SEGMENTATION OF LIGHT FIELD IMAGE WITH THE STRUCTURE TENSOR

Hao Sheng, Senyou Deng, Shuo Zhang, Chao Li, Zhang Xiong

State Key Laboratory of Software Development Environment, School of Computer Science and Engineering, Beihang University, Beijing 100191, P.R. China

ABSTRACT

We propose a segmentation model for light field images based on superpixels segmentation and graph-cuts algorithm. Unlike traditional images, which do not offer information for different directions, a light field image encodes space data which can be computed on its epipolar plane images (EPI) with some effective methods. In our work, we analyze the structure of EPI and research the computational process of disparity using EPI. On this basis, we present a new method for computing disparity using the modified structure tensor on EPIs. We further apply the computed disparity labels by fusing RGB images and disparity labels to obtain more detailed over-segmentation. Meanwhile, the modified structure tensor algorithm is used to get more accurate image boundaries, which plays a role in computing disparity features. All these processes are applied in an interactive segmentation model. Our experiments on public data sets demonstrate that the proposed light field image segmentation achieves a higher performance compared with state-of-the-art methods.

Index Terms— Light field, disparity, the structure tensor, image fusion, interactive segmentation

1. INTRODUCTION

With the marketization of the light field camera, e.g. Lytro and Raytrix, it is more convenient for people to gain the light field images [1],[2]. Because the light field images is accessed by the camera arrays [3] or the professional light field camera, it can capture the scene from different directions, and this particularity can bring additional information which allows new applications, e.g. image segmentation. As an important pre-processing procedure, image segmentation is a critical foundation of 3D modeling, image editing and object recognition [4]. However, the existing methods are either just appropriate for traditional image or very intricate in computing on light field images.

For light field images, due to the planar sampling, 3D points are projected onto lines in the light field called epi-polar-plane images. In recent researches, it is shown that robust disparity reconstruction is possible by analyzing this line structure [5],[6]. In contrast to traditional stereo matching, no correspondence search is required, and floating-point precision disparity data can be reconstructed at a very small cost. So Wanner *et al.* performed an interactive segmentation on 4D light fields [7]. They achieved a large improvement than the results on single view by using ray space features in the multi-label assignment process. In their work, they not only have access to the color of a pixel and information about the neighboring image texture but also assume that disparity is readily available as a possible additional feature in segmentation model.

Disparities turn out to be a highly effective feature for increasing the prediction quality of segmentation. Light fields are ideally suited for image segmentation. One reason is that geometry is an inherent characteristic of a light field, and thus we can use disparity as a very helpful additional feature. While this has already been realized in related work on multi- view co-segmentation [8],[9] or segmentation with depth or motion cues, which is in many aspects similar to disparity [10].

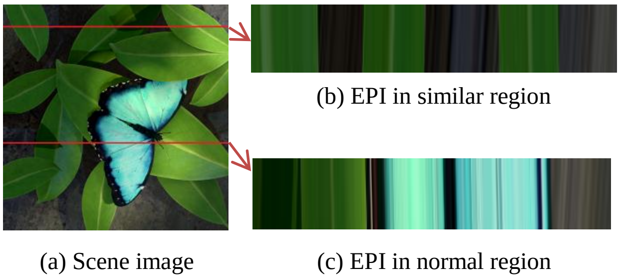


Figure 1. EPI in different regions will induce different computing results of EPI line. It is difficult to calculate the right slope in similar regions compared to normal regions.

Currently, there are some disparity estimation methods for light field images, such as stereo matching [11],[12] and the structure tensor & EPI based approaches [13],[14],[15]. All the disparity estimation methods encounter own problems. The main problem of the former is the correspondence of pixels. Though researchers have proposed some methods to improve stereo matching results, the problem is still unsolved. The other one avoids the matching problem partly, but there are some wrong disparity values which are beyond the reasonable scope, especially in the similar color regions. After analyses, it turns out that Wanner et al.'s simple computing formulations on EPI images are flawed. When the color value in a large region is nearly constant, it is difficult for the original formulations to find the right EPI line, see Figure 1. This problem can be solved by restricting the matching range, so we add a regular term on the original formulation. Meanwhile, this modification is reasonable because of the principle of structure tensor [16]. The modified formulation is precisely the eigenvector which is along the EPI line and indicates the tangent direction.

On the other hand, we obtain another eigenvector which is across the edge and it indicates the normal direction. Obviously, every eigenvector has its corresponding eigenvalue, so we get two eigenvalues by modified formulations. Then we can use the relation between the two eigenvalues to acquire the image boundary. The boundary information can be used in computing disparity features. Thus, we have additional segmentation features and experimental results show that it improves the segmentation accuracy in some degree.

After modifying the formulation, we get a more accurate disparity label of light field images, and the label includes the space information of objects. In traditional methods, researchers only use RGB information to complete the image over-segmentation, now we segment the disparity label obtained by our modified structure tensor & EPI method into superpixels. In superpixel level, the RGB superpixels and the disparity superpixels are merged together, called combined superpixels which include color, texture information and space information. By using combined superpixels, we can get more detailed over-segmentation results to avoid wrongly and missing classifying. At last, we perform an interactive segmentation model on light field images. Our model mainly implements a multi-label assignment graph-cut based on superpixels. With the disparity label and image boundary well applied, our model achieves comparable results to state-of-the-art methods.

2. DISPARITY LABELS

To obtain better segmentation results in light field images, the computation and utilization of disparity is one of keys. Wanner *et al.* acquired globally consistent depth labels of 4D light fields with the structure tensor based on the EPI method. Meanwhile, they also have proved that the intensity of the light field should not change along the EPI line, provided that the objects in the scene are Lambertian. Thus, if we want to estimate the depth for a point, we can try to find the slope of EPI lines in slices corresponding to points. On a basis of this proof, we propose a modified computing method with structure tensor on EPI. Then we can use the disparity labels in next step.

2.1. The modified structure tensor

In order to remove the adverse effect caused by unreasonable disparity values, we improved this kind of computing method by modifying the computing formulations of structure tensor. While the structure tensor we used is defined as:

$$T_\sigma = \begin{bmatrix} g_x^2 * G_\sigma & g_x g_y * G_\sigma \\ g_y g_x * G_\sigma & g_y^2 * G_\sigma \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ T_{12} & T_{22} \end{bmatrix}, \quad (1)$$

where G_σ is a Gaussian function with standard deviation σ, g_x and g_y are horizontal and vertical components of the gradient vector at each pixel respectively. Since matrix T_σ is symmetric and positive semi-definite, it has two orthogonal eigenvectors as follows:

$$\mathrm{V} = \begin{pmatrix} T_{22} - T_{11} + \sqrt{(T_{22} - T_{11})^2 + 4T_{12}^2} \\ -2T_{12} \end{pmatrix}, \quad (2)$$

$$V^\perp = \begin{pmatrix} 2T_{12} \\ T_{22} - T_{11} + \sqrt{(T_{22} - T_{11})^2 + 4T_{12}^2} \end{pmatrix}, \quad (3)$$

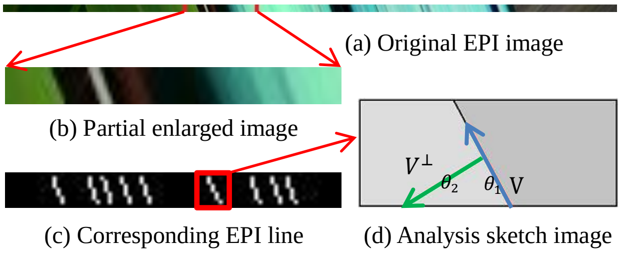
The relationship of the above two orthogonal eigenvectors with the EPI line is shown in Figure 2. According to the figure, we just need arbitrary one of the two eigenvectors to compute the incline angle of the EPI line because the intersection angle of the two eigenvectors with x-positive axis is complementary, like θ_1 and θ_2. Here we can compute the slope of the EPI line by using the intersection angle. Based on this inference, we choose V to complete the computation of disparity. In order to obtain visual disparity value, we swap T_{22} and T_{11} in (2). Comparing to the original formulation, we add a regular term ($\sqrt{(T_{22} - T_{11})^2 + 4T_{12}^2}$) to control the matching range.

The corresponding eigenvalues for each eigenvector are as follows:

$$\mathscr{v} = \frac{1}{2}\left(T_{22} + T_{11} - \sqrt{(T_{22} - T_{11})^2 + 4T_{12}^2}\right), \quad (4)$$

$$\mathscr{v}^\perp = \frac{1}{2}\left(T_{22} + T_{11} + \sqrt{(T_{22} - T_{11})^2 + 4T_{12}^2}\right), \quad (5)$$

Apparently the eigenvalue $\mathscr{v}$ is smaller than $\mathscr{v}^\perp$. Based on the two eigenvalues, edge structure can be determined as this inequation: $\mathscr{v}^\perp \gg \mathscr{v} \approx 0$. For edge points, the eigenvector V corresponding to the smaller eigenvalue is along the edge (tangent direction), while the eigenvector $V^\perp$ is across the edge (normal direction). Now we apply the edges inequation to obtain a better boundary image for light field images. For ease of calculation, we propose a reasonable deformation on the inequation like this: $\mathscr{v}^\perp - \mathscr{v} \geq \mathscr{p}_1$ && $\mathscr{v} \leq \mathscr{p}_2$, where $\mathscr{p}_1$ and $\mathscr{p}_2$ are control thresholds. While we can get the values of $\mathscr{v}^\perp$ and $\mathscr{v}$ for each pixel, and only when the values meet the above inequations, the pixel is viewed as a boundary point. For all pixels, we can get a boundary image after this procedure.


(a) Original EPI image

(b) Partial enlarged image

(c) Corresponding EPI line

(d) Analysis sketch image

Figure 2. The principle of using eigenvector to compute the slope of EPI lines, and then disparity labels can be computed. In (d), blue arrow is V, green arrow is $V^\perp$ and the black oblique line is one of EPI lines. The (c) is invisible in our model and the aim of the image here is just for facilitating understanding.

2.2. Disparity labels calculation

Based on the modified structure tensor and the principle of using slope to compute disparity that Wanner *et al.* have proved, we can get the computing formulation of disparity, as follows:

$$d = \frac{f}{Z} = \frac{\Delta x}{\Delta s} = \tan(\theta_1), \quad (6)$$

where Δx is the increment of a pixel in image, while Δs is the increment in space, and f is focal length, Z is depth distance.

By using the above formulation, we can acquire the disparity labels for each pixel in horizontal and vertical direction, respectively. Then, we can proceed to compute disparity label for the whole image by using the reliability formulation as follows:

$$\mathscr{r} = \frac{(T_{22} - T_{11})^2 + 4T_{12}^2}{(T_{22} + T_{11})^2}, \quad (7)$$

In order to increase the robustness of computing disparity labels map, we change the employment mechanism of $\mathscr{r}$ that if the two disparity values are both reasonable, we choose the one that corresponds to the higher reliability; if one of the two values is unreasonable then we directly choose the reasonable one instead of comparing reliability. By the proving experimentation, we show that neither of disparity values is unreasonable. Thus, this can also verify that our modified formulations can remove the unreasonable disparity values in disparity label. So the new disparity labels we computed have a higher credibility than the original one.

3. THE FUSION OF RGB AND DISPARITY LABELS

The traditional RGB images contain many visual features, such as color, texture, scale, etc. But these features may be very similar even the same in some special scenes, it induces many difficulties in image segmentation even though the segmentation is just regarded as a data preprocessing procedure. For example, in our over-segmentation step, we want to segment the image into superpixels with the SLIC [17] or Meanshift [18] method. The purpose is to reduce the amounts of pixels by viewing a superpixel as the smallest processing unit instead of a pixel. But with the difficulty caused by the similarity of simple RGB image in special regions, the over-segmentation results may have some non-ideal superpixels, such as a superpixel should be divided into two parts because the superpixel contains two objects' regions. This dividing procedure is difficult in RGB image because of the similarity. Thus, the importance of disparity labels image becomes obvious.

When a region has similar features in RGB, it can also be

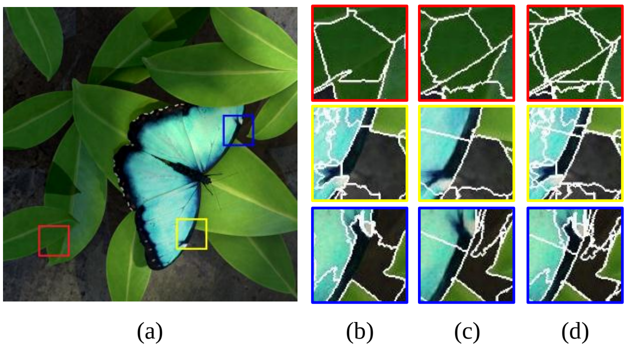


Figure 3. The superpixels obtained by over-segmentation. In order to clearly show the syncretic process, The superpixel size is set to 50. ((a) Original scene image; (b) RGB superpixels; (c) Disparity labels superpixels; (d) Combined superpixels)

segmented by the disparity labels information with the following formulation:

$$\mathcal{S} = \mathcal{C}\big(sp_i, i \in (1, \dots, n)\big) \cup \mathcal{D}\left(sp_j, j \in (1, \dots, m)\right), \qquad (8)$$

where $\mathcal{C}$ indicates color superpixels, $\mathcal{D}$ indicates disparity labels superpixels, $\mathcal{S}$ indicates combined superpixels and sp indicates a superpixel. m and n are the number of superpixels in disparity and color, respectively. This procedure can be seen as a union operation. So the segmentation results are more detailed, and the above problem obtained a better solution, see Figure 3. Each superpixel only contains a region in one object.

4. INTERACTIVE SEGMENTATION MODEL

To perform segmentation in light field images and verify the above theories, we propose an interactive segmentation model in this section. The model contains four steps: disparity estimation, the two-channel fusion of RGB and disparity labels, feature expression and segmentation labels assignment. The previous two steps are presented in above sections. The segmentation labels assignment step contains two substeps: local merging and global assignment. The former is similar to the greedy merging process in [19]. However, the merging rules and definition of superpixel similarity in our method is different, the latter exploits a pairwise MRF (Markov Random Field) to model the multi-label assignment problem.

4.1. Feature expression with image boundary

To perform light field image segmentation, an appropriate method for feature extraction or expression is necessary. While the better the extracting method is, the better segmentation results we can have. In section 2.1, we have defined the formulation for edges by using the structure tensor. After this procedure, we can get an image boundary mask map which contains some pixels computed as image boundary pixel. The image boundary can be used as additional information in computing disparity features.

In the feature expression step, we select five remarkable features based on superpixel and an image boundary feature. The five features are F_{sift}, F_{color}, $F_{disparity}$, $F_{orientation}$ and F_{size}. F_{sift} represents the texture features, which is obtained by SIFT (Scale-invariant Feature Transform) descriptor. F_{color} represents the color distribution using clustered color space. $F_{disparity}$ donates the distribution of disparity labels. $F_{orientation}$ donates the distribution of disparity gradient orientation. While F_{size} is a single number that equals to the proportion of a superpixel and F_{size} just work in local merging to control the merging range in case oversize superpixels.

The previous four features are varied in their form of expression, so we apply BOW (Bag of Words) method to calculate their own histogram. Then we can receive a multi-dimensional vector set, and each of them represents a corresponding superpixel. After this, we can use these multi-dimensional vectors to calculate the similarity between adjacent superpixels by using χ^2 distance of multi-dimensional vectors. The similarity value can be used in the next step.

4.2. Segmentation labels assignment

The last step contains two substeps: local merging and global assignment. The former performs a local merging for the most similar adjacent superpixel pair, for that too many superpixels may increase the data scale and reduce the efficiency of the next calculation. Additionally, most objects that users want to segment would not be too small. So we greedily and iteratively merge the adjacent pair that has the highest similarity until the value is smaller than threshold, or merge the superpixel whose size is smaller than threshold into its most similar neighbor.

In the latter substep, we formulate the estimation of superpixel assignment problem as a multi-label MRF optimization problem. In the optimization model, each superpixel is regarded as a node in MRF and edges is taken as the adjacent relationship. The energy function is defined as:

$$\begin{aligned} \mathrm{E}(L) &= \sum DC(i, l_i) + \alpha \sum SC(i,j), \qquad (i,j) \in N, \\ \mathrm{DC}(i, l_i) &= S_{unary}\big(sp_i, Region_{l_i}\big), \\ SC(i,j) &= S_{pair}\big(sp_i, sp_j\big), \end{aligned} \qquad (9)$$

where $\mathrm{L} = (l_1, .., l_n)$ is labels of all superpixels and n is the number of superpixels. l_i takes a value in $\{1,2,..,m\}$, the set of region labels are obtained by seed input image. N represents the set which contains all superpixel pairs. DC and SC measure the data cost and smooth cost in MRF. S_{unary} denotes the relationship between superpixel and region label and S_{pair} denotes two superpixels belong to one label. We use similarity value computed in section 4.1 to calculate DC and SC. At last, we use graph cuts [20] to complete the optimization process, which is proved to be an efficient method. Then we can get the final segmentation results of light field images by giving the parameter α(The value is 0.08).

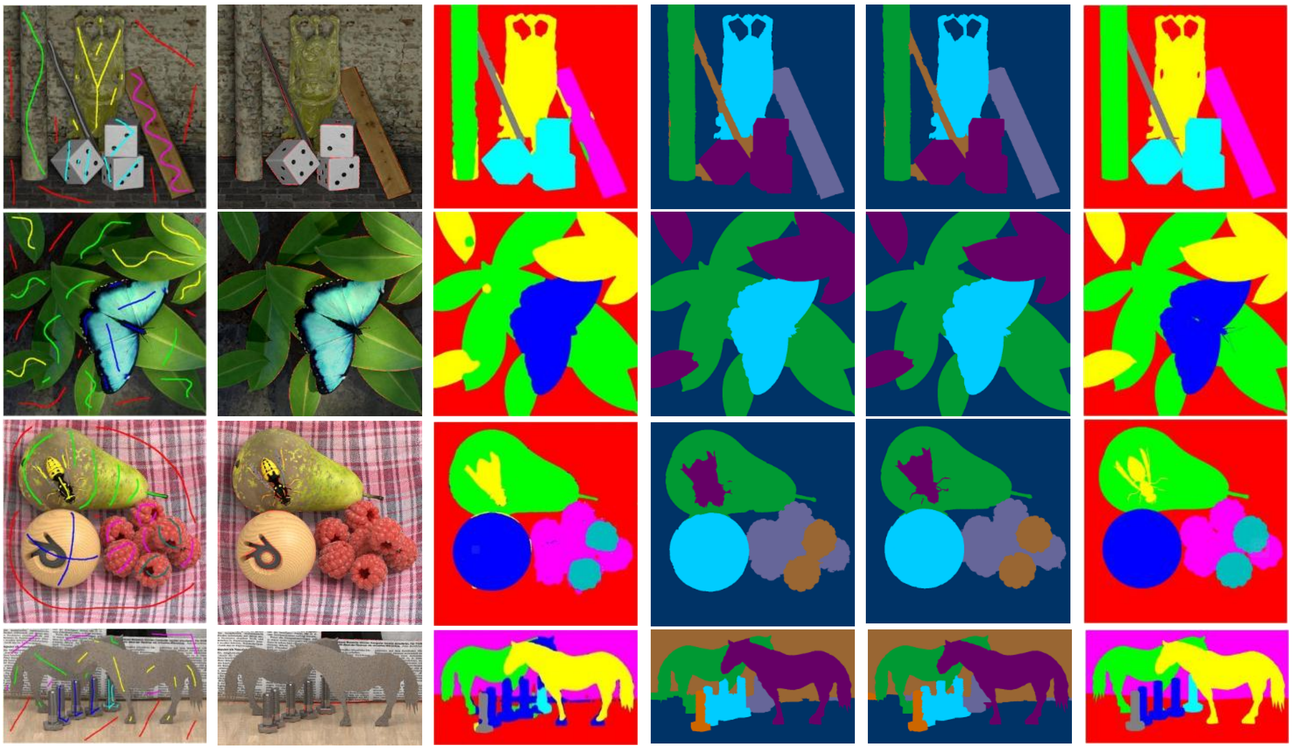
5. EXPERIMENT

In this section, we present the experiment results for our modified interactive segmentation model. We perform the experiment on the data sets that provided by Wanner *et al.* [21]. The data sets contain raw light field images, interactive seed images, ground truth of depth and segmentation labels.

Data Sets	Neither	+Boundary	+Fusion	Both
Buddha	95.9	**97.8**	**97.8**	**98.9**
Horses	93.8	**94.8**	**95.7**	**98.1**
Papillon	97.2	**98.3**	**98.8**	**99.1**
StillLife	96.8	**97.1**	**97.7**	**99.0**

Table 1. The evaluation of the usefulness of boundary information and the process of fusion. The table value donates the percentage(%) of accuracy. "+" represents only apply this information in segmentation model.

As is shown in Table 1, we firstly demonstrate the usefulness of the image boundary and the process of superpixels' fusion in section 2 and 3. Compared with the segmentation results without them, the segmentation operation only use arbitrary one of the two procedures brings stable improvement and the fusion operation is more effective. When both of them



(a) Seed image (b) Boundary image (c) Wanner *et al.*'s (d) Ours without (b) (e) Ours with (b) (f) Groundtruth label

Figure 4. Our segmentation results on different light field image data sets. (a) is the input seed image; (b) is the boundary image obtained by structure tensor; (c) is the contrast experiment results; (d) and (e) are our performance under different conditions while (e) is the best final segmentation results; (f) is the ground truth label of each data set.

are added to our model, we obtained the best performance.

Secondly, we compare our final segmentation results with the state-of-the-art results calculated by Wanner *et al.*[7] which are based on [13], as shown in the left column of Table 2. In Wanner *et al.*'s data, the second column has used the disparity features compared with the first column, so the accuracy rate is higher. In our segmentation process, we also used the disparity information as a valid segmentation feature and we have obtained a higher accuracy rate. In [22], Wanner *et al.* have proved that their model in [13] is far too slow to be useable in any practical application while our model just needs a few minutes. The visible segmentation results are exhibited in Figure 4. As shown in the picture, the segmentation results with the two procedures have an evident improvement on the edge of objects.

Data sets	Segmentation Accuracy			Segmentation Model(+GT)	
	Wanner *et al.*	Wanner *et al*(+D)	Ours (+D)	Wanner *et al.*	Ours
Buddha	96.3	98.8	**98.9**	**99.1**	99.0
Horses	94.8	97.7	**98.1**	**99.2**	98.6
Papillon	93.0	98.9	**99.1**	99.5	**99.7**
StillLife	99.0	98.9	**99.0**	**99.2**	99.1

Table 2. Comparison of our best performance and the usefulness of segmentation model with Wanner *et al.*.

At last, we compare our segmentation model to Wanner *et al.*'s ray space model by using the ground truth depth, which is shown in the right column of Table 2. Though our accuracy rate is lower than Wanner *et al.*, the difference is very small. The model of Wanner *et al.* has applied strong and complex optimization which is time-consuming [22]. On the contrary, our model is much more time-efficient even including the process of computing disparity labels. Thus, we propose a competitive model for light field image segmentation.

6. CONCLUSION

In this paper, we utilize the disparity labels computed during the modified EPI & structure tensor method in light field images to obtain a better segmentation result. Based on the analysis of the globally consistent depth labeling, we successfully recognize the unreasonable measured disparity values that induced by the ill-suited structure tensor formulations. So we modified the computing formulations to obtain a more accurate disparity label. Then the disparity label is segmented into superpixels in order to accomplish the fusion with RGB image. Meanwhile, we also acquired an image boundary by using modified formulations, and this new feature helps us to achieve better performance in our further interactive segmentation model. This model has remarkable performance in image segmentation. Besides, it is fairly efficient and flexible. But we believe that there is still a large space for improvement on our model by using stronger optimization algorithm. More importantly, our modified formulations and computed disparity labels may be widely applied in different studies of light field image process.

7. ACKNOWLEDGEMENTS

This study was partially supported by the National Natural Science Foundation of China (No. 61370122) and the National High Technology Research and Development Program of China (No. 2013AA01A603). Supported by the Programme of Introducing Talents of Discipline to Universities and the Open Fund of the State Key Laboratory of Software Development Environment under grant #SKLSDE-2015ZX-21. Thank you for the support from HAWKEYE Group.

8. REFERENCES

[1] R. Ng, M. Levoy, M. Bre ́dif, G. Duval, M. Horowitz, and P. Hanrahan, “Light Field Photography with a Hand-Held Plenoptic Camera,” Technical Report CSTR 2005-02, Stanford Univ., 2005.

[2] C. Perwass and L. Wietzke, “The Next Generation of Photography,” www.raytrix.de, 2010.

[3] V. Vaish, B. Wilburn, N. Joshi, and M. Levoy, “Using Plane + Parallax for Calibrating Dense Camera Arrays,” Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2004.

[4] Sheng H, Zhang S, Liu X, Xiong Z. “Relative location for light field saliency detection”[C]. Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on:IEEE, 2016.

[5] J. Berent and P. Dragotti. Segmentation of epipolar-plane image volumes with occlusion and disocclusion competition. In IEEE 8th Workshop on Multimedia Signal Processing, pages182–185, 2006.

[6] A. Criminisi, S. 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, 97(1):51–85, 2005.

[7] Sven Wanner, Christoph Straehle, and Bastian Goldluecke, “Globally consistent multi-label assignment on the ray space of 4d light fields,” in Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on.2013, pp. 1011–1018, IEEE.

[8] A. Kowdle, S. Sinha, and R. Szeliski. Multiple view object cosegmentation using appearance and stereo cues. In Proc. European Conference on Computer Vision, 2012.

[9] Mario Sormann, Christopher Zach, and Konrad F. Karner, “Graph cut based multiple view segmentation for 3d reconstruction,” in 3rd International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT 2006), 14-16 June 2006, Chapel Hill, North Carolina, USA, 2006, pp. 1085–1092.

[10] A. Stein, D. Hoiem, and M. Hebert. Learning to find object boundaries using motion cues. In Proc. International Conference on Computer Vision, 2007.

[11] Zhan Yu, Xinqing Guo, Haibing Ling, Andrew Lumsdaine, and Jingyi Yu, “Line assisted light field triangulation and stereo matching,” in Computer Vision (ICCV), IEEE International Conference on, 2013, pp.2792–2799.

[12] Sheng H, Zhang S, Zhu G, et al. “Guided integral filter for light field stereo matching”[C] IEEE International Conference on Image Processing. IEEE, 2015.

[13] Sven Wanner and Bastian Goldluecke, “Globally consistent depth labeling of 4d light fields,” in IEEE Conference on Computer Vision and Pattern Recognition(CVPR). IEEE, 2012, pp. 41–48.

[14] Li J, Li Z N. “Continuous Depth Map Reconstruction from Light Fields.”[J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2013, 24(11):1-6.

[15] Zhang S, Sheng H, Li C, et al. “Robust depth estimation for light field via spinning parallelogram operator”[J]. Computer Vision & Image Understanding, 2016, 145:148-159.

[16] Baghaie, Ahmadreza, and Z. Yu. “Structure tensor based image interpolation method.” AEU - International Journal of Electronics and Communications 69.2(2014):515–522.

[17] Radhakrishna, Achanta, et al. “SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. ” Pattern Analysis & Machine Intelligence IEEE Transactions on 34.11(2012):2274-2282.

[18] Comaniciu, Dorin, and P. Meer. “Mean shift: A robust approach toward feature space analysis.” IEEE Transactions on Pattern Analysis and Machine Intelligence 2002:603--619.

[19] Pekka Rantalankila, Juho Kannala, and Esa Rahtu,“Generating object segmentation proposals using global and local search,” in 2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, OH, USA, June 23-28, 2014, 2014, pp. 2417–2424.

[20] Yuri Y Boykov and M-P Jolly, “Interactive graph cuts for optimal boundary and region segmentation of objects in nd images,” in Computer Vision, 2001. ICCV 2001.Proceedings. Eighth IEEE International Conference on. vol. 1, pp. 105–112, IEEE.

[21] Sven Wanner, Stephan Meister, and Bastian Goldluecke, “Datasets and benchmarks for densely sampled 4d light fields,” in Vision, Modeling & Visualization. The Eurographics Association, 2013, pp. 225–226.

[22] Wanner S, Goldluecke B. Variational Light Field Analysis for Disparity Estimation and Super-Resolution.[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 36(3):606-19.