its precision, multiple edge cues are proposed to precisely detect occlusion boundaries.

Firstly, a fact to be known is that there always exist edges between an occluder and an occluded region, which is ascribed to lighting changes in-between. Thus the following lemma is given.

Lemma I. An occlusion boundary set OB_s is a proper subset of an edge set EG_s.

The edge set is approximately constructed in our work for efficiency, i.e., a union of edge points and edge lines,

$$EG_s \simeq EG_{point} \cup EG_{region} \quad (7)$$

where EG_{point} denotes the edge points that are acquired by an edge detector, and EG_{region} indicates the edges from a region/superpixel detector [Stutz et al., 2018]. Note that the region size is set to a smaller value so as to be not much larger than the objects in the scene. Additionally, a small region used in a superpixel detector could boost the edge accuracy.

The occlusion boundaries that belong to the occlusion boundary set OB_s are taken from the approximated edge set. Firstly, for the view R_{s^*,t^*}, a disparity edge map DEM is computed from a relatively reliable local disparity map LDM using a canny edge detector [Canny, 1986], and an edge map EM is intersected by the canny edge map CEM and the superpixel edge map SEM. Then we calculate an intersection of DEM and EM to get an initial occlusion boundary map OBM^i. Furthermore, the disparity variance in a 10x10 window and the difference operator are computed as masks to update the difference between OBM^i and themselves in order to remove edge point outliers,

$$\begin{aligned} EM^u &= M_{disp} * (EM - OBM^i) \\ DEM^u &= M_{\nabla} * (DEM - OBM^i) \end{aligned} \quad (8)$$

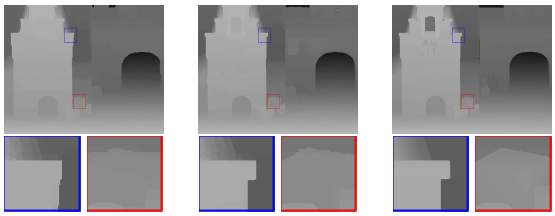
where M_{disp} and M_{∇} denote the disparity mask and the difference mask respectively. If the pixel has a disparity variance beyond a threshold φ that is adaptive to the disparity range, M_{disp} is assigned 1, otherwise 0. Similarly, if the pixel has a difference beyond a specified threshold $\nabla(= 0.05)$, M_{∇} is assigned as 1, otherwise 0. Finally, a union of multiple maps are used to produce the occlusion boundary map $OBM = OBM^i \cup DEM^u \cup EM^u$ with a high precision.

3.3.3 Score-Volume Recomputation

The score volume recomputation consists of two steps: 1) computing the disparity bound, 2) score-volume computation, targeting the improvement of the occluded pixel disparity estimation.

Disparity Bound The new upper bound ub and the lower bound lb in disparity are determined by the disparities of pixels in their neighborhood beforehand. The upper and lower bound are assigned to the maximum and minimum disparity of neighboring pixels respectively.

Score-Volume Computation The procedure in the previous score-volume computation is reused here, but there exist two differences. The first difference is that a disparity bound, i.e, a half-closed interval $[lb, ub)$, is utilized for computing the occluded pixel score $OccS_p(d)$ for the pixel p of the view R_{s^*,t^*} at a candidate disparity d. The second difference is that the visible views Ω_{vis} for photo-consistency are selected. More specifically, the relative location of the occluded pixel to the occlusion boundary from OBM (with rare negative occlusion boundaries) is used to simply select



w/o occ with occ Ground truth

Figure 5: Comparisons between without (w/o) and with occlusion detection results (occ) in the energy function. It demonstrates that the proposed occlusion-aware energy function contributes to a higher accuracy (a lower RMSE 0.099) without over-smoothing the sharp edges.

the visible views.

At the end of the occlusion handling flow, the occlusion boundary map with a high accuracy can be extracted and the score of the occluded pixel will be improved, which are beneficial to the following optimization step.

3.4 Optimization

Our disparity estimation is optimized by minimizing a Markov Random Field-based energy function, as given in Eq. 9.

$$E = \lambda * \sum_{p} E_{data}(p, d(p)) + \sum_{q \in N_p} E_{smooth}(p, q, d(p)) \tag{9}$$

where N_p is a 4-neighborhood of the pixel p of the view R_{s^*,t^*}, q represents one of the neighboring pixels and $d(p)$ denotes a disparity that is mapping to an integer. Herein λ (= 10) is introduced to balance the ratio of the data term and the smoothness term.

The data term in the energy function is built by weighting the score $\widetilde{S}$ and the occlusion score $OccS$,

$$E_{data}(p, d(p)) = \kappa - \alpha * \widetilde{S}_p(d) - (1-\alpha) * OccS_p(d) \tag{10}$$

where E_{data} measures the photo-consistency for the pixel p, α is a weighting coefficient (= 0.6) and κ is a large constant $(= 10)$.

The smoothness term is computed by a weighted neighboring function,

$$E_{smooth}(p, q, d(p)) = w_{p,q} * min(|d(p) - d(q)|, \Gamma) \tag{11}$$

$$w_{p,q} = exp(-\frac{||R_{s^*,t^*,p} - R_{s^*,t^*,q}||^2}{\psi^2} - \frac{|OB_p - OB_q|}{\phi^2} - \frac{|OP_p - OP_q|}{\phi^2}) \tag{12}$$

where Γ represents a truncated threshold that is set to 10; ψ and ϕ is set to 1/9 and 1 respectively; OB is an occlusion boundary mask from the occlusion boundary map OBM and OP is an occluded pixel mask from the occluded pixel map OPM that are enforced as constraints. If an occlusion boundary exists in-between two pixels or one of two neighbouring pixels is an occluded pixel, the strength of smoothness will be reduced. Besides, the color in the view R_{s^*,t^*}, is encoded as a constraint in which two pixels with different colors will decrease smoothness. To solve the proposed occlusion-aware energy function, the graph cut algorithm [Kolmogorov and Zabih, 2002] is used. As a consequence, the proposed occlusion metrics in the energy function especially help a lot to avoid over-smoothing, hence preserving sharp edges, see Fig. 5.

4 EXPERIMENTAL RESULTS

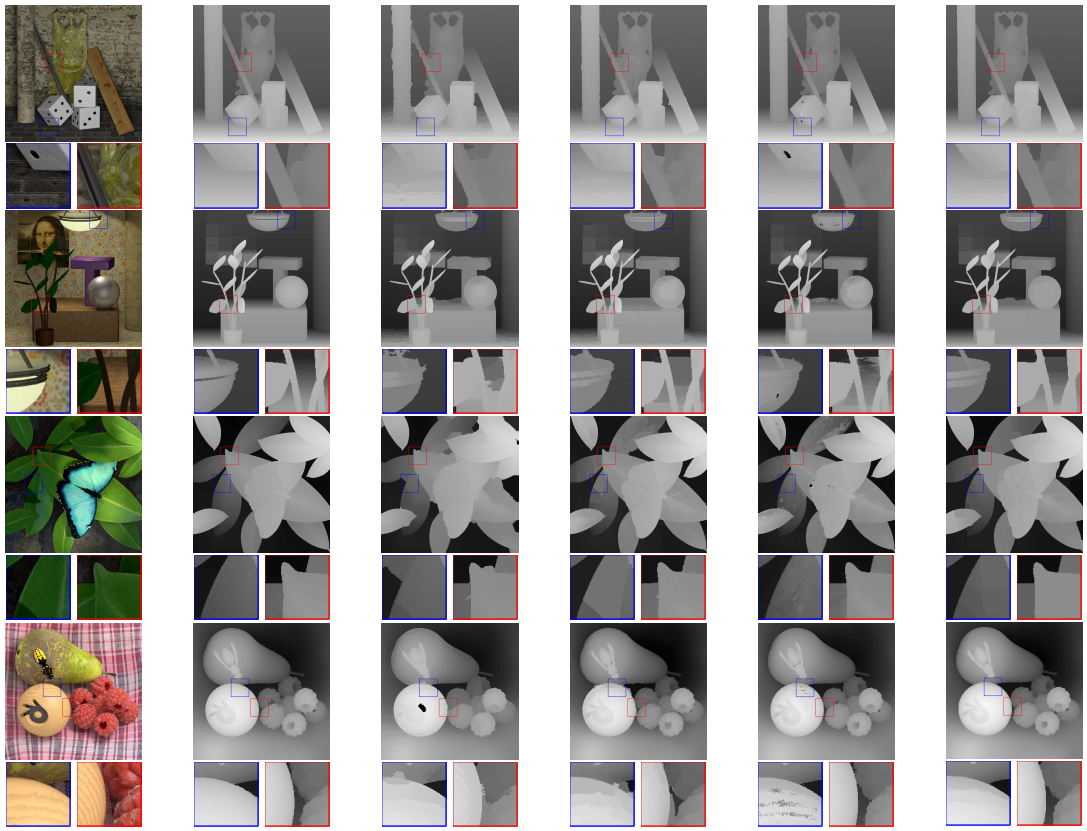
We present the results of the proposed approach that are evaluated on light field datasets, which are composed of synthetic datasets and natural datasets. In order to validate the accuracy and scalability, the experimental results are compared with several state-of-the-art references, PSD [Jeon et al., 2015], OADE [Wang et al., 2015], SPO [Zhang et al., 2016], and an Enhanced Depth Estimation Reference Software eDERS [Senoh et al., 2018]. Note that the results from the state-of-the-art references are generated by utilizing their public code under default settings, except for the number of labels and the disparity range. For validations onto both datasets, two metrics are adopted: a direct metric RMSE for synthetic datasets with available ground truth, and an indirect metric for natural datasets without ground truth (i.e, view synthesis quality using the estimated depth maps).

4.1 Synthetic Dataset

A popular synthetic dataset HCI [Wanner et al., 2013] with ground truth is used for qualitative and quantitative comparisons. Note that the number of labels and the disparity range used into the state-of-the-art works are set to the same values with the proposed method for the sake of better comparisons. The HCI dataset includes 9x9 densely-sampled light field views with a low resolution and has quite a low disparity range, i.e., less than 8 pixels. To well estimate the disparities, 101 disparity labels (a label is less than 8/100 pixel) are employed in all four approaches. Table 1 illustrates that we achieve the highest accuracy of disparity maps on this dataset when compared with PSD, SPO and OADE. Herein, the RMSE for the central view of light fields is calculated as done in [Chen et al., 2014, Wang et al., 2015, Zhu et al., 2017] thanks to the given ground truth. Fig. 6 shows our visual comparisons against the ground truth and the three references. From this comparison, we clearly observe that our framework produces the closest disparity maps to the ground truth with good disparity discontinuity preservations.

Dataset	Buddha	Buddha2	Horses	Medieval	MonasRoom	Papillon	StillLife	Average
PSD	0.109	**0.071**	0.151	0.125	0.084	0.248	0.294	0.154
OADE	0.098	0.109	0.146	0.115	0.088	0.108	0.199	0.123
SPO	0.076	0.101	0.113	**0.094**	0.075	**0.081**	0.119	0.094
Proposed	**0.057**	**0.071**	**0.072**	0.099	**0.072**	0.088	**0.103**	**0.080**

Table 1: The Root Mean Squared Error (RMSE), the lowest value in bold black means the highest accuracy.

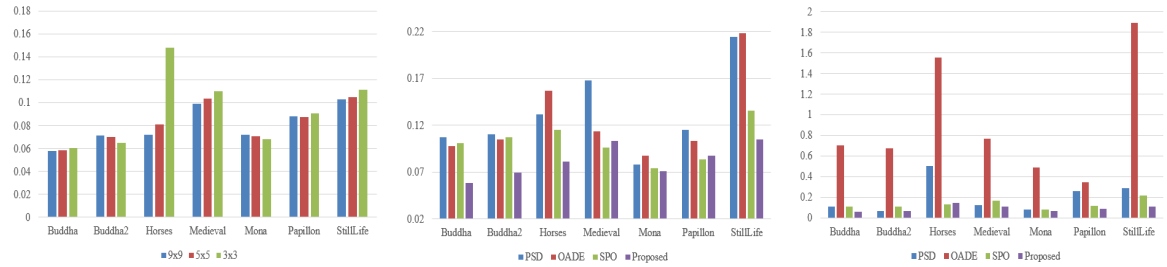


(a) Central View (b) Ground truth (c) PSD (d) OADE (e) SPO (f) Proposed

Figure 6: Disparity estimation results on the HCI dataset. From the top to the bottom, it corresponds to the scene 'Buddha', 'MonasRoom', 'Papillon', 'StillLife' respectively. Our disparity maps seem less noisy than SPO and less over-smoothed than PSD and OADE at occlusion boundary regions, see close-ups of 'StillLife', 'Buddha', etc.

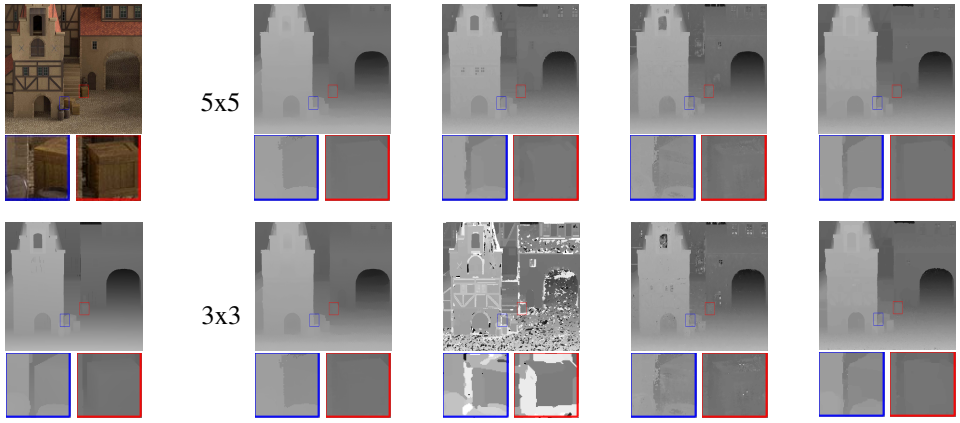
4.2 Density and Disparity Range

When the light field is sparsely-sampled with a large disparity range, it might pose a challenge for the state-of-the-art methods. Hence, we explore the performance of the proposed method on such light fields, which are obtained by skipping a multiple of 2 views from the 9x9 views in both angular directions (i.e., the 5x5 and 3x3 light fields in the paper). Similar to Sec 4.1, the RMSE is calculated for the 5x5 and 3x3 light fields. The computed RMSE is firstly made comparisons with that in Sec 4.1. We can see from Fig. 7 (a) that the errors seem almost unchanged except in the scene 'Horses'. Furthermore, we compare the proposed results with the state-of-the-art references, which is shown in Fig. 7 (b) and (c). It demonstrates that the proposed method, in contrast, mostly achieves the lowest errors and exhibits the robustness to the density and disparity range of the light field. Our method, meanwhile, get at least 14.9%, 19.6%, 29.3% RMSE gains on average in the 9x9, 5x5 and 3x3 light fields respectively. Fig. 8, Fig. 9 and Fig. 10 illustrate the visual comparision results on the 'Medieval', 'Papillon' and 'StillLife' scene respectively. We observe that the quality of the disparity map from OADE [Wang et al., 2015] degrades gradually with a smaller number of light field views, whereas the PSD [Jeon et al., 2015] and SPO [Zhang et al., 2016] decrease a bit but more than that of the proposed method. Moreover, the proposed method does not behave more smoothed as PSD [Jeon et al., 2015] or more noised as SPO [Zhang et al., 2016]. Therefore



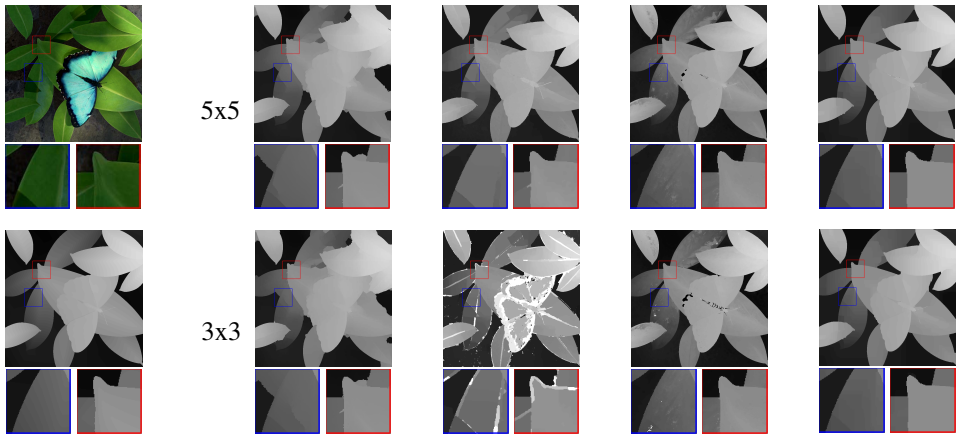
(a) Proposed (b) 5x5 (c) 3x3

Figure 7: The RMSE of the proposed framework in the 9x9, 5x5 and 3x3 light fields is shown in (a). (b) and (c) show the RMSE comparisons between the proposed and the state-of-the-art references in the 5x5 and 3x3 light fields respectively. The lowest value means the highest accuracy.



(a) Central view and Ground truth (b) PSD (c) OADE (d) SPO (e) Proposed

Figure 8: Disparity estimation results on 'Medieval'. Our disparity map is robust around the edges of the wall and/or the box.



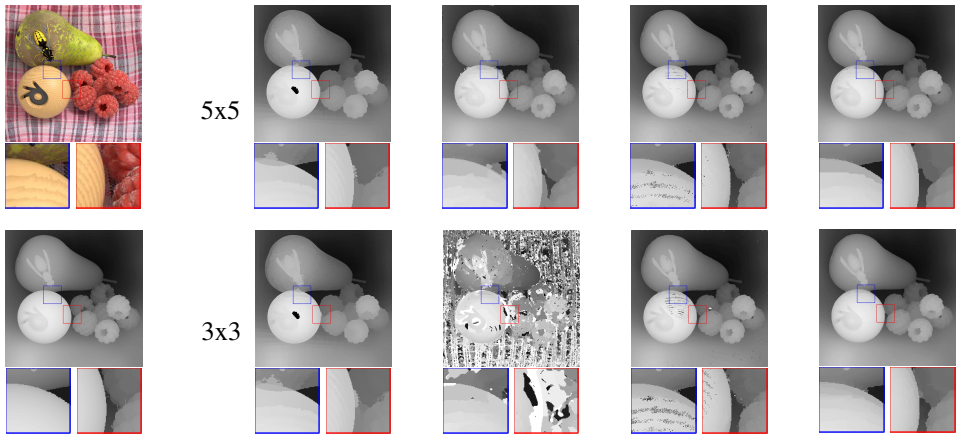
(a) Central view and Ground truth (b) PSD (c) OADE (d) SPO (e) Proposed

Figure 9: Disparity estimation results on 'Papillon'. Our disparity map is achieved with a preciser disparity discontinuity and without noise, see the edges of the leaves.

our method is scalable to the density and disparity range of the light field.

4.3 Occlusion Boundary

Since the accurate occlusion detections were integrated into our global optimization, the occlusion boundary map *OBM* extracted from the final disparity map *FDM* has a significantly high precision. Table 2 gives our performance against the-state-of-the-art methods on the HCI dataset (9x9 light fields) using the common metric Precison-Recall [Sundberg et al., 2011]. An edge detector is used for extracting the proposed and the ground truth occlusion boundary. From the quantitative value, we learn that the precisest occlusion boundaries on av-



(a) Central view and Ground truth (b) PSD (c) OADE (d) SPO (e) Proposed

Figure 10: Disparity estimation results on 'StillLife'. Our disparity map is robust around the surface of the ball.

Dataset	Buddha	Buddha2	Horses	Medieval	MonasRoom	Papillon	StillLife	Average
OADE	0.6632	0.7515	0.7617	0.6043	0.7469	0.7965	0.6181	0.7086
PSD	0.5536	0.7355	0.7354	0.5719	0.6831	0.6089	0.5145	0.6290
SPO	0.6927	0.8330	0.7642	**0.6894**	0.7449	**0.8352**	**0.7115**	0.7530
Proposed	**0.7719**	**0.8480**	**0.8409**	0.6240	**0.7786**	0.7818	0.6950	**0.7629**

Table 2: The Precision-measure of occlusion boundaries, the highest value means the highest accuracy.

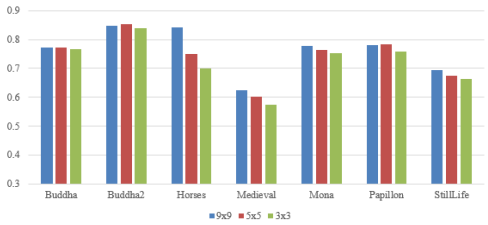


Figure 11: The precisions of the proposed occlusion boundary results onto the 9x9, 5x5 and 3x3 light fields respectively.

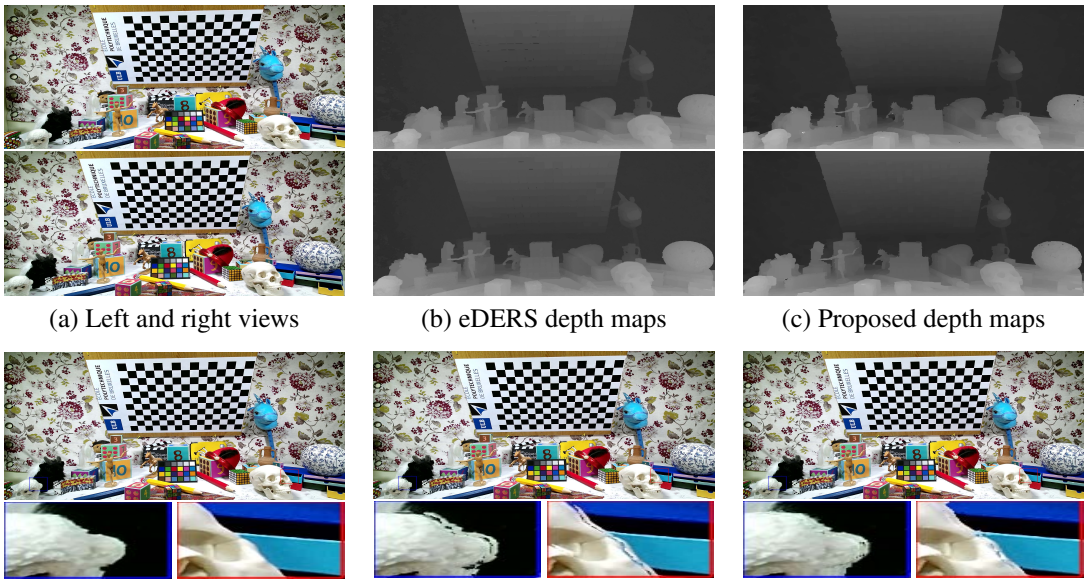
erage are obtained by the proposed work. In addition, the precison values for 5x5 and 3x3 views are also calculated. Note that the 5x5 and 3x3 light fields are also obtained by skipping a multiply of 2 views in both angular directions, similar to Sec 4.2. In Fig. 11, when the number of light field views is reduced, the precision of the occlusion boundary decreases by a very small value, illustrating that the proposed method is also scalable to occlusions in multiple densities and disparity ranges of light fields.

4.4 Natural dataset

In addition to synthetic datasets, the challenging natural datasets ULB Unicorn [Bonatto et al., 2017] and Technicolor Painter [Sabater et al., 2017], which have a larger baseline (35 and 70 mm resp.) for objects at a distance of 0.5 to 4m and a fewer number of views (5x5 and 4x4 views resp.), are evaluated. Moreover, in these datasets, there exists a larger disparity range, i.e., [16-76] and [30, 90] in pixels respectively. Since these two datasets lack ground truth disparities, the view synthesis results generated from view synthesis reference software [Stankiewicz et al., 2013] are used for evaluations, apart from visual comparisons on depth maps that are simply converted from disparity maps. For the view synthesis, two depth maps from two views are required. As the OADE, PSD and SPO are used to predict the disparities for the central view of light fields, we compare our technique with another state-of-the-art technique eDERS [Senoh et al., 2018] (using the same number of disparity labels 241). The experimental results show that the better synthesized/virtual maps are produced from our technique, especially at occlusion regions. Fig. 12 shows that the synthesized map using the proposed depth maps looks much cleaner, see the close-ups in (e) and (f). In Fig. 13, (b) and (c) clearly show that our method correctly estimates the wooden stand and the chair disparities whereas this fails in eDERS. Furthermore, the synthesized map gets more benefits from the proposed depth maps than from the eDERS depth maps, see the close-ups in (e) and (f).

5 CONCLUSIONS

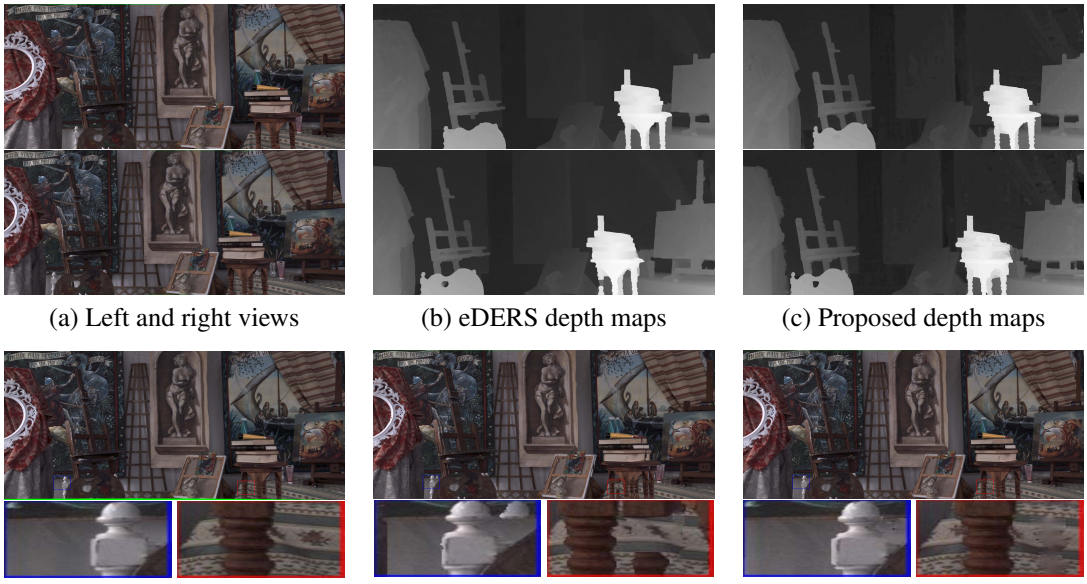
An occlusion-aware framework via multiple edge cues and score updates is proposed for disparity estimation in light fields. Through a variety of evaluations, the proposed method achieves a higher accuracy of disparity estimation on both synthetic and natural datasets when compared with the state-of-the-art approaches. Moreover, the fidelity of the disparity map is still kept even in a sparse light field with a large disparity range.



(a) Left and right views (b) eDERS depth maps (c) Proposed depth maps

(d) Reference/Central view (e) eDERS synthesized map (f) Proposed synthesized map

Figure 12: Depth map and view synthesis result comparisons on the ULB Unicorn dataset. From the top to bottom in (a-c), they correspond to the left and right camera view respectively.



(a) Left and right views (b) eDERS depth maps (c) Proposed depth maps

(d) Reference/Central view (e) eDERS synthesized map (f) Proposed synthesized map

Figure 13: Depth map and view synthesis result comparisons on the Technicolor Painter dataset. From the top to bottom in (a-c), they correspond to the left and right camera view respectively.

6 ACKNOWLEDGMENTS

This work is supported by the China Scholarship Council (CSC) and by Innoviris, the Brussels Institute for Research and Innovation Belgium, under contract number 2015-DS-39a, 3DLicorneA.

7 REFERENCES

[Bonatto et al., 2017] Bonatto, D., Schenkel, A., Lenertz, T., Li, Y., and Lafruit, G. (2017). ULB High Density 2D/3D Camera Array data set, version 2.

[Canny, 1986] Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6):679–698.

[Chen et al., 2014] Chen, C., Lin, H., Yu, Z., Kang, S. B., and Yu, J. (2014). Light field stereo matching using bilateral statistics of surface cameras. In *Computer Vision and Pattern Recognition*

(CVPR), 2014 IEEE Conference on, pages 1518–1525. IEEE.

[He et al., 2010] He, K., Sun, J., and Tang, X. (2010). Guided image filtering. In *European conference on computer vision*, pages 1–14. Springer.

[Huang et al., 2017] Huang, J., Chen, Z., Ceylan, D., and Jin, H. (2017). 6-DOF VR videos with a single 360-camera. In *Virtual Reality (VR), 2017 IEEE*, pages 37–44. IEEE.

[Jeon et al., 2015] Jeon, H.-G., Park, J., Choe, G., Park, J., Bok, Y., Tai, Y.-W., and So Kweon, I. (2015). Accurate depth map estimation from a lenslet light field camera. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1547–1555.

[Kim et al., 2013] Kim, C., Zimmer, H., Pritch, Y., Sorkine-Hornung, A., and Gross, M. H. (2013). Scene reconstruction from high spatio-angular resolution light fields. *ACM Trans. Graph.*, 32(4):73–1.

[Kolmogorov and Zabih, 2002] Kolmogorov, V. and Zabih, R. (2002). Multi-camera scene reconstruction via graph cuts. In *European conference on computer vision*, pages 82–96. Springer.

[MPEG-I, 2017] MPEG-I (2017). Coded representation of immersive media. *MPEG-I Part 3: immersive video, ISO/IEC JTC1/SC29 NP 23090-3*.

[Ng et al., 2005] Ng, R., Levoy, M., Brédif, M., Duval, G., Horowitz, M., and Hanrahan, P. (2005). Light field photography with a hand-held plenoptic camera. *Computer Science Technical Report CSTR*, 2(11):1–11.

[Sabater et al., 2017] Sabater, N., Boisson, G., Vandame, B., Kerbiriou, P., Babon, F., Hog, M., Gendrot, R., Langlois, T., Bureller, O., Schubert, A., et al. (2017). Dataset and pipeline for multi-view lightfield video. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1743–1753. IEEE.

[Scharstein and Szeliski, 2002] Scharstein, D. and Szeliski, R., "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International journal of computer vision*, vol. 47, no. 1-3, pp. 7–42, 2002.

[Senoh et al., 2018] Senoh, T., Yamamoto, K., Tetsutani, N., and Yasuda, H. (2018). Enhanced DERS for Quad Reference Views (eDERS). *ISO/IEC JTC1/SC29/WG11 MPEG2018 m41955*.

[Stankiewicz et al., 2013] Stankiewicz, O., Wegner, K., Tanimoto, M., and Domanski, M. (2013). Enhanced view synthesis reference software (VSRS) for free-viewpoint television. *ISO/IEC JTC*, 1:533–541.

[Stutz et al., 2018] Stutz, D., Hermans, A., and Leibe, B. (2018). Superpixels: an evaluation of the state-of-the-art. *Computer Vision and Image Understanding*, 166:1–27.

[Sundberg et al., 2011] Sundberg, P., Brox, T., Maire, M., Arbeláez, P., and Malik, J. (2011). Occlusion boundary detection and figure/ground assignment from optical flow. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 2233–2240. IEEE.

[Wang et al., 2015] Wang, T.-C., Efros, A. A., and Ramamoorthi, R. (2015). Occlusion-aware depth estimation using light-field cameras. In *Computer Vision (ICCV), 2015 IEEE International Conference on*, pages 3487–3495. IEEE.

[Wanner and Goldluecke, 2012] Wanner, S. and Goldluecke, B. (2012). Globally consistent depth labeling of 4D light fields. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 41–48. IEEE.

[Wanner and Goldluecke, 2014] Wanner, S. and Goldluecke, B. (2014). Variational light field analysis for disparity estimation and super-resolution. *IEEE transactions on pattern analysis and machine intelligence*, 36(3):606–619.

[Wanner et al., 2013] Wanner, S., Meister, S., and Goldluecke, B. (2013). Datasets and benchmarks for densely sampled 4d light fields. In *VMV*, pages 225–226. Citeseer.

[Wei and Quan, 2005] Wei, Y. and Quan, L. (2005). Asymmetrical occlusion handling using graph cut for multi-view stereo. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 902–909. IEEE.

[Zhang et al., 2010] Zhang, C., Wang, L., and Yang, R. (2010). Semantic segmentation of urban scenes using dense depth maps. In *European Conference on Computer Vision*, pages 708–721. Springer.

[Zhang et al., 2016] Zhang, S., Sheng, H., Li, C., Zhang, J., and Xiong, Z. (2016). Robust depth estimation for light field via spinning parallelogram operator. *Computer Vision and Image Understanding*, 145:148–159.

[Zhu et al., 2017] Zhu, H., Wang, Q., and Yu, J. (2017). Occlusion-model guided antiocclusion depth estimation in light field. *IEEE Journal of Selected Topics in Signal Processing*, 11(7):965–978.