# CONSISTENCY CONSTRAINED RECONSTRUCTION OF DEPTH MAPS FROM EPIPOLAR PLANE IMAGES

*Ziling Huang* *, *Chia-Wen Lin* *, *Hao-Chiang Shao* **, *Xiangsheng Huang* #

* National Tsing Hua Univ., Taiwan ** Fu Jen Catholic Univ., Taiwan # Chinese Academy of Science, China

## ABSTRACT

In this paper, we propose a method of reconstructing the depth map of a set of multiview images from the epipolar plane images (EPIs) of multiview Images. Our method involves two steps: finding support points and estimating depth. First, we propose to include a consistency term and a smoothness term in the objective function for edge point detection, where the consistency term is used to identify edge points and the smoothness term is applied to mitigate false edge detection due to light density variations caused by viewpoint changes. Then, based on the detected edge points, a depth map can be estimated by solving a energy minimization problem, in which a line uniformness term and a matching error term are introduced to ensure the line traces estimated from EPIs for depth estimation match the colors of edge points well. The depths of non-edge points are then estimated by introducing an additional prior term. In order to speed up our algorithm, the depth estimation problem is aggregated by a winner-take-all strategy. Experiments show that our method outperforms the state-of-the-art schemes in reconstructing depth map with fine details.

***Index Terms***— Depth reconstruction, disparity estimation, epipolar plane images, multiview image processing.

## 1. INTRODUCTION

In recent years, light field (LF) images [1] [2] have drawn much attention as a powerful approach for disparity/depth estimation as LF multiview images provide additional information for disparity estimation. There are several ways to acquire LF images, such as camera arrays [3], lenslet arrays and coded aperture techniques [4]. In this paper, we use the dataset in [5] collected by mounting a consumer digital single-lens reflex (DLSR) camera on a motorized linear stage.

There are multiple approaches for disparity estimation, such as monocular approaches [6] which usually require additional prior information like object shapes, and binocular approaches [7] which find correspondences between the left-view and right-view images and can often lead to inaccurate depth estimation due to limited information. Another way is multi-view stereo matching [8] that can more easily find the correspondences between multi-view images since it can provide additional views of images. In this paper, we propose to reconstruct depth image from epipolar plane images (EPIs) as an extension of multi-view stereo matching [5, 9–13].

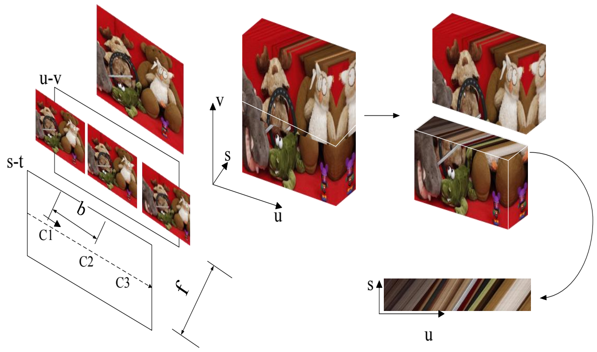

**Fig. 1**. Illustration of producing EPIs from multiview images, where the number of EPIs is equal to the height of the original image and the height of EPIs is equal to the number of original images.

For EPI image processing, the method proposed in [9] is among the first to extract EPIs to reconstruct depth maps. However, it is not robust for real-world scenarios due to light intensity variations, resulting in noisy and sparse depth maps. By dividing an EPI into several tubes, the method proposed in [10] can obtain a more compact representation of a 3D light field. The methods in [11] [12] calculate local structure tensors to estimate an initial depth map, then apply total variation optimization to construct more precise depth. These methods, however, consume high computation and memory costs, making it impractical for processing high-resolution images. The method in [12] filters out occluded pixels based on a bilateral metric on the surface metric. Although it can do a good job for occluded parts, it is constrained to small field view and cannot be applied to real scenes. The coarse-to-fine method in [5] finds support points in different levels and estimates their depths so as to estimate depth precisely. However, it consumes high computation cost and may not well reconstruct depth details in some areas.

Our method also reconstructs depth maps from EPIs which are captured densely along a linear path equidistantly as illustrated in Fig. 1, where $C_x$ $(x = 1, 2, ...)$ indicate the views from which the pictures are taken. Then, the pictures are piled up in order along $s$. We randomly select a fixed

978-1-5386-4658-8/18/$31.00 ©2019 IEEE

point on line $v$, denoted by $v^*$, then the $s-u$ plane constitutes an EPI image. As show in (1), the depth of every pixel corresponds to the slope of liner trace in the EPI. By estimating the slope (i.e., the disparity) $d$ of linear trace, we can calculate the depth $z$ for each pixel by

$$z = \frac{f * b}{d} \tag{1}$$

where $f$ is the camera focal length, $d$ is the disparity between a pair of adjacent images, and $b$ is the metric distance for each pair of adjacent images. Because $b$ and $f$ are already known, our main task is to estimate $d$ (i.e., the slope of a line trace).

For convenience, we denote a light ray as $\mathbf{r} = L(u, v, s)$, where $v$ indexes the EPI images and $(u, s)$ represents a point in an EPI image.

## 2. DEPTH RECONSTRUCTION METHOD

We propose a new approach to reconstruct a depth map from EPI images, which are produced by sampling a densely captured multiview image cube along a line (say, the center line) as shown in Fig. 1. Our method involves two steps: finding support points and estimating depth. To find support points, we propose a cross detector constrained with a smoothness term to reliably detect edge pixels as support points and avoid wrong detections due to light intensity variations at the same corresponding points between various view-point images. For depth estimation, we first estimate the depth on the detected support points by using a photo-consistency term and color-entropy. Then, based on the estimated depths of nearest support points, we can estimate the depth values of non-support points accordingly.

### 2.1. Finding support points

As shown in Fig. 2, we propose a two-step scheme to identify support points. In the first step, we determine a support point by the cross detector. As can be observed from the EPIs in 2, there are two types of edges in EPIs: the horizontal edges and the vertical edges. The horizontal edges are also the edges appearing in the raw RGB images. In certain EPI images, the edge slope, however, can be too small to detect because horizontal edges in the RGB image may not be sharp enough. Therefore, it is hard to find the edges just solely using the horizontal edge detector. In order to address this problem, we utilize the following cross detector proposed in our previous work [13]:

$$C_{conf}^{i}(u,s) = \sum_{(u',s')\in(V(u,s)\cup H(u,s))} \left\| E(u,s) - E(u',s') \right\|^2 \tag{2}$$

The cross detection function is also called the confidence term, where superscript $i$ represents the $i$-th EPI, $V(u,s)$ and $H(u,s)$ are the vertical and horizontal neighborhoods of pixel $(u,s)$, respectively, and $E(u,s)$ represents the color of pixel $(u,s)$. For the cross detection results, we define matrix $C_{raw}$ with entries indicating these detected edge points:

$$C_{raw} = \begin{pmatrix} 0 & 1 & ... & 0 & 0 \\ 0 & 1 & ... & 1 & 1 \\ ... & ... & ... & ... & ... \\ 1 & 0 & ... & 0 & 0 \end{pmatrix} \tag{3}$$

where the $i$-th column represents the $i$-th pixel in an EPI, the $n$-th rows means the $n$-th EPI, and the binary values in the matrix signify whether there are edge pixels detected, where a value of 1 indicating an edge pixel, and 0 otherwise. Note, a single row in $C_{raw}$ represents whether there exist edges in the center line of an EPI.

Although we can determine edge points using the cross detector, there can still be some wrongly detected edge pixels caused by unavoidable light intensity variations, because the multi-view cameras take pictures from the same scene in different views making the received light intensity different. To mitigate this influence, we propose a new smoothness term. As shown in Fig. 2(a), true edges are continuous and usually can be successfully detected by the cross detector from serial EPIs. In contrast, should the edges be produced by light intensity variations, it is unlikely to consistently find the corresponding edges in the next EPIs. We therefore define the following smoothness term to filter out the wrong edges:

$$C_{\text{smooth}} = \underset{(u,s)\in(V_{i-1}=1\cup V_{i+1}=1)}{\kappa} \left\| E(u,s) - E(u^*,s^*) \right\|^2 \tag{4}$$

where $i$ is the index of EPIs.

Equation (4) means if there is an edge in the $i$-th EPI, we set an $1 \times 3$ smoothness window for the $(i+1)$-th EPI to calculate the degree of confidence with the edges in the $i$-th EPI. We set a threshold $\varepsilon$, if $C_{\text{smooth}} < \varepsilon$, then $\kappa(\cdot) = 1$, meaning that the edges in the $i$-th EPI exist confidently. Should edges in $C_{\text{smooth}}$ not satisfy (4), we set $\kappa(\cdot) = 0$.

### 2.2. Depth estimation for support and non-support points

After identifying support points, we first estimate the depths of support points, and then estimated the depths of the remaining pixels based on the depths of support points.

Before depth estimation, we select fix $s$ on EPI image, where $s$ means the $s$-th image to be reconstructed. The disparity $d_N$ of a 3D scene point between two images is quantized into $N$ levels, where $N = 256$. As a result, for a support point $(u^*, s^*)$ in an EPI image as shown in Fig. 3, we define the pixel set $R$ along the red line as a function of $d_N$:

$$R(u^*, d_N) = \{E(u^* + (s - s^*)\, d_N, s)\, | s = 1, 2, ...\} \tag{5}$$

There are 256 kinds of pixel set $R$, since the value of $d_N$ ranges from 0 to 255. In order to find the best-match $d_N$ for the support point, we define a photo inconsistency cost $D_{\text{incon}}$ as follows:

$$D_{\text{incon}} = \alpha D_{\text{line}} + \beta D_{\text{match}} + \gamma D_{\text{prior}} \tag{6}$$

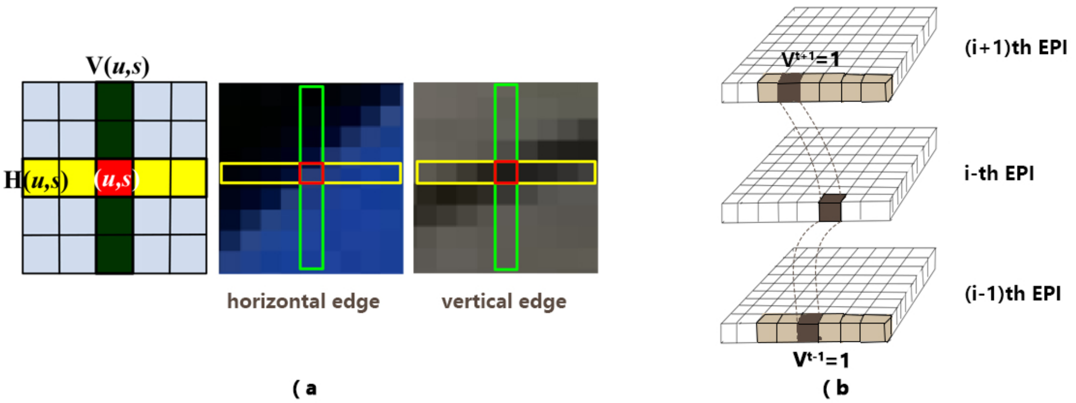

**Fig. 2**. Illustrations of finding support points using the cross corner detector. There may exist wrong support points due to light intensity variations caused by viewpoint change. For example, in EPI $i$, we can detect a support point; however, in EPI $i+1$, the support point is missing because EPI $i+1$ is not influenced by light in that view. In the original image, there exists no corner at all. In order to solve this problem, we add a smoothness term based on the assumption that all corner points' colors are continuous, as shown in (b).

where $D_{\text{line}}$ represents the line uniformness distortion, $D_{\text{match}}$ the line color-matching distortion, $D_{\text{prior}}$ the disparity estimation distortion for a non-support point based on the priors of its nearest support points, and $\alpha$, $\beta$, and $\gamma$ the weights for the three terms, respectively. In our method, the depth of a support point is estimated by minimizing the first two distortion terms, $D_{\text{line}}$ and $D_{\text{match}}$ (i.e., $\gamma = 0$), whereas the depth of a non-support points are estimated based on all the three terms.

The first term $D_{\text{line}}$ represents the line uniformness distortion. It means that in a single line trace in an EPI image, as illustrated in Fig. 3 all the corresponding pixels representing the same point in a 3D scene are taken from different viewpoints, making the color in the same light trace slightly different due to the viewpoint change. In order to make depth estimation more accurate, we treat all pixels in a candidate line trace $R$ in a whole rather than individual pixels. We define the $D_{\text{line}}$ term based on the entropy of the pixel values on line trace $R$ as follows:

$$D_{\text{line}} = -\sum_{(u,s)\in R} Pr\left\{E\left(u,s\right)\right\}\log Pr\left\{E\left(u,s\right)\right\}, \qquad (7)$$

which is used to measure the uniformness degree of pixels on line trace $R$. The more uniform the pixel colors in a line trace, the smaller $D_{\text{line}}$.

The second term $D_{\text{match}}$ is used to find the line that best matches a support point $E\left(u^*,s^*\right)$ in pixel color by calculating the color matching error between the support point and all the pixels in set $R$ as follows:

$$D_{\text{match}} = \varphi\left(\frac{\sum\limits_{(u,s)\in R}\left|E\left(u,s\right)-E\left(u^*,s^*\right)\right|}{|R|}\right), \qquad (8)$$

where

$$\varphi(s) = 1 - e^{-\frac{s^2}{2\sigma^2}}. \qquad (9)$$

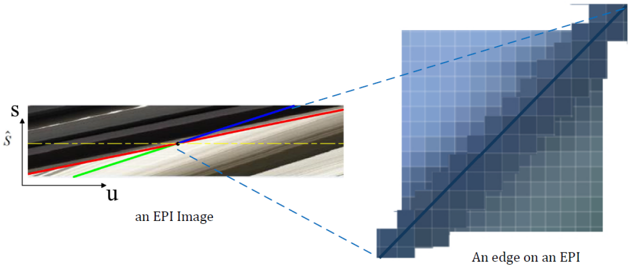

**Fig. 3**. Illustration of uniformness of pixel values along an edge.

As for the term $D_{\text{prior}}$, when we estimate the depth for a support point, we set $\gamma = 0$, meaning this term is ignored. After all the depths for the support points have been estimated, we estimate the location where depth changes gradually or remains unchanged. The depths of support points are horizontally propagated to the less detailed non-support points to obtain a dense reconstruction.

Denote the disparity in a non-support pixel with unknown depth as $d\left(u,s\right)$, assuming that $d\left(u,s\right)$ satisfies a Gaussian distribution centered at the estimated disparity $d_{est}\left(u,s\right)$ and variance $\epsilon$.

$$D_{\text{prior}} = -\log\left\{\phi + \exp\left(-\frac{\left(d-d_{est}\right)^2}{2\epsilon^2}\right)\right\}, \qquad (10)$$

where, $\phi$ is a positive value to avoid taking logarithm on zero.

In a region without any support points, its depth changes gradually or stays unchanged. Therefore for a pixel $E\left(u,s^*\right)$ with unknown depth, we find the two nearest support points $E_{left}\left(u_{left},s^*\right)$ and $E_{right}\left(u_{right},s^*\right)$, and linearly interpolate the expected disparity $E\left(u,s^*\right)$ as follows:

$$\mu(u,s^*) = \frac{u-u_{left}}{u_{right}-u_{left}}d\left(u_{right},s^*\right)+\frac{u_{right}-u}{u_{right}-u_{left}}d\left(u_{left},s^*\right) \qquad (11)$$

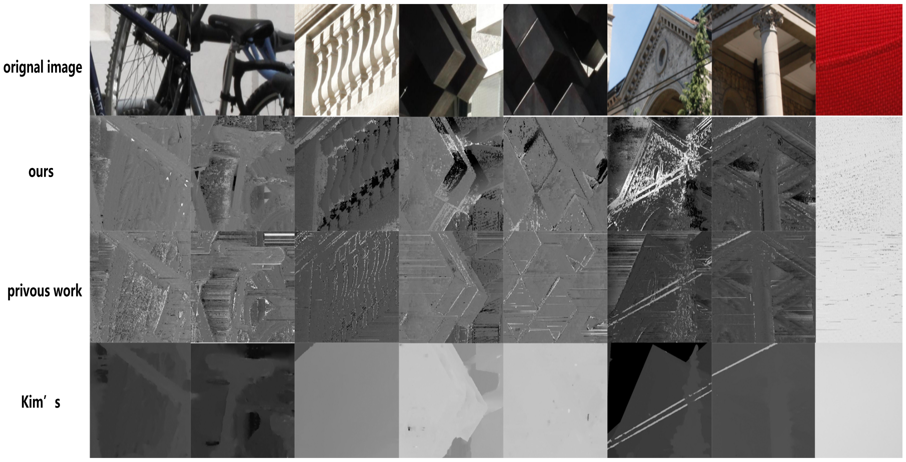

**Fig. 4**. Comparison of our algorithm with our previous work [13] and Kim et al.’s method [5]. Our method can well perverse more details.

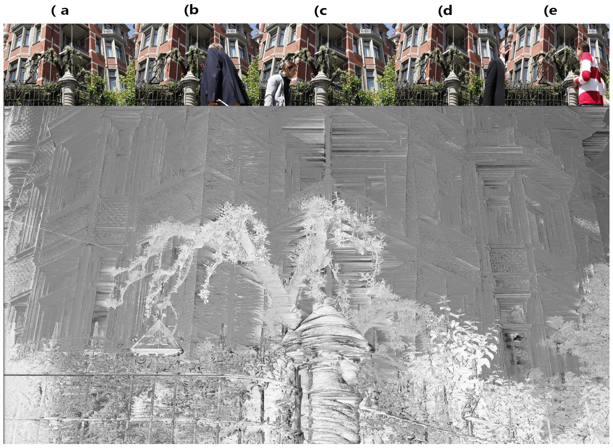

**Fig. 5**. Example of successful depth reconstruction even if there exist occlusions in the original multiview images.

In this step, we can search the disparity in a small range to save time because the depth changes gradually or stays unchanged for non-support points: $[d_{left} - \xi_0, d_{right} + \xi_0]$.

In total, we estimate the depth by minimizing the term $D_{\text{incon}}$. At first, we set $\gamma$ to zero and estimate the depths of support points. In this step, we use the winner-take-all strategy for computational efficiency. After all the depths in the support points are obtained, we can estimate the depths in less detailed regions.

## 3. EXPERIMENTAL RESULTS

In order to evaluate the performance of our method, we compare our method with our previous work [13] and the method proposed in [5], which are state-of-the-arts for reconstructing depth maps from EPIs, on the light field dataset proposed in [5].

We first compare the result in the whole. As shown in Fig. 4, where the first row shows the original images, the second, third, and fourth rows show the results obtained by our algorithm, our previous work [13] and the method proposed in [5], respectively. As show in the first column of Fig. 4, our algorithm can reconstruct the edges of tires and axles of a bike. However, our previous work [13] can just find edges of tires, whereas the result of method proposed in [5] shows blurry edges of tires. As shown in the third column of Fig. 4, we can find that for our algorithm, the edges of the statue are clear and their depths are different. The method proposed in [5] fails to reconstruct the edges and different depths. In the final column, there are small gaps in the sofa and their depth is different. The result shows that only our algorithm can successfully find these gaps and reconstruct their depth precisely.

Another challenging problem in depth reconstruction is that occlusion will influence the result of the depth map estimation. Fig. 5 shows that a number of images are occluded by pedestrians, which causes information loss. However, the result shows that our algorithm can reconstruct depth map precisely even if the original maps are occluded by other objects.

## 4. CONCLUSION

In this paper, we proposed a novel method to reconstruct depth images from EPIs. The proposed smoothness term can mitigate the influence of light variations so as to find true support points. Besides, we also proposed a uniformness term in the cost function to estimate pixelwise depth map. By using these two trick, we find the depth in the map is more reliable and the edge is more precise than the compared methods. Experimental results show that our method can reconstruct more accurate and detailed depth maps.

## 5. REFERENCES

[1] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, “The lumigraph,” in *Proc. ACM SIGGRAPH*, 1996.

[2] M. Levoy and P. Hanrahan, “Light field rendering,” in *Proc. ACM SIGGRAPH*, 1996.

[3] 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. Graphics*, vol. 24, pp. 765–776, 2005.

[4] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin, “Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing,” *ACM Trans. Graphics*, vol. 26, 2007.

[5] C. Kim, H. Zimmer, Y. Pritch, A. Sorkine-Hornung, and M. Gross, “Scene reconstruction from high spatio-angular resolution light fields,” *ACM Trans. Graphics*, pp. 1–12, 2013.

[6] J. Michels, A. Saxena, and A. Y. Ng, “High speed obstacle avoidance using monocular vision and reinforcement learning,” in *Proc. Int Conf. Mach. Learn.*, Bonn, Germany, Aug. 2005, pp. 593–600.

[7] K. N. Ogle, *Researches in binocular vision*, 1950.

[8] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, “A comparison and evaluation of multi-view stereo reconstruction algorithms,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, New Yotok, USA, June 2006.

[9] R. C. Bolles, H. H. Baker, and D. H. Marimont, “Epipolar-plane image analysis: An approach to determining structure from motion,” *Int. J. Comput. Vis.*, vol. 1, no. 1, pp. 7–55, 1987.

[10] A. Criminisi, S. B. Kang, R. Swaminathan, R. Szeliski, and P. Anandan, “Extracting layers and analyzing their specular properties using epipolar-plane-image analysis,” *Comput. Vis. Image Understand.*, vol. 97, pp. 51–55, 2005.

[11] S. Wanner and B. Goldlueck, “Variational light field analysis for disparity estimation and super-resolution,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 3, pp. 606–619, Mar 2014.

[12] S. Wanner, C. Straehle, and B. Goldluecke, “Globally consistent multi-label assignment on the ray space of 4d light fields,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Portland, OR, USA, June 2013.

[13] Z. Huang and X. Huang, “A semi-global matching method for large-sacle light field images,” in *Proc. IEEE Conf. Acoust. Speech Signal Proc.*, Shanghai, China, Mar. 2016.