

Surface Cameras from Shearing for Disparity Estimation on a Lightfield

Nuno Barroso Monteiro¹
nmonteiro@isr.tecnico.ulisboa.pt
João Pedro Barreto²
jpbar@isr.uc.pt
José Gaspar¹
jag@isr.tecnico.ulisboa.pt

¹ Institute for Systems and Robotics
University of Lisbon
Portugal

² Institute for Systems and Robotics
University of Coimbra
Portugal

Abstract

Disparity estimation from lightfields is usually based on multi-view stereo geometry, epipolar plane image geometry, or on testing some disparity hypotheses using shearing. Recently, the concept of surface camera image has been used to improve disparity estimation. In this work, we introduce the idea of considering a surface camera image as a generalization of shearing and evaluate the capabilities of using surface camera images in disparity estimation.

1 Introduction

Plenoptic cameras are capable of discriminating the contribution of each light ray emanating from a particular point. The collection of rays captured by these cameras is called a lightfield [8].

In a plenoptic camera, a point in the object space is projected into multiple points in the image sensor. The multiple projections allow to recover disparity assuming no particular position for the cameras, *e.g.* using multi-view stereo [1], or assuming the cameras define a linear path, *e.g.* using the epipolar plane image (EPI) geometry [6]. For the EPI analysis, one can consider gradient based approaches using standard image gradient operators [4] or structure tensors [12]. Nonetheless, these approaches limit the disparity range that can be estimated accurately to one pixel [6]. Shearing of the lightfield [6] increase the disparity range while maintaining the gradient operators constant.

Other strategies test predefined disparity hypothesis by shearing the lightfield and evaluating correspondence and defocus cues on the resulting lightfield [11]. These methods assume lambertian surfaces free of occlusion. Recently, the concept of surface camera images (SCams) [3] has been introduced to identify types of surfaces (lambertian or specular) and occlusions that allow to adapt the metrics used to evaluate correspondences. Although shearing and SCams have been presented as alternative methods, we show that these methods are related and that the SCam is a generalization of the shearing operation.

2 Surface Camera Images as a Generalization

A SCam is a camera that collects rays that intersect at an arbitrary point in the object space [13]. These rays can emanate from different points if the camera's projection center is located on free space (camera A of Figure 2) or is located on a surface point which is partially occluded (camera B of Figure 2). On the other hand, the rays emanate from a common point if the projection center of the camera is defined on a surface point (camera C of Figure 2).

Considering the lightfield in the object space $L_{\Pi}(s, t, u, v)$ acquired by a plenoptic camera with the plane Ω in focus (Figure 1), one can obtain a SCam with projection center at point (q, r) of plane Γ . $L_{\Pi}(s, t, u, v)$ collects rays $\Psi = [s, t, u, v, 1]^T$ that are parameterized using a point (s, t) and a direction (u, v) defined on a plane Π in metric units [10]. Assuming that the plane Γ is at a distance d from the plane Π , one can re-parameterize the lightfield captured by the plenoptic camera relatively to the plane Γ [2], $L_{\Gamma}(q, r, u, v)$, by

$$\Psi_{\Gamma} = \mathbf{D} \Psi, \quad (1)$$

where

$$\mathbf{D} = \begin{bmatrix} 1 & 0 & d & 0 & 0 \\ 0 & 1 & 0 & d & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad (2)$$

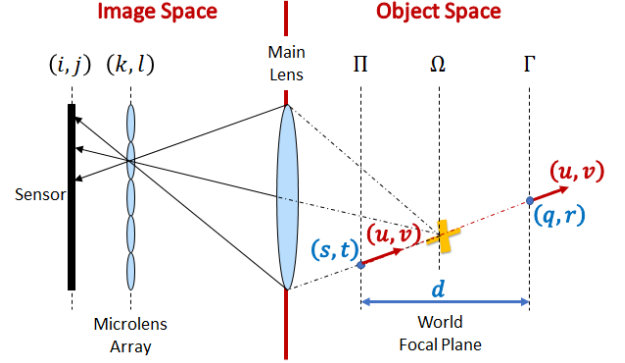


Figure 1: Geometry of a plenoptic camera. On the left, the lightfield in the image space is parameterized using pixels and microlenses indexes. On the right, the lightfields in the object space are parameterized using a point and a direction, and can be parameterized on an arbitrary plane regardless of the original plane Ω in focus.

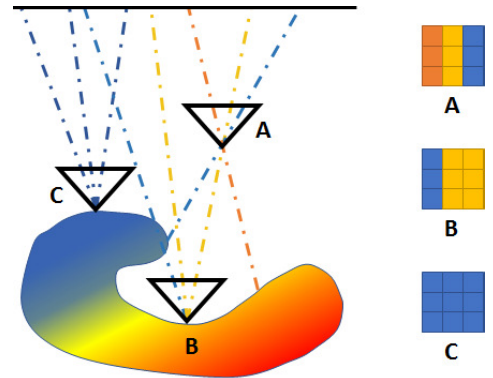


Figure 2: SCams considering different intersection points for the captured rays. The projection center of Camera A does not correspond to a surface point and, therefore, the camera collects rays that emanate from arbitrary surface points. Camera B collects rays that emanate from different points of the surface due to occlusion. Camera C collects rays that emanate from the same surface point. Adapted from Yu *et al.* [13].

and $\Psi_{\Gamma} = [q, r, u, v, 1]^T$ correspond to rays parameterized by a point (q, r) and a direction (u, v) on plane Γ . Notice that the directions remain unchanged with the re-parameterization. Mapping the lightfield in the object space $L_{\Pi}(s, t, u, v)$ to the lightfield in the image space $L(i, j, k, l)$ by the intrinsic matrix \mathbf{H} introduced by Dansereau *et al.* [5], one obtains

$$\Psi_{\Gamma} = \mathbf{D} \mathbf{H} \Phi, \quad (3)$$

with

$$\mathbf{H} = \begin{bmatrix} h_{si} & 0 & h_{sk} & 0 & h_s \\ 0 & h_{tj} & 0 & h_{tl} & h_t \\ h_{ui} & 0 & h_{uk} & 0 & h_u \\ 0 & h_{vj} & 0 & h_{vl} & h_v \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad (4)$$

where $\Phi = [i, j, k, l, 1]^T$ correspond to rays that are parameterized by pixels and microlenses indexes. The new intrinsic matrix $\mathbf{H}_{\Gamma} = \mathbf{D} \mathbf{H}$ allows to relate the lightfield in the object space $L_{\Gamma}(q, r, u, v)$ and the lightfield

in the image space $L(i, j, k, l)$. The mapping between the lightfields allows to define a constraint to identify the rays that intersect at an arbitrary point of the plane Γ . Let Φ_a and Φ_b be two rays with the same coordinates (q, r) on plane Γ , by taking their difference one defines a constraint on the lightfield coordinates to define a SCam

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \mathbf{H}_{ij}^{qr} \begin{bmatrix} i - i_r \\ j - j_r \end{bmatrix} + \mathbf{H}_{kl}^{qr} \begin{bmatrix} k - k_r \\ l - l_r \end{bmatrix}, \quad (5)$$

where (i_r, j_r, k_r, l_r) are reference coordinates to enforce the constraint, and $\mathbf{H}_{(\cdot)}^{qr}$ corresponds to 2×2 sub-matrices of \mathbf{H}_Γ obtained from selecting the entries of the first two rows, denoted by qr , and selecting either the entries of the 1st and 2nd columns, denoted by ij , or the 3rd and 4th columns, denoted by kl . Using the constraint (5) and assuming that we want to maximize the number of rays (see Section 4.1 of [10]) that define a SCam, one obtains a sampling on (i, j) for disparities lower or equal than one

$$k = k_r + \beta_{ik}(i - i_r) \wedge l = l_r + \beta_{jl}(j - j_r), \quad (6)$$

and a sampling on (k, l) for disparities greater than one

$$i = i_r + \beta_{ik}^{-1}(k - k_r) \wedge j = j_r + \beta_{jl}^{-1}(l - l_r). \quad (7)$$

The parameters $\beta_{ik} = -\frac{h_{vi}+d}{h_{sk}+d} \frac{h_{ui}}{h_{ik}}$ and $\beta_{jl} = -\frac{h_{vj}+d}{h_{tl}+d} \frac{h_{vj}}{h_{vl}}$ correspond to the disparities considered on the viewpoint images¹ for a point at depth d [10]. The sampling defined in equation (6) correspond to the sampling performed during the shearing operation defined by Tao *et al.* [11] considering $\beta_{ik} = \beta_{jl} = 1 - \frac{1}{\alpha}$.

The SCam defines a camera with an arbitrary position for the projection center but one is interested on cameras whose projection centers lie on the scene surfaces. These cameras collect rays that originate at the same surface point. On the other hand, in shearing, one wants to shear the EPIs in order to have the rays that originate at a given surface point in the same microlens, i.e. the information present in the microlenses of the sheared lightfield correspond to the SCams. Furthermore, the SCams are a generalization of the shearing operation on the lightfield by considering the sampling on (k, l) for disparities greater than one. Remember that shearing considers the sampling on (i, j) regardless of the disparity being evaluated.

3 Results

In this section, we compare the disparity maps obtained from evaluating different correspondence cues for the Greek dataset of the 4D Lightfield Benchmark [7]. The dataset is not fully analyzed, but instead a small region with 80.56% of pixels with absolute disparity values greater than one. The dataset is illustrated in Figure 3.a-b. Notice that a dense disparity map is obtained considering a disparity estimation framework that comprises several steps like filtering and refining the disparity cost volume [11] or a densification strategy like Total Variation regularization [9]. Nonetheless, in this work, we only evaluate the quality of the initial disparity map.

The disparity maps are obtained by testing different disparity hypothesis and evaluating the metrics that define the correspondence cues [3, 11] on the sheared lightfield and on the SCams. The occlusion is handled using the strategy defined in Chen *et al.* [3]. The results are exhibited on Figure 3.c, and the disparity errors are summarized in Table 1. The errors in this table discard pixels in homogeneous regions. Table 1 shows that the analysis using SCams provides more accurate results for pixels with ground truth disparity greater than one. For example, using the correspondence cue [3] (denoted as CNS), the disparity estimation improves by 64.19% for pixels with disparity greater than one.

Cues	Shearing Disparity Error		SCams Disparity Error	
	$ Disparity > 1$	$ Disparity \leq 1$	$ Disparity > 1$	$ Disparity \leq 1$
Correspondence [3, 11]	3.9819	1.4138	3.9811	1.4322
Correspondence CNS [3]	5.2353	1.7707	3.3607	3.2936

Table 1: Disparity errors obtained for the Greek dataset of the 4D Lightfield Benchmark [7] by evaluating correspondence cues [3, 11].

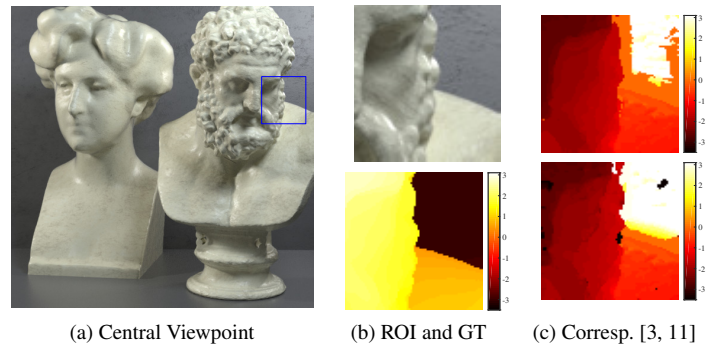


Figure 3: Greek dataset [7] used to evaluate the disparity map obtained using shearing and SCams. (a) central viewpoint image. (b) selected region and corresponding ground truth disparity map. (c) disparity maps from correspondence cue [11] applied to a subset of unoccluded rays [3] on the sheared lightfield (first row) and on the SCams (second row).

4 Conclusions

In this work, we defined the constraint that allows to obtain a SCam for a plenoptic camera using the camera model defined by Dansereau *et al.* [5], and showed that a SCam is a generalization of the shearing operation on the lightfield by introducing a sampling on the microlenses coordinates (k, l) . The capabilities of the SCams were evaluated on the Greek dataset of the 4D Lightfield Benchmark [7] using different correspondence cues [3, 11]. The results obtained suggest that the SCams are more accurate for pixels with disparity greater than one.

Acknowledgements

Work partially supported by the FCT project UID / EEA / 50009 / 2013.

References

- [1] Edward H Adelson and John Y. A. Wang. Single lens stereo with a plenoptic camera. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (2): 99–106, 1992.
- [2] Clemens Birklbauer and Oliver Bimber. Panorama light-field imaging. In *Computer Graphics Forum*, volume 33, pages 43–52. Wiley Online Library, 2014.
- [3] Can Chen, Haiting Lin, Zhan Yu, Sing Bing Kang, and Jingyi Yu. Light field stereo matching using bilateral statistics of surface cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1518–1525, 2014.
- [4] Don Dansereau and Len Bruton. Gradient-based depth estimation from 4d light fields. In *Circuits and Systems, 2004. ISCAS'04. Proceedings of the 2004 International Symposium on*, volume 3, pages III–549. IEEE, 2004.
- [5] Donald G Dansereau, Oscar Pizarro, and Stefan B Williams. Decoding, calibration and rectification for lenselet-based plenoptic cameras. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1027–1034, 2013.
- [6] Maximilian Diebold and Bastian Goldluecke. Epipolar plane image refocusing for improved depth estimation and occlusion handling. 2013.
- [7] Katrin Honauer, Ole Johannsen, Daniel Kondermann, and Bastian Goldluecke. A dataset and evaluation methodology for depth estimation on 4d light fields. In *Asian Conference on Computer Vision*, pages 19–34. Springer, 2016.
- [8] Marc Levoy and Pat Hanrahan. Light field rendering. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 31–42. ACM, 1996.
- [9] Nuno Barroso Monteiro, Joao Pedro Barreto, and José Gaspar. Dense light-field disparity estimation using total variation regularization. In *International Conference Image Analysis and Recognition*, pages 462–469. Springer, 2016.
- [10] Nuno Barroso Monteiro, Simão Marto, João Pedro Barreto, and José Gaspar. Depth range accuracy for plenoptic cameras. *Computer Vision and Image Understanding*, 2018.
- [11] Michael W Tao, Sunil Hadap, Jitendra Malik, and Ravi Ramamoorthi. Depth from combining defocus and correspondence using light-field cameras. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 673–680, 2013.
- [12] Sven Wanner and Bastian Goldluecke. Variational light field analysis for disparity estimation and super-resolution. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(3):606–619, 2014.
- [13] Jingyi Yu, Leonard McMillan, and Steven Gortler. Scam light field rendering. In *Computer Graphics and Applications, 2002. Proceedings. 10th Pacific Conference on*, pages 137–144. IEEE, 2002.

¹A viewpoint or sub-aperture image is obtained by selecting and combining the rays that reach the same pixel of each microlens, i.e. by selecting the pixel (i, j) of each microlens (k, l) .