4D Light Field Superpixel and Segmentation*

Hao Zhu, Qi Zhang, Qing Wang
School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, P.R. China
qwang@nwpu.edu.cn

Abstract

Superpixel segmentation of 2D image has been widely used in many computer vision tasks. However, limited to the Gaussian imaging principle, there is not a thorough segmentation solution to the ambiguity in defocus and occlusion boundary areas. In this paper, we consider the essential element of image pixel, i.e., rays in the light space, and propose light field superpixel (LFSP) segmentation to eliminate the ambiguity. The LFSP is first defined mathematically and then a refocus-invariant metric named LFSP self-similarity is proposed to evaluate the segmentation performance. By building a clique system containing 80 neighbors in light field, a robust refocus-invariant LFSP segmentation algorithm is developed. Experimental results on both synthetic and real light field datasets demonstrate the advantages over the state-of-the-arts in terms of traditional evaluation metrics. Additionally the LFSP self-similarity evaluation under different light field refocus levels shows the refocus-invariance of the proposed algorithm.

1. Introduction

Superpixel segmentation is the key fundamental to connect pixel-based low-level vision to object-based high-level understanding, which aims at grouping similar pixels into larger and more meaningful regions to increase the accuracy and speed of post processing [21]. To accomplish a good over-segmentation, previous works [22, 8, 26, 12, 27, 1, 15] have built various grouping methods to model the proximity, similarity and good continuation [21] in the classical Gestalt theory [9]. However, in traditional imaging system (ideal pinhole model and thin lens model), there inevitably exist ambiguities in object boundaries where the light rays emitted from different objects are accumulated, including vignette, occlusions. These ambiguities may cause image degradation to disturb superpixel segmentation and further to decrease the accuracy of object segmentation and recognition.

*The work was supported in part by NSFC under Grant 61531014, Grant 61401359 and Grant 61272287.

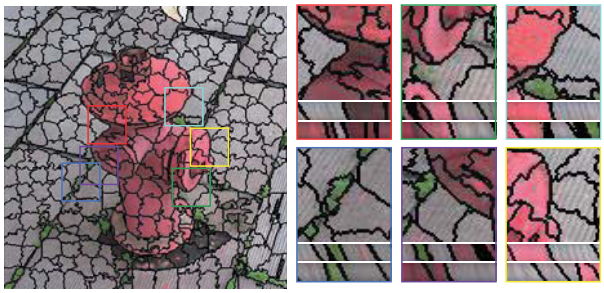


Figure 1. Light field superpixel segmentation on real scene data. The left image is a 2D slice of LFSP segmentation in the central view. For each region in the right images, the first row shows the close-up and the second and third rows are the corresponding segmentations on horizontal and vertical EPIs respectively.

To overcome the ambiguity in traditional superpixel segmentation, we introduce the light field superpixel segmentation. The light field [13] records the scene information both in angular and spatial spaces, forming a 4D function named $L(u, v, x, y)$. The light field data can benefit superpixel segmentation on two aspects. First, since each ray is recorded in light field, the ambiguity in object boundaries can be well analyzed. Second, the multi-view nature of the light field enables the bottom-up grouping not only in the color and position but also in the structure.

However, 4D light field segmentation is still a challenging task. As mentioned in [11], light field segmentation faces two major difficulties. First, each segmentation in light field ought to be propagated coherently to preserve the redundance of the 4D data. Second, although the depth is embedded in the multi-view images, it is still unavailable, inconvenient and imperfect to segment the full 4D data.

In this paper, we explore superpixel segmentation on 4D light field. We show that the LFSP can represent the proximity regions better, especially in object boundaries (in Section 3). The traditional superpixel is just a 2D slice of LFSP by fixing the angular dimensions. Additionally, the angular segmentation by fixing the spatial dimensions in LFSP coincides with the light field occlusion theory in [28]. On the basis of LFSP definition, we analyze the characteristics of

light field and propose the LFSP self-similarity to evaluate the segmentation result.

In Section 4, we define a clique system containing 80 neighbors in light field, and embed a 2D disparity map into the energy function to produce refocus-invariant LFSP segmentation. In Section 5, extensive experimental results are provided both on synthetic data and real scenes captured by Lytro [17]. Quantitative and qualitative comparisons show the effectiveness and robustness of the proposed algorithm.

The main contributions of the work include,

1) The definition of the light field superpixel.

2) A robust refocus-invariant superpixel segmentation algorithm in 4D light field.

2. Related Works

2.1. Light Field in Computer Vision

Unlike conventional imaging systems, light field cameras [17, 20] can record the intensity of objects in a higher angular dimension, and have benefited many problems in computer vision, such as depth and scene flow estimation [30, 28, 23], saliency detection [14], super resolution [3] and material recognition [29]. Light field can generate depth map [30, 24, 28] from multiple cues such as epipolar lines, defocus and correspondence. Compared with traditional multi-view based stereo matching methods, light field based methods can provide a high quality sub-pixel depth map, especially in occlusion boundaries. In this work, the algorithm developed by Wang *et al*. [28] is utilized to generate depth map for LFSP segmentation.

For light field segmentation, only a few of approaches have been proposed, especially most of them are interactive. Wanner *et al*. [32] proposed GCMLA (globally consistent multi-label assignment) for light field segmentation, where the color and disparity cues of input seeds are used to train a random forest, which is used to predict the label of each pixel. However, the method can only segment the central view of light field. Mihara *et al*. [19] improved the GCMLA by building a graph in the 4D space. A '4-neighboring system' in light field is defined and the 4D segmentation is optimized using the MRF. Hog *et al*. [10] exploited the light field redundancy in the ray space by defining free rays and ray bundles. A simplified graph-based light field is constructed, which greatly decreases the computational complexity. Xu *et al*. [34] segmented the 4D light field automatically. By defining the LF-linearity and occlusion detector in light field, a color and texture independent algorithm for transparent object segmentation is proposed.

Compared with previous segmentations, our work focuses on a smaller unit – the superpixel in light field, and it is the basis for many computer vision tasks [21, 14, 36, 4, 2].

2.2. Superpixel Segmentation

Superpixel segmentation of 2D image has been researched for years and many excellent algorithms have been proposed. Shi *et al*. [22] treated the image as a 2D graph using contour and texture cues. They proposed the normalized cuts to globally optimize the cost function. Felzenszwalb *et al*. [8] improved the efficiency of normalized cuts using an efficient graph cuts method. Liu *et al*. [16] introduced an entropy rate term and balance term into a clustering objective function to preserve jagged object boundaries. Achanta *et al*. [1] adapted a k-means clustering algorithm to seek the cluster centers iteratively. Li *et al*. [15] mapped the traditional color and position features into a higher spectral space to produce more compact and uniform superpixels.

All these works are built on the traditional 2D image and are not suitable for 4D LFSP segmentation. Although the 4D light field can be treated as a serial of 2D images and each image can be segmented using these algorithms, ignoring the connection between these images not only cuts off the segmentation consistency but also increases the running time (Fig.7(d)). In contrast to previous independent segmentation algorithms, we treat the 4D light field as a whole and improve the accuracy and running time of LFSP using the angular coherence in light field.

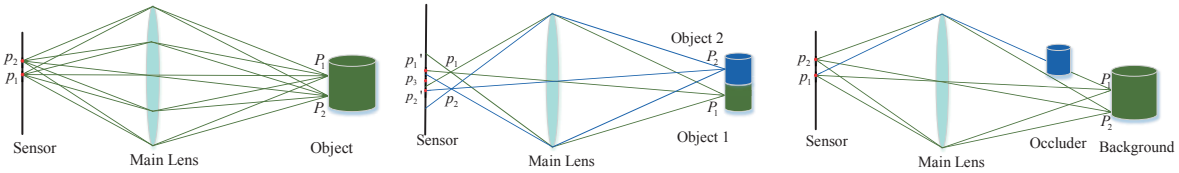
3. LFSP Definition and Evaluation Metric

In this section, we first present the definition of light field superpixel (LFSP). Then the difference and features of LFSP compared with traditional 2D superpixel are analyzed. Finally, we propose the evaluation metric of LFSP.

3.1. LFSP Definition

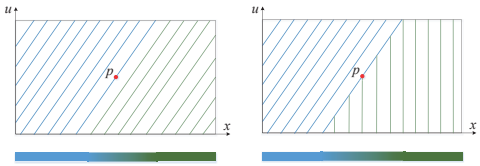
The superpixel algorithms model the proximity, similarity and continuation of the object in the 2D image. We ray-trace the points in the superpixel from the 2D image to the 3D space (see Fig.2(a)). In the propagation, each point spreads into multiple light rays and reaches the object in the real world. The LFSP contains all rays here.

The inverse propagation mentioned above can only model the all-in-focus and non-occlusion situations, however the following two conditions are hard to achieve actually. First, when the camera focusing on a different depth (Fig.2(b)), there exist defocus blurs on the sensor and the original clear boundary is blurred. Since the boundary pixel both suffers rays emitted from different objects, it is ambiguous to segment it. Second, for the occlusion points (Fig.2(c)), when the camera is focused on the background, a part of light rays emitted from the background point are occluded by the occluder so that the convergent point on the imaging sensor is a mixture of these rays – part from the background and part from the occluder, which makes it difficult to segment the pixel. The light field camera records



(a) Normal case (b) Defocus case (c) Occlusion case

Figure 2. (a) All rays emitted from P_1, P_2 to p_1, p_2 are contained in the LFSP. (b) The rays emitted from P_1, P_2 converge to p_1, p_2, forming two defocus areas centered at p_1', p_2' respectively. p_3 suffers from rays emitted from both P_1 and P_2. (c) There is an occluder between the background and the main lens, and part of the rays emitted from the green point P_1 are occluded by the blur occluder. There is an ambiguity in the segmentation for these mixed points here.



(a) Defocus case (b) Occlusion case

Figure 3. The top row shows the light ray intensity distributions in defocus and occlusion cases respectively in the EPI of light field. The bottom row shows the corresponding pixel intensity distributions in traditional 2D image.

all rays emitted from the real world such that the defocus and occlusion cases can be well segmented in the ray space using the LFSP.

Based on the above-mentioned analyses, we give the definition of LFSP as follows.

Definition 1. The LFSP is a light ray set which contains all rays emitted from a proximate, similar and continuous surface in the 3D space.

Mathematically, supposing R is a proximate, similar and continuous surface in the 3D space and the recorded light field is $L(u, v, x, y)$, the LFSP $s_R(u, v, x, y)$ is defined as,

$$s_R(u, v, x, y) = \bigcup_{i=1}^{|R|} L(u_{P_i}, v_{P_i}, x_{P_i}, y_{P_i}), \tag{1}$$

where $L(u_{P_i}, v_{P_i}, x_{P_i}, y_{P_i})$ is the recorded light field from i-th point P_i of the surface R. $|\cdot|$ denotes the number of elements in the set.

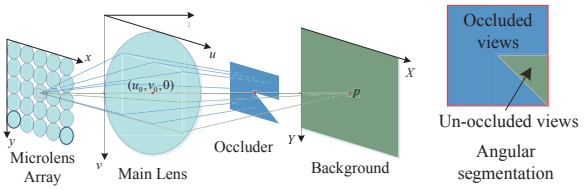
Ambiguity Elimination

The LFSP eliminates the defocus and occlusion ambiguity essentially. In Fig.3, the object boundary is blurred in traditional 2D image (the bottom row) since all rays are accumulated in a same point. However, since all rays are recorded in the light field, the object boundary are obviously in light ray space and can be well analyzed (the top row).

Limiting Cases

The definition above describes the general 4D LFSP and it can be reduced to 2D spatial or angular case by taking appropriate limits. First, considering fixing the angular dimensions $(u, v) \to (u^*, v^*)$, the 4D LFSP reduces to a 2D superpixel segmentation s^{u^*, v^*} in the view (u^*, v^*).

Then, if the spatial dimensions (x, y) are fixed, the 4D LFSP reduces to an angular segmentation. When the light field is refocused to the corresponding depth, this segmentation is a reference to determine the occlusion (see Fig.4). If all points in $s_R(u, v, x^*, y^*)$ share a same label, there is no occlusion here and all views can be used to improve depth estimation. If $s_R(u, v, x^*, y^*)$ is segmented into two or more regions, the views sharing the same label with the central view are the unoccluded views and others are occluded views. It coincides with the light field occlusion theory developed in [28].



(a) (b)

Figure 4. The limiting case when fixing the spatial dimensions. p is an occlusion boundary point. (a) The light field is refocused to the background, and only a few of views capture p. The green rays belong to the background LFSP and the blue rays belong to another LFSP. (b) The angular segmentation by fixing the spatial dimensions of p. It can be seen that the blue and green regions are occluded and unoccluded views respectively.

3.2. Evaluation metric

From the definition of LFSP, it is noticed each ray in the LFSP ought to be **refocus-invariant**, *i.e.* the label of each ray should be unchangeable during the refocus operation, since the point in 3D space is unchangeable and the light

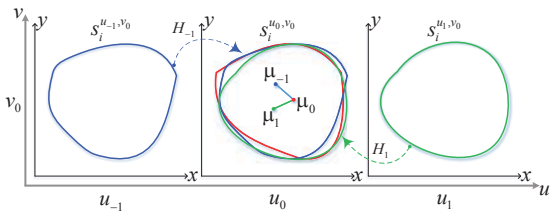


Figure 5. An illustration of the self-similarity. The 2D slice of i-th LFSP in the view (u_{-1}, v_0), (u_0, v_0) and (u_1, v_0) are labeled in blue, red, green respectively. Then $s_i^{u_{-1},v_0}$ and $s_i^{u_1,v_0}$ are projected to the central view, and μ_{-1}, μ_1 are the projected centers. μ_0 is the center of $s_i^{u_0,v_0}$.

field is unchanged.

Existing evaluation metrics for superpixel segmentation focus on the boundary adherence, such as the under-segmentation error (UE), boundary recall (BR) and achievable segmentation accuracy (ASA) [16]. There is no proper metric for the refocus-invariance. To measure the refocus-invariance, we propose the LFSP self-similarity.

The self-similarity of the i-th LFSP SS_i is defined as,

$$SS_i = \frac{1}{N_{uv}-1}\sum_{u,v}\left\|\mu_{H(s_i^{u,v},d,u,v,u_0,v_0)} - \mu_{s_i^{u_0,v_0}}\right\|_2, \quad (2)$$

where N_{uv} is the angular sampling number of light field. $s_i^{u,v}$ is the 2D slice of i-th LFSP in the view (u, v) and (u_0, v_0) is the central view of light field. μ_s denotes the position center of superpixel s and $H(s_i^{u,v}, d, u, v, u_0, v_0)$ projects the 2D superpixel $s_i^{u,v}$ from the view (u, v) to (u_0, v_0) according to the ground truth disparity map d. For each pixel $p = (u, v, x, y)^{\mathrm{T}} \in s_i^{u,v}$, the projected coordinate $p' = (u_0, v_0, x', y')^{\mathrm{T}}$ is defined as (in homogeneous coordinate)

$$\begin{pmatrix} u_0 \\ v_0 \\ x' \\ y' \\ 1 \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & 0 & 0 & 0 & u_0 \\ 0 & 0 & 0 & 0 & v_0 \\ -d(p) & 0 & 1 & 0 & u_0 d(p) \\ 0 & -d(p) & 0 & 1 & v_0 d(p) \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}}_{H(s_i^{u,v},d,u,v,u_0,v_0)} \begin{pmatrix} u \\ v \\ x \\ y \\ 1 \end{pmatrix} \quad (3)$$

We also give an intuitive explanation of the above definition. For a light field (Fig.5) with 1×3 angular resolution, the slices $s_i^{u_{-1},v_0}$ and $s_i^{u_1,v_0}$ of i-th LFSP are projected to the central view according to the ground truth disparity. The new centers of the projected $s_i^{u_{-1},v_0}$ and $s_i^{u_1,v_0}$ are denoted as μ_{-1} and μ_1 respectively, and the center of $s_i^{u_0,v_0}$ is μ_0. The mean of $\|\mu_1 - \mu_0\|_2$ and $\|\mu_{-1} - \mu_0\|_2$ is the self-similarity of the i-th LFSP.

For a full segmentation in the 4D light field, the LFSP self-similarity SS is defined as the mean of all SS_i,

$$SS = \frac{1}{K}\sum_{i=1}^{K} SS_i, \quad (4)$$

where K is the number of LFSP.

From the definition, the LFSP self-similarity is measured as pixel unit and a low SS value implies a high refocus-invariance. Apart from this, since the disparity changes with the refocus level, the LFSP self-similarity can measure the refocus-invariance of the LFSP segmentation accurately.

4. Refocus-invariant LFSP algorithm

The essence of refocusing is shearing the pixels in each view[24]. To make the LFSP refocus-invariant, the disparity should be removed in the position distance measurement.

Since the disparity map for full 4D light field is hard to estimated, in the proposed algorithm, a 2D disparity map d^{u_0,v_0} for the central view (u_0, v_0) of light field is obtained using [28]. To propagate the disparity from (u_0, v_0) to other views, the LFSP is modeled as a slanted plane in the disparity space. Supposing $\pi_i = (A_i, B_i, C_i)$ assigns a plane function to the i-th LFSP s_i, the disparity of each pixel $p = (u, v, x, y) \in s_i$ can be computed using

$$\hat{d}(p, \pi_i) = \frac{A_i x + B_i y + C_i}{1 + A_i(u - u_0) + B_i(v - v_0)}. \quad (5)$$

The detailed proof can be found at [6].

Then the energy function for 4D LFSP segmentation is defined as follows,

$$\begin{aligned} &E(s, \pi, o) \\ &= \sum_{u,v}\sum_{p}\left(E_c(p, s_{s(p)}^{u,v}) + \lambda_p E_p(p, s_{s(p)}^{u,v})\right) + \lambda_d \sum_{p} E_d(p, \pi_{s(p)}) \\ &\quad + \lambda_s \sum_{(i,j)\in N_{seg}} E_s(\pi_i, \pi_j, o_{i,j}) + \lambda_b \sum_{(p,q)\in N_{80}} E_b(s(p), s(q)), \end{aligned} \quad (6)$$

where s is the segmentation in full 4D light field and $s^{u,v}$ is the 2D slice of 4D LFSP in the view (u, v). $s(p)$ denotes the label that s assigns to pixel p. o records the connection types between two neighboring LFSPs.

In Eqn.(6), the terms E_c, E_p and E_d measure the color, position and disparity distance between the pixel p and the superpixel center respectively. The term E_s measures the connectivity between two LFSPs in the disparity space. The last but the most important, the term E_b measures the 2D slice shape and **the connectivity between each 2D slice superpixel** $s^{u,v}$, which is the core idea to make the LFSP refocus-invariant.

The color, position and disparity energy terms are defined as follows.

$$\begin{aligned} E_c(p, s_{s(p)}^{u,v}) &= \left\|L(p) - c_{s_{s(p)}^{u,v}}\right\|_2^2 \\ E_p(p, s_{s(p)}^{u,v}) &= \left\|p - \mu_{s_{s(p)}^{u,v}}\right\|_2^2 \\ E_d(p, \pi_{s(p)}) &= \left\|d^{u_0,v_0}(p) - \hat{d}(p, \pi_{s(p)})\right\|_2^2, \end{aligned} \quad (7)$$

where $c_{s_i^{u,v}}$ and $\mu_{s_i^{u,v}}$ denote the color and position centers of the 2D slice $s_i^{u,v}$ respectively. $L(p)$ denotes the color of

pixel p (the CIE-Lab color space is used here). The disparity term only works for the central view image.

The smoothness term encourages the slanted planes of neighboring LFSPs (N_{seg}) to be similar. Like the usage in [35], it contains three types of LFSP boundaries, *i.e.*, the occlusion, hinge and co-planar, and is defined as,

$$E_s(\pi_i, \pi_j, o_{i,j}) = \begin{cases} 0 & o_{i,j} = occ \\ \frac{1}{|\mathcal{B}_{i,j}|} \sum_{p \in \mathcal{B}_{i,j}} (\hat{d}(p, \pi_i) - \hat{d}(p, \pi_j))^2 & o_{i,j} = hi \\ \frac{1}{|s_i \cup s_j|} \sum_{p \in s_i \cup s_j} (\hat{d}(p, \pi_i) - \hat{d}(p, \pi_j))^2 & o_{i,j} = co, \end{cases} \tag{8}$$

where $\mathcal{B}_{i,j}$ is the set of boundary pixels between s_i and s_j.

There are two major functions in the boundary term, corresponding to two different types of neighboring system in light field, *i.e.*, the spatial and angular. Additionally these two types of neighboring systems are mixed to control the full shape of 4D LFSP. For a 4D point $p = (u, v, x, y)$ in the light field, supposing its disparity is $\hat{d}(p)$, there are 8 pixels in its spatial and angular neighboring system respectively,

$$\begin{aligned} N_{spa}(p) &= \begin{cases} (u, v, x \pm 1, y + 1) \\ (u, v, x \pm 1, y - 1) \\ (u, v, x, y \pm 1) \\ (u, v, x \pm 1, y) \end{cases} \\ N_{ang}(p) &= \begin{cases} (u \pm 1, v + 1, x \pm \hat{d}(p), y + \hat{d}(p)) \\ (u \pm 1, v - 1, x \pm \hat{d}(p), y - \hat{d}(p)) \\ (u \pm 1, v, x \pm \hat{d}(p), y) \\ (u, v \pm 1, x, y \pm \hat{d}(p)) \end{cases} . \end{aligned} \tag{9}$$

Apart from N_{spa} and N_{ang}, there is also a mixed neighboring system N_{mix} containing 64 points in both spatial and angular domains simultaneously (see supplementary material [6]). Fig.6 gives an illustration of these neighboring systems.

In total, there are 80 points (N_{80}) in p's neighboring system. The boundary term is defined as,

$$E_b(s(p), s(q)) = \begin{cases} 0 & s(p) = s(q) \\ E_{pen_s} & s(p) \neq s(q), N_{pq}\ is\ spatial \\ E_{pen_a} & s(p) \neq s(q), N_{pq}\ is\ angular \\ E_{pen_m} & s(p) \neq s(q), N_{pq}\ is\ mixed, \end{cases} \tag{10}$$

where the penalty E_{pen_s} in the spatial neighboring system encourages 2D slice $s^{u,v}_{s(p)}$ to be regular, preferring straight boundaries. The penalty E_{pen_a} in the angular neighboring system encourages the 2D slice of LFSP to be 'regular' in the epipolar plane, *i.e.* the pixels in a same epipolar line share the same LFSP label. It is the core to connect each 2D spatial slices of LFSP. Since the disparity is removed here, this term makes the LFSP to be refocus-invariant. The third penalty E_{pen_m} in the mixed system encourages the spatial 2D slice of the LFSP in other views to be regular.

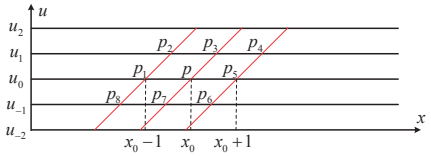


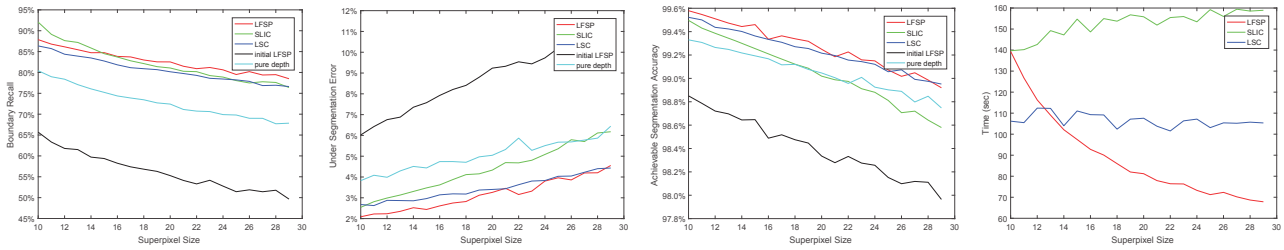
Figure 6. An illustration of neighboring system in light field. In this EPI expression (the red lines are epipolar lines), for a pixel $p = (u_0, x_0)$, p_1, p_5 are the spatial neighbors, p_3, p_7 are the angular neighbors, and p_2, p_4, p_6, p_8 are the mixed neighbors.

Reviewing the full energy function, the refocus-invariance is guaranteed since (1) the 2D slices of LFSP in different views are segmented independently just using the local 2D image information (E_c, E_p, E_d); and (2) the angular penalty (E_{pen_a}) in the boundary term encourages similar slices to connect together according to the disparity. The initial depth estimation always follows the step of light field refocusing and provides a good disparity map for building angular neighboring system. Based on these designs, the proposed LFSP algorithm is refocus-invariant.

The Block Coordinate Descent (BCD) algorithm [35] is used to solve the Eqn.(6). The full LFSP algorithm is described in the Algo.1. First, a 2D depth map d^{u_0,v_0} is obtained using [28]. Then an initial segmentation for 4D light field is obtained. Finally, the final result is optimized by minimizing the Eqn.(6). Since the BCD algorithm only guarantees to converge to a local optima, a good initial value is in need. First, an initial superpixel segmentation s^{u_0,v_0} of the central view is obtained by embedding the disparity map d^{u_0,v_0} into the SLIC framework. Then the position and disparity center of each superpixel $\mu_{s_i^{u_0,v_0}}$ and $\bar{d}_{s_i^{u_0,v_0}}$ are calculated and are used to project s^{u_0,v_0} to 4D light field by Eqn.(3) (lines 3-9). For each non-label pixel in other views, it is assigned as the nearest pixel's label (lines 10-14).

5. Experimental Results

We compare the proposed LFSP segmentation with the state-of-the-art superpixel segmentation algorithms including the SLIC [1] and LSC [15]. Noting that, the results of SLIC come from the vlfeat [25] library, and the code of LSC comes from the author's website. All three algorithms are evaluated both on synthetic data and real scene light field. For synthetic data, the HCI benchmark light field datasets [31] are used, which consist of 4 light fields with ground truth depth and segmentation. Each data includes a 9×9 (angular resolution) light field. The real scene data are captured by a consumer light field camera Lytro. The 4D light field data are extracted using the LFToolbox [7]. The quantitative evaluation contains the UE, BR, ASA [16], running time and LFSP self-similarity. All evaluations are conducted on the synthetic data since the ground truth disparity and segmentation are not available in real scene data, and



(a) BR (b) UE (c) ASA (d) Time

Figure 7. Quantitative evaluation of different superpixel segmentation algorithms.

Input: The 4D light field $L(u, v, x, y)$
Output: The 4D LFSP segmentation s.

$d^{u_0,v_0} = DepthEstimation(L)$
$s^{u_0,v_0} = SLIC(L(u_0, v_0, x, y), d^{u_0,v_0})$
for $i = 1$ **to** $|s^{u_0,v_0}|$ **do**
  $\mu_{s_i^{u_0,v_0}} = \frac{1}{|s_i^{u_0,v_0}|} \sum_{p \in s_i^{u_0,v_0}} p$
  $\bar{d}_{s_i^{u_0,v_0}} = \frac{1}{|s_i^{u_0,v_0}|} \sum_{p \in s_i^{u_0,v_0}} d^{u_0,v_0}(p)$
  for *each view* (u, v) **do**
    $s_i^{u,v} = H(s_i^{u_0,v_0}, \bar{d}_{s_i^{u_0,v_0}}, u_0, v_0, u, v)$
  end
end
for *each view* (u, v) **do**
  for *non-label pixel* p **do**
    $s(p) = \arg\min_{s(q)} \|p - q\|_2, q \in s^{u,v}$
  end
end
$E(s, \pi, o) = \sum_{u,v} \sum_p (E_c + \lambda_p E_p) + \lambda_d \sum_p E_d + \lambda_s \sum E_s + \lambda_b \sum E_b$
$s = \arg\min_s E(s, \pi, o)$

Algorithm 1: The LFSP segmentation algorithm.

there is no light field segmentation benchmark in real scene data like the classical Berkeley segmentation database [18].

Unless otherwise stated, the same parameters are set for all experiments, *i.e.*, $\lambda_p = 100$, $\lambda_d = 5$, $\lambda_s = 0.01$, $\lambda_b = 5$, $E_{pen_s} = E_{pen_m} = 1$ and $E_{pen_a} = 8$. λ_p balances the weights between the position and color distance, which is further divided by the superpixel size. A larger λ_p leads to a more well-shaped superpixel. λ_d measures the role of initial disparity. Since the state-of-the-art depth algorithms [30, 24, 28] always over-smooth the occlusion boundaries, it is not recommended to assign a large value to λ_d. λ_s controls slanted plane functions and it is mainly decided by the initial disparity map. Due to the same over-smoothing reason, it is suggested to assign a small value to adjust the plane function more stable. For the boundary term, we believe the angular consistency is more important than the spatial consistency, since the pixels in a same EPI line describe the same point instead of different points, and the (E_{pen_a}) ought to be assigned as a large value. For others (E_{pen_s}, E_{pen_m}), small values are assigned trying to encourage straight boundaries. λ_b balances the boundary adherence and the shape. The boundary adherence decreases with the increase of λ_b.

5.1. Synthetic Scenes

Fig.7 shows quantitative results which are average values in the HCI segmentation datasets. It can be seen that the proposed LFSP algorithm (red lines in Fig.7) shows competitive results over the state-of-the-art algorithms (green and blues lines in Fig.7) in all three traditional metrics. The qualitative results are shown in Fig.10 (see [6] for more results.), from which we can see that the LFSP segmentation can produce more regular superpixels in occlusion boundary areas.

Fig.7(d) shows the running time of different algorithms. Noting that, all algorithms are evaluated on the same desktop computer with a 3.4 GHz i7 CPU. Each view of light field contains 589824 pixels (768×768 or 576×1024). The time of SLIC and LSC are the sum of the algorithm conducted on each view image of light field. It can be seen that our un-optimized Matlab/C implementation shows great advantages over previous works with the increasing of superpixel size, since the light fields are treated as a whole instead of multiple independent images in the proposed LFSP algorithm, and the BCD algorithm just iteratively optimizes the boundary pixels in the LFSP segmentation.

Additionally, to evaluate the influence of initial value (the LFSP segmentation without BCD optimization) in the proposed algorithm, the BR, UE, ASA of initial segmentation are also plotted in Fig.7 (the black lines), showing the effectiveness of the proposed optimization function. In the forth column of Fig.10, the segmentation results in the 4D space are partly shown. For each local region, the first row shows the initial results and the second row shows the optimized results. It can be seen that the proposed LFSP optimization can correct the errors in initial value and preserve the occlusion boundaries well.

Since a 2D disparity map is contained, to evaluate its in-

fluence, we first segment the central view of light field using the disparity map (the cyan lines in Fig.7(a),7(b),7(c)) and compare it with the optimized segmentation. The proposed LFSP algorithm outperforms the results using disparity map only, since the disparities in occlusion boundaries are hard to estimate due to the under-sampling in the angular space [33], and the existed algorithms tend to over-smooth the occlusion boundaries [5, 28]. As a result, the influence of disparity maps with different qualities for the LFSP is worthy of research in the future work.

Apart from these traditional evaluation metrics (BR, UE, ASA), we also evaluate the LFSP using the LFSP self-similarity. Since there is no previous work on light field superpixel segmentation, it is hard and unfair to compare it with the traditional 2D superpixel segmentation. We refocus the light field for 4 times (the refocus levels $1 - \frac{1}{\alpha}$ are $0, 0.5, 1, 1.5$) and segment it. Then the LFSP self-similarity on each segmentation is plotted in Fig.8. It can be seen that the curves always maintain at a low level and all values are smaller than 1 pixel, which shows well refocus-invariance of the proposed LFSP algorithm. Additionally, the curves are very close to each other, implying the stableness of the algorithm.

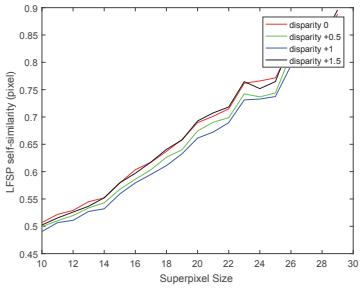
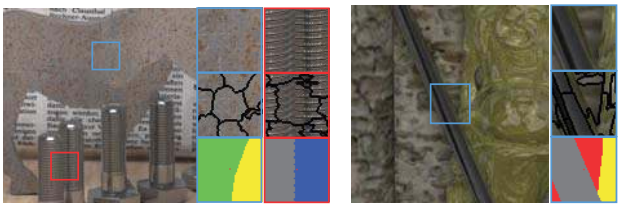


Figure 8. The LFSP self-similarity evaluations for the proposed LFSP algorithm at different refocus levels.

5.2. Real Scenes

Fig.1, 11 show experimental results on real scene light fields, captured by a Lytro camera (The superpixel size is set as 20 here). Due to the low SNR of Lytro camera, the SLIC and LSC can not produce reliable results from single image of central view . However, due to the introduction of angular neighboring system, the proposed LFSP algorithm can produce more convincing results. It can be seen that the 2D spatial slice of LFSP is more regular and has a better boundary adherence over SLIC and LSC. Apart from this, the occlusion boundaries in EPI space are also preserved well. The segmentation boundaries can always cling the occlusion boundaries or remain the same direction with EPI lines, showing a good LFSP self-similarity of the proposed LFSP algorithm. In Fig.12, the segmentation under different refocus levels $(-0.5, 0, 0.5)$ are demonstrated. Obviously, the occlusion boundaries can always be preserved well for different refocus levels.



(a) Similar and Non-lambertian areas (b) Non-Lambertian objects

Figure 9. Limitations. The upper insets the close-ups of blue/red rectangles, the middle inserts the segmentation, and the bottom show the ground truth labels. (a) In blue rectangle, there are two horses here, however it is hard to distinguish them due to the similar textures. In red rectangle, the boundaries of screws are hard to be distinguished due to the inter-reflection. (b) Part of metal tube reflects the background, showing similar textures with the background. It is difficult to connect this reflective region with others.

5.3. Limitations

The algorithm cannot handle situations where the background and foreground share similar textures, or non-Lambertian objects (see Fig.9). If the background and foreground share similar textures, it is difficult to segment them well using existing cues. If the objects do not satisfy the Lambertian assumption, the color or depth cues are not reliable so that the boundaries of these objects can not be segmented well. These two difficulties are also not solved well in traditional algorithms.

6. Conclusions and Future Work

In the paper, we first propose the definition of light field superpixel. The LFSP is defined in the 4D space and can essentially eliminate the defocus and occlusion ambiguities in traditional 2D superpixel. We then propose a refocus-invariant LFSP segmentation algorithm. By embedding the 2D disparity map into superpixel segmentation and a clique system with 80 (spatial, angular and mixed) neighbors in the 4D light field, the proposed algorithm outperforms the state-of-the-arts in traditional evaluation metrics and achieves a good refocus-invariance. In the future, we will evaluate the influence of disparity maps with different quality for the LFSP segmentation and explore the LFSP on more challenging non-Lambertian surfaces.

Acknowledgement

We thank Xianqiang Lv for his helps on real scene light field collection and anonymous reviewers for their valuable suggestions.

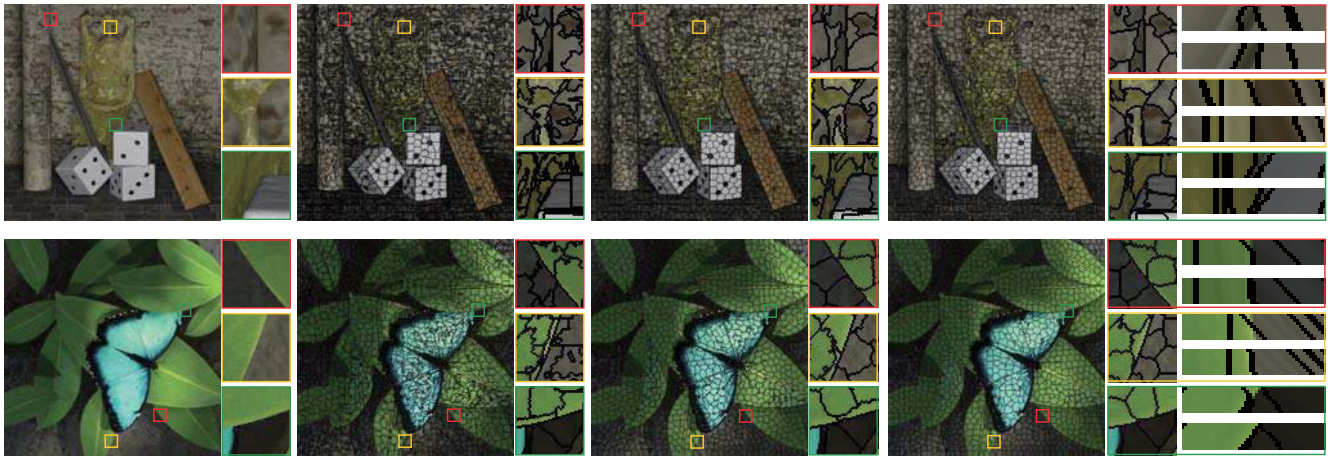


Figure 10. The segmentation results on synthetic light field data Buddha and Papillon (the superpixel size is 20). The first column shows the central view of input light field. The second to fourth columns show the results from SLIC, LSC and the proposed LFSP. For each region in our results, the first row shows the initial segmentation in EPI space, and the second row shows the optimized segmentation results.

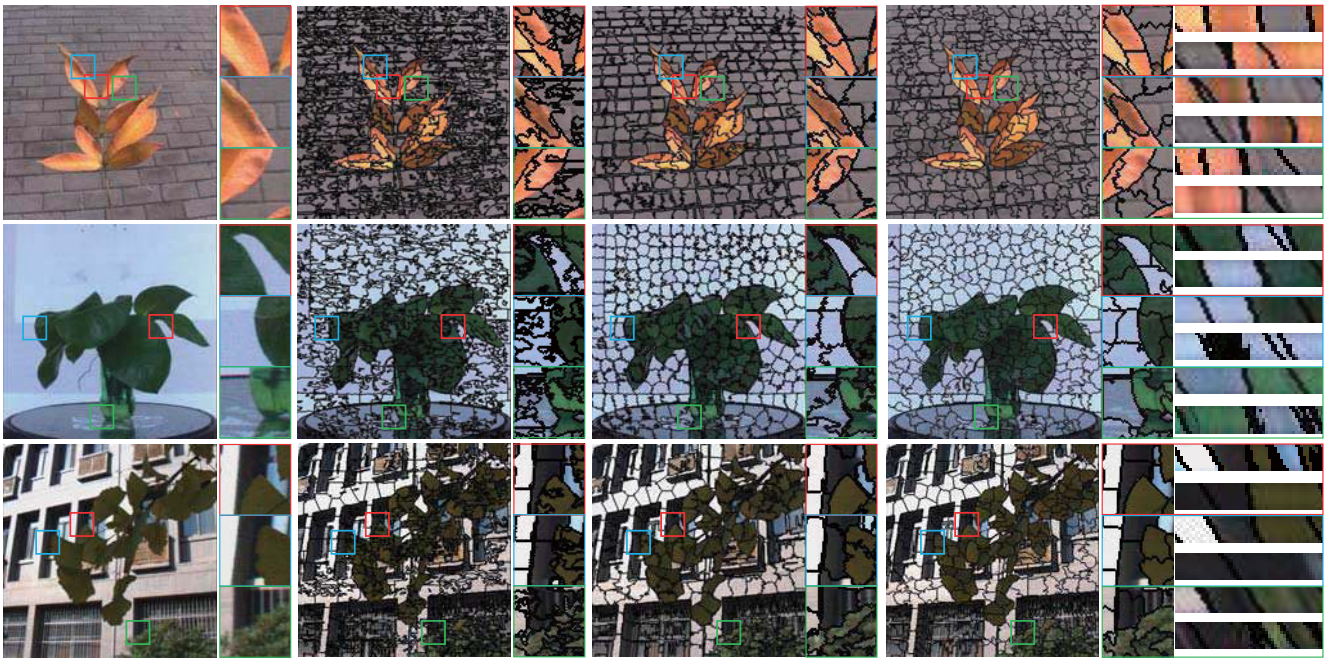


Figure 11. The segmentation results on real light field data (the superpixel size is 20). The first column shows the central view of input light field. The second to fourth columns show the results from SLIC, LSC and the proposed LFSP. For each region in our results, the first and second rows show the segmentation in the horizontal and vertical EPIs respectively.

Figure 12. The segmentation results under different refocus levels. The first column shows the central view of input light field. The second to fourth columns show the results under different refocus levels $(-0.5, 0, 0.5)$.

References

[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2274–2282, 2012.

[2] L. Baraldi, F. Paci, G. Serra, L. Benini, and R. Cucchiara. Gesture recognition in ego-centric videos using dense trajectories and hand segmentation. In *CVPR Workshops*, pages 688–693, 2014.

[3] T. E. Bishop and P. Favaro. The light field camera: Extended depth of field, aliasing, and superresolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(5):972–986, 2012.

[4] J. Chang, D. Wei, and J. W. Fisher. A video representation using temporal superpixels. In *CVPR*, pages 2051–2058, 2013.

[5] C. Chen, H. Lin, Z. Yu, S. Bing Kang, and J. Yu. Light field stereo matching using bilateral statistics of surface cameras. In *CVPR*, pages 1518–1525, 2014.

[6] CVPG. Computer vision and computational photography group. `http://www.npu-cvpg.org/publication`.

[7] D. G. Dansereau, O. Pizarro, and S. B. Williams. Decoding, calibration and rectification for lenselet-based plenoptic cameras. In *CVPR*, pages 1027–1034, 2013.

[8] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181, 2004.

[9] Gestalt. Gestalt principles. `http://facweb.cs.depaul.edu/sgrais/gestalt_principles.htm`.

[10] M. Hog, N. Sabater, and C. Guillemot. Light field segmentation using a ray-based graph structure. In *ECCV*, pages 35–50. Springer, 2016.

[11] A. Jarabo, B. Masia, A. Bousseau, F. Pellacini, and D. Gutierrez. How do people edit light fields? *ACM Transactions on Graphics*, 33(4):146–1, 2014.

[12] A. Levinshtein, A. Stere, K. N. Kutulakos, D. J. Fleet, S. J. Dickinson, and K. Siddiqi. Turbopixels: Fast superpixels using geometric flows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(12):2290–2297, 2009.

[13] M. Levoy and P. Hanrahan. Light field rendering. In *SIGGRAPH*, pages 31–42. ACM, 1996.

[14] N. Li, J. Ye, Y. Ji, H. Ling, and J. Yu. Saliency detection on light field. In *CVPR*, pages 2806–2813. IEEE, 2014.

[15] Z. Li and J. Chen. Superpixel segmentation using linear spectral clustering. In *CVPR*, pages 1356–1363. IEEE, 2015.

[16] M. Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa. Entropy rate superpixel segmentation. In *CVPR*, pages 2097–2104. IEEE, 2011.

[17] Lytro. Lytro redefines photography with light field cameras. `http://www.lytro.com`, 2011.

[18] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *ICCV*, volume 2, pages 416–423. IEEE, 2001.

[19] H. Mihara, T. Funatomi, K. Tanaka, H. Kubo, Y. Mukaigawa, and H. Nagahara. 4d light field segmentation with spatial and angular consistencies. In *ICCP*, pages 1–8, 2016.

[20] Raytrix. ∞ raytrix. `http://www.raytrix.de`, 2012.

[21] X. Ren and J. Malik. Learning a classification model for segmentation. In *ICCV*, pages 10–17. IEEE, 2003.

[22] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[23] P. P. Srinivasan, M. W. Tao, R. Ng, and R. Ramamoorthi. Oriented light-field windows for scene flow. In *ICCV*, pages 3496–3504, 2015.

[24] M. W. Tao, S. Hadap, J. Malik, and R. Ramamoorthi. Depth from combining defocus and correspondence using light-field cameras. In *ICCV*, pages 673–680, 2013.

[25] A. Vedaldi and B. Fulkerson. Vlfeat: An open and portable library of computer vision algorithms. In *ACM Multimedia*, pages 1469–1472. ACM, 2010.

[26] A. Vedaldi and S. Soatto. Quick shift and kernel methods for mode seeking. In *ECCV*, pages 705–718. Springer, 2008.

[27] O. Veksler, Y. Boykov, and P. Mehrani. Superpixels and supervoxels in an energy optimization framework. In *ECCV*, pages 211–224. Springer, 2010.

[28] T. C. Wang, A. A. Efros, and R. Ramamoorthi. Depth estimation with occlusion modeling using light-field cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 2170–2181, 2016.

[29] T.-C. Wang, J.-Y. Zhu, E. Hiroaki, M. Chandraker, A. A. Efros, and R. Ramamoorthi. A 4d light-field dataset and cnn architectures for material recognition. In *ECCV*, pages 121–138. Springer, 2016.

[30] S. Wanner and B. Goldluecke. Variational light field analysis for disparity estimation and super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(3):606–619, 2014.

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

[32] S. Wanner, C. Straehle, and B. Goldluecke. Globally consistent multi-label assignment on the ray space of 4d light fields. In *CVPR*, pages 1011–1018, 2013.

[33] Z. Xiao, Q. Wang, G. Zhou, and J. Yu. Aliasing detection and reduction in plenoptic imaging. In *CVPR*, pages 3326–3333, 2014.

[34] Y. Xu, H. Nagahara, A. Shimada, and R.-i. Taniguchi. Transcut: Transparent object segmentation from a light-field image. In *ICCV*, pages 3442–3450. IEEE, 2015.

[35] K. Yamaguchi, D. McAllester, and R. Urtasun. Efficient joint segmentation, occlusion labeling, stereo and flow estimation. In *ECCV*, pages 756–771. Springer, 2014.

[36] J. Yang and H. Li. Dense, accurate optical flow estimation with piecewise parametric model. In *CVPR*, pages 1019–1027, 2015.