

# VARIATIONAL DISPARITY ESTIMATION FRAMEWORK FOR PLENOPTIC IMAGES

Trung-Hieu Tran, Zhe Wang, Sven Simon

Department of Parallel Systems  
University of Stuttgart, Stuttgart, Germany  
trung.hieu.tran@ipvs.uni-stuttgart.de

## ABSTRACT

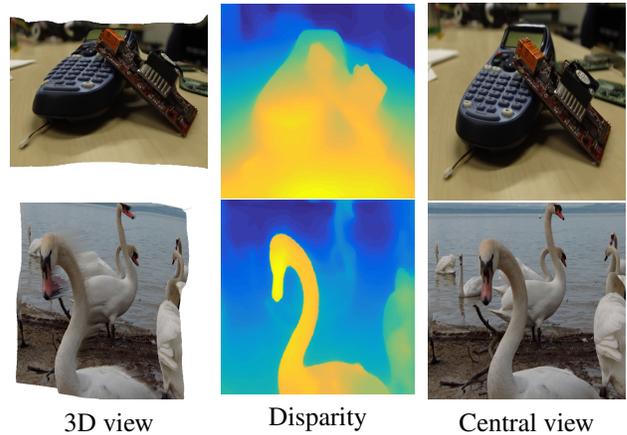
This paper presents a computational framework for accurately estimating the disparity map of plenoptic images. The proposed framework is based on the variational principle and provides intrinsic sub-pixel precision. The light-field motion tensor introduced in the framework allows us to combine advanced robust data terms as well as provides explicit treatments for different color channels. A warping strategy is embedded in our framework for tackling the large displacement problem. We also show that by applying a simple regularization term and a guided median filtering, the accuracy of displacement field at occluded area could be greatly enhanced. We demonstrate the excellent performance of the proposed framework by intensive comparisons with the Lytro software and contemporary approaches on both synthetic and real-world datasets.

**Index Terms**— light-field, correspondence, disparity, plenoptic, variational framework

## 1. INTRODUCTION

The last decade saw the dramatically increasing attention of the research community to light-field photography. Light-field acquisition provides a richer content capturing method compared to traditional photography for not only acquiring the spatial but also the directional information of the scene. Various approaches were provided for capturing light-field such as multi-camera array [1], programmatically moving camera (also known as gantry setup), and microlens-array camera [2]. Among them the lenslet-based or plenoptic camera [2–4] provides the most convenient and efficient way to acquire light-field. On the one hand, it is far less hardware intensive compared to multi-camera array. On the other hand, unlike gantry which captures only a still scene, it could capture both still and dynamic sceneries.

In order to take the most benefits from captured light-field, the disparity map estimation is the most important task that must be considered. Accurate estimation of disparity map will enable the better visualization and manipulation of light-field data. However, it still remains as a challenging problem



**Fig. 1:** Real world result. *top* scene from our laboratory captured light-field. *bottom* scene from EPFL dataset [8].

for plenoptic images. Compared with conventional stereo images, plenoptic sub-aperture images possess a very narrow-baseline configuration with typical disparity range between -1 and 1 pixel. For computing an accurate result, discrete label-based approaches [5–7] require the previous knowledge of sub-pixel disparity unit for initial disparity map estimation and then requires exhaustive adjustment of regularization and refinement parameters in order to acquire a smooth and precise solution. This fact, therefore, limits the use of these approaches for practical applications.

In this paper, we present a computational framework for robust light-field disparity estimation. We formulate the problem using a continuous variational framework that allows an intrinsic sub-pixel precision and introduce a joint light-field motion tensor for combining different advanced data terms. A coarse-to-fine warping strategy is embedded in our framework for dealing with the problem of large displacement in sub-aperture images at directional borders. In addition, we also show that by applying a guided median filter appropriately, the accuracy of displacement field at occluded areas could be greatly enhanced. Fig. 1 demonstrates the results of our proposed framework on Lytro Illium light-field data.

In the following, we provide a brief introduction on the mathematical representation of light-field and literature review on related works. Our light-field variational framework

is then formulated and presented in Section 2. The post-processing technique for sharpening disparity discontinues will be discussed in Section 3. In Section 4, we provide the results of our intensive experiments on both synthetic dataset and real world dataset. Both quantitative and qualitative comparisons with state-of-the art approaches are also provided in this section.

### 1.1. Light-field

Light-field is a 4D parameterization of the plenoptic function, it could be visually described as a ray indexed by its intersection with two parallel planes.

$$L : \Omega \times \Pi \rightarrow \mathbb{R}, (\mathbf{x}, \theta) \rightarrow L(\mathbf{x}, \theta) \quad (1)$$

where  $\mathbf{x} = (x, y)^\top$  and  $\theta = (s, t)^\top$  denote coordinate pairs in the image plane  $\Omega \subset \mathbb{R}^2$  and in the lens plane  $\Pi \subset \mathbb{R}^2$ . For lenslet-based light-field camera,  $\Omega$  and  $\Pi$  correspond to the spatial and directional coordinates respectively.

Given a light-field  $L$ , spatial information could be obtained from one direction by keeping the directional component  $\theta$  unchanged and varying over all spatial domain  $\Omega$ . Such spatial information gives us a sub-aperture image (or a view) of the captured scene. The number of sub-aperture images depends on the configuration of lenslet-based camera with a consideration of the trade-off between spatial and angular resolution [4]. For example, contemporary cameras such as Lytro Ilium provide up to  $15 \times 15$  sub-aperture images with the same amount of reduction in spatial domain. These dense multiple overlapping sub-aperture images show a clear relation to multi-stereo problems or could be referred as a correspondence matching problem. This point of view motivates us to propose this computational framework.

### 1.2. Related works

Bishop and Favaro [9] constructed an image formation for lenslet-based light-field camera and embedded it in their iterative method for multi-view stereo problems. Wanner et al. [10, 11] computed a depth map through estimating the vertical and horizontal slopes in epipolar planes. They then proposed a global optimization scheme using a variational framework for the final depth map. Yu et al. [12] explored geometric structures of 3D lines in ray space and computed disparity maps through line matching between the sub-aperture images. Tao et al. [5] used both correspondence and defocus cue to estimate disparity maps with intention to complement the disadvantages of each other. Wang et al. [6] proposed a depth estimation algorithm that treats occlusion explicitly. Jeon et al. [7] proposed the method based on the phase shift theorem to deal with narrow baseline multi-view images. In [5–7], the proposed approaches came up with a discrete label depth map, and a multi-label optimization method was applied in order to regularize and refine the estimated disparity map.

Compared with these label-based approaches, continuous modeling approaches [13, 14] have advantages because it could provide an intrinsic sub-pixel precision and do not require previous knowledge on sub-pixel displacement units. In [13], Heber and Pock proposed a method that is based on active wave-front sampling and the variational principle. In [14], a method is proposed to estimate disparity maps using the low-rank structure regularization to align the sub-aperture images.

Compared with previous works, especially variational approaches [11, 13], our contribution is two-fold. Firstly, we introduce a light-field motion tensor that allows us to take advantages of different constancy assumptions and explicitly define the contribution of different color channels. Secondly, we proposed to enhance the discontinues of disparity field at occluded areas by appropriately applying guided median filtering. Experimental results show the great accuracy achievement when compared with previous approaches.

## 2. PROPOSED FRAMEWORK

In this section, we formulate the proposed light-field disparity estimation framework, which is based on a variational model. We begin with the general form of a variational problem which is usually written as follows.

$$\operatorname{argmin}_{\omega} E(\omega) = \int_{\Omega} \mathbf{D}(L, \omega) + \alpha \mathbf{S}(L, \omega) d\mathbf{x} \quad (2)$$

where  $\mathbf{D}$  refers to a data term,  $\mathbf{S}$  refers to a smoothness term or a regularization term and  $\alpha > 0$  denotes a smoothness weight. The data function  $\mathbf{D}$  penalizes the deviations of estimated solution  $\omega$  from the true solution with respect to the constraints on the data  $L$ . Such constraints are known as the constancy assumption in optical flow literature [15],[16]. The regularization function  $\mathbf{S}$  takes into account the neighborhood information and guarantees the smoothness of the solution.

We introduce unit displacement  $\omega : (x, y) \rightarrow \mathbb{R}$  which denotes the disparity between the pixels in the central view and in its direct east neighbor view. Based on  $\omega$  we could define the constancy assumption across the light-field  $L(\mathbf{x}, \theta)$ .

$$L(\mathbf{x}, \theta_0) = L(\mathbf{x} + \kappa \theta_i \omega, \theta_i) \quad (3)$$

Here,  $\theta_0$  denotes the central view index and  $\theta_i$  denotes an arbitrary view index within the directional domain  $\Pi$ . A parameter  $\kappa = \kappa_{\Omega} \kappa_{\Pi} = \begin{bmatrix} \kappa & 0 \\ 0 & 1 \end{bmatrix}$ ,  $\kappa \in \mathbb{R}$ , compensates the relative difference between the horizontal and vertical components of directional ( $\kappa_{\Pi}$ ) and spatial ( $\kappa_{\Omega}$ ) coordinates. Without loss of generality and to simplify the exposition, we will absorb  $\kappa$  into  $\theta$ . Following sections will discuss in detail the various settings of the data term and the smoothness term.

## 2.1. Data term

*Intensity constancy assumption* The Eq. 3 actually presents the constancy assumption of pixel intensity that assumes the intensity of corresponding pixels between two sub-aperture images is the same. Supposes  $\omega$  is small, the first-order Taylor expansion gives us the following

$$0 = \omega \theta_i^T \nabla_2 L + |\theta_i| L_\theta = \omega^T \nabla_{\theta_i} L \quad (4)$$

where  $\nabla_2 = (\sigma_x, \sigma_y)^T$  denotes the spatial gradient operator,  $L_\theta$  denotes the directional derivative with direction  $\theta_i$ ,  $\omega = (\omega, 1)^T$  and  $\nabla_{\theta_i} L = (\theta_i^T \nabla_2 L, |\theta_i| L_\theta)^T$

Using Eq. 4, we could derive a data term that takes into account the intensity constancy assumption as follows.

$$\begin{aligned} \mathbf{D}_g(L, \omega) &= \sum_{c=1}^3 \Psi_g \left( \sum_{\theta_i \in \Pi} (L(\mathbf{x} + \theta_i \omega, \theta_i) - L(\mathbf{x}, \theta_0))^2 \right) \\ &\approx \sum_{c=1}^3 \Psi_g \left( \sum_{\theta_i \in \Pi} \omega^T \mathbf{J}_{g,i}^c \omega \right) = \sum_{c=1}^3 \Psi_g(\omega^T \mathbf{J}_g^c \omega) \end{aligned} \quad (5)$$

where  $\mathbf{J}_{g,i}^c = \nabla_{\theta_i} L^c \nabla_{\theta_i}^T L^c$  denotes a light-field motion tensor for a single color channel, and  $L^c$  denotes one color channel of the captured light-field  $L$ . We define a joint light-field motion tensor  $\mathbf{J}_g^c$  for all views  $\theta_i \in \Pi$  as in Eq. 5.  $\Psi_g(s)$  is a positive defined robustification function that helps in reducing the outliers. Here we choose  $\Psi_g(s) = \sqrt{s + \epsilon_g}$ , with  $\epsilon_g > 0$  serves as a small regularization parameter, which also allows the derivative of  $\Psi_g$  available when  $s = 0$ . This  $L_1$  norm is known to be better in handling outliers caused by noise and occlusions.

*Gradient constancy assumption* In addition to intensity, it also makes sense to assume that the gradient of corresponding pixels is also unchanged. In the same manner, the gradient constancy assumption data term could be described as follows.

$$\begin{aligned} \mathbf{D}_G(L, \omega) &= \sum_{c=1}^3 \Psi_G \left( \sum_{\theta_i \in \Pi} (\omega^T \mathbf{J}_{G,x,i}^c \omega + \omega^T \mathbf{J}_{G,y,i}^c \omega) \right) \\ &= \sum_{c=1}^3 \Psi_G(\omega^T \mathbf{J}_G^c \omega) \end{aligned} \quad (6)$$

where  $\mathbf{J}_{G*,i}^c = \nabla_{\theta} L_*^c \nabla_{\theta}^T L_*^c$  and  $L_*^c$  define the derivative of the light-field  $L$  on color channel  $c$ .

*Color space* Both the *Red Green Blue* (RGB) and the *Hue Saturation Value* (HSV) color spaces contain different characteristics that could be exploited for providing a constancy assumption. Both of them are experimented in our framework. Notes that, we apply the joint robustification in the case of RGB because of the mutual relationship between these three channels. For HSV, the separate  $L_1$  norms are used instead,

since each color channel in this case contain information that is not encoded in other channels [16].

Combining both constancy assumptions, we have the final form of the data function. For the RGB color space:

$$\mathbf{D}(L, \omega) = \Psi_g \left( \sum_{c=1}^3 \omega^T \mathbf{J}_g^c \omega \right) + \gamma \Psi_G \left( \sum_{c=1}^3 \omega^T \mathbf{J}_G^c \omega \right) \quad (7)$$

And in the case of the HSV color space.

$$\mathbf{D}(L, \omega) = \sum_{c=1}^3 \Psi_g(\omega^T \mathbf{J}_g^c \omega) + \gamma \sum_{c=1}^3 \Psi_G(\omega^T \mathbf{J}_G^c \omega) \quad (8)$$

Parameter  $\gamma$  allows us to adjust the importance of gradient constancy data term. In following sections, we focus on HSV data term and the RGB data term could be simply derived with a similar manner.

## 2.2. Smoothness term

The light-field motion tensor possesses a strong energy in crowd texture area. However, it is weak in homogeneous area and could lead to wrong estimations. The smoothness term in this case plays a crucial role because it spreads the information through the neighborhood and provides a smooth disparity field. There are various regularization functions that could fit into our framework. The most well known one is  $L_2$  Total Variation (TV -  $L_2$ ) that penalizes the variation of  $\omega$  in two spatial directions using a quadratic function.

$$S(\omega) = |\nabla_2 \omega|^2 \quad (9)$$

Despite providing a smooth solution, this regularizer also blurs the disparity edges and reduces the accuracy of the solution. One solution to it is an isotropic image-driven smoothness term defined with a weighting function  $g(\mathbf{x})$

$$S(\omega) = g(\mathbf{x}) |\nabla_2 \omega|^2 \quad (10)$$

Here,  $g(\mathbf{x}) = \frac{1}{\sqrt{|\nabla_2 L(\mathbf{x}, \theta_0)|^2 + \epsilon}}$  allows reducing smoothness at image edge therefore results in a sharper disparity edge. One problem with this regularization term is the over-segmentation especially in dense texture area. For that reason, we choose the non-quadratic penalizer also known as TV -  $L_1$  that allows piece-wise smoothness in disparity field.

$$\mathbf{S}(\omega) = \Psi_s(|\nabla_2 \omega|^2) \quad (11)$$

where  $\Psi_s(s) = \sqrt{s + \epsilon_s}$ ,  $\epsilon_s > 0$ . It also notes that more sophisticated regularizers such as joint image- and flow-driven [16] or an-isotropic diffusion tensor [13, 17] are also possible in our framework.

### 2.3. Optimization

Combining both data and smoothness terms, we have a final global variational energy function.

$$\begin{aligned} E(\omega) = & \int_{\Omega} \sum_{c=1}^3 \Psi_g(w^T \mathbf{J}_g^c w) + \gamma \sum_{c=1}^3 \Psi_G(w^T \mathbf{J}_G^c w) \\ & + \alpha \Psi_s(|\nabla_2 \omega|^2) dx \end{aligned} \quad (12)$$

In order to estimate the solution, we apply Euler-Lagrange equation:

$$0 = \bar{\mathbf{J}}_{11} \omega + \bar{\mathbf{J}}_{12} - \alpha \operatorname{div}([\Psi_s]'. \nabla_2 \omega) \quad (13)$$

with the Neumann boundary condition  $n^T [\Psi_s]' \nabla_2 \omega = 0$ .  $\bar{\mathbf{J}}_{11}$ , and  $\bar{\mathbf{J}}_{12}$  refer to the two elements in the first row of joint light-field motion tensor defined as the following.

$$\bar{\mathbf{J}} = \sum_{c=1}^3 \left( [\Psi_g^c]' \mathbf{J}_g^c + \gamma [\Psi_G^c]' \mathbf{J}_G^c \right) \quad (14)$$

where  $[\Psi_*]'$  is the derivative of robustification function  $\Psi(s)$ .  $[\Psi_*]' = \Psi'_*(s) = \frac{1}{2\sqrt{s+\epsilon_*}}$ .

### 2.4. Discretization

We follow the discretization strategy from optical flow literature [15], [16] for spatial derivative. The directional derivative is computed as follows

$$L_{\theta} = \frac{L(\mathbf{x}, \theta_i) - L(\mathbf{x}, \theta_0)}{|\theta_i|} \quad (15)$$

For each spatial discrete position  $(i, j) \in \Omega$  we have

$$\begin{aligned} 0 = & [\bar{\mathbf{J}}_{11}]_{i,j}^k \omega_{i,j}^{k+1} + [\bar{\mathbf{J}}_{12}]_{i,j}^k \\ & - \alpha \sum_{l \in x,y} \sum_{(i,j) \in N_l} \frac{[\Psi_s']_{i,j}^k + [\Psi_s']_{i,j}^k}{2} \left( \frac{\omega_{i,j}^{k+1} - \omega_{i,j}^k}{h_l^2} \right) \end{aligned} \quad (16)$$

Here,  $N_x = \{(i-1, j), (i+1, j)\}$ ,  $N_y = \{(i, j-1), (i, j+1)\}$  and  $h_x, h_y$  denote the unit distances for spatial derivative. The Eq. 16 contains nonlinear term  $\bar{\mathbf{J}}_*, \Psi'_s$ . In order to solve it as a linear system of equation we apply lagged non-linearity method where  $\bar{\mathbf{J}}_*, \Psi'_s$  are taken from the old time step ( $k$ ) while computing  $\omega$  for the current time step ( $k+1$ ). By stacking Eq. 16 for all pixels  $(i, j)$  in spatial domain, we have a sparse  $N \times N$ -system of equations that could be solved using Gauss-Seidel iterative method. Here,  $N = N_x \times N_y$  with  $N_x$  and  $N_y$  being spatial sampling resolution.

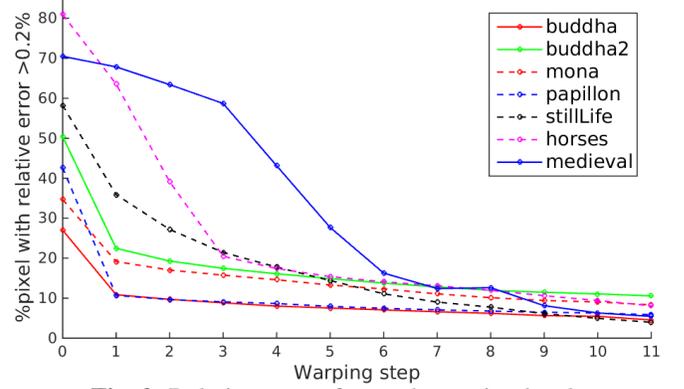


Fig. 2: Relative error after each warping level.

### Algorithm 1 Warping strategy with coarse-to-fine

**Require:**  $\alpha, \gamma, \sigma, \eta, l$

```

1: procedure DEPTH(L)
2:   initialize  $R_i, L_i$ 
3:    $\omega = 0$ 
4:   for ( $i = l; i \geq 0; i = i - 1$ ) do
5:      $\hat{\omega} = \text{upscale}(\omega, R_i)$ 
6:      $\hat{L}_i = \text{warp}(L_i, \hat{\omega})$ 
7:      $\mathbf{J}_g = \text{lfmogray}(\hat{L}_i)$ 
8:      $\mathbf{J}_G = \text{lfmograd}(\hat{L}_i)$ 
9:      $\omega = \hat{\omega} + \text{itersol}(\mathbf{J}_g, \mathbf{J}_G, \alpha, \gamma)$ 
return  $\omega$ 

```

### 2.5. Warping strategy

As a condition for the linearization, we assume the displacement is small. This assumption will not always hold true, especially when we come to the directional position  $\theta_i$  that is far away from  $\theta_0$ . In order to overcome this problem, we apply a coarse-to-fine warping technique [15]. The outline of our implementation is shown in Algorithm. 1.

The input parameters include the smoothness weight  $\alpha$ , the weight of gradient constancy data term  $\gamma$ , the down-sampling factor  $\eta$ , the standard deviation for Gaussian presmoothing  $\sigma$  and the number of warping steps  $l$ . We first initialize the set of down-sample light-field  $L_i$  and resolution  $R_i$  for each warping stage  $i = 1, \dots, l$ . To downscale light-field, we first convoluted each sub-aperture image with Gaussian kernel and then sampled using bicubic interpolation. This is for avoiding alias during the down-sampling. We then begin with the coarsest level  $l$  with the resolution of sub-aperture image down sample to  $R_l = \eta^l (N_x, N_y)^T$ . For each warping stage, the displacement from previous stage is up-scaled (upscale) to the current resolution. The light-field at the current level is warped to the central sub-aperture view (warp). The light-field motion tensors are then computed for intensity constancy (lfmogray) and gradient constancy (lfmograd) assumption. The disparity map  $\omega$  at current warping level is computed using the iterative solver (itersol).

Fig. 2 shows the changes in the percentage of erroneous

pixels with relative error larger than 0.2% after each warping stage. The results are computed for Synthetic light-field dataset [18] with  $l = 11$  and  $\eta = 0.8$ . The graph shows great improvements on the disparity accuracy after each warping level.

### 3. POST PROCESSING

Occlusion is one of the main problem in stereo matching. It happens when a patch of pixels exists in one view but vanishes in the others because of occluders. The problem is even more serious in the case of multi-stereo like light-field when the amount of vanish is varied from view to view. Our light-field motion tensors encode the information from all possible views and therefore contain least precise energy in occluded area. Since we compute the disparity map with respect to the central sub-aperture image, it makes sense to perform a post-processing to sharpen disparity field with respect to this reference view. For this purpose we propose a simple approach that employs the guided median filtering [19].

The procedure contains two steps. Firstly, we search for occluded areas where there is a high possibility of erroneous pixels. Secondly, we apply guided median filtering on the disparity map with respect to these occluded areas. We notice that the necessary condition for occlusion is depth discontinuity, and therefore we propose a simple occlusion detection based on computed displacement field.

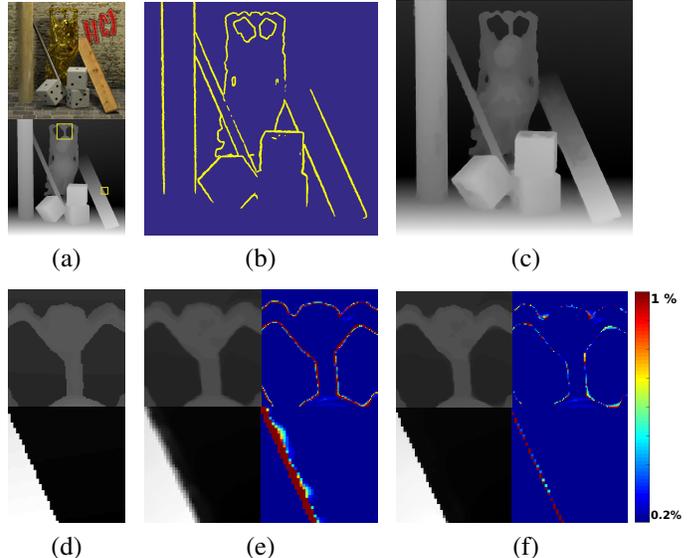
$$P_{occ} = f(B_r * |\nabla_2 \omega|^2) \quad (17)$$

where  $f(\cdot)$  is a binary marking function that marks the response above some threshold and  $B_r *$  denotes the convolution with a box kernel with size  $r$  to expand suspected occlusion areas. An example of marked occluded areas is shown in Fig. 3(b). We then apply a median filter with a central sub-aperture image  $I_c = L(\mathbf{x}, \theta_0)$  as a guide. This allows us to have a sharper and more precise disparity discontinues as could be seen in Fig. 3(e),(f).

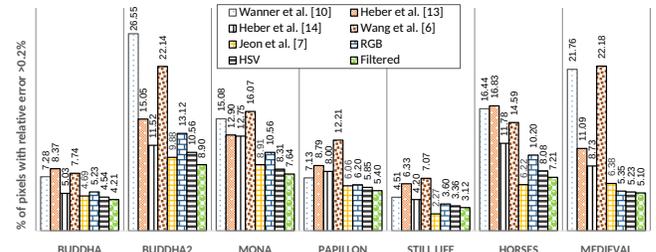
### 4. EXPERIMENTAL RESULTS

The performance of the proposed variational framework was evaluated using both synthetic and real world datasets. The synthetic 4D Light-field dataset [18] was used for quantitative comparisons with related work. All the real world dataset was captured using lenslet based light-field camera Lytro. We used both light-field data provided by EPFL [8] and data captured by our Lytro Illium camera.

We implemented our proposed framework on MATLAB. Source code of this implementation can be found at [https://github.com/hieuttcse/variational\\_plenoptic\\_disparity\\_estimation](https://github.com/hieuttcse/variational_plenoptic_disparity_estimation). The computation ran on Intel i5 2.4Ghz CPU with 8GB RAM and required 4 minutes for the *Lytro* dataset and 9 minutes for the Synthetic dataset. We applied the



**Fig. 3:** Impact of proposed post-processing method on estimated disparity field. The result is computed for *buddha2* scene from Synthetic dataset [10]. (a) *top*: the central sub-aperture image. *bottom*: the ground truth with two interest areas. (b) Estimated occlusion area. (c) Computed disparity map. (d) Zoom in ground truth for two interest areas. (e),(f) estimated disparity and relative error before and after guided filtering respectively.



**Fig. 4:** Relative error comparison on synthetic dataset [18]

Gauss-Seidel iteration solver with successive over-relaxation levels set to 1.88 which allowed the solution to converge faster, around 100 iterations. The two parameters  $\alpha, \gamma$  are adjusted for each light-field data. The other parameters were selected as  $l = 11, \eta = 0.8, \sigma = 0.5, \epsilon_* = 0.001^2$ . We notice that the most time consuming tasks are warping ( $\text{warp}(L_i, \hat{\omega})$ ) and computing light-motion tensor ( $\text{lfmogray}(\hat{L}_i), \text{lfmograd}(\hat{L}_i)$ ). These tasks could be dramatically speeded up on high-parallelism platform such as GPU or FPGA.

#### 4.1. Synthetic scene

For quantitative evaluation, we compared our work with both continuous modelling approaches [10, 13, 14] and discrete label-based approaches [6, 7]. Since Wang et al. [6] did not report their results on synthetic dataset, we run their provided code for this comparison. For the others, we used their best

results reported from related papers [7, 13, 14].

The relative error is adopted as a common metric for comparison. For each synthetic light-field data, we computed the percentage of pixels with a relative depth error of more than 0.2%. As mentioned by Heber [13], this is the smallest meaningful accuracy level since the depth discretization of the provided ground truth is too low. We reported the result for both RGB and HSV color space and the post-processing result of the best solution.

Fig. 4 shows the relative error results of the previous works and ours. It can be seen from the bar chart that our proposed framework outperforms previous continuous modelling approaches [10, 13, 14] in term of accuracy. Compared with the RGB color space, the computation on the HSV color space provides more precision due to the separate robustification of the data term. The guided median filtering further improves the accuracy by sharpening the disparity discontinues. We notice that the approach of Wang et al. [6] is very sensitive to estimated occlusion results, and tends to have a wrong disparity estimation at strong texture areas. Their approach therefore is less accurate when compared to the others.

Compared with the work of Jeon et al. [7], our approach provides better results for the five out of seven synthetic light-field data. The two less accurate results (*still life, horses*) show some insights about the limitation of our approach that will be discussed at the end of this section. However, advantages of our approach are as follows. Firstly, our approach does not require the previous knowledge of disparity range as well as spatial disparity unit for each disparity label due to the continuous formation. Without this knowledge, discrete label-based approaches [6],[7] need to increase the number of labels and reduce the spatial unit in order to sufficiently cover the disparity range. It consequently increases the computation time. In addition, a number of well-turned parameters for each data, around 9 parameters, are critical in their approach [7] in order to guarantee a smooth disparity field without outliers. Our framework, in contrary, requires mainly adjustment on only two parameters  $\gamma$  and  $\alpha$ .

## 4.2. Real scene

In this section, we provided some qualitative results on real world dataset obtained by using *Lytro* camera. The images are captured using Ilium version which provides a spatial resolution of around  $434 \times 625$  and a directional resolution of  $15 \times 15$ . Due to the vignetting impact of microlens, the effective directional resolution is limited to only 193 views (85.7%). Fig. 1 presents the results of our proposed approach for different scenes. Along with the central sub-aperture image, there is also a rendered 3D view as well as a color-coded disparity map. Fig. 5 shows the disparity map for two different scenes computed respectively by the *Lytro* software, the methods of Wang et al.[6], Jeon et al.[7] and ours. The results for [6, 7] are computed using their provided codes. It can be

seen from the figure that our proposed approach provides solutions with more details and fewer outliers.

## 4.3. Limitation and future work

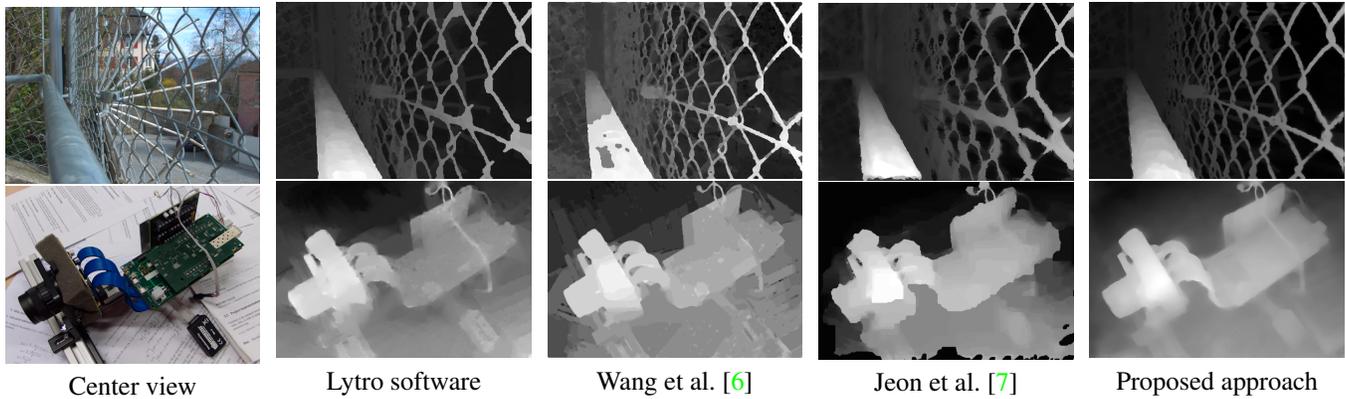
The accuracy of our approach is limited by the error at depth discontinues where occlusion takes place. While guided median filtering does improve the accuracy at these areas, it could also introduce “halo” artefact [19] when the texture borders in guided image are weak. One solution for this problem is taking into account the occlusion border [6] to neglect the artefact. Another solution is to employ more sophisticated smoothness terms that could better preserve the depth discontinues such as image- and flow-driven regularizer [16] or Total Generalized Variation (TGV) [17]. These tasks are listed in our plan for future work.

## 5. CONCLUSION

In this paper we proposed a variational computation framework for the disparity estimation problem targeting very narrow-baseline multi-stereo data from plenoptic images. We introduced a light-field motion tensor that allows different constancy assumptions to be applied. We embedded coarse-to-fine warping strategy to our framework in order to overcome the problem of large displacement and proposed an effective post-processing technique for further enhancing the accuracy at occluded areas. The experimental results show our competitive performance on both challenging synthetic and real world light-field dataset.

## 6. REFERENCES

- [1] Bennett Wilburn, Neel Joshi, Vaibhav Vaish, Eino-Ville Talvala, Emilio Antunez, Adam Barth, Andrew Adams, Mark Horowitz, and Marc Levoy, “High performance imaging using large camera arrays,” *ACM Transactions on Graphics (TOG)*, vol. 24, no. 3, pp. 765, jul 2005.
- [2] E.H. H Adelson and J.Y.A. Y A Wang, “Single lens stereo with a plenoptic camera,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 99–106, 1992.
- [3] Ren Ng, Marc Levoy, Gene Duval, Mark Horowitz, and Pat Hanrahan, “Light Field Photography with a Hand-held Plenoptic Camera,” *Computer Science Technical Report*, vol. 2, no. 11, pp. 1–11, 2005.
- [4] Andrew Lumsdaine and Todor Georgiev, “The focused plenoptic camera,” in *2009 IEEE International Conference on Computational Photography (ICCP)*. apr 2009, pp. 1–8, IEEE.
- [5] Michael W Tao and Pratul P Srinivasan, “Depth from Shading , Defocus , and Correspondence Using Light-Field Angular Coherence,” *Cvpr*, pp. 1940–1948, 2015.
- [6] Ting-Chun Wang, Alexei A. Efros, and Ravi Ramamoorthi, “Occlusion-Aware Depth Estimation Using Light-Field Cameras,” in *2015 IEEE International Conference on Computer Vision (ICCV)*. dec 2015, pp. 3487–3495, IEEE.



**Fig. 5:** Comparison on real-world dataset. *top* scene from EPFL dataset [8]. *bottom* scene from our captured light-field.

- [7] Hae Gon Jeon, Jaesik Park, Gyeongmin Choe, Jinsun Park, Yunsu Bok, Yu Wing Tai, and In So Kweon, “Accurate depth map estimation from a lenslet light field camera,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. jun 2015, vol. 07-12-June, pp. 1547–1555, IEEE.
- [8] Martin Rerabek and Touradj Ebrahimi, “New Light Field Image Dataset,” in *8th International Conference on Quality of Multimedia Experience (QoMEX)*, 2016, pp. 1–2.
- [9] Tom E. Bishop and Paolo Favaro, “The Light Field Camera: Extended Depth of Field, Aliasing, and Superresolution,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 5, pp. 972–986, may 2012.
- [10] Sven Wanner and Bastian Goldluecke, “Globally consistent depth labeling of 4D light fields,” in *2012 IEEE Conference on Computer Vision and Pattern Recognition*. jun 2012, pp. 41–48, IEEE.
- [11] Sven Wanner and Bastian Goldluecke, “Variational Light Field Analysis for Disparity Estimation and Super-Resolution,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 3, pp. 606–619, mar 2014.
- [12] Zhan Yu, Xinqing Guo, Haibing Ling, Andrew Lumsdaine, and Jingyi Yu, “Line assisted light field triangulation and stereo matching,” *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2792–2799, 2013.
- [13] Stefan Heber, Rene Ranftl, and Thomas Pock, “Variational shape from light field,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2013, vol. 8081 LNCS, pp. 66–79.
- [14] Stefan Heber and Thomas Pock, “Shape from light field meets robust PCA,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8694 LNCS, no. PART 6, pp. 751–767, 2014.
- [15] Andrés Bruhn, Joachim Weickert, Timo Kohlberger, and Christoph Schnörr, “A multigrid platform for real-time motion computation with discontinuity-preserving variational methods,” *International Journal of Computer Vision*, vol. 70, no. 3, pp. 257–277, 2006.
- [16] Henning Zimmer, Andrés Bruhn, and Joachim Weickert, “Optic flow in harmony,” *International Journal of Computer Vision*, vol. 93, no. 3, pp. 368–388, 2011.
- [17] Rene Ranftl, Stefan Gehrig, Thomas Pock, and Horst Bischof, “Pushing the limits of stereo using variational stereo estimation,” in *2012 IEEE Intelligent Vehicles Symposium*. jun 2012, number 1, pp. 401–407, IEEE.
- [18] Sven Wanner, Stephan Meister, and Bastian Goldluecke, “Datasets and Benchmarks for Densely Sampled 4D Light Fields,” *Vision, Modeling & Visualization*, pp. 225–226, 2013.
- [19] Kaiming He, Jian Sun, and Xiaoou Tang, “Guided Image Filtering,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1397–1409, jun 2013.