DISPARITY-GUIDED DEMOSAICKING OF LIGHT-FIELD IMAGES

Mozhdeh Seifi, Neus Sabater, Valter Drazic and Patrick Pérez

Technicolor, 975 avenue des Champs Blancs, 35576 Cesson-Sévigné, France

ABSTRACT

Light-field imaging has been recently introduced to mass market by the hand held plenoptic camera Lytro. Thanks to a microlens array placed between the main lens and the sensor, the captured data contains different views of the scene from different view points. This offers several post-capture applications, e.g., computationally changing the main lens focus. The raw data conversion in such cameras is however barely studied in the literature. The goal of this paper is to study the particularly overlooked problem of demosaicking the views for plenoptic cameras such as Lytro. We exploit the redundant sampling of scene content in the views, and show that estimated disparities from the mosaicked data can guide the demosaicking, resulting in minimum artifacts compared to the state of art methods. Besides, by properly addressing the demultiplexing step, we take the first step towards light-field super-resolution with negligible computational overload.

Index Terms— plenoptic camera, multi-frame demosaicking, disparity estimation.

1. INTRODUCTION

Capturing the scene's light-field has been an old interest in the field of computational photography [1, 2]. However, the recent release of hand held plenoptic cameras as Lytro[1] has introduced the potentials of light-field imaging to the mass market. By placing a microlens array between the main lens and the sensor, a plenoptic camera captures the direction of the light bundles that enter the camera, in addition to their intensity and color. Captured data is then demultiplexed to provide the light-field, a matrix of horizontally and vertically aligned views from slightly different points of view over the scene. With the light-fields, a number of natural applications have risen such as depth estimation [3, 4, 5] or post-capture refocusing [6]. However, the angular resolution of the plenoptic cameras comes at the price of lower spatial resolution of images. But promising super-resolution methods for plenoptic images have already been proposed [7, 3, 8, 4].

Among the state of art post-processing methods of the plenoptic data, only very few address the very first steps regarding raw data conversion: (i) *demosaicking*, that aims to recover the color content of the scene from the mosaicked captured raw data (discussed only by [9, 7]) and (ii) *demultiplexing*, that consists in reordering raw image pixels based on microlenses positions in order to recover the matrix of views (discussed only by [10, 11, 12] for Lytro).

Most of the works in the literature propose to first demosaick the raw data and then demultiplex to recover the views, but we shall show that this leads to color artifacts on the views. Indeed, by construction, neighbor pixels in a plenoptic raw image contain different angular information (each pixel under a microlens corresponds to a different view). So, demosaicking the raw plenoptic image as it was a conventional image wrongly mixes angular information. Indeed, classical algorithms interpolate neighbor color values creating the so-called *view cross-talk* artifacts. Furthermore, it has been shown in [12] that disparity estimation from views obtained from the demosaicked raw image is prone to tremendous errors. For this reason, we build on the work in [12] in which the raw image is demultiplexed without demosaicking and we study how to recover the full RGB views. This means that demosaicking is done on the views and not on the raw data. Note that the demultiplexing step (pixel reordering) transforms the Bayer pattern on the raw data in a new color pattern on the views (see Fig. 1, Fig. 3-(1) and [12] for more details). On this new color pattern classical demosaicking algorithms poorly recover highly textured areas. Hence, in this paper we propose a new demosaicking algorithm specifically designed for plenoptic data and inspired by multi-frame demosaicking approaches [13]. The goal is to increase the chromatic resolution of one target view, exploiting the redundant sampling of the scene using the other low-resolution views. In particular, our strategy is to estimate and use pixel disparities to guide our demosaicking algorithm.

It should be noted that most of the state of art methods demosaick raw data before demultiplexing. The only exceptions are [9, 7] who work with focused plenoptic cameras [14]. In contrary to unfocused plenoptic cameras, the focused ones have the microlens array placed at a distance different from the microlens focal lengths, increasing the complexity of view demultiplexing. Therefore [9, 7] demosaick final full-aperture images of the scene, benefiting from their high resolution. However, view demosaicking without the burden of calculating the full-resolution image is not discussed in such works. In addition, none of the mentioned works explicitly address the quinqunx sampling of the light-field.

[1] http://www.lytro.com

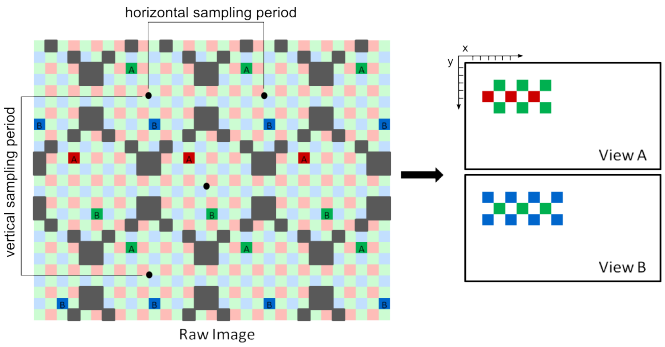

Fig. 1. Demultiplexing as in [12]. The matrix of views has as many views as number of pixels per microlens, but only two views are shown here for the sake of visualization.

2. DISPARITY-GUIDED DEMOSAICKING OF THE LIGHT-FIELD DATA

For conventional cameras, demosaicking methods exploit the regularity of the color-sampling pattern to locally estimate the color content of a captured scene [15]. However, demosaicking views with color patterns as in Fig. 3-(1) is not straightforward considering the loss of the local chromatic content in relatively large neighborhoods. So, given this color pattern, we discard using the well-known methods for conventional cameras. Instead, for each pixel of the target view we propose to gather all color information from all the pixels imaging the same point in the scene (see a schema of the proposed demosaicking approach in Fig. 2).

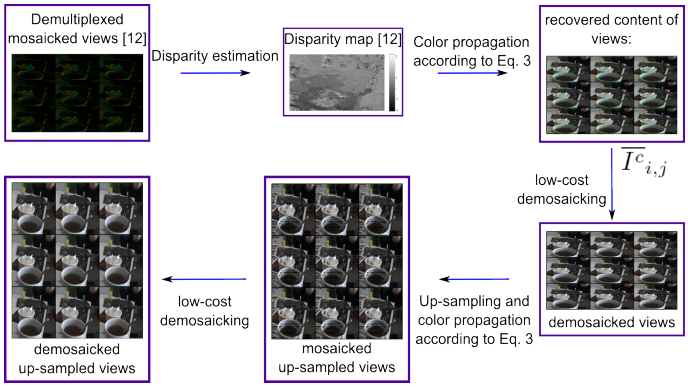

Fig. 2. Schema of the proposed demosaicking approach.

After demultiplexing, the sampled ligh-field is represented by a matrix of views of size $n \times m$ (see Fig. 1). Let $I_{i,j}$ be the view in position (i, j) in the matrix of views and $I_{k,j}$ another view in the same row of the matrix of views. Then, the disparity map between the two views is calculated with an adapted block-matching method to plenoptic images as in [12]. For a given point $\mathbf{x} = (x, y)$ of $I_{i,j}$, its corresponding point in $I_{k,j}$ is $\mathbf{x}' := \mathbf{x} + d_{ik}(\mathbf{x}) = (x + d^x_{ik}(\mathbf{x}), y)$ where $d_{ik}(\mathbf{x}) = (d^x_{ik}(\mathbf{x}), 0)$ is the disparity function. Note that the disparity second coordinate is null because the two views are in the same row. Similarly, when the two views are in the same column the disparity first coordinate is null.

Now, the estimated disparity at a certain point is considered as reliable if the disparity of the corresponding point is similar in absolute value. More precisely, $d_{ik}(\mathbf{x})$ is reliable if

$$|d^x_{ik}(\mathbf{x}) + d^x_{ki}(\mathbf{x}')| \leqslant \varepsilon\,. \tag{1}$$

In the binocular stereovision literature, this condition is known as the right-left coherence check [16]. In the plenoptic framework, we estimate the disparity using several pairs of views from the same row. So, the *Number of Reliable Estimations* (NRE) of $\mathbf{x}$ is the number of reliable disparities over all the possible view pairs in each row:

$$\text{NRE}(\mathbf{x}) = \sum_{\substack{k=1\ldots m \\ k\neq i}} \chi\left[d_{ik}(\mathbf{x})\right]\,, \tag{2}$$

where χ is the characteristic function (equal to 1 if Eq. (1) stands and 0 otherwise).

Thus, points with high NRE (highly reliable disparities) are used to propagate color channel values among different views of the light-field. It can be easily shown that, in a plenoptic camera, the horizontal disparity of one pixel on $I_{i,j}$ with respect to $I_{i+a,j}$ is equal to its vertical disparity with respect to $I_{i,j+a}$ with $a \in \mathbb{N}$. The color propagation is then performed both horizontally and vertically in the matrix of views. In particular, if the value of the color channel c of pixel $\mathbf{x}$ on view $I_{i,j}$ is missing, then it is recovered if at least one of the corresponding pixels on the other views contain information in channel c (i.e., it does not correspond to the empty pixels inserted to address the quinqunx light-field sampling). Formally, we initialize for each color channel c the intermediate recovered view $\overline{I}^c_{i,j} = I^c_{i,j}$. Then, we recover color information from the other views as follows

$$\overline{I}^c{}_{i,j}(\mathbf{x}) = \begin{cases} \emptyset & \text{if } NRE(\mathbf{x}) \leqslant \tau \\ \underset{\substack{u=1\ldots n \\ v=1\ldots m}}{\text{Mean}} \left\{ I^c_{u,v}(\mathbf{x}') \notin \emptyset \right\} & \text{otherwise} \end{cases} \tag{3}$$

where $\mathbf{x}' = \mathbf{x} + \left(d^x_{iu}(\mathbf{x}), d^y_{jv}(\mathbf{x})\right)$ denote the corresponding points in the other views. In this case, the disparities d^x_{iu} and d^y_{jv} are the horizontal and vertical reliable disparities. Remark that points $\mathbf{x}$ such that $\text{NRE}(\mathbf{x}) \leqslant \tau$ do not have any new color assigned. Therefore, the intermediate recovered view $\overline{I}_{i,j}$ has considerably more colors than $I_{i,j}$ but it is still not a full RGB image.

It should be noted that non-recovered colors in $\overline{I}_{i,j}$ mostly belong to homogeneous areas since block-matching methods are manly accurate in textured areas. Also, a plenoptic camera like Lytro provides views with very small baselines, meaning

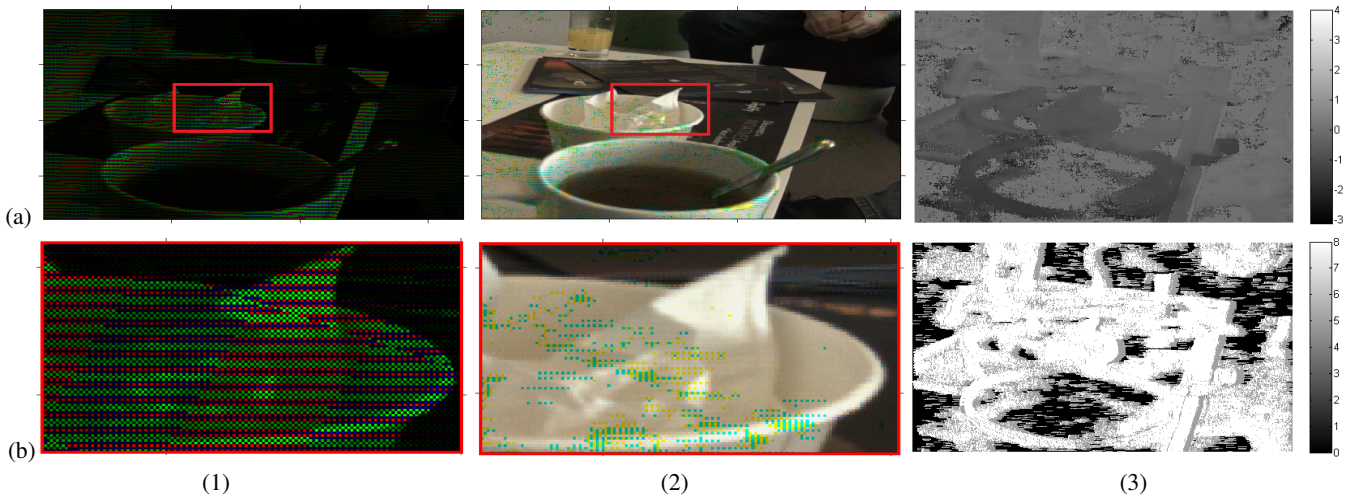

Fig. 3. Disparity-guided color recovery. (1) Top: The mosaicked view $I_{4,4}$. Bottom: zoom-in of the red rectangle. (2) Top: the disparity-guided color propagation view $\overline{I}_{4,4}$ in which more than 90% of the missing color values are recovered. Visible green-yellow pixels correspond to the still missing color channel values(see Fig. 4(a) for final results). Bottom: zoom-in of the red rectangle. (3) Top: Disparity map of the red rectangle. Bottom: NRE map corresponding to the same region.

that the possible occlusions are negligible and small NRE regions are exclusively poor textured areas. However, recovery of missing color values in homogeneous areas can be safely performed on the views $\overline{I}_{i,j}$, thanks to the lack of high frequency content. In this manner, after performing color propagation according to Eq. 3, a further low-complexity demosaicking method recovers the missing colors without introducing color artifacts. In particular, a bilinear color value interpolation of neighboring pixels on the same view is used.

Up to this point, the horizontal and vertical sampling periods of the views are assumed to be equal. However, from the estimation of microlens centers, the vertical sampling period is estimated to be in average twice the horizontal sampling period (see Fig. 1). To properly address the sampling period ratios, the views shall be vertically up-sampled. Considering the lack of chromatic information on mosaicked demultiplexed views as in Fig. 3-(1), such an up-sampling step does not appear beneficial for the task of disparity estimation. But, in the demosaicked views, a fast up-sampling step inserts empty rows, and exploits again the estimated disparities to recover the missing values.

The disparity maps and the reliability maps are vertically interpolated using a median filter of size 3×4. To reduce aliasing artifacts of disparity map interpolation, the reliability threshold is increased to 7, resulting in the recovery of more than 70% of the pixels of the empty rows.

In the next section, it is shown that using the described approach creates up-sampled views without introducing color artifacts. This is because the chromatic content in the light-field is sufficient for the first steps towards super-resolution.

3. EXPERIMENTAL RESULTS

Although our method is generic for any plenoptic data, we focus on the images captured by Lytro to show the performance of the approach. In particular, we use the toolbox of [17] to decode the light-field from the native file format of Lytro. For all our results, the parameter of disparity reliability ε is set to 0.1 pixels and the number of reliable estimation threshold τ is set to 3 reliable estimations. In this section we compare our method with the method in [10] (algorithm implemented by the authors) and the results after demosaicking directly the raw plenoptic image as a conventional image.

As explained beforehand, we use the algorithm in [12] for demultiplexing raw data and estimating disparities from mosaicked views. Fig. 3-(1) shows one of the views after demultiplexing and a zoom-in of the same view. Given the color pattern observed in the views, it is clear that state of art demosaicking methods are not well adapted to our data. The estimated disparity and reliability map (NRE) of the red rectangle region are shown in Fig. 3-(3) respectively. Essentially, points with small NRE correspond to poorly textured areas. Finally, Fig. 3-(2) show the intermediate recovered view $\overline{I^c}_{i,j}$ after applying our disparity-guided color propagation (Eq. 3), and a zoom-in of the red rectangle. 49 views (out of 100 views) in the light-field have been used. It can be seen through these images that high frequency content of the scene is recovered from the light-field. Besides, in average more than 90% of the missing color values, including the empty pixels, are recovered.

Fig. 4-(a,d) shows the final results of our algorithm in which the vertical/horizontal sampling period ratio is corrected for two different examples. The zoom-ins on the right

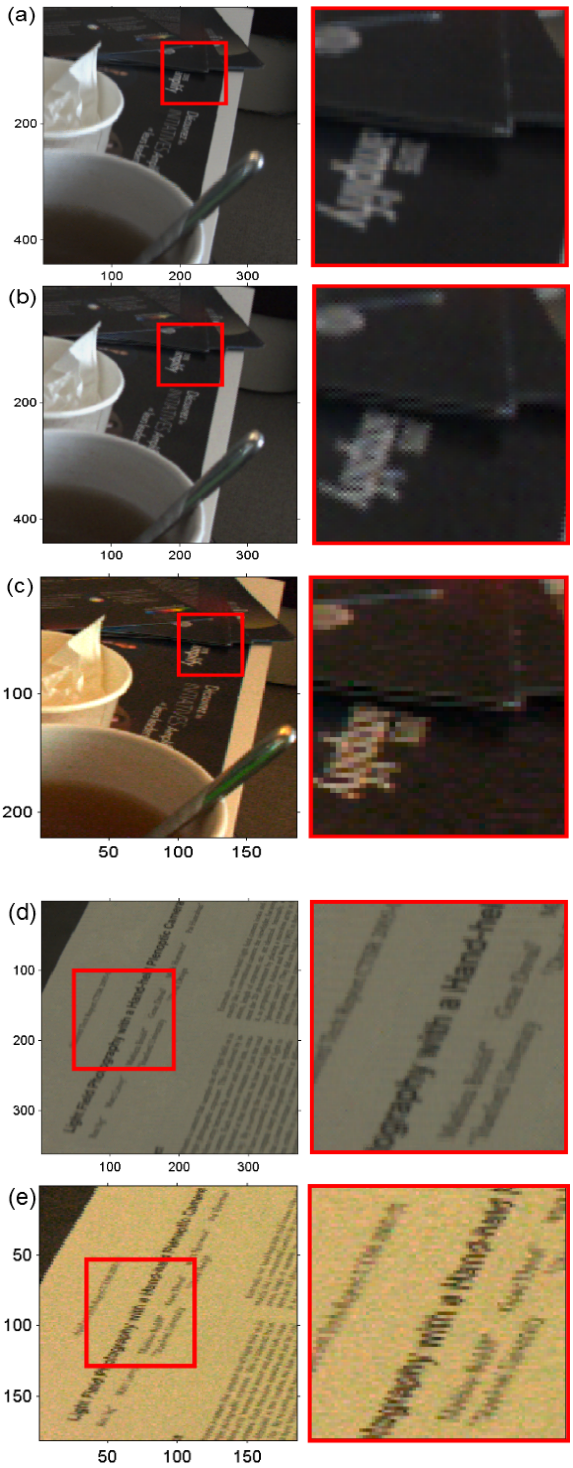

Fig. 4. (a,d) Final results of our method. (b) Resulting views after demosaicking the raw data and then demultiplexing. (c,e) Results using the algorithm in [10].

point out parts of the views where it is difficult to recover all the color information as high-frequancies.

In Fig. 4-(c,e) we compare our results with the method in [10]. Note that in our results letters are easily readable while important artifacts appear on the results from [10]. Besides, our method takes into account the microlenses ligh-field sampling avoiding aliasing on the views. As a consequence, our results are upsampled by a factor of two with respect to the results in [10]. Finally, Fig. 4-(b) shows the same view obtained by demosaicking the raw data as it was a conventional image and then demultiplexing with the same algorithm (instead of demultiplexing and then demosaick as we propose). As mentioned in the introduction, the results suffer from view cross-talk and aliasing.

4. CONCLUSION

Plenoptic cameras capture the light-field of a scene in a single snapshot, providing interesting post-capture possibilities. Demosaicking of the captured data is mostly overlooked in the literature. It is shown in this paper that to avoid color artifacts mainly produced by view crosstalk, the raw data requires to be demultiplexed into views of the scene, without being demosaicked. A previously block-matching method for plenoptic data is used to estimate pixel disparities. Then reliable estimated disparities are used to demosaick views, exploiting the redundant sampling of the scene. The results do not contain color artifacts, compared to the state of art methods. Thanks to accurate view demultiplexing and sub-pixel accuracy of the estimated disparities, the spatial resolution of the demosaicked views are higher than the state of art methods by a factor of 2, without bearing the complexity of additional super-resolution steps. Circumventing the fattening effect of the block-matching method, and achieving higher factors of super-resolution are left as future works.

5. REFERENCES

[1] G. Lippmann, "Epreuves reversibles donnant la sensation du relief.," *J. Phys. Theor. Appl.*, vol. 7(1),, pp. 821–825., 1908.

[2] E. H. Adelson and J. Y. Wang, "Single lens stereo with a plenoptic camera," *TPAMI*, vol. 14(2), pp. 99–106, 1992.

[3] T. E. Bishop and P. Favaro, "The light field camera: Extended depth of field, aliasing, and superresolution," *TPAMI*, vol. 34(5), pp. 972–986, 2012.

[4] S. Wanner and B. Goldluecke, "Variational light field analysis for disparity estimation and super-resolution," *TPAMI*, 2013 (to appear).

[5] M.W. Tao, S. Hadap, J. Malik, and R. Ramamoorthi, "Depth from combining defocus and correspondence using light-field cameras," in *ICCV*, 2013.

[6] R. Ng, M. Levoy, M. Brdif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held

plenoptic camera," Tech. Rep., Computer Science Technical Report CSTR, 2(11)., 2005.

[7] T. Georgiev, G. Chunev, and A Lumsdaine, "Superresolution with the focused plenoptic camera," in *SPIE Electronic Imaging*, 2011.

[8] F. Perez, A. Perez, M. Rodriguez, and E. Magdaleno, "Fourier slice super-resolution in plenoptic cameras.," in *ICCP*, 2012.

[9] Z. Yu, J. Yu, A. Lumsdaine, and T. Georgiev, "An analysis of color demosaicing in plenoptic cameras," in *CVPR*. IEEE, 2012, pp. 901–908.

[10] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Decoding, calibration and rectification for lenselet-based plenoptic cameras," in *CVPR*, 2013.

[11] D. Cho, M. Lee, S. Kim, and Y.-W. Tai, "Modeling the calibration pipeline of the lytro camera for high quality light-field image reconstruction," in *ICCV*, 2013.

[12] N. Sabater, V. Drazic, M. Seifi, G. Sandri, and P. Perez, "Light-field demultiplexing and disparity estimation," 2014.

[13] S. Farsiu, M. Elad, and P. Milanfar, "Multiframe demosaicing and super-resolution of color images," *IEEE Transactions on Image Processing*, vol. 15, no. 1, pp. 141–159, 2006.

[14] A. Lumsdaine and T. Georgiev, "The focused plenoptic camera," in *ICCP*, 2009.

[15] X. Li, B. Gunturk, and L. Zhang, "Image demosaicing : A systematic survey," in *Proceedings of SPIE, the International Society for Optical Engineering*. 2008, pp. 68221J.1–68221J.15, Society of Photo-Optical Instrumentation Engineers.

[16] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *IJCV*, vol. 47(1-3), pp. 7–42, 2002.

[17] *http://code.behnam.es/python-lfp-reader/*.