Surface Cameras from Shearing for Disparity Estimation on a Lightfield

Nuno Barroso Monteiro
nmonteiro@isr.tecnico.ulisboa.pt
João Pedro Barreto
jpb@isr.uc.pt
José Gaspar
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 \( L_\Omega(s,t,u,v) \) acquired by a plenoptic camera with the plane \( \Omega \) in focus (Figure 1), one can obtain a SCam with projection center at point \((q,r)\) of plane \( \Gamma \). \( L_\Omega(s,t,u,v) \) collects rays \( \Psi=[s,t,u,v] \) that are parameterized using a point \((s,t)\) and a direction \((u,v)\) defined on a plane \( \Omega \) in metric units [10]. Assuming that the plane \( \Gamma \) is at a distance \(d\) from the plane \( \Omega \), one can re-parameterize the lightfield captured by the plenoptic camera relatively to the plane \( \Gamma \) [2], \( L_{\Gamma}(q,r,u,v) \), by

\[
\Psi_{\Gamma} = D \Psi,
\]

where

\[
D = \begin{bmatrix}
1 & 0 & d & 0 & 0 \\
0 & 1 & 0 & d & 0 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix},
\] (2)

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

\[
\Psi_{\Gamma} = D H \Phi ,
\] (3)

with

\[
H = \begin{bmatrix}
h_{\text{u}1} & 0 & h_{\text{u}2} & 0 & h_{\text{u}3} \\
0 & h_{\text{v}1} & 0 & h_{\text{v}2} & h_{\text{v}3} \\
h_{\text{u}4} & 0 & h_{\text{u}5} & 0 & h_{\text{u}6} \\
0 & h_{\text{v}4} & 0 & h_{\text{v}5} & h_{\text{v}6} \\
0 & 0 & 0 & 0 & 1
\end{bmatrix},
\] (4)

where \( \Phi = [i,j,k,l] \) correspond to rays that are parameterized by pixels and microlenses indexes. The new intrinsic matrix \( H_{\Gamma} = D H \) 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 \( \Gamma \). Let \( \Phi_1 \) and \( \Phi_2 \) be the two rays with the same coordinates \((q,r)\) on plane \( \Gamma \), by taking their difference one defines a constraint on the lightfield coordinates to define a SCam

\[
0 = H^r_{ij} [i - i'] + H^r_{jl} [j - j'],
\]

where \((i', j', k, l)\) are reference coordinates to enforce the constraint, and \(H^r_{ij}\) corresponds to \(2 \times 2\) sub-matrices of \(H^r\) obtained from selecting the entries of the first two rows, denoted by \(qr\), and selecting either the entries of the \(1^{\text{st}}\) and \(2^{\text{nd}}\) columns, denoted by \(ij\), or the \(3^{\text{rd}}\) and \(4^{\text{th}}\) 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

\[
k = k_0 + \beta_k(i - i_0) \land l = l_0 + \beta_l(j - j_0),
\]

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

\[
i = i_0 + \beta_i^{-1}(k - k_0) \land j = j_0 + \beta_j^{-1}(l - l_0).
\]

The parameters \(\beta_k = -\frac{h_{ij} - d}{h_{ij} - d_{ij}}\) and \(\beta_l = -\frac{h_{ij} + d}{h_{ij} + d_{ij}}\) correspond to the disparities considered on the viewpoint images \([3]\) for a point at depth \(d\) [10]. The sampling defined in equation (6) correspond to the sampled performing during the shearing operation defined by Tao et al. [11] considering \(\beta_k = \beta_l = 1 - \frac{1}{d}\).

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% 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.

![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).](image)

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