Dense Lightfield Disparity Estimation using Total Variation Regularization

Nuno Barroso Monteiro[1,2], João Pedro Barreto[2], and José Gaspar[1]

[1] Institute for Systems and Robotics, Univ. of Lisbon, Portugal
{nmonteiro,jag}@isr.tecnico.ulisboa.pt
[2] Institute for Systems and Robotics, Univ. of Coimbra, Portugal
jpbar@isr.uc.pt

Abstract. Plenoptic cameras make a trade-off between spatial and angular resolution. The knowledge of the disparity map allows to improve the resolution of these cameras using superresolution techniques. Nonetheless, the disparity map is often unknown and must be recovered from the lightfield captured. Hence, we focus on improving the disparity estimation from the structure tensor analysis of the epipolar plane images obtained from the lightfield. Using an hypercube representation, we formalize a data fusion problem with total variation regularization using the Alternating Direction Method of Multipliers. Assuming periodic boundary conditions allowed us to integrate the full 4D lightfield efficiently using the frequency domain. We applied this methodology to a synthetic dataset. The disparity estimations are more accurate than those of the structure tensor.

Keywords: Lightfield, Disparity, Denoising, ADMM, Total Variation

1 Introduction

The plenoptic cameras allow to capture the direction and contribution of each ray to the total amount of light captured on an image. These cameras sample the plenoptic function.

The original 7D plenoptic function can be simplified into a 4D lightfield, the lumigraph $\mathrm{L}(u, v, x, y)$ [7]. The lightfield is parameterized describing a ray by its intersection with two planes. In this parameterization a light ray intersects the first plane at coordinates (u, v) and then intersects a second plane at (x, y). The first pair of coordinates define the location of the ray and the second pair of coordinates allows to define the direction of the ray.

The image sensors are flat, therefore, this 4D space must be mapped into a 2D space, limiting the spatial and angular resolution [6]. Superresolution techniques are useful to overcome this limitation. These techniques can be applied if the images are shifted by sub-pixel amounts between successive images [8] or if there is *a priori* knowledge of the camera geometry to obtain this sub-pixel accuracy [5]. In plenoptic cameras we do not know the correspondence between the image views since the depth map, and consequently the disparity map, of the scene

is unknown. Therefore, the knowledge of the disparity map is a requirement for superresolution techniques. In this work, we will focus on improving and recovering the disparity map from the lightfield captured.

2 Related Work

The problem of depth or disparity estimation is widely studied in computer vision. In the recent years, with the appearance of commercial versions this subject has been studied more intensively in plenoptic cameras.

Adelson and Wang [1] applied their lenticular array setup to estimate depth using the different views provided by the sensor by observing the vertical and horizontal parallax between the different views. Other strategies consider that the object becomes blurred according to their distance from the plane of focus [9]. More recent approaches, estimate depth from the epipolar plane images (EPI) obtained from the lightfield using the structure tensor analysis [15, 13, 4]. In these images, points in space are projected onto lines which can be more robustly detected than point correspondences.

Wanner et al. [15] proposed a global optimization framework that is based on the regularization and integration of the disparity maps obtained from the EPIs. This framework can be preceded of a labeling scheme to impose visibility constraints that imply a discretization of the disparity values. This step is computationally expensive and the discretization reduces the accuracy of the disparity estimation. Hence, Wanner et al. [13] considered a more efficient approach by performing a fast denoising scheme of a disparity estimation that combines the disparities obtained from two slices with orthogonal directions according to the coherence measurement of the structure tensor. In this method, the occlusion is handled by merging the information of horizontal and vertical EPIs. Although the results are good, the allowed disparity range is small and the occlusion boundaries are noisy. Diebold et al. [4] proposed a method that allows to extend the allowed disparity range using a refocusing scheme. The refocus of the EPIs to virtual depth layers allows to accommodate the orientation of the lines in the EPIs to the allowed disparity range. The results are then integrated into a global disparity map using a metric for occlusion handling that is based on a coherence measurement.

In this work, we propose a variational method to regularize and integrate the 2D EPIs of the full 4D lightfield without the discretization of the disparity values. This approach is computed efficiently by considering periodic boundary conditions that allows us to use Fast Fourier Transforms (FFTs).

3 Disparity Estimation from Lumigraph

Let us consider the 4D lightfield or lumigraph introduced in section 1. The lumigraph $\mathrm{L}(u, v, x, y)$ maps an intensity value to the light ray whose direction is defined by the intersection with the main lens plane at (u, v) and the image

plane at (x, y) where u,v are the viewpoint coordinates, and x,y are the image coordinates.

The sampling of the lightfield and the configuration of the plenoptic camera satisfies the conditions presented by Bolles et al. [3] for constructing an EPI (Fig. 1). In the EPI, a point in space is projected onto a line with a slope $\Delta x/\Delta u$ proportional to its disparity. Considering the geometry defined by the EPI, the disparity d of a point in space is given by $d = f/z = \Delta x/\Delta u$, where z is the depth of the point, and f is the distance between the main lens plane and the image plane. To compute the slopes of the lines in the EPIs, we will use the structure tensor analysis similarly to Diebold et al. [4] and Wanner et al. [13].

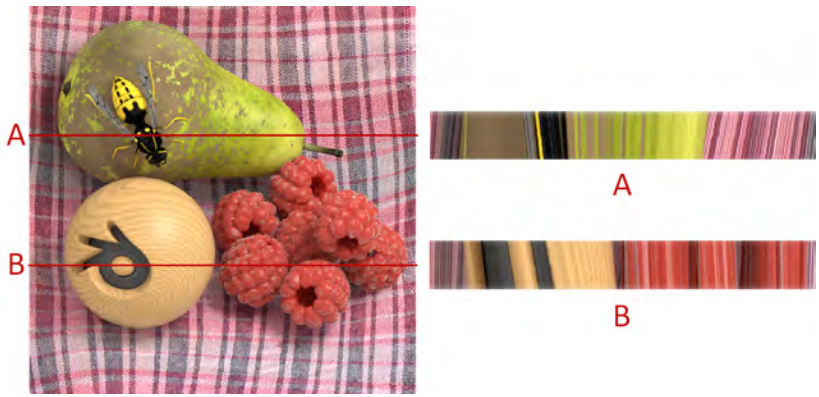

Fig. 1. Epipolar plane images obtained from slicing and stacking the sequence of viewpoint images at position A and B for the *Still* dataset [14].

4 Disparity Estimation Regularization

The EPI analysis allows to retrieve disparity information from a static scene for each pixel. The approaches used, normally, depend on gradient computations that increase the noise of the original image. Therefore, denoising using regularization is a useful approach to obtain a more accurate disparity estimation.

Let us consider the lumigraph $\mathbf{L} \in \mathbb{R}^{n_u \times n_v \times n_x \times n_y}$, where n_u,n_v are the horizontal and vertical angular resolutions, and n_x,n_y, are the horizontal and vertical spatial resolutions. For each ray in the lumigraph we can assign a disparity value performing an analysis of the EPIs. Therefore, the disparity values have the same dimensionality of the lumigraph, i.e., $\mathbf{D} \in \mathbb{R}^{n_u \times n_v \times n_x \times n_y}$.

Let us consider an alternative representation for the disparity structure, an hypercube $\mathbf{H}$. The hypercube is a set of datacubes $\mathrm{C}_{v_k}(u, x, y) = \mathrm{D}(u, v_k, x, y)$ that are obtained fixing one of the angular coordinates. These datacubes $\mathbf{C}_{(\cdot)}$

consist on a vertical stack of the disparity observations obtained from the EPIs (Fig. 2). Another observation of the hypercube can be obtained by fixing u. Although they represent the same object, the disparity observations may differ due to the nature of the structure tensor.

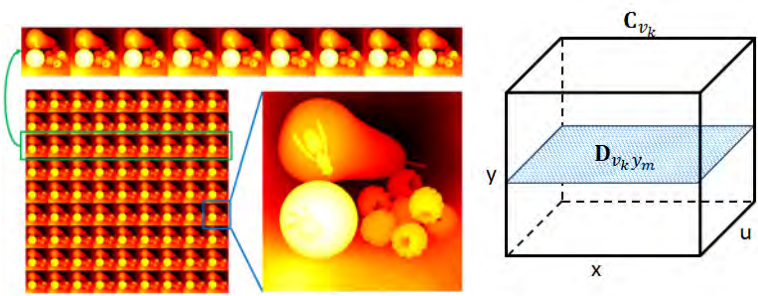

Fig. 2. Disparity hypercube representation. (A) Hypercube and datacube (green) represented as matrices, and image (blue) as viewed from a given viewpoint. (B) Datacube structure for fixed $v = v_k$.

For an easier notation, let us consider that the disparity observations from a datacube $\mathbf{C}_{v_k}$ can be represented as a two-dimensional matrix, where each line corresponds to the disparity retrieved from each of the pixels of the EPI, lexicographically ordered (Fig. 2). Let the matrix representing the observed disparity be $\mathbf{Y}_{v_k} \in \mathbb{R}^{n_y \times (n_x \times n_u)}$. Assuming that our observations are only affected by i.i.d. additive noise $\mathbf{W}_{v_k} \in \mathbb{R}^{n_y \times (n_x \times n_u)}$, we can model our disparity observations as $\mathbf{Y}_{v_k} = \mathbf{Z}_{v_k} + \mathbf{W}_{v_k}$, where $\mathbf{Z}_{v_k} \in \mathbb{R}^{n_y \times (n_x \times n_u)}$ are the real disparities from the datacube. For simplicity, we assume that the boundary conditions are periodic. This allows to compute convolutions and matrix inversions using FFTs. The previous observation model can be generalized for the hypercube $\mathbf{H}$ by including the datacubes $\mathbf{C}_{v_k}, k = 1, \ldots, n_v$ (Fig. 2).

$$\mathbf{Y}_v = \mathbf{M}_v \mathbf{Z} + \mathbf{W}_v \tag{1}$$

with $\mathbf{Y}_v, \mathbf{Z}, \mathbf{W}_v \in \mathbb{R}^{(n_y \times n_v) \times (n_x \times n_u)}$ resulting from the vertical stacking of the matrices $\mathbf{Y}_{v_k}, \mathbf{Z}_{v_k}$, and $\mathbf{W}_{v_k}$ for $k = 1, \ldots, n_v$, respectively. $\mathbf{M}_v \in \mathbb{R}^{(n_y \times n_v) \times (n_y \times n_v)}$ correspond to a uniform subsampling of $\mathbf{Z}$. This allows to include disparity observations from several viewpoints while considering the same virtual camera motion to obtain the EPIs. Similarly, we can obtain the observation model for the hypercube observation obtained by fixing u, $\mathbf{Y}_u = \mathbf{Z}^T \mathbf{M}_u^T + \mathbf{W}_u$.

These structures (Fig. 2) represent a repeating sequence of disparity images that differ in a small number of pixels due to the different viewpoint coordinates. Since we obtain natural like images we propose to apply an isotropic total

variation regularizer proposed by Rudin et al. [11] to promote sharp discontinuities at edges. This type of regularizer has already been used in the context of lightfield analysis [13]. This leads to an optimization problem that is similar to the one presented by Simões et al. [12] in the context of hyperspectral cameras superresolution:

$$\hat{\mathbf{Z}} = \arg\min_{\mathbf{Z}} \frac{1}{2}\|\mathbf{Y}_v - \mathbf{M}_v\mathbf{Z}\|_F^2 + \frac{\lambda_u}{2}\left\|\mathbf{Y}_u^T - \mathbf{M}_u\mathbf{Z}\right\|_F^2 + \lambda_r \mathrm{TV}\left(\mathbf{Z}\mathbf{D}_h, \mathbf{Z}\mathbf{D}_v\right) \quad (2)$$

where $\|\cdot\|_F = \sqrt{tr\left[(\cdot)(\cdot)^T\right]}$ corresponds to the Frobenius norm, TV corresponds to the isotropic total variation regularizer [11], and $\mathbf{D}_h$ and $\mathbf{D}_v$ are matrices that allow to compute the horizontal and vertical discrete differences considering periodic boundary conditions, respectively. In this optimization problem, the first two terms are data-fitting terms while the last term is the regularizer. The data terms should explain the observed disparities considering the observation models for $\mathbf{Y}_v$ and $\mathbf{Y}_u$. The weights λ_u and λ_r allows to control the contribution of each of the terms.

The solution for this optimization problem is obtained using an Alternating Direction Method of Multipliers (ADMM) instance, the Split Augmented Lagrangian Shrinkage Algorithm (SALSA) [2]. Thus, the optimization variable $\mathbf{Z}$ is split into auxiliary variables using the variable splitting technique. The optimization problem (2) is now defined by:

$$\begin{aligned} \hat{\mathbf{Z}} = \arg\min_{\mathbf{Z}} \tfrac{1}{2}\|\mathbf{Y}_v - \mathbf{M}_v\mathbf{V}_1\|_F^2 + \tfrac{\lambda_u}{2}\left\|\mathbf{Y}_u^T - \mathbf{M}_u\mathbf{V}_2\right\|_F^2 + \lambda_r \mathrm{TV}\left(\mathbf{V}_3, \mathbf{V}_4\right) \\ \text{subject to} \quad \mathbf{V}_1 = \mathbf{Z}, \quad \mathbf{V}_2 = \mathbf{Z}, \quad \mathbf{V}_3 = \mathbf{Z}\mathbf{D}_h, \quad \mathbf{V}_4 = \mathbf{Z}\mathbf{D}_v \end{aligned} \quad (3)$$

Considering $\mathrm{f}\left(\mathbf{V}\right) = \frac{1}{2}\|\mathbf{Y}_v - \mathbf{M}_v\mathbf{V}_1\|_F^2 + \frac{\lambda_u}{2}\left\|\mathbf{Y}_u^T - \mathbf{M}_u\mathbf{V}_2\right\|_F^2 + \lambda_r \mathrm{TV}\left(\mathbf{V}_3, \mathbf{V}_4\right)$, $\mathbf{V} = \left[\mathbf{V}_1^T\ \mathbf{V}_2^T\ \mathbf{V}_3^T\ \mathbf{V}_4^T\right]^T$, and $\mathbf{G} = \left[\mathbf{I}\ \mathbf{I}\ \mathbf{D}_h^T\ \mathbf{D}_v^T\right]^T$, the problem has the following Augmented Lagrangian [10]:

$$\mathcal{L}(\mathbf{Z}, \mathbf{V}, \mathbf{A}) = \mathrm{f}\left(\mathbf{V}\right) + \frac{\mu}{2}\left\|\mathbf{G}\mathbf{Z}^T - \mathbf{V} - \mathbf{A}\right\|_F^2 \quad (4)$$

where μ is a positive constant called the penalty parameter. Now, we are able to apply SALSA. Considering as input the observations $\mathbf{Y}_v$ and $\mathbf{Y}_u$, the parameters λ_u, λ_r and μ, and the initializations for $\mathbf{V}^{(0)}$ and $\mathbf{A}^{(0)}$, we will solve the following optimizations at each iteration k until a stopping criterion is satisfied:

$$\begin{aligned} \mathbf{Z}^{(k+1)} &= \arg\min_{\mathbf{Z}} \quad \mathcal{L}(\mathbf{Z}, \mathbf{V}^{(k)}, \mathbf{A}^{(k)}) \\ \mathbf{V}^{(k+1)} &= \arg\min_{\mathbf{V}} \quad \mathcal{L}(\mathbf{Z}^{(k+1)}, \mathbf{V}, \mathbf{A}^{(k)}) \end{aligned} \quad (5)$$

The algorithm described above has a matrix $\mathbf{G}$ with full column rank (due to the presence of identity matrix $\mathbf{I}$), and the function $\mathrm{f}\left(\cdot\right)$ is a sum of closed, proper, and convex functions. Therefore, the conditions for the convergence of SALSA established in [2] are met.

5 Experimental Results

The methodology described in the previous sections was applied to the *Still* synthetic dataset provided by the HCI Heidelberg Group [14] (Fig. 2). The results of the regularization with one and two data terms are compared with the disparity estimates from the structure tensor. The ground truth provided with the synthetic dataset is used to determine the PSNR values. Since the ground truth values correspond to depth measurements and our values are disparity measurements, we converted the depth ground truth to disparity. The results are depicted in Fig. 3.

The structure tensor analysis was performed assuming an equal contribution for each of the color channels. Also, the smoothing of the image and the components of the structure tensor was obtained by applying Gaussian distributions with standard deviation 0.8 and 3.2, respectively. To compute the derivatives we considered the 3×3 Sobel mask. For the optimization problem, we consider an equal contribution for each of the data terms ($\lambda_u = 1$), and the penalty parameter μ to be fixed and equal to 1 since it only affects the convergence speed and not the convergence. The parameter λ_r was fine tuned by performing a denoising of the disparity ground truth and selecting the one that provides the highest PSNR.

From Fig. 3, we can see that the hypercube obtained from the structure tensor analysis presents a noticeable decay on accuracy in the peripheral viewpoints. Focusing on a specific viewpoint (peripheral or central), we can conclude that the depth accuracy also depends on the region of the image. The disparity estimates are less accurate on homogeneous regions, which in the EPI represent regions of constant intensity between the lines that we want to detect. A small change of intensity in these regions lead to disparity estimates that change rapidly and have high variability. This is more noticeable on the 3D representation of the disparity estimates (Fig. 3.B and 3.D).

The hypercube after the regularization has reduced noise and the accuracy remains almost the same from the central to the peripheral viewpoints, which confirms the statement of section 4. Indeed, if we compare the 3D representations of the disparity estimates we can see that the noise is significantly reduced. This is confirmed by the increased PSNR after regularization (from 8.65 dB to 10.76 dB). The noise can be further reduced by considering the additional data term of the optimization problem (2) (PSNR of 11.00 dB).

Furthermore, the formulation allows to select specific disparity measurements for each of the observations of the hypercube through the matrices $\mathbf{M}_v$ and $\mathbf{M}_u$. Therefore, we performed the same analysis but now considering only the disparity observations with higher coherence values from each hypercube observation. In this scenario, a compromise between the two hypercube observations will only occur for disparities with similar coherence values between the two observations instead of achieving this compromise for all disparities. Hence, this approach leads to an increase in the PSNR value for 11.38 dB.

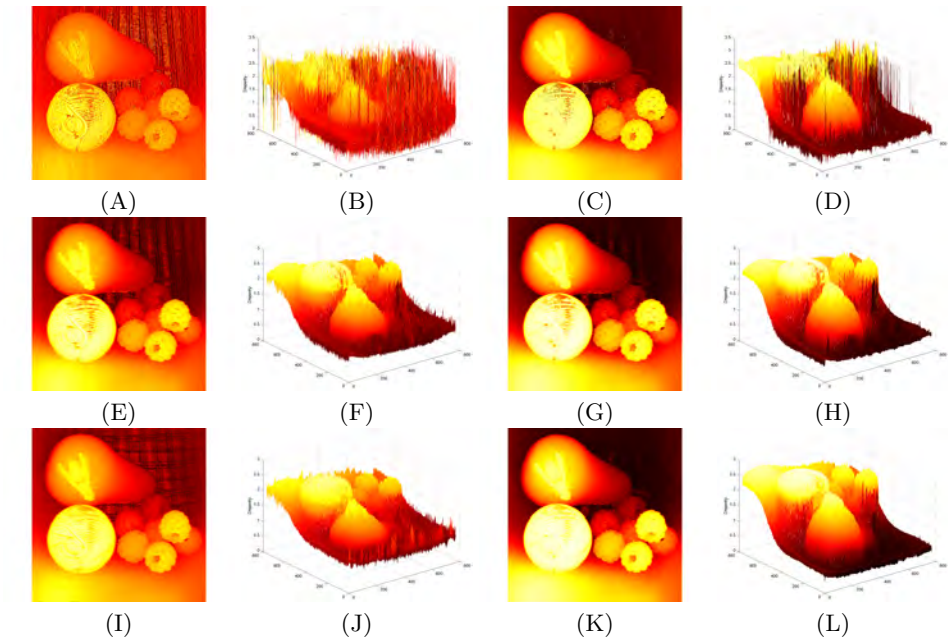

Fig. 3. Disparity estimation using the structure tensor analysis (top) and after regularization with one (middle) and two data terms (bottom). The columns correspond to the disparity estimates using a 2D representation (1^{st} and 3^{rd} columns) and a 3D representation (2^{nd} and 4^{th} columns) to highlight the noise. Peripheral viewpoints are depicted on the 1^{st} and 2^{nd} columns. Central viewpoints are depicted on the 3^{rd} and 4^{th} columns.

6 Conclusions

In this work, we formalize a data fusion problem which uses the full 4D lightfield. The optimization problem was solved by resorting to an ADMM instance that provides good results with few iterations. Furthermore, we considered simplifications to the boundary conditions that allowed to use FFTs in the computations. Therefore, the algorithm is computationally efficient. This methodology was applied to a synthetic dataset and allows to obtain estimates that are more accurate and robust to noise. The results are improved when we consider only the disparities with higher coherence values in each of the hypercube observations.

The formulation for the data fusion problem can still be improved. In the formulation, we are considering the full 4D lightfield information for the disparities but we are not considering blurring in the observation models. Additionally, the initial disparity estimations can also be improved. The disparity observations were obtained from the structure tensor analysis of the EPIs, but we did not consider refocusing or occlusion handling [4] to improve the disparity observations.

Acknowledgments

We would like to thank Prof. José Bioucas-Dias for the discussions that allowed us to improve our work. This work has been supported by the Portuguese Foundation for Science and Technology (FCT) project [UID / EEA / 50009 / 2013] and by the CMU-Portugal Project Augmented Human Assistance (AHA) [CMUP-ERI / HCI / 0046 / 2013]. Nuno Barroso Monteiro is funded by FCT PhD grant PD/BD/105778/2014.

References

1. Adelson, E.H., Wang, J.Y.A.: Single lens stereo with a plenoptic camera. IEEE Transactions on Pattern Analysis & Machine Intelligence (2), 99–106 (1992)
2. Afonso, M.V., Bioucas-Dias, J.M., Figueiredo, M.A.: An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems. Image Processing, IEEE Transactions on 20(3), 681–695 (2011)
3. Bolles, R.C., Baker, H.H., Marimont, D.H.: Epipolar-plane image analysis: An approach to determining structure from motion. International Journal of Computer Vision 1(1), 7–55 (1987)
4. Diebold, M., Goldluecke, B.: Epipolar plane image refocusing for improved depth estimation and occlusion handling (2013)
5. Georgiev, T., Lumsdaine, A.: Superresolution with plenoptic camera 2.0. Adobe Systems Incorporated, Tech. Rep (2009)
6. Georgiev, T., Zheng, K.C., Curless, B., Salesin, D., Nayar, S., Intwala, C.: Spatio-angular resolution tradeoffs in integral photography. Rendering Techniques 2006, 263–272 (2006)
7. Gortler, S.J., Grzeszczuk, R., Szeliski, R., Cohen, M.F.: The lumigraph. In: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques. pp. 43–54. ACM (1996)
8. Ng, M.K., Yau, A.C.: Super-resolution image restoration from blurred low-resolution images. Journal of Mathematical Imaging and Vision 23(3), 367–378 (2005)
9. Ng, R., Levoy, M., Brédif, M., Duval, G., Horowitz, M., Hanrahan, P.: Light field photography with a hand-held plenoptic camera. Computer Science Technical Report CSTR 2(11), 1–11 (2005)
10. Nocedal, J., Wright, S.: Numerical optimization. Springer Science & Business Media (2006)
11. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena 60(1), 259–268 (1992)
12. Simões, M., Bioucas-Dias, J., Almeida, L.B., Chanussot, J.: A convex formulation for hyperspectral image superresolution via subspace-based regularization. Geoscience and Remote Sensing, IEEE Transactions on 53(6), 3373–3388 (2015)
13. Wanner, S., Goldluecke, B.: Variational light field analysis for disparity estimation and super-resolution. Pattern Analysis and Machine Intelligence, IEEE Transactions on 36(3), 606–619 (2014)
14. Wanner, S., Meister, S., Goldluecke, B.: Datasets and benchmarks for densely sampled 4d light fields. In: VMV. pp. 225–226. Citeseer (2013)
15. Wanner, S., Straehle, C., Goldluecke, B.: Globally consistent multi-label assignment on the ray space of 4d light fields. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1011–1018 (2013)