SILHOUETTE ENHANCEMENT IN LIGHT FIELD DISPARITY ESTIMATION USING THE STRUCTURE TENSOR

Rui Lourenco [⋆†], *Pedro A. A. Assuncao* [⋆†], *Luis M.N. Tavora*[⋆], *Rui Fonseca-Pinto* [⋆†] *and Sergio M. M. Faria* [⋆†]

[⋆]Instituto de Telecomunicações, Portugal
[†]Instituto Politécnico de Leiria / ESTG, Leiria, Portugal
e-mails: {rui.lourenco, amado, sergio.faria}@co.it.pt, {luis.tavora, rui.pinto}@ipleiria.pt

ABSTRACT

This paper presents a method to improve disparity maps computed from light field images using structure tensor methods, which tend to expand the borders of the occluding objects, enlarging their silhouette. The proposed method relies on the fact that such regions of the silhouette are defined by a mismatch between the disparity map edges, computed by the structure tensor, and those of the corresponding epipolar plane images (EPI) representing the light field. The proposed silhouette improvement method determines a correspondence between EPI edges and the disparity map edges, identifying the erroneous silhouette regions. The disparity map is corrected by using neighbouring values computed with high structure tensor reliability. The achieved results show that the disparity map is improved both around object edges and overall, reducing the MSE by 47.9% in comparison with other methods also based on the structure tensor.

Index Terms— Light field, disparity, depth, structure tensor

1. INTRODUCTION

Light field technology has been rapidly evolving in recent years providing dense representations of visual scenes by capturing information not only from light intensity but also light ray directionally. This type of visual information enables new computational methods for focusing and changing the depth-of-field, which are traditional functions of the optical domain [1]. Since light fields are comprised of a high number of dense views, it is possible to exploit interesting data consistency properties, which do not always exist in generic multiview imaging. Depth reconstruction is often necessary in different types of computer vision applications, which can benefit from the regular data structure of the light fields. To this aim, several approaches have been investigated in the past. Among the most relevant methods, one may find focus stack based methods, such as [2], [3] and [4], and methods based on multi-view stereo correspondence, such as [5].

Among the various depth reconstruction methods, those based on the structure tensor (ST) approach are particularly relevant due to their low computational complexity and reasonable accuracy [6]. These methods rely on specific characteristics of the so-called epipolar plane images (EPI), which are created by projecting 3D points of a visual scene onto different views and then taking the images defined by corresponding pixels in either horizontal or vertical directions. The high density of views comprising light fields, results in EPIs mostly containing lines of different widths, whose slope line is related to the depth of the point in the space. Such characteristic transforms the problem of depth estimation into a line and slope detection on EPIs. In this context, the ST is an efficient method to detect linear symmetry in EPIs with good accuracy and low complexity.

However, this method systematically fails in the presence of occluded regions, and the resulting effect is an enlargement of the objects' silhouette in the depth map. This paper presents a novel method to improve the reconstruction of such depth maps, by enhancing the silhouettes obtained through the ST based approach. In the proposed method, the enlarged silhouettes are first detected using epipolar geometry combined with edge detection, and then a structural inpainting method, specifically devised to reconstruct the disparity map, is used. The performance of this approach was evaluated and the results show that consistent gains (i.e., depth maps with lower MSE) are achieved in comparison with those obtained from the ST method on its own and they are also competitive with other state-of-the-art algorithms.

The paper is organised as follows: Section 2 presents a brief summary on light field and epipolar images, structure tensor and the silhouette enlargement problem. In section 3 the proposal for silhouette enhancement is described, and Section 4 presents the experimental results. Finally, Section 5 provides a summary of the main conclusions of this work.

This work was supported by the Fundação para a Ciência e Tecnologia, Portugal, under Dermopleno project in the scope of R&D Unit 50008, through national funds and where applicable co-funded by FEDER – PT2020 partnership agreement.

2. DEPTH ESTIMATION WITH THE STRUCTURE TENSOR

2.1. Light Field and Epipolar Images

A 4D light field represents the light flow through unobstructed space from a static scene with fixed illumination, usually denoted by a 4-dimensional function $L(s,t,x,y)$, which can be seen as a 2D array of different views of the scene [7]. The spatial position of each viewpoint is represented by the index of the different views (s,t), while the pixel coordinates within each view are defined by (x,y). Since the early days of stereo vision, (e.g., see the paper by Bolles et al [8]) EPIs are known for containing relevant geometric information about the 3D scene. In the case of dense views from light fields, the slope of an EPI line corresponds to the disparity of the 3D point, and therefore the depth of all visual objects in the scene, relative to the position of the acquisition system in a 3D space, can be computed from EPI lines.

Several different approaches have been devised to take advantage of the existing linear symmetry in EPIs. Bolles [8] extracted salient lines by finding zero crossings with a Difference of Gaussians filter. Criminisi *et al.* [9] extracted EPI regions by using photo consistency. Tao et al. [4] used both focus and correspondence cues to achieve a disparity estimation. Michael Strecke et al [10] improved on Tao's work by jointly optimizing the disparity estimation and the normal map estimation, achieving better results. Shuo Zhang et al. [11] presented a novel approach by using parallelograms with different rotations to estimate the disparity.

2.2. The Structure Tensor

The ST, was introduced independently by Bigün *et al.* [12] and Harris *et al.* [13], as a robust mathematical tool for the analysis of structure orientation in images. More recently Wanner [14] proposed its use to characterize the EPIs' regular structures, by means of a 2×2 matrix ($\mathcal{T}$) calculated on a pixel basis, as given by equation 1,

$$\mathcal{T} = \begin{bmatrix} w * K_x^2 & w * (K_x K_y) \\ w * (K_x K_y) & w * K_y^2 \end{bmatrix} = \begin{bmatrix} J_{xx} & J_{xy} \\ J_{xy} & J_{yy} \end{bmatrix} \quad (1)$$

where w is an averaging window function and $K_{\rho=x,y}$ is the gradient of the smoothed image K in the ρ^{th} direction. Usually, the operator w is a Gaussian kernel with an outer scale σ_o. K is a filtered version of the original epipolar image I obtained by means of a Gaussian averaging at an inner scale σ_i (i.e. $K = G_{\sigma_i} * I$).

As a symmetric and positive semi-definite matrix conveying gradient-based information, $\mathcal{T}$ has two orthogonal eigenvectors, which can be used to estimate the direction of the EPI lines. Accordingly, at each point, the disparity (d) can be estimated from equation 2, as demonstrated in [15]:

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

Furthermore, the coherence of d itself can be estimated by means of a reliability measure (r), defined as

$$r = \frac{(J_{yy} - J_{xx})^2 + 4J_{xy}^2}{(J_{xx} + J_{yy})^2}, \quad (3)$$

which varies from zero, for isotropic structures, to one, for well defined orientations [14]. Not the least, determining T for each EPI pixel of a light field is equivalent to calculate a mean estimation of d for all points, with its inherent averaging process weighted by the window function w. Overall, an important characteristic of the ST approach, is to allow dense depth estimations in quite smooth (or even uniform) regions of the EPI.

2.3. Silhouette Enlargement

When capturing a scene with various objects along a given direction, only those located closer to the camera plane will be definitely free from occlusions. On the EPI representation of a light field, the edges of such objects will be represented by complete and well defined lines spanning over all views. The opposite stands for objects partially occluded, or even for texture patterns on the background plane, where the corresponding EPI lines might be missing for some views.

Recalling that the ST approach actually computes, for every single point, the gradients averaged over a given spatial region (i.e., window w), in regions where depth is discontinuous such weighting process leads to wrong estimations. This is because the weighted averaging performed by the windowing function w is equivalent to spreading high magnitude gradients of a well defined EPI line over their neighbouring pixels. Hence, this results in intrinsic blurring of the EPI, yielding inaccurate depth maps with dilated object boundaries, i.e. silhouette enlargement.

An attempt to overcome this silhouette enlargement artefact, a penalty measure for the reliability estimation (r) was proposed in [16]. However, this has proven to be highly sensitive to the texture in the surrounding regions, providing only a minimal penalty when the algorithm fails in regions of uniform texture. A more robust and reliable method to is presented in Section 3.

3. PROPOSED METHOD FOR SILHOUETTE ENHANCEMENT

The method herein proposed for silhouette enhancement involves determining the edges of the disparity map, obtained from Equation 2, and then searching for their corresponding edges in the original EPI representation. Whenever the edges

of the disparity map do not match the corresponding ones in the EPI, a correction procedure is applied using the neighbouring values with high reliability estimations. Additionally, a Laplacian matting-based optimization is applied to smooth objects' contours. The proposed method is schematically presented in Figure 1, and its main steps are further described as follows.

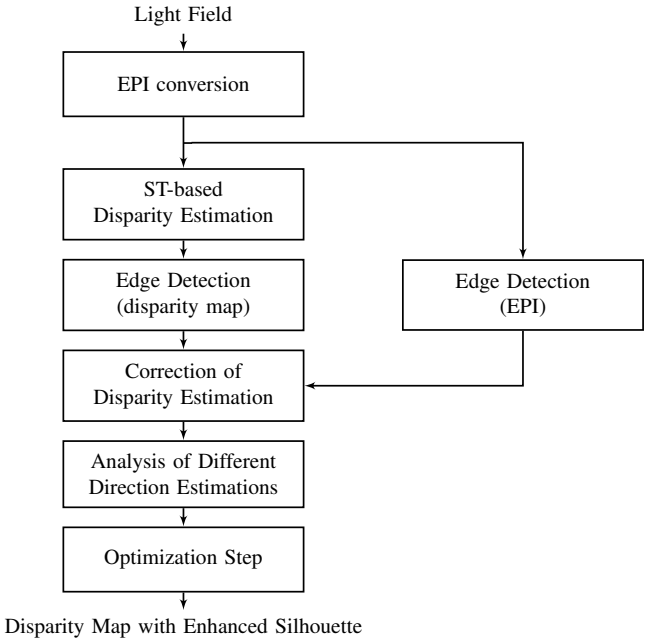


Fig. 1: Silhouette Enhancement process.

ST-based Disparity Estimation

The light field, $L(s,t,x,y)$ is split onto its EPIs $I(x,s)$ and $I(y,t)$, for the horizontal and the vertical directions, respectively (for simplicity, the method is solely described for horizontal EPIs). At this stage an ST-based disparity estimation, $d(x,s)$, is determined from Equation 2 for each point of every EPI. The reliability coefficient, $r(x,s)$ is also estimated from Equation 3. For the sake of efficiency, d and r only consider the central line of the EPI, representing by $d(x)$ and $r(x)$ the disparity and reliability of the central view, respectively.

Edge Detection

This step involves the determination of edges on both the EPI and disparity map. The EPI edges are determined using the Canny Edge algorithm, which has shown to be more accurate and less parameter sensitive than the ST using gradient information. To reduce the effect of local noise, a median filter is applied.

Regarding the disparity map, the edges were computed by using an algorithm based on the Laplacian, which was found to be less sensitive to noise in the initial estimation and to produce less false positives in slanted surfaces. It was also found that this method is able to overcome the limitations of the Canny Edge detector in dealing with images where very high and very low gradient edges occur.

Correction of Disparity Estimation

The correction of disparity is carried out by first determining the EPI region affected by the silhouette enlargement. As represented in Figure 2, starting from an edge point at the EPI representation (P_2), a search is performed on a window of width a, in order to find out the corresponding edge on the disparity map (P'_2). The dimension of the search region (a) is related to the size of the window (w) used in the outer Gaussian averaging function (Eq. 1).

After identifying P_2 and P'_2, the search is extended beyond the window a, in order to find a new region ($\mathcal{R}_3$ in Figure 2) with a lower disparity than that of P'_2 . When it occurs, the region with disparity values to be corrected (width Δs) is already clearly determined. If either (P'_2) or a new region (R_3) are not identified, then P_2 is not to be treated as a true edge. This may occur, for instance, when P_2 is associated to a local texture feature on the occluding object.

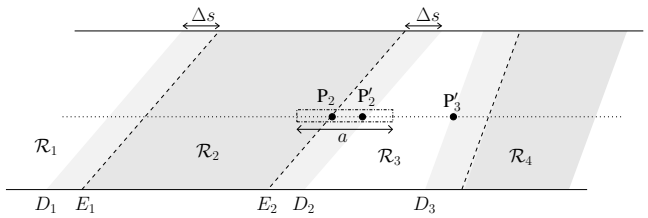


Fig. 2: Silhouette enlargement: the original EPI lines (E_1; E_2) and the corresponding lines on the disparity map (D_1; D_2). Working out the region (width Δs) involves starting from an EPI edge (P_2) and search over window of size a to determine the position of P'_2.

After identifying an enlarged region, its disparity values are corrected by substituting them by the mean disparity of the occluded region (R_3), which requires determination of its far end edge (P'_3 on D_3). To reduce the influence of noise, only points with a high reliability indicator (r) are considered to calculate this mean value. After exhaustive testing $r = 0.7$ was chosen, but small variations of this value do not have critical impact on the performance.

Analysis of vertical and horizontal disparity estimations

The method described above for correcting disparity estimation was only presented for horizontal EPIs ($I(x,s)$), but the whole algorithm also includes its used in all vertical EPIs ($I(y,t)$). Therefore, for each point (x,y) two disparity values are computed. The selection of best disparity estimation relies on the disparity value with higher degree of reliability (i.e. the highest r).

Optimization Step

The previously described edge correction strategy is highly dependent on the success of the edge detection step. At points where the algorithm fails to find accurate edges, the disparity map will exhibit jagged edges. To correct these, a further optimization is applied by considering a Laplacian-based image matting correction [17]. In brief, starting from an initial disparity map estimation $\widetilde{\boldsymbol{d}}$, an optimal solution $\boldsymbol{d}$ is achieved by minimizing the energy function:

$$J(\boldsymbol{d}) = \boldsymbol{d}^T L \boldsymbol{d} + \lambda(\boldsymbol{d}–\widetilde{\boldsymbol{d}})^T C(\boldsymbol{d}–\widetilde{\boldsymbol{d}}). \quad (4)$$

with $\boldsymbol{d}$ and $\widetilde{\boldsymbol{d}}$ represented as $N \times 1$ vectors, where N represents the number of pixels in each view (i.e. $N = Height \times Width$), and L represents the so called affinity matrix, whose $N \times N$ elements can be understood as a measure of the affinity between each pixel and the rest of the pixels on the image. Furthermore, λ is a global weight of the data term determining the strength of the smoothing operation ($\lambda = 5$, in [17]) and C is a diagonal, $N \times N$, matrix with reliability values (r) previously determined.

4. RESULTS

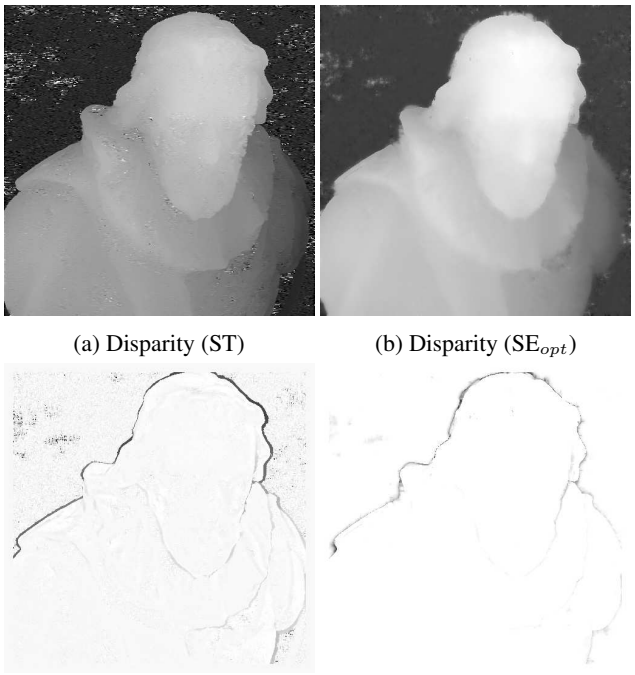
The performance of the proposed method was evaluated by using the 4D light field dataset made available in [18], which provides images with 9×9 views. The ST was implemented with $\sigma_i = 0.75$ and $\sigma_o = 1.5$, which follows from [14]. Some authors optimize these parameters to match the ground truth values, but such approach is not representative of general situations, since the ground truth is not always available.

The results, presented in Table 1, show the Mean Square Error (MSE) of the estimated disparity for the original ST approach proposed in [6], alongside with the proposed Silhouette Enhancement (SE) over ST, and the matting-based optimization applied after SE (SE$_{opt}$). The MSE reduction achieved by the SE$_{opt}$ method over other ST-based methods, EPI2 [6] and Li [16], is also shown in percentage - column $\Delta(\%)$ in Table 1.

Table 1: MSE of Disparity Estimation ($\times 100$)

	ST	SE	SE$_{opt}$	EPI2	Li	Δ (%)
Boxes	15.94	14.59	8.23	10.93	NA	-48.4
Cotton	4.17	2.65	0.64	4.32	NA	-84.7
Dino	1.61	0.924	0.62	2.08	NA	-61.5
Sideboard	2.92	2.096	1.30	4.65	NA	-55.4
Buddha	0.84	0.79	0.55	NA	0.64	-14.1
Mona Lisa	1.36	0.78	0.49	NA	0.73	-32.9

As shown in the Table, the proposed algorithms, SE and SE$_{opt}$, consistently improve the disparity estimation, with MSE reductions up to 81.7%, when compared with the other ST-based methods EPI2 and Li. On average, the MSE reduction is approximately 49.5%. The proposed method was also compared to other state-of-the-art methods, like LISAD [19], where it is also able to achieve lower MSE, for instance, 20% reduction for the Mona Lisa image. Furthermore, an informal subjective assessment of the resulting depth maps has also shown that the proposed methods achieve better quality than the other ST-based methods.



(a) Disparity (ST) (b) Disparity (SE$_{opt}$)

(c) Disparity difference (ST) (d) Disparity difference (SE$_{opt}$)

Fig. 3: Disparity map and difference to the ground truth.

A visual representation of the improvements is shown in Figure 3 for the central view of the Cotton light field. The Figure shows the disparity maps reconstructed with the proposed method (SE$_{opt}$) and with the original ST-based methods, as well their differences relative to the ground truth. The better quality of the silhouette and the whole disparity map shown Figure 3b, is quite evident. This is also confirmed by the smaller differences shown in Figure 3d.

5. CONCLUSIONS

The method proposed in this paper to enhance light field disparity estimation achieved consistently better quality in comparison with other ST-based methods. An average reduction of 47.9% of MSE, with a peak of 84.7%, is obtained in comparison with other reference methods. The overall improvements obtained with this method are mostly due to enhanced silhouettes and optimisation of low reliability disparity values. In the light of these results, the proposed method proofs to be a reliable, simple and fast to enhance disparity maps by reducing the silhouette effects.

6. REFERENCES

[1] Josef Bigun, *Vision with Direction*, Springer, 2006.

[2] Haiting Lin, Can Chen, Sing Bing Kang, and Jingyi Yu, "Depth Recovery from Light Field Using Focal Stack Symmetry," *Proc. International Conference on Computer Vision*, 2015.

[3] W. Williem, I. K. Park, and K. M. Lee, "Robust light field depth estimation using occlusion-noise aware data costs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2018.

[4] Michael W. Tao, Sunil Hadap, Jitendra Malik, and Ravi Ramamoorthi, "Depth from combining defocus and correspondence using light-field cameras," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 673–680, 2013.

[5] H. G. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y. W. Tai, and I. S. Kweon, "Accurate depth map estimation from a lenslet light field camera," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 1547–1555.

[6] S. Wanner and B. Goldluecke, "Globally consistent depth labeling of 4d light fields," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, June 2012, pp. 41–48.

[7] Marc Levoy and Pat Hanrahan, "Light field rendering," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, 1996.

[8] Robert C. Bolles, H. Harlyn Baker, and David H. Marimont, "Epipolar-plane image analysis: An approach to determining structure from motion," *International Journal of Computer Vision*, 1987.

[9] Antonio Criminisi, Sing Bing Kang, Rahul Swaminathan, Rick Szeliski, and P. Anandan, "Extracting layers and analyzing their specular properties using epipolar-plane-image analysis," *Computer Vision and Image Understanding (CVIU)*, vol. 97, pp. 51–85, 2005.

[10] M. Strecke, A. Alperovich, and B. Goldluecke, "Accurate depth and normal maps from occlusion-aware focal stack symmetry," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2529–2537.

[11] 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, 2016.

[12] Josef Bigün, "Optimal orientation detection of linear symmetry," Tech. Rep., Department of Electrical Engineering Linköping University, 1986.

[13] Chris Harris and Mike Stephens, "A combined corner and edge detector," in *In Proc. of Fourth Alvey Vision Conference*, 1988, pp. 147–151.

[14] Sven Wanner, *Orientation Analysis in 4D Light Fields*, Ph.D. thesis, Universitat Heidelberg, 2014.

[15] S. Wanner and B. Goldluecke, "Variational light field analysis for disparity estimation and super-resolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 3, pp. 606–619, 2014.

[16] J. Li and Z. N. Li, "Continuous depth map reconstruction from light fields," in *IEEE International Conference on Multimedia and Expo (ICME)*, July 2013, pp. 1–6.

[17] 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, 2008.

[18] Katrin Honauer, Ole Johannsen, Daniel Kondermann, and Bastian Goldluecke, "A Dataset and Evaluation Methodology for Depth Estimation on 4D Light Fields," pp. 19–34, 03 2017.

[19] I. Tosic and K. Berkner, "Light field scale-depth space transform for dense depth estimation," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2014, pp. 441–448.