

EDGE RECONSTRUCTION METHOD TO IMPROVE DEPTH ESTIMATION FROM LIGHT FIELDS

Rui Lourenco[†], Pedro A. A. Assuncao^{†*}, Luis M. N. Tavora^{*},
Lucas A. Thomaz[†], Rui Fonseca-Pinto^{†*}, Sergio M. M. Faria^{†*}

[†]Instituto de Telecomunicações, Portugal

^{*}ESTG, Politécnico de Leiria, Leiria, Portugal

e-mails: {rui.lourenco, amado, lucas.thomaz, sergio.faria}@co.it.pt, {luis.tavora, rui.pinto}@ipleiria.pt

ABSTRACT

Light fields capture both the intensity and angular information of light rays emanating from objects in a given visual scene. The combined spatial information allows the estimation of dense depth maps through the use of different methods currently available in related literature. One such method is based on the structure tensor, which globally yields good results but fails to find coherent depth values in occlusion areas. In this paper we propose a method to obtain an accurate depth representation of object boundaries that are located in such occlusion areas. The proposed method is based on the reconstruction of an edge map comprised of depth transitions, while avoiding texture gradients with no depth variation. This reconstructed edge map representation is used to highly enhance the quality of disparity maps obtained from the structure tensor. The achieved results show a significant reduction of the mean square error for a wide variety of light fields. Informal subjective analysis of the obtained disparity maps also reveals a significant reduction in artifacts near the object boundary regions.

Index Terms — Light Field, Disparity Estimation, Plenoptic Imaging

1. INTRODUCTION

Nowadays, depth information is used in various computer vision applications, such as automatic measurements and quality control in different types of industries [1], 3D analysis and computerized effects in imaging and video, and also for diagnostic of medical conditions [2]. With the advent of light field images, depth estimation has become a common processing step in several applications based on this type of multidimensional visual information.

Light fields, represented as 4D signals, comprise information not only about the intensity but also the directionality of light rays, which must be processed prior to practical use. In order to estimate the disparity from 4D light fields, several methods are described in the literature. A few notable examples include the spinning parallelogram operator (SPO) algorithm [3] that utilizes a novel operator to detect the orientation of symmetries in each point of an epipolar plane image (EPI), the EPINET method [4], which makes use of light field specific data augmentation techniques to train a neural network in disparity estimation, and the structure tensor method [5] that provides a fast and accurate disparity estimation.

The most impacting limitation of the structure tensor method is the systematic inaccuracy of depth estimation in occlusion regions of the light fields. Since this problem is not addressed by the structure tensor's reliability metric, multiple techniques have been developed to mitigate this issue. Wanner et al. [5] use both Total Variation Denoising and a global optimization method to obtain an enhanced disparity map, built from disparity maps and obtained from horizontal and vertical epipolar plane images. In spite of its sophisticated algorithm, this method does not completely eliminate the inaccuracy around border edges. The work of Li *et al.* [6], on the other hand, tries to mitigate this limitation by developing a penalty metric for the reliability of the structure tensor based on epipolar geometry and using such indicator in a closed form optimization based on the Laplacian matting matrix [7]. However, this method is not effective in some specific image regions of the light fields. More recently [8], the authors of this paper took advantage of the systematic nature of this occlusion error to identify the erroneous areas and inpaint them with values from correct regions. While this technique achieves better results in terms of objective metrics, the detection of relevant image regions is heavily dependent on the quality of the utilized edge map, with visible artifacts appearing when the edge detection algorithm fails.

While in [8], the correction is done simply based on the detected disparity and image edges, as described in Subsection 3.1, in the present work, the inaccuracy of the structure tensor in occlusion areas is further addressed by proposing an algorithm to improve the light field edge map, described in Subsection 3.2. To this goal, an edge map is created including only object boundaries and avoiding sharp transitions in the texture of the image that are not represented in the disparity map (i.e., rapid changes in texture without a correspondence in depth). Such edge map is then used to improve a structure-tensor-based disparity estimation by increasing accuracy in occlusion areas and reducing the amount of artifacts.

The remainder of the paper is organized as follows: Section 2 describes the structure tensor and explains its limitations near the occlusion areas. Section 3 details the algorithm to generate accurate edge maps and how they are used to improve the accuracy of the disparity maps obtained with structure tensor. In Section 4, experimental results obtained with the proposed method are presented, showing a significant improvement when compared to previous structure-tensor-based methods and comparable results to other state-of-the-art methods. Finally, in Section 5, the overall contributions of the work are analyzed along with possibilities for further research.

This work is funded by FCT/MEC through national funds, under project PlenoISLA PTDC/EEI-TEL/28325/2017 and when applicable co-funded by FEDER – PT2020 partnership agreement under the project UID/EEA/50008/2019.

2. THE STRUCTURE TENSOR

The structure tensor is defined for a given pixel (x, y) , as 2×2 matrix \mathbf{T}_{xy} , computed as:

$$\mathbf{T}_{xy} = \begin{bmatrix} I_x^2 * G_\sigma & I_x I_y * G_\sigma \\ I_x I_y * G_\sigma & I_y^2 * G_\sigma \end{bmatrix} = \begin{bmatrix} J_{xx} & J_{xy} \\ J_{xy} & J_{yy} \end{bmatrix}, \quad (1)$$

where I_x and I_y represent the smoothed directional discrete derivatives of an image I , and G_σ represents a Gaussian kernel with scale σ .

In [9], it is shown that the structure tensor has two orthogonal eigenvectors, where the direction of the eigenvector with the highest associated eigenvalue is the direction of linear symmetry. When applying the structure tensor to an EPI, the direction of linear symmetry will correspond to the orientation of the structures and, therefore, it is proportional to the disparity [5].

While using regular numeric values to find the eigenvectors of \mathbf{T}_{xy} provides a reliable estimate of the disparity, it is more computationally efficiency to analytically solve for the eigenvectors of \mathbf{T}_{xy} , obtaining:

$$d = \frac{J_{yy} - J_{xx} + \sqrt{(J_{yy} - J_{xx})^2 + 4J_{xy}^2}}{2J_{xy}}, \quad (2)$$

where d is the disparity estimation for a pixel in the EPI image I . Furthermore, the coherence of d itself can be estimated by means of a reliability measure r , which provides a metric for how closely the window around the pixel represents a linearly symmetric function. Pixels near strong edges have a high reliability value, while pixels in constant areas of the image have a lower reliability value.

By definition, the structure tensor \mathbf{T}_{xy} includes convolutions with a Gaussian kernel (G_σ , see (1)) using a window centred around each pixel. In occlusion regions, one has to take into account that the linear symmetry is associated with an object edge. Yet, this symmetry will be calculated for the entire window and assigned to the central pixel of said window. This central pixel might not coincide exactly with the object edge. In such case, the orientation of the edge in the EPI (i.e. its disparity) is incorrectly assigned to a pixel that does not represent the occluding object. This situation occurs every time a window includes an object edge that does not cross the central pixel of said window. In practice, this leads to objects having larger silhouettes in the disparity map than in the original image.

3. EDGE RECONSTRUCTION METHOD

The proposed edge reconstruction method comprises three main steps. Firstly, an initial map of object boundaries in the central view is estimated based on EPI edge detection. Secondly, this initial edge map is improved using an edge reconstruction algorithm. Finally, the improved map of object boundaries is used to detect and correct erroneous regions of the disparity map.

3.1. Initial Object Boundary Estimation

The goal of this first step is to find the initial binary edge map $m_{init}(x, y)$ that gives the position of object boundaries in the central view of the light field. To do so, the light field $L(s, t, x, y)$ is split into its various EPIs in the vertical and horizontal directions,

$$I_v(t, y) = L(s_{mid}, t, x_i, y), \quad (3)$$

$$I_h(s, x) = L(s, t_{mid}, x, y_j), \quad (4)$$

where x_i and y_j refer to the i^{th} line or j^{th} column of each view of the light field with i_{max} lines, and j_{max} columns respectively. The total number of views in the horizontal and vertical directions is defined as s_{max} and t_{max} , $t_{mid} = \text{round}(\frac{t_{max}}{2})$ and $s_{mid} = \text{round}(\frac{s_{max}}{2})$. This implies that only the crosshair views of the light field are used.

The remainder of this section will be limited to explain the procedure for the horizontal EPIs $I_h^{(j)}(s, x)$. The method for the vertical EPIs $I_v^{(i)}(t, y)$ is considered to be equivalent. The structure tensor is used to calculate the disparity $d^{(j)}(s, x)$ and the reliability $r^{(j)}(s, x)$ for every EPI $I_h^{(j)}(s, x)$ for all j such that $0 \leq j \leq j_{max}$. In order to differentiate texture-only edges from object boundaries, all edges in $I_h^{(j)}(s, x)$ are matched with an edge in $d^{(j)}(s, x)$ whenever a correspondence exists.

In order to accomplish this, it is necessary to detect all texture and disparity edges. For each $I_h^{(j)}(s, x)$, a texture edge map $e_t^{(j)}(s, x)$ is obtained using the Canny edge detector. For the corresponding disparity $d^{(j)}(s, x)$, gradient-based methods, such as the Canny edge detector, would provide false-positive results in surfaces approximately orthogonal to the camera plane as they would provide high disparity gradients. Thus, the Laplacian $\nabla^2 d^{(j)}(s, x)$ proves to be a better tool to compute the disparity edge maps of $d^{(j)}(s, x)$. In this way, a binary map $e_d^{(j)}(s, x)$ is created such that

$$e_d^{(j)}(s, x) = \begin{cases} 0 & \text{if } \nabla^2 d^{(j)}(s, x) \leq th \\ 1 & \text{if } \nabla^2 d^{(j)}(s, x) > th \end{cases}, \quad (5)$$

where th refers to a predefined threshold.

Since the algorithm's goal is to obtain an improved disparity map for the center view of the light field alone, to reduce the computational complexity in further steps, only the central lines of $e_t^{(j)}(s, x)$ and $e_d^{(j)}(s, x)$ are kept (those corresponding to the center view), i.e., the EPI edge line $l_t^{(j)}(x) = e_t^{(j)}(s_{mid}, x)$ and the disparity edge line $l_d^{(j)}(x) = e_d^{(j)}(s_{mid}, x)$.

As previously discussed, the disparity map $d^{(j)}(s, x)$ is inaccurate in occlusion regions causing the EPI edges indicated in $l_t^{(j)}(x)$ not to coincide with those in $l_d^{(j)}(x)$. This complicates the matching between EPI edges and disparity edges.

To solve the edge matching problem, let us define for all x_0 such that $l_t^{(j)}(x_0) = 1$, a vicinity region N :

$$N = \{x : |x - x_0| < w\}, \quad \forall x \in \mathbb{N}, \quad (6)$$

where w is a constant defined based on the size of the Gaussian kernel G_σ used to construct \mathbf{T}_{xy} . Then a disparity edge can be considered a matching candidate if:

$$\exists x'_0 \in N \mid l_d^{(j)}(x'_0) = 1. \quad (7)$$

Then, matching candidates are considered valid object boundaries if they respect the following disparity constraint,

$$d^{(j)}\left(s_{mid}, x'_0 + \frac{(x_0 - x'_0)}{|x_0 - x'_0|}\right) < d^{(j)}(s_{mid}, x_0), \quad (8)$$

which imposes a disparity in the occluding region that is larger than the disparity in the occluded region. If this constraint is fulfilled, then the map of object boundaries $m_{init}(x_0, y_j) = 1$, otherwise $m_{init}(x_0, y_j) = 0$. Additionally, to be used in the improvement step, a map of the edges of the center view disparity, $m_d(x, y)$ is created by concatenating all disparity edge lines $l_d^{(j)}(x)$ with $1 < j < j_{max}$.

3.2. Edge Map improvement

Due to limitations of both edge detectors, $m_{init}(x, y)$ still has broken edge segments. This leads to visible artifacts in the disparity, when such edge map is used to enhance the silhouettes of objects in disparity maps. In order to create the final map $m(x, y)$ with more accurate object boundaries, the following method was devised:

1. Very short edges are eliminated (i.e., shorter than 4 pixels).
2. Edges of the center view $L(s_{mid}, t_{mid}, x, y)$ are computed using the Canny detector and compared to $m_{init}(x, y)$.
3. Edges that share a significant number of points in both maps are merged.
4. Finally, the remaining broken edges segments are reconstructed.

To reconstruct these broken edges, the spatial position of the end points for each line are calculated and the euclidean distance between all pairs of endpoints is stored in a symmetric matrix \mathbf{M} . Additionally, the directional estimate is assigned to each endpoint, based on the four closest points.

The objective is to draw a straight line, using the Bresenham line-drawing algorithm, connecting endpoints that should constitute a single, complete boundary of an object. Here, the challenge lies in estimating which endpoint pairs should be connected. Since closer unconnected endpoints are more likely to belong to the same edge, an iterative algorithm is used to choose the unconnected endpoint pair with the smallest euclidean distance, which is equivalent to finding the smallest value in \mathbf{M} (excluding its principal diagonal).

For each endpoint pair, a decision must then be made to connect or not the pair with a straight line. Figure 1 shows an example of an endpoint pair. The black dashed lines represent the estimated direction associated to the endpoints. Variables α and α' represent the angles between these lines for each endpoint and the horizontal direction (light gray dashed lines). The colored dashed line represents the straight line defined to connect both endpoints. While, β is the angle between this line and the horizontal direction.

If the difference between α and α' is larger than 90° the endpoint pair is ignored and no line is drawn. Similarly, if either $|\alpha - \beta|$ or $|\alpha' - \beta|$ are larger than a predefined threshold, the endpoints are considered unconnected, since this indicates that the respective directions are significantly different from the straight line. In all other cases a connecting line is drawn in $m_i(x, y)$ between the two endpoints. After reconstructing all possible broken edges throughout the whole disparity map, an improved map $m(x, y)$, with more accurate object boundaries is obtained.

Figure 2 compares details in the $m(x, y)$ before and after edge improvement for the Cotton image of the HCI dataset [10]. There is an accurate reconstruction of incomplete edges around the shoulder area of the statue also shown in the Figure.

As real object boundaries constitute closed forms and do not include discontinuities, it is valid to consider that any discontinuity in $m(x, y)$ corresponds to some sort of error in the edge detection process. The results in Section 4 confirm this.

3.3. Improving the disparity

After the previous reconstruction process, the spatial position of the object boundary edges in $m(x, y)$, is compared with the corresponding position of the disparity edges in $m_d(x, y)$. The region between two corresponding edges is the one that needs to be corrected.

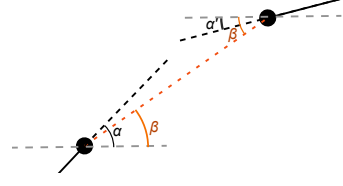
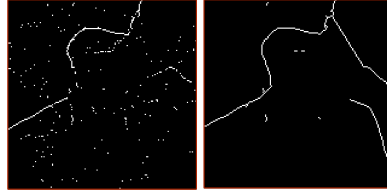


Figure 1: Example of a broken edge. Black dashed lines show the predicted extension of the edges. If α and α' are close to β , then the two edges are connected through the colored dashed line.



(a) Cotton Image Center View



(b) Map before im- (c) Map after im-
improvement improvement

Figure 2: Edge reconstruction detail on the Cotton image.

For this puprose, the process used to find matching candidates, defined in Subsection 3.1, is repeated for each line $m(x, y_j)$. In this case, for every x_0 , such that $m(x_0, y_j) = 1$, a neighborhood N is defined as in (6). Then a disparity edge can be considered a matching candidate if:

$$\exists x'_0 \in N \mid m_d(x'_0, y_j) = 1. \quad (9)$$

Additionally, the following disparity constraint must hold for any two matching edges:

$$d\left(x'_0 + \frac{(x'_0 - x_0)}{|x'_0 - x_0|}, y_j\right) < d(x_0, y_j), \quad (10)$$

where $d(x, y)$ is the disparity map obtained by concatenating all lines $d^{(j)}(s_{mid}, x)$ with $1 < j < j_{max}$. If this constraint is not fulfilled, then the matching is considered impossible, and the disparity edge for $x = x'_0$ is discarded. If the constraint is fulfilled, then the edges are considered matching and the region E is defined:

$$E = \begin{cases} \{x : x \leq x'_0\}, & \text{if } x, x'_0 \geq x_0 \\ \{x : x \geq x'_0\}, & \text{if } x, x'_0 < x_0 \end{cases} \quad (11)$$

After this step, the disparity values to be corrected are defined as the vector $d(e, y_j) \forall e \in E$, which contains the disparities of the occluding object, rather than those of the background region. Thus the corrected disparity map is obtained as,

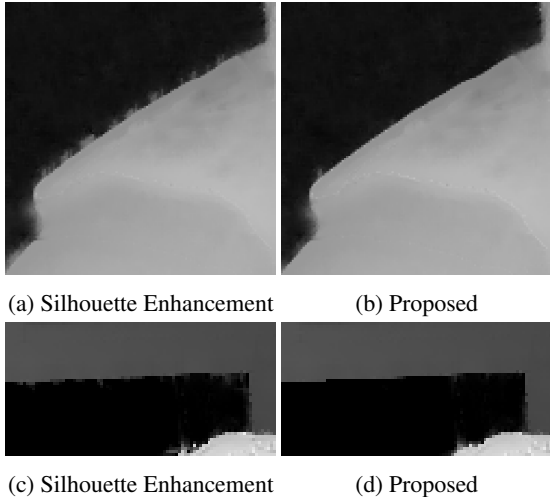


Figure 3: Edge details on disparity maps taken from the Cotton and Dino images with and without edge improvement.

$$d(e, y_j) = \text{median}(d(b, y_j)), \quad (12)$$

where $d(b, y_j) \forall b \in B$ contains values of the disparity of the first P pixels (from the disparity edge position) of the background region $B = \{x_0, \dots, x_0 + (P - 1)\}$.

Finally, the optimization based on the Laplacian Matting Matrix [7] (as applied in [6] and [8]) is used to reduce the noise from the final disparity estimation.

4. EXPERIMENTAL RESULTS

The proposed method was implemented in MATLAB and evaluated on the HCI training dataset [10]. The results for the SPO [3] and ST [5] methods are taken from the HCI Light Field Benchmark website [10]. Table 1 shows the impact in terms of mean square error (MSE), comparing the algorithm with state-of-the-art methods and with the method previously proposed by the authors (ST-SI) in [8]. One can see a clear improvement relative to the structure tensor-based methods presented in [5] and [8]. From Figure 3 one can confirm that the proposed algorithm is able to improve the disparity map by eliminating border artifacts. The use of a more accurate edge map ensures that the inpainting occurs across the whole edge without discontinuities.

Table 1: Comparison between state-of-the-art methods and the proposed method (Prop), in terms of Mean Square Error ($\times 100$).

	SPO [3]	ST[5]	ST-SI[8]	Prop
Boxes	10.374	10.928	9.31	8.353
Cotton	1.329	4.318	0.79	0.479
Dino	0.254	2.076	0.63	0.572
Sideboard	0.932	4.658	1.21	1.164

When compared with state-of-the-art methods, it is clear that the proposed method presents a performance superior to the more traditional ST approach in [5]. When compared to the SPO framework, the proposed method proves to be superior for some images, namely those with visible and well defined edges, such as Boxes and Cotton, while obtaining inferior results in images with larger number of relatively small features, such as Dino and Sideboard.

While neural-network based methods exist and show promising results, algorithms such as [4] include these LFs in their training dataset, thus they cannot be fairly compared with other methods.

5. CONCLUSIONS AND FUTURE WORK

In this paper, a method is proposed to generate improved maps with accurate delineation of object boundaries in light fields. Such maps are then used to improve the performance of a structure tensor-based disparity estimation methods.

This improvement is notable both in terms of the subjective analysis of the results, where a significant reduction in annoying artifacts and an increase in border accuracy can be observed, as well as through an objective metric, the mean square error, where the algorithm performs very competitively in relation to the state-of-the-art methods. While this work focus on disparity estimation improvements, future work can make use of the accurate object boundary map to help in object segmentation and region-of-interest determination.

6. REFERENCES

- [1] “Raytrix-3d optical inspection,” <https://raytrix.de/inspection/>, Accessed: 2018-11-30.
- [2] Steven McDonagh, Robert B Fisher, and Jonathan Rees, “Using 3D information for classification of non-melanoma skin lesions,” in *Medical Image Understanding and Analysis*, Dundee, United Kingdom, July 2008, pp. 164–168.
- [3] Shuo Zhang, Hao Sheng, Chao Li, Jun Zhang, and Zhang Xiong, “Robust depth estimation for light field via spinning parallelogram operator,” *Computer Vision and Image Understanding*, vol. 145, pp. 148–159, April 2016.
- [4] C. Shin, H.-G. Jeon, Y. Yoon, I. S. Kweon, and S. J. Kim, “Epinet: A fully-convolutional neural network using epipolar geometry for depth from light field images,” in *IEEE Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, USA, June 2018, pp. 4748–4757.
- [5] S. Wanner and B. Goldluecke, “Globally consistent depth labeling of 4d light fields,” in *IEEE Conf. on Computer Vision and Pattern Recognition*, Providence, USA, June 2012, pp. 41–48.
- [6] J. Li and Z. N. Li, “Continuous depth map reconstruction from light fields,” in *IEEE International Conf. on Multimedia and Expo*, San Jose, USA, July 2013, pp. 1–6.
- [7] Anat Levin, Dani Lischinski, and Yair Weiss, “A Closed Form Solution to Natural Image Matting,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 228–242, February 2008.
- [8] R. Lourenco, P. A. A. Assuncao, L. M. N. Tavora, R. Fonseca-Pinto, and S. M. M. Faria, “Silhouette enhancement in light field disparity estimation using the structure tensor,” in *IEEE International Conf. on Image Processing*, Athens, Greece, October 2018, pp. 2580–2584.
- [9] Josef Bigun, “Optimal orientation detection of linear symmetry,” in *IEEE First International Conf. on Computer Vision*, London, Great Britain, June 1987, pp. 433–438.
- [10] O. Johannsen, K. Honauer, and B. Goldluecke et. al., “A taxonomy and evaluation of dense light field depth estimation algorithms,” in *IEEE Conf. on Computer Vision and Pattern Recognition Workshops*, Honolulu, USA, July 2017, pp. 1795–1812.