

Sheared EPI Analysis for Disparity Estimation from Light Fields

Takahiro SUZUKI^{†a)}, Nonmember, Keita TAKAHASHI^{†b)}, and Toshiaki FUJII^{†c)}, Members

SUMMARY Structure tensor analysis on epipolar plane images (EPIs) is a successful approach to estimate disparity from a light field, i.e. a dense set of multi-view images. However, the disparity range allowable for the light field is limited because the estimation becomes less accurate as the range of disparities become larger. To overcome this limitation, we developed a new method called sheared EPI analysis, where EPIs are sheared before the structure tensor analysis. The results of analysis obtained with different shear values are integrated into a final disparity map through a smoothing process, which is the key idea of our method. In this paper, we closely investigate the performance of sheared EPI analysis and demonstrate the effectiveness of the smoothing process by extensively evaluating the proposed method with 15 datasets that have large disparity ranges.

key words: epipolar plane image, light fields, disparity estimation

1. Introduction

Depth (disparity) estimation from images has attracted much research interest for many years. One of the most common conventional configurations is stereo matching using two images [1]–[5], for which many sophisticated techniques have been developed. Recent advances in image acquisition techniques [6]–[11] has brought about a new trend in this research field: depth estimation from a set of multi-view images or a light field. The most straightforward approach is called multi-view stereo (MVS) [12]–[16], where the classical stereo matching methods for two images are extended directly to a set of multi-view images; the basic idea is to find corresponding points across the images.

Another approach is to analyze the structure of an epipolar plane image (EPI) that is obtained from the light field [17]–[24]. This approach is based on the fact that an EPI consists of many line patterns, and the slopes of those lines are directly related to the depth information. Waner and Goldluecke [20], [22] have applied structure tensor analysis to EPIs. Their method has been proven to be fast and accurate when the light field is sufficiently dense, i.e., the disparity range between the neighboring viewpoints is sufficiently small. However, we found that the accuracy of their method is quite limited for relatively sparse light fields where neighboring images have non-small disparities.

To overcome this limitation while keeping computational cost low, we have developed a method called sheared EPI analysis [25]*, where EPIs are transformed with several shear values before the structure tensor analysis. Then, the results of analysis obtained with different shear values are integrated into a final disparity map through a smoothing process, which is the key idea of our method.

We have found that an idea similar to ours had been presented as EPI refocusing [21]. However, our study has several nontrivial differences. First, our method is applied to 2-D viewpoint arrangements, while that of Diebold and Goldluecke [21] was limited to 1-D ones. Second, our method includes a smoothing process during the integration process, which significantly increases the accuracy of disparity estimation. Moreover, we extensively evaluated 15 datasets that have large disparity ranges to closely investigate the effectiveness of shearing EPIs and the smoothing process. We also prove that our method is comparably accurate to and much faster than a multi-view stereo method.

2. Epipolar Plane Image Analysis

2.1 Outline of Epipolar Plane Image Analysis

We assume that a set of multi-view images, such as that shown in Fig. 1 (a), is given. These images constitute a 4-D light field $l(s, t, x, y)$, where (s, t) denotes a viewpoint and (x, y) denotes a pixel position. A 2-D subspace of the 4-D light field with a fixed (s, x) or (t, y) is called an epipolar plane image (EPI). For example, $l^{t^*y^*}(x, s) = l(s, t^*, x, y^*)$ is an EPI on (x, s) plane where t and y are fixed to t^* and

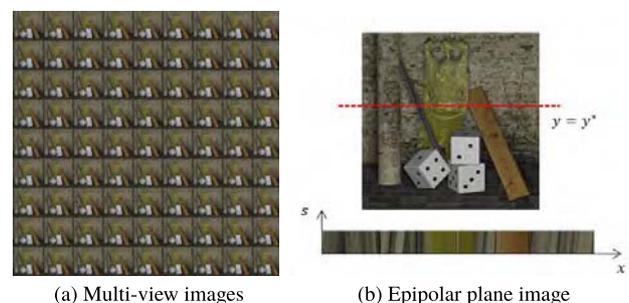


Fig. 1 Multi-view images and EPI

Manuscript received December 15, 2016.

Manuscript revised March 29, 2017.

Manuscript publicized June 14, 2017.

[†]The authors are with Graduate School of Engineering, Nagoya University, Nagoya-shi, 464–8603 Japan.

a) E-mail: suzuki@fujii.nuee.nagoya-u.ac.jp

b) E-mail: keita.takahashi@nagoya-u.jp

c) E-mail: fujii@nuee.nagoya-u.ac.jp

DOI: 10.1587/transinf.2016PCP0004

*This paper is an extension of a conference paper [25]. We have included more thorough experimental validations in the present paper to fully reveal the effectiveness of our method.

y^* , respectively, as shown in Fig. 1 (b). The EPI consists of many lines, each of which is a trace of an object point, and its direction (slope) corresponds to the depth of the object point. Therefore, analyzing the line direction is equivalent to estimating depth [17]–[23].

On the basis of this idea, Wanner and Goldluecke [20], [22] have developed a depth estimation method using structure tensor analysis. A structure tensor on an EPI $l(x, s)$ is defined as

$$J(x, s) = \begin{bmatrix} G * (l_x l_x) & G * (l_x l_s) \\ G * (l_x l_s) & G * (l_s l_s) \end{bmatrix}, \quad (1)$$

where $G*$ denotes convolution with a Gaussian filter kernel. In this paper, the filter size is fixed to the default value of [26]: a 3×3 kernel with $\sigma = 1$. Symbols l_x and l_s denote gradients of the EPI along x and s directions. The dominant gradient direction $\theta(x, s)$ and its confidence $c(x, s)$ (coherence in [20], [22]) can be obtained from principle component analysis of matrix J .

$$\theta(x, s) = \frac{1}{2} \arctan \left(\frac{2J_{12}(x, s)}{J_{11}(x, s) - J_{22}(x, s)} \right) \quad (2)$$

$$c(x, s) = \frac{\sqrt{(J_{11}(x, s) - J_{22}(x, s))^2 + 4J_{12}^2(x, s)}}{J_{11}(x, s) + J_{22}(x, s)}, \quad (3)$$

where larger $c(x, s)$ means more confidence. The disparity $d(x, s)$ is given by $d(x, s) = \tan \theta(x, s)$. For simplicity, we describe these processes as a function $\text{EPIAnaly}(\cdot)$. For fixed t^* and y^* , this function is written as

$$(\theta^{t^*y^*}(x, s), c^{t^*y^*}(x, s)) = \text{EPIAnaly}(l^{t^*y^*}(x, s)). \quad (4)$$

To obtain a disparity map from a specific viewpoint (s^*, t^*) , we perform EPI analysis for both the horizontal and vertical directions and combine the results in accordance with the point-wise confidence.

$$d^{s^*t^*}(x, y) = \begin{cases} d^{t^*y^*}(x, s^*) & c^{t^*y^*}(x, s^*) \geq c^{s^*x}(y, t^*) \\ d^{s^*x}(y, t^*) & \text{otherwise.} \end{cases} \quad (5)$$

The disparity map $d^{s^*t^*}(x, y)$ is further refined using a fast denoising or a more sophisticated global optimization. The former is given as:

$$\hat{d}^{s^*t^*}(x, y) = \arg \min_{d(x, y)} E(d(x, y), d^{s^*t^*}(x, y)) \quad (6)$$

$$E(\alpha, \alpha_0) = \int_{\Omega \subset \mathcal{R}^2} h |D_{xy} \alpha| + \frac{1}{2\rho} |\alpha - \alpha_0| d\Omega \quad (7)$$

$$h(x, y) = 1 - \max(c^{t^*y^*}(x, s^*), c^{s^*x}(y, t^*)) \quad (8)$$

where Ω is the 2-D pixel domain, ρ is a smoothing strength, and D_{xy} is a 2-D derivative operator. The first term of Eq. (7) penalizes non-smoothness on the disparity map and its strength is controlled by the per pixel confidence of the initial disparity value; less confident disparities are more strongly smoothed.

2.2 Comparison to 4-D Analysis

A light field has originally a 4-D structure. If a pixel (x, y) at a specific viewpoint (s_0, t_0) has a disparity d ,

$$l(s_0, t_0, x, y) \simeq l(s, t, x - d(s - s_0), y - d(t - t_0)) \quad (9)$$

should be satisfied except for occluded regions. Several methods use this 4-D constraint directly to obtain accurate disparity values from the light field [11], [27]–[29]. This constraint can be used for multi-view stereo matching where the point correspondence is evaluated across all the images arranged in a 2-D grid, or to derive defocus cues by taking the average of disparity-compensated multi-view images. Generally, these methods are computationally heavy due to the complexity of analysis that is performed on the 4-D space.

Meanwhile, the EPI analysis presented in [20], [22] works in 2-D; only the 2-D subspaces, i.e. (s, x) and (t, y) planes, extracted from the original 4-D light field, are analyzed to obtain disparities. This analysis is computationally much more simpler than the 4-D methods. This fast 2-D analysis is used as the basic building block of our method. As a result, our method can achieve a good performance in terms of the trade-off between the accuracy and computational cost, as will be shown in Sect. 4.

3. Proposed Method

3.1 Sheared EPI Analysis

EPI analysis has been proven to be fast and accurate for a dense light field where the range of disparities among the viewpoints is small [20], [22]. However, this does not hold true for the images that have a larger range of disparities. For example, Fig. 2 (b) shows a disparity map estimated from 3×3 viewpoint images where disparities between neighboring viewpoints are from -2.54 to 4.86 pixels. This disparity map is quite erroneous compared with the ground truth in Fig. 2 (a).

To identify the reason for this erroneous result, we closely observe several EPIs in Fig. 3, where the original EPIs are visualized in the row of $\delta = 0$. As mentioned earlier, the direction of each line corresponds to its disparity. If the disparity is near zero, the line direction is almost vertical. However, as the disparity diverges from zero, the line becomes more slanted and finally separates into discontinuous dots. In such cases, the dominant gradient direction obtained by the structure tensor analysis no longer corresponds correctly to the line direction. In fact, the most erroneous parts in Fig. 2 (b) originally have disparities that diverge from zero.

To fix these errors, we introduce shear transforms to EPI analysis. Specifically, a sheared EPI is written as

$$l_{\delta}^{t^*y^*}(x, s) = l^{t^*y^*}(x + s\delta, s), \quad (10)$$

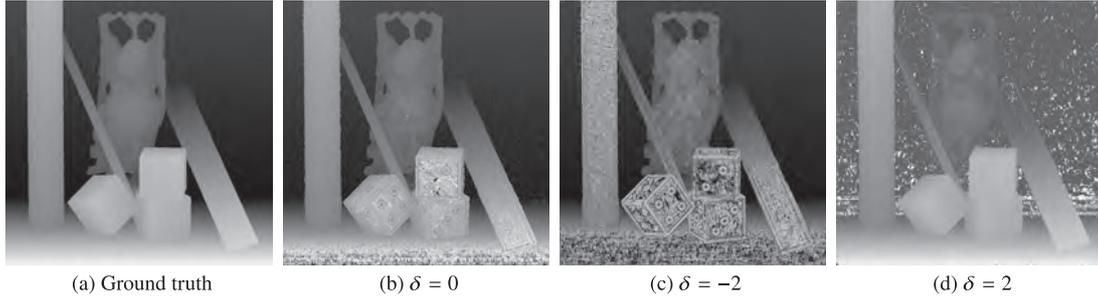


Fig. 2 Disparity maps estimated by EPI analysis with different shear values

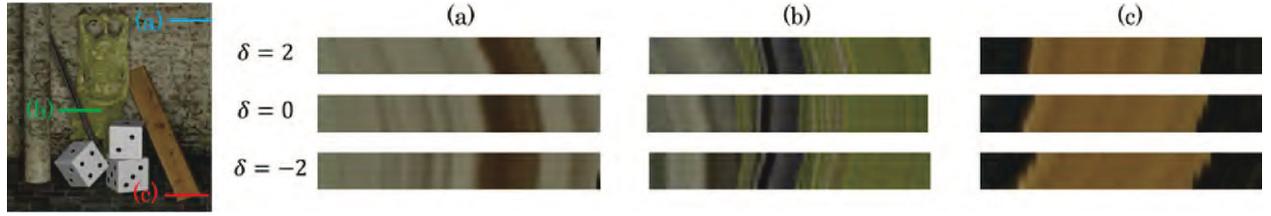


Fig. 3 Sheared EPIs

where δ is the amount of shear, which is called the shear value in this paper. Examples of sheared EPIs with $\delta = 2$ and -2 are shown in Fig. 3. It can be observed that thanks to the shear transforms, the directions of several lines become closer to vertical. These directions can now successfully be estimated by the structure tensor analysis.

Disparity estimation from a sheared EPI is formulated using Eq. (4) as

$$(\theta_{\delta}^{s^*y^*}(x, s), c_{\delta}^{t^*y^*}(x, s)) = \text{EPIAnaly}(l_{\delta}^{t^*y^*}(x, s)) \quad (11)$$

$$d_{\delta}^{s^*y^*}(x, s) = \delta + \tan \theta_{\delta}^{t^*y^*}(x, s), \quad (12)$$

where the shear value δ is compensated for in Eq. (12). Similarly to Eq. (5), a disparity map from a specific viewpoint (s^*, t^*) is obtained as

$$d_{\delta}^{s^*t^*}(x, y) = \begin{cases} d_{\delta}^{t^*y^*}(x, s^*) & c_{\delta}^{t^*y^*}(x, s^*) \geq c_{\delta}^{s^*x}(y, t^*) \\ d_{\delta}^{s^*x}(y, t^*) & \text{otherwise.} \end{cases} \quad (13)$$

Here, $d_{\delta}^{s^*t^*}(x, y)$ denotes a disparity map obtained through a shear transform with δ . Such disparity maps with $\delta = -2$ and 2 are presented in Figs. 2 (c) and 2 (d). We observed that the shear transforms partly improve disparity estimation; with $\delta = 2$, the nearer regions become accurate, while with $\delta = -2$, the farther regions become accurate. These results validate the effectiveness of introducing shear transforms to EPI Analysis.

3.2 Integration of Multiple Disparity Maps

Shear transform of EPIs helps improve the accuracy of disparity estimation where and only where the original disparities are close to the shear value δ . To cover a large range of disparities in a target light field, we need to perform sheared EPI analysis several (N) times while changing the value of

δ and combine the results. Here, we describe how to integrate multiple disparity maps obtained with different shear values.

The n -th shear value is denoted as δ_n ($n = 1, \dots, N$). First, for each δ_n , we obtain a disparity map $d_{\delta_n}^{s^*t^*}(x, y)$ using Eq. (13) and a confidence map $c_{\delta_n}^{s^*t^*}(x, y)$ using

$$c_{\delta_n}^{s^*t^*}(x, y) = \max(c_{\delta_n}^{t^*y^*}(x, s^*), c_{\delta_n}^{s^*x}(y, t^*)). \quad (14)$$

Next, we integrate N disparity maps, $d_{\delta_n}^{s^*t^*}(x, y)$ ($n = 1, \dots, N$). For this purpose, the most confident value of δ for each (x, y) is selected as

$$\delta^{s^*t^*}(x, y) = \arg \max_{\delta \in \{\delta_n\}} c_{\delta}^{s^*t^*}(x, y), \quad (15)$$

which is called a δ map. A straightforward method of integration is to use it directly as

$$d^{s^*t^*}(x, y) = d_{\delta^{s^*t^*}(x, y)}^{s^*t^*}(x, y), \quad (16)$$

which is the same as in the work of Diebold and Goldluecke [21] except for the arrangement of viewpoints. However, this straightforward integration results in a noisy disparity map. We found that this noisiness comes from the noisiness of the δ map visualized in Fig. 4 (a). The true δ map is expected to be locally smooth because it is ideally a quantized version of the continuous disparity map and the continuous disparity map is locally smooth for a natural scene in general. Therefore, we smooth the δ map with weights in accordance with the per pixel confidence.

$$\hat{\delta}^{s^*t^*}(x, y) = \text{round}_{\delta_n} \left(\underset{\delta(x, y)}{\text{argmin}} E(\delta(x, y), \delta^{s^*t^*}(x, y)) \right) \quad (17)$$

$$E(\alpha, \alpha_0) = \int_{\Omega \subset \mathbb{R}^2} g|D_{xy}\alpha| + \frac{1}{2\lambda}|\alpha - \alpha_0|d\Omega, \quad (18)$$

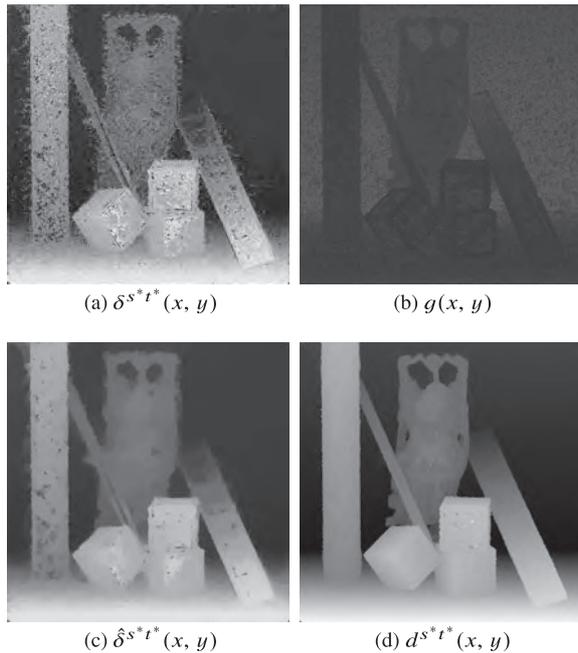


Fig. 4 Visualization of proposed method

where Ω is the 2-D pixel domain, λ is a smoothing strength (when $\lambda = 0$, no smoothing is performed), and D_{xy} is a 2-D derivative operator. Symbol g denotes a weighting function described as

$$g(x, y) = 1 - \frac{\max_{\delta \in \{\delta_n\}} c_{\delta}^{s^*t^*}(x, y)}{\sum_{\delta \in \{\delta_n\}} c_{\delta}^{s^*t^*}(x, y)}. \quad (19)$$

Function $g(x, y)$ takes larger values where the corresponding disparities are less confident as visualized in Fig. 4 (b). Larger values of $g(x, y)$ lead to a stronger smoothing effect for the pixel (x, y) . Finally, the smoothed δ map $\hat{\delta}^{s^*t^*}(x, y)$, which is shown in Fig. 4 (c), is used to integrate N disparity maps as

$$d^{s^*t^*}(x, y) = d_{\hat{\delta}^{s^*t^*}(x, y)}^{s^*t^*}(x, y). \quad (20)$$

As shown in Fig. 4 (d), the disparity map obtained by Eq. (20) is quite accurate. It is further refined by the additional denoising given by Eqs. (6)–(8), where ρ was fixed to 5 throughout this paper and the weight function h was defined as

$$h(x, y) = 1 - \max_{\delta \in \{\delta_n\}} c_{\delta}^{s^*t^*}(x, y). \quad (21)$$

4. Experiments

We examined the performance of our method and the effect of several parameters using light field datasets obtained from other studies [30], [31]. All of the original datasets have 9×9 viewpoints, but except for a few datasets the disparity ranges are too small to show difference between the original EPI analysis and our method. Therefore, from

Table 1 Light field datasets from [30], [31]

Datasets	Disparity range	Views
buddha	[-2.54, 4.86]	3×3
buddha2	[-2.15, 3.74]	3×3
couple	[-2.20, 3.76]	3×3
cube	[-2.42, 2.39]	3×3
horses (a)	[-8.92, 2.93]	3×3
horses (b)	[-2.97, 0.98]	9×9
maria	[-1.69, 1.29]	3×3
medieval (a)	[-8.10, 2.72]	3×3
medieval (b)	[-2.70, 0.91]	9×9
monasRoom	[-2.38, 2.15]	3×3
papillon	[-3.51, 2.68]	3×3
pyramid	[-2.23, 2.36]	3×3
statue	[-2.79, 1.61]	3×3
stillLife (a)	[-8.12, 7.87]	3×3
stillLife (b)	[-2.71, 2.62]	9×9

each dataset, we selected 3×3 viewpoint images with constant intervals and used them as an input light field. As for the datasets *horses*, *medieval*, and *stillLife*, we also used the original 9×9 viewpoints because these datasets have originally large disparity ranges. The input data and the disparity ranges between the neighboring viewpoints are listed in Table 1. To implement our proposed method, we used *cocolib* and light field suite software available online [26], [31]. The *cocolib* software includes a fast implementation of the continuous minimization for Eqs. (7) and (18). For each dataset, the disparity map at the central viewpoint was evaluated against the ground truth.

In Sects. 4.1–4.5, we evaluated the performance of our method. In Sects. 4.1–4.3, we present the effects of the three parameters: the shear range Δ , the shear step τ , and the smoothing strength λ (in Eq. (18)). For example, when $\Delta = [-3, 5]$, the candidate shear values are given as $\delta_n = \{-3, -2, -1, 0, 1, 2, 3, 4, 5\}$ with $\tau = 1$, and $\delta_n = \{-2, 1, 4\}$ with $\tau = 3$. The performance of our method also depends on the target disparity ranges D and disparity-map denoising, which are analyzed in Sects. 4.4 and 4.5. For the experiments described in Sects. 4.1–4.5, we used the *buddha* dataset. The disparity-map denoising was omitted in Sects. 4.1–4.4 but was enabled in Sects. 4.5 and 4.6. In Sect. 4.6, we compare our method with other methods over all datasets and discuss the trade-off between the accuracy and computational cost for disparity estimation.

4.1 Effect of the Shear Range Δ

First, we explain how the shear range Δ affects the performance of our method. In this experiment, the shear range Δ was changed while the target disparity range D was fixed. The shear step τ was fixed to 1, and the smoothing strength λ was optimized for each condition. We compared the cases with and without δ -map smoothing. The performance was measured by using the accuracy of disparity estimation (PSNR against the ground truth) and is summarized in Fig. 5.

It can be seen that the δ -map smoothing significantly

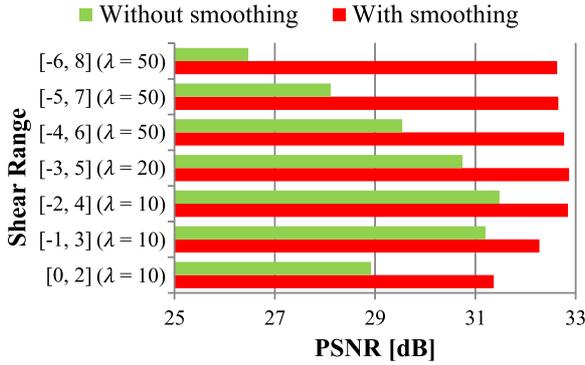


Fig. 5 PSNR and Shear range Δ

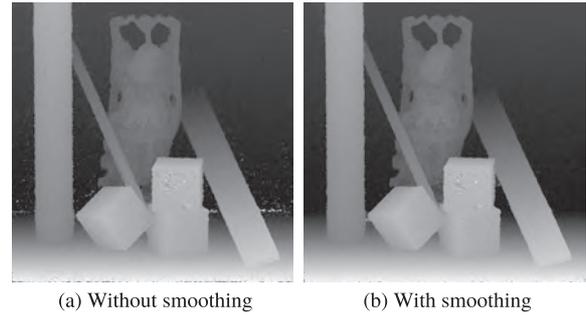


Fig. 8 Disparity maps with $\Delta = [0, 2]$

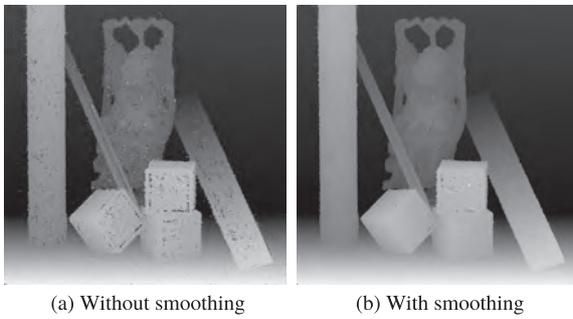


Fig. 6 Disparity maps with $\Delta = [-6, 8]$

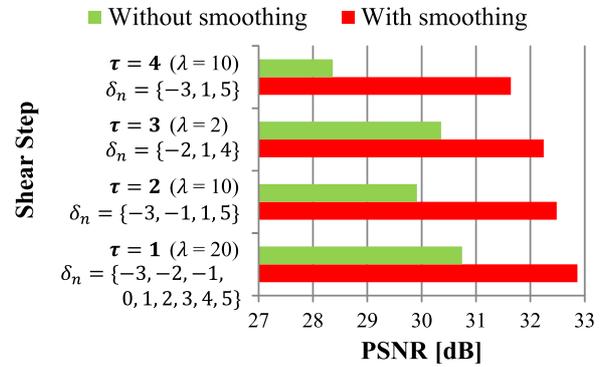


Fig. 9 PSNR with different shear steps τ

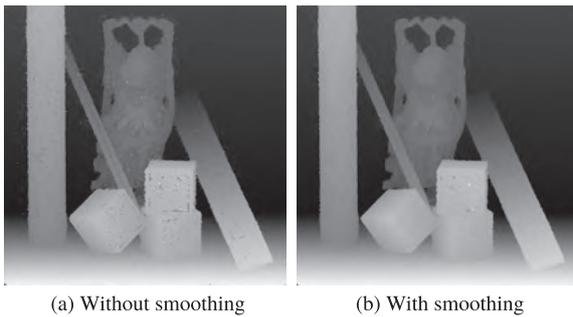


Fig. 7 Disparity maps with $\Delta = [-3, 5]$

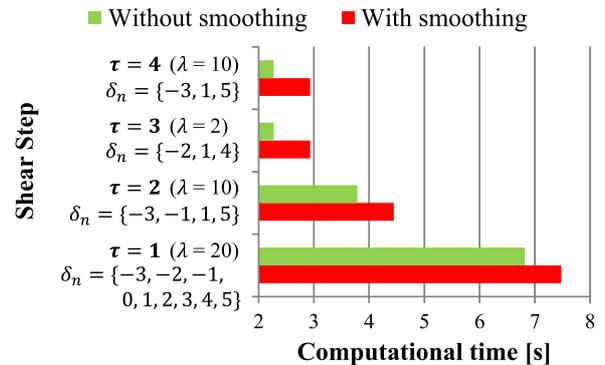


Fig. 10 Computational time with the shear steps τ

improves the accuracy. We can also see that the accuracy becomes low when the shear range Δ is much narrower than the target disparity range D . This indicates that the shear range Δ should cover the target disparities D to fully exploit the potential of the sheared EPI analysis. Meanwhile, making the shear range Δ wider than the target disparity range D also negatively affects the accuracy. However, this negative effect is greatly mitigated by using the δ -map smoothing. Therefore, if we use the δ -map smoothing, Δ does not need to be selected strictly. Several resulting disparity maps are presented in Figs. 6–8.

4.2 Effect of the Shear Step τ

Second, we examined the effect of the shear step τ on the performance of our method. Here, we fixed the shear range

Δ to $[-3, 5]$ and varied the shear step τ , where the smoothing strength λ is optimized for each condition. The accuracy and computational time were evaluated and are presented in Figs. 9 and 10, respectively. As expected, increasing the shear step τ results in a lower accuracy. However, decreasing the shear step τ results in more candidate shear values, which leads to a longer computational time. Meanwhile, the computational time for the δ -map smoothing is moderate and independent of the shear step τ . Therefore, in terms of the trade-off between the accuracy and the computational time, using δ -map smoothing is often more beneficial than using a smaller shear step τ .

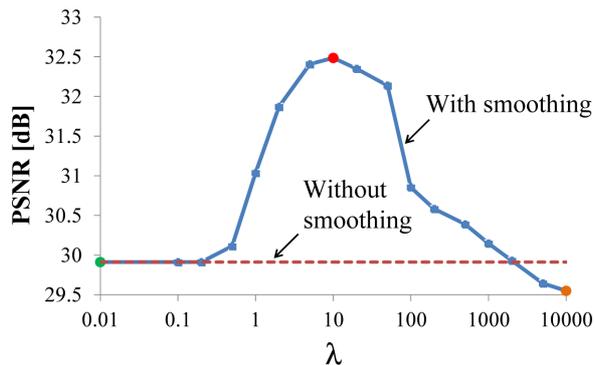


Fig. 11 PSNR and smoothing strength λ

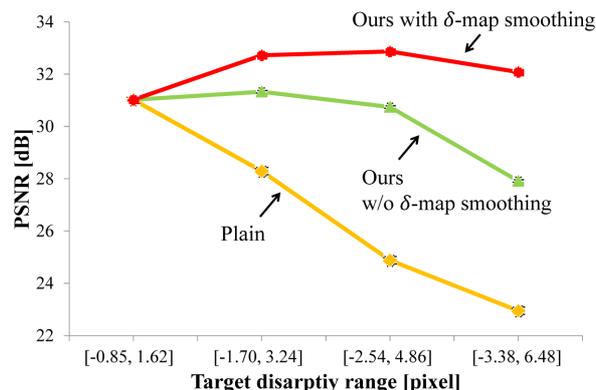


Fig. 14 PSNR and the target disparity range

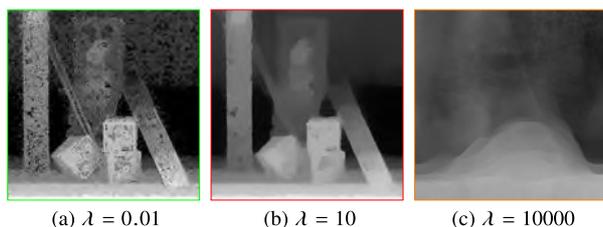


Fig. 12 δ maps with different smoothing strengths λ

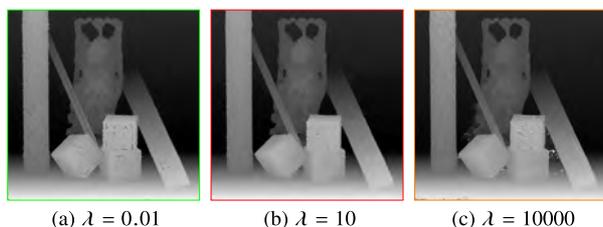


Fig. 13 Disparity maps with different smoothing strengths λ

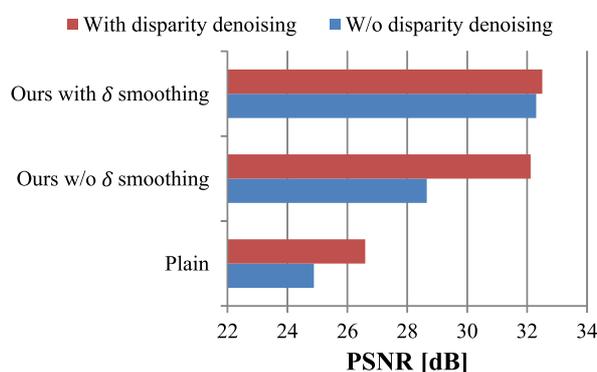


Fig. 15 Effect of disparity-map denoising

4.3 Effect of the Smoothing Strength λ

To see how the smoothing strength λ affects the performance of our method, we show a performance curve along λ in Fig. 11, where Δ and τ were fixed to $[-3, 5]$ and 2, respectively. Moreover, the δ maps and the disparity maps with $\lambda = 0.01, 10,$ and 10000 are represented in Figs. 12 and 13, respectively. As shown in those figures, the performance of our method depends greatly on the value of λ ; when λ is too small, the δ map is still noisy; meanwhile, when λ is too large, the information on the δ map is lost due to smoothing being too strong. In this case, the best performance was obtained with $\lambda = 10$.

4.4 Effect of the Target Disparity Range D

Next, we analyzed how the target disparity range D affects the performance. The disparity range was controlled by changing the viewpoint intervals of input images because the disparity range is proportional to the viewpoint interval. The viewpoint intervals were varied from 1 to 4. The

shear range Δ was changed in accordance with the viewpoint interval while the shear step τ was kept to 2. Therefore, as the target disparity range increased, the number of the candidate shear values also increased. Specifically, we used $\delta_n = \{0\}$ for the target range $D = [-0.85, 1.62]$, $\delta_n = \{-2, 0, 2, 4\}$ for $D = [-1.70, 3.24]$, $\delta_n = \{-3, -1, 1, 3, 5\}$ for $D = [-2.54, 4.86]$, and $\delta_n = \{-4, -2, 0, 2, 4, 6\}$ for $D = [-3.38, 6.48]$.

We compared our method with the original EPI analysis [20], [22] denoted as “plain,” and our method without δ -map smoothing in Fig. 14. The plain EPI analysis significantly deteriorated as the disparity range increased, prevention of which was the original motivation of our study. In contrast, our method can maintain or even increase the accuracy with a large disparity range, especially when combined with δ -map smoothing.

4.5 Effect of Disparity-Map Denoising

Now, the disparity-map denoising given by Eqs. (6)–(8) was applied to the bare outputs of plain/sheared EPI analysis. The results are presented in Fig. 15, where the denoising improved all the results. Especially, our method without δ -map smoothing was significantly improved by the denoising but it was still below our method with δ -map smoothing. Consequently, we can conclude that our method should be combined with both the δ -map smoothing and disparity-map de-

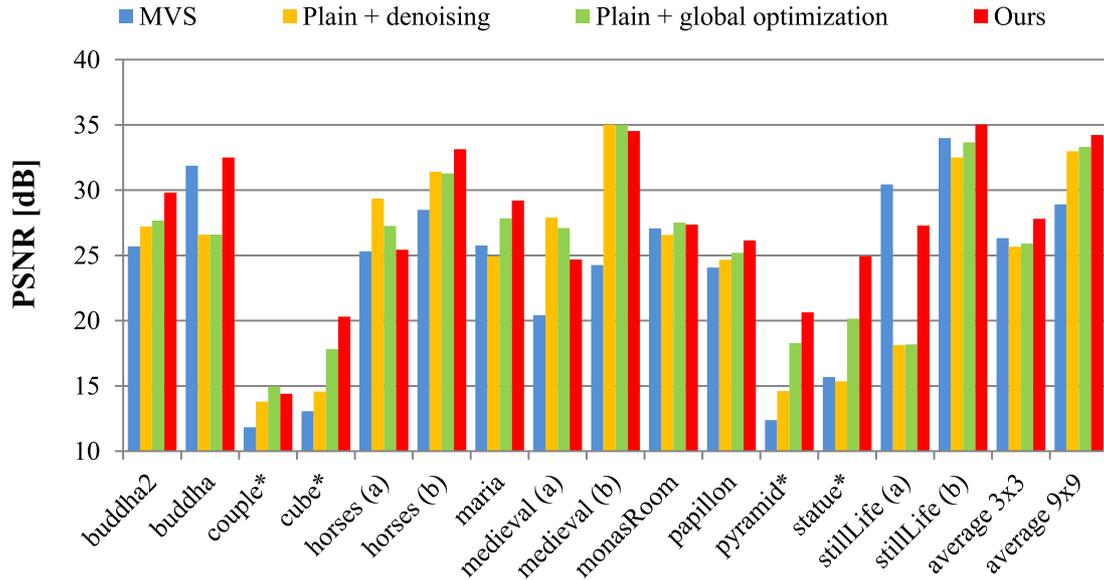


Fig. 16 Disparity map accuracy over 15 datasets

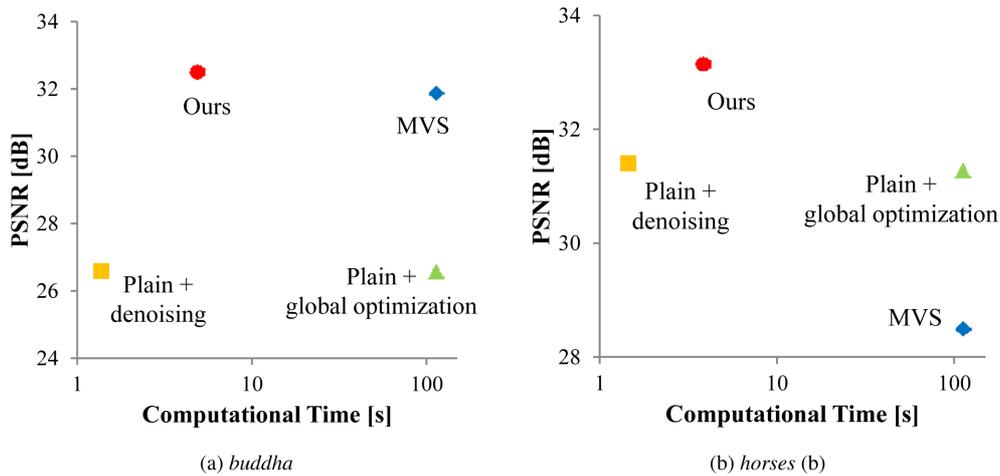


Fig. 17 The trade-off between the accuracy and computational time

noising.

4.6 Comparison with Other Methods

We compared our method with three other methods as follows. The first one, “plain + denoising”, is the original EPI analysis without shear transform combined with the disparity-map denoising. The second, “plain + global optimization”, is similar to the first one but combined with a more sophisticated global optimization. The above two are the proposed method in Wanner et.al. [22]. The third one, “MVS”, is a multi-view stereo, the detail of which is described as Eqs. (11)–(13) in [22]. This method is categorized as 4-D analysis methods mentioned in Sect. 2.2. The implementation of this method is also available online [26]. Our method was configured as follows. The shear step τ was fixed to 2 and the smoothing strength λ was fixed to 10.

Table 2 Shear range Δ for each dataset

Datasets	Δ	Datasets	Δ
buddha	$[-3, 5]$	medieval (b)	$[-3, 1]$
buddha2	$[-2, 4]$	monasRoom	$[-3, 3]$
couple	$[-2, 4]$	papillon	$[-4, 2]$
cube	$[-3, 3]$	pyramid	$[-3, 3]$
horses (a)	$[-8, 2]$	statue	$[-2, 2]$
horses (b)	$[-3, 1]$	stillLife (a)	$[-8, 6]$
maria	$[-1, 1]$	stillLife (b)	$[-3, 3]$
medieval (a)	$[-8, 2]$		

The shear range Δ are summarized in Table 2. Both the δ -map smoothing and disparity-map denoising were enabled to show the best performance of our method.

The PSNR values of the estimated disparity maps against the ground truth are presented in Fig. 16. The average PSNR values for the datasets with 3×3 viewpoints and those with 9×9 viewpoints are also reported. For sev-

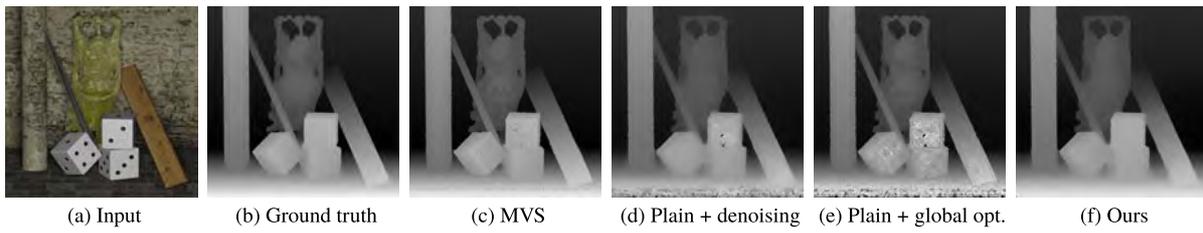


Fig. 18 Results from buddha dataset

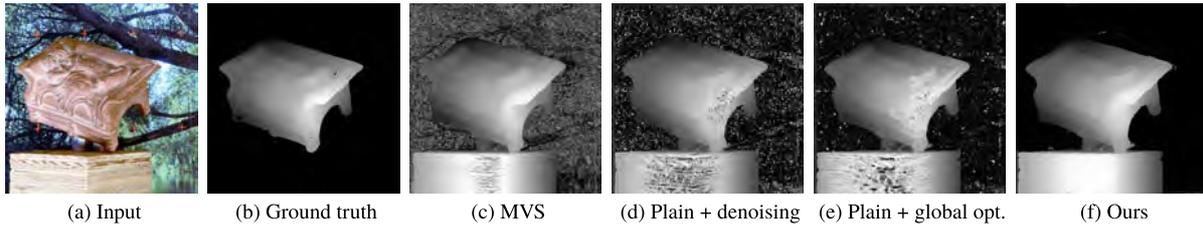


Fig. 19 Results from couple dataset

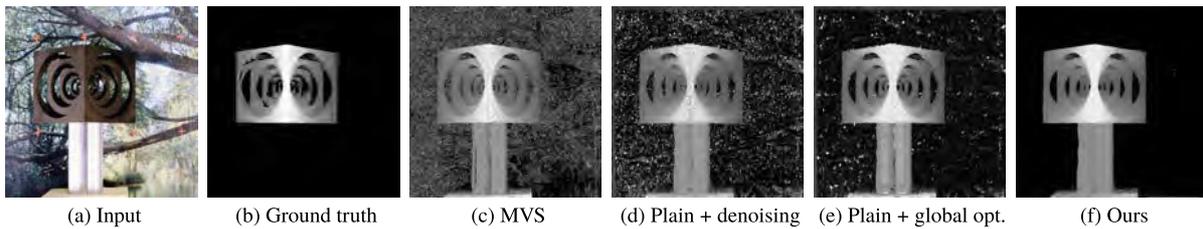


Fig. 20 Results from cube dataset

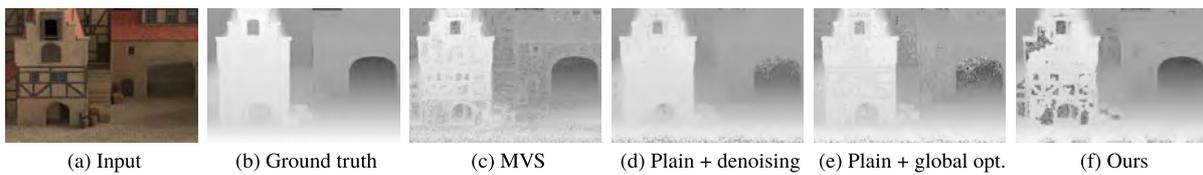


Fig. 21 Results from medieval (a) dataset

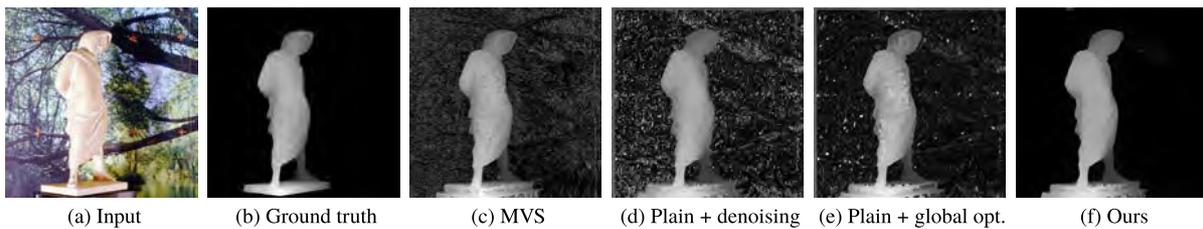


Fig. 22 Results from statue dataset

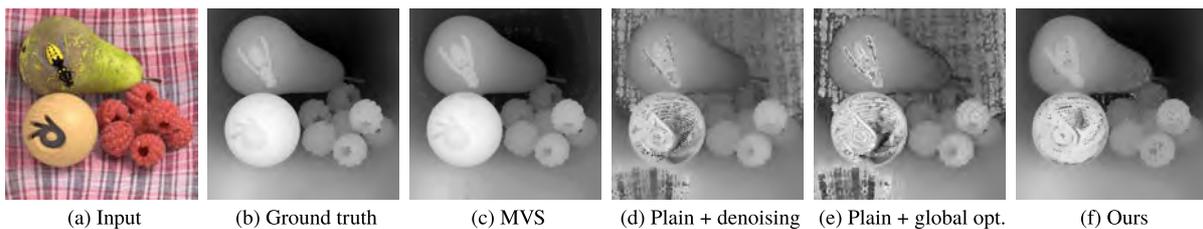


Fig. 23 Results from stillLife (a) dataset

eral datasets among them, the input images, the ground truth disparity maps, and the estimated disparity maps using the four methods mentioned above are shown in Figs. 18–23.

As can be seen from Fig. 16, our method consistently achieved better quality than the other methods in almost all the datasets, which shows the effectiveness of our method. The disparity maps produced by our method are also visually compelling as shown in Figs. 18–23. Note that for the datasets *couple*, *cube*, *maria*, *pyramid*, and *statue*, the ground truth disparities are obviously missing for the part of the stand; therefore, the PSNR values for those datasets are not trustworthy[†]. However, as seen from Figs. 19, 20, and 22, our method obviously achieved better quality than the other methods with those datasets.

We can also observe from Fig. 16 that the methods using EPI analysis (ours, plain + denoising, and plain + global optimization) performed better for the datasets with 9×9 viewpoints than for those with 3×3 viewpoints. This reflects the fact that the datasets with 9×9 viewpoints have smaller disparity ranges than those with 3×3 viewpoints. As the disparity range decreases, the advantage of our method over the other two methods reduces, as was also shown in Fig. 14.

Finally, we show the trade-off between the accuracy and computational cost of disparity estimation among the four methods, all of which were implemented using GPU. We used a desktop computer equipped with GeForce GTX 970. As clearly seen from Fig. 17, “plain + denoising” needed extremely short computation time thanks to the simple algorithm that works in the 2-D EPI domains combined with a fast denoising method. This simple algorithm was used as the building block of our method. The computation time required for our method increases as the target disparity range D increases because the computational cost is proportional to the number of candidate shear values. The δ -map smoothing requires additional computation time. However, the total computation time for our method is still moderate, and at the cost of this increased computation, our method achieves better accuracy. Meanwhile, 2-D EPI analysis with the sophisticated optimization (plain + global optimization) and 4-D light field stereo matching (MVS) required significant amounts of time due to their complexities, but did not always yield accurate results.

5. Conclusion

Aiming to estimate an accurate disparity map even from a non-dense light field, we proposed sheared EPI analysis where EPIs are transformed with several shear values before the structure tensor analysis and the results of analysis are integrated into a final disparity map. We carefully examined how the parameters of our method affect the result to fully exploit the performance of our method. Moreover, we demonstrated the effectiveness of our method by compar-

ing our method with three other methods over 15 datasets. Experimental results proved that our method achieves significant accuracy especially when combined with the δ -map smoothing step. Moreover, using the simple and efficient EPI analysis as the building block, our method requires much less computational cost compared to the complex multi-view stereo method. In the future work, our method will be further improved by combining it with appropriate handling of occlusions [21]–[23] and weakly textured regions.

Acknowledgments

This work was supported by JSPS Kakenhi under Grant 15H05314.

References

- [1] D. Scharstein and R. Szeliski, “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.
- [2] H. Hirschmüller, “Accurate and efficient stereo processing by semi-global matching and mutual information,” *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.807–814, 2005.
- [3] H. Hirschmüller and D. Scharstein, “Evaluation of cost functions for stereo matching,” *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.1–8, 2007.
- [4] D. Scharstein and C. Pal, “Learning conditional random fields for stereo,” *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.1–8, 2007.
- [5] M. Bleyer, C. Rhemann, and C. Rother, “Patchmatch stereo – stereo matching with slanted support windows,” *British Machine Vision Conference*, pp.1–11, 2011.
- [6] C. Zhang and T. Chen, “A self-reconfigurable camera array,” *Proc. ACM SIGGRAPH 2004 Sketches*, p.151, 2004.
- [7] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, “Light field photography with a hand-held plenoptic camera,” *Stanford Tech Report*, vol.2, pp.1–11, 2005.
- [8] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, “High performance imaging using large camera arrays,” *ACM Trans. Graph.*, vol.24, no.3, pp.765–776, 2005.
- [9] Y. Taguchi, K. Takahashi, and T. Naemura, “Real-time all-in-focus video-based rendering using a network camera array,” *3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video*, pp.241–244, 2008.
- [10] K. Marwah, G. Wetzstein, Y. Bando, and R. Raskar, “Compressive light field photography using overcomplete dictionaries and optimized projections,” *ACM Trans. Graph.*, vol.32, no.4, 2013.
- [11] H.-G. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y.-W. Tai, and I.S. Kweon, “Accurate depth map estimation from a lenslet light field camera,” *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.1547–1555, 2015.
- [12] M. Okutomi and T. Kanade, “A multiple-baseline stereo,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.15, no.4, pp.353–363, 1993.
- [13] S. Kang, R. Szeliski, and J. Chai, “Handling occlusions in dense multi-view stereo,” *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.103–110, 2001.
- [14] C. Strecha, R. Fransens, and L.V. Gool, “Combined depth and outlier estimation in multi-view stereo,” *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.2394–2401, 2006.
- [15] S.M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, “A comparison and evaluation of multi-view stereo reconstruction algorithms,” *Proc. IEEE Conference on Computer Vision and Pattern*

[†]The missing parts are included in the PSNR values for these datasets because we treated the data as they were. However, these untrustworthy PSNR values, which are marked with *, are not included in the average PSNR values reported in Fig. 16.

- Recognition, pp.519–528, 2006.
- [16] N. Campbell, G. Vogiatzis, C. Hernández, and R. Cipolla, “Using multiple hypotheses to improve depth-maps for multi-view stereo,” *Proc. European Conference on Computer Vision*, pp.766–779, 2008.
- [17] R.C. Bolles, H.H. Baker, and D.H. Marimont, “Epipolar-plane image analysis: An approach to determining structure from motion,” *International Journal of Computer Vision*, vol.1, no.1, pp.7–55, 1987.
- [18] A. Criminisi, S.B. Kang, R. Swaminathan, R. Szeliski, and P. Anandan, “Extracting layers and analyzing their specular properties using epipolar-plane-image analysis,” *Computer Vision and Image Understanding*, vol.97, no.1, pp.51–85, 2005.
- [19] J. Berent and P. Dragotti, “Segmentation of epipolar-plane image volumes with occlusion and disocclusion competition,” *Proc. IEEE Workshop on Multimedia Signal Processing*, pp.182–185, 2006.
- [20] S. Wanner and B. Goldluecke, “Globally consistent depth labeling of 4D light fields,” *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.41–48, 2012.
- [21] M. Diebold and B. Goldluecke, “Epipolar plane image refocusing for improved depth estimation and occlusion handling,” *Annual Workshop on Vision, Modeling and Visualization*, pp.145–152, 2013.
- [22] S. Wanner and B. Goldluecke, “Variational light field analysis for disparity estimation and super-resolution,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.36, no.3, pp.606–619, 2014.
- [23] J. Li, M. Lu, and Z.-N. Li, “Continuous depth map reconstruction from light fields,” *IEEE Trans. Image Process.*, vol.24, no.11, pp.3257–3265, 2015.
- [24] I. Tosic and K. Berkner, “Light field scale-depth space transform for dense depth estimation,” *Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp.441–448, 2014.
- [25] T. Suzuki, K. Takahashi, and T. Fujii, “Disparity estimation from light fields using sheared EPI analysis,” *Proc. IEEE International Conference on Image Processing*, pp.1444–1448, 2016.
- [26] B. Goldluecke, “Cocolib / light field suite,” <http://sourceforge.net/p/cocolib/home/Home/>, 2016.
- [27] M.W. Tao, S. Hadap, J. Malik, and R. Ramamoorthi, “Depth from combining defocus and correspondence using light-field cameras,” *Proc. IEEE International Conference on Computer Vision*, pp.673–680, 2013.
- [28] M.-J. Kim, T.-H. Oh, and I.S. Kweon, “Cost-aware depth map estimation for lytro camera,” *Proc. IEEE International Conference on Image Processing*, pp.36–40, 2014.
- [29] T.-C. Wang, A.A. Efros, and R. Ramamoorthi, “Occlusion-aware depth estimation using light-field cameras,” *Proc. IEEE International Conference on Computer Vision*, pp.3487–3495, 2015.
- [30] S. Wanner, S. Meister, and B. Goldluecke, “Datasets and benchmarks for densely sampled 4D light fields,” *Annual Workshop on Vision, Modeling and Visualization*, pp.225–226, 2013.
- [31] Heidelberg Collaboratory for Image Processing, “Datasets and Benchmarks for Densely Sampled 4D Light Fields,” http://klimt.iwr.uni-heidelberg.de/HCI/Research/LightField/lf_benchmark.php, 2017.



Takahiro Suzuki received a B.E. degree in electrical and electronic engineering and information engineering from Nagoya University in 2015 and is studying for an M.E. degree in electrical engineering and computer science from Nagoya University, which he is due to finish in 2017. His research interests are depth estimation from a light field. After finishing his study, he will be working at Nippon Steel & Sumitomo Metal Corporation.



Keita Takahashi received his B.E., M.S., and Ph.D. degrees in information and communication engineering from the University of Tokyo, Japan, in 2001, 2003, and 2006. He was a project assistant professor at the University of Tokyo from 2006–2011 and was an assistant professor at the University of Electro-Communications, Japan from 2011–2013. He is currently an associate professor at Graduate School of Engineering, Nagoya University, Japan. His research interests include light field

processing, image-based rendering, computational photography, and 3D display. He is a member of IEEE SPS & CS, ITE, IPSJ, and IEICE.



Toshiaki Fujii received B.E., M.E., and Dr.E. degrees in electrical engineering from the University of Tokyo in 1990, 1992, and 1995. From 1995, he has been with the Graduate School of Engineering, Nagoya University apart from 2008 to 2010, when he was with the Graduate School of Science and Engineering, Tokyo Institute of Technology. He is currently a professor in the Graduate School of Engineering, Nagoya University. His current research interests include multi-dimensional signal processing,

multi-camera systems, multi-view video coding and transmission, free-viewpoint television, and their applications.