

On plenoptic sub-aperture view recovery

Mozhdeh Seifi, Neus Sabater, Valter Drazic, Patrick Pérez
 Technicolor, 975 avenue des Champs Blancs,
 35576 Cesson-Sévigné, France
 Email: firstname.lastname@technicolor.com

Abstract—Light field imaging is recently made available to the mass market by Lytro and Raytrix commercial cameras. Thanks to a grid of microlenses put in front of the sensor, a plenoptic camera simultaneously captures several images of the scene under different viewing angles, providing an enormous advantage for post-capture applications, *e.g.*, depth estimation and image refocusing. In this paper, we propose a fast framework to re-grid, denoise and up-sample the data of any plenoptic camera. The proposed method relies on the prior sub-pixel estimation of micro-images centers and of inter-views disparities. Both objective and subjective experiments show the improved quality of our results in terms of preserving high frequencies and reducing noise and artifacts in low frequency content. Since the recovery of the pixels is independent of one another, the algorithm is highly parallelizable on GPU.

I. INTRODUCTION

Light field imaging by plenoptic cameras is gaining a lot of popularity in the field of computational photography due to the rich and costless capture of angular and spatial information of the scene, made possible thanks to a microlens array placed between the main lens and the sensor. As a result, such cameras have novel post-capture processing capabilities (*e.g.*, depth estimation [1]–[3], superresolution [1], [2], refocusing [4], and light field editing [5]).

A very first step in the pipeline of light field image processing called the *view demultiplexing* converts the 2D raw image to the 4D light field addressing the spatial and angular coordinates of the sampled signal. The demultiplexing process consists in estimating the origin of the angular coordinate system in the image of each microlens, *i.e.*, the μ -image center, and then reorganizing the raw data in a way that all pixels capturing the light rays with a certain angle of incidence are stored in the same image, creating the *sub-aperture views*. These demultiplexed images form the matrix of views. In practice, due to translational and rotational offsets between the microlens grid and the sensor, the μ -image centers are estimated as non-integer sensor coordinates. To obtain the matrix of views, the state of art algorithms either round the estimations, which introduces aliasing or re-sample the μ -images in the raw data, which mixes the angular information and creates view cross-talk.

The problem of light field denoising is separately addressed by [6] through filtering in the frequency domain, targeting low-light imagery. Their proposed hyperfan filter with the Gaussian roll-off effectively removes the noise, at the cost of increasing the blur or ringing artifacts in the light field. In [7], gray-scale light fields degraded by additive noise are denoised using the

estimated disparities. In [8], Poisson noise of camera array data is circumvented employing the principal component analysis and tensor analysis of image patches found through disparities. None of the mentioned denoising works however handle the sub-pixel sampling of the signal correctly.

In this paper, we propose to exploit both the sub-pixel estimation of the μ -image centers *and* the sub-pixel estimation of disparities to spatially re-grid¹, denoise and up-sample the views. This allows us to address correctly the spatial sampling of the signal and thus to reduce the aliasing in the matrix of views.

II. DISPARITY-GUIDED VIEW RECOVERY

The inputs of the proposed view recovery framework are the initial demultiplexed matrix of views obtained from demosaicked raw data, the sub-pixel misalignment of the μ -image centers, and the pixel correspondences on the views (for example from disparity estimation). Let I^i be a demultiplexed view containing N non-negative real-valued intensities in 3 color channels and positioned in the angular coordinate $\mathbf{i} = (i_1, i_2) \in \mathbb{Z}^2$ of the matrix of views. We will denote $I^i(\mathbf{x}, c)$ the intensity of this view at the 2D pixel-coordinate $\mathbf{x} \in \mathbb{N}^2$ in color channel c . For a second view I^u , let $\mathbf{a}^{i,u} = a(\mathbf{i} - \mathbf{u}) \in \mathbb{Z}^2$ be its baseline to the first one (a is constant), and let $\mathbf{d}^i(\mathbf{x}) \in \mathbb{R}^2$ denote the disparity of pixel \mathbf{x} in view \mathbf{i} . The 2D coordinates of the corresponding pixel in view \mathbf{u} read $\mathbf{x}' = \lfloor \mathbf{x} + \mathbf{a}^{i,u} \odot \mathbf{d}^i(\mathbf{x}) \rfloor$, \odot denoting element-wise multiplication and $\lfloor \cdot \rfloor$ denoting the rounding function. Our aim is to recover a better estimation of the intensity at pixel \mathbf{x} in view \mathbf{i} using the intensities of its correspondents in all views.

The sub-pixel misalignment of each μ -image center is inherited by all of the pixels that are demultiplexed from that μ -image into the different views. Besides, accurate estimations of disparities have most probably sub-pixel accuracy, due to the intended aliasing in the light field sampling introduced by manufacturing [3], [4]. To correctly import the corresponding pixels from the other views within each pixel of a target view, we convolve the imported samples with a triangle filter to approximate the ideal anti-aliasing *sinc* filter.

The 2D misalignment for \mathbf{x} is defined as follows:

$$\epsilon(\mathbf{x}) := \begin{cases} \epsilon^m & \text{if } I^i(\mathbf{x}, c) \in \mathbb{R}, \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

¹Re-gridding in this context corresponds to the resampling of the non-regularly sampled data into a regular and integer grid of pixel coordinates.

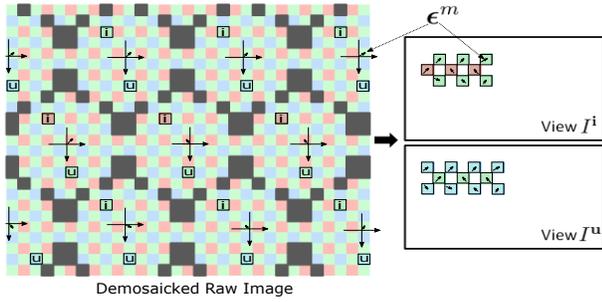


Fig. 1. An example of pixel reordering from the demosaicked raw image to extract two different raw views [9]. Pixels with the same relative position (same angular information) w.r.t. the μ -image centers are assembled in the same view. In general, the number of recovered views is equal to the number of pixels per microlens but only two views are illustrated here for visualization. Color is only shown for illustrative purposes and corresponds to sensor color on original Bayer pattern. The origin of the angular coordinate system on each μ -image is represented by crossing arrows.

where m identifies the μ -image used to demultiplex \mathbf{x} to I^i , and $\epsilon^m \in \mathbb{R}^2$ denotes the corresponding sub-pixel misalignment of the μ -image center (see Fig. 1). The second condition in Eq. 1 addresses the pixels that do not contain chromatic information, e.g., the empty pixels that are introduced during the demultiplexing step in [9]. The proximity promoting weight $w^{i,u}(\mathbf{x}, \mathbf{x}')$ is then defined for pixel \mathbf{x} and its corresponding pixel \mathbf{x}' in I^u , as follows:

$$w^{i,u}(\mathbf{x}, \mathbf{x}') = 1 - \|\epsilon(\mathbf{x}') + \mathbf{a}^{i,u} \odot \mathbf{d}^i(\mathbf{x}) - [\mathbf{a}^{i,u} \odot \mathbf{d}^i(\mathbf{x})]\|_2 \quad (2)$$

These weights are normalized over all of the corresponding pixels \mathbf{x}'' on all views \mathbf{u}' that are contributing to the recovery of \mathbf{x} :

$$\hat{w}^{i,u}(\mathbf{x}, \mathbf{x}') = \frac{\Gamma^u(\mathbf{x}') w^{i,u}(\mathbf{x}, \mathbf{x}')}{\sum_{\mathbf{u}'} \Gamma^{\mathbf{u}'}(\mathbf{x}'') w^{i,\mathbf{u}'}(\mathbf{x}, \mathbf{x}'')} \quad (3)$$

where $\Gamma^u(\mathbf{x}') \in \{0, 1\}$ is equal to 1 only if \mathbf{x}' contains chromatic information. Besides, note that for $\mathbf{i} = \mathbf{u}$, we have $\mathbf{x}' = \mathbf{x}$ and $w^{i,i}(\mathbf{x}, \mathbf{x}) = 1 - \|\epsilon(\mathbf{x})\|_2$. Then, the recovered color information from all of the views are obtained as follows:

$$\tilde{I}^i(\mathbf{x}, c) = \sum_{\mathbf{u}} \hat{w}^{i,u}(\mathbf{x}, \mathbf{x}') I^u(\mathbf{x}', c). \quad (4)$$

Finally, if the difference in vertical and horizontal sampling is not corrected in the initially demultiplexed views (e.g., [9]), our view recovery method performs an additional up-sampling step on the denoised/re-gridded views provided by (4). Particularly, the intensity of pixel \mathbf{x}' in \tilde{I}^u is used to recover the intensity at pixel \mathbf{x} in a new, up-sampled view \check{I}^i , where $\mathbf{x} = \lfloor \mathbf{k} \odot [\mathbf{x}' + \mathbf{a}^{u,i} \odot \mathbf{d}^u(\mathbf{x}')] \rfloor$. In this formulation $\mathbf{k} \in \mathbb{R}^2$ denotes the horizontal and vertical sampling rates, e.g., $\mathbf{k} = (\sqrt{3}, 1)$ in [9] taking into account the quinqux arrangement of the microlenses in Lytro. Eqs. 2–4 are then used to recover \check{I}^i accordingly, with all the mis-alignments $\epsilon(\mathbf{x}), \epsilon(\mathbf{x}')$ replaced by $\mathbf{0}$, since \tilde{I}^i is already resampled into an integer-coordinate grid.

It shall be noted that linear demosaicking of raw data (that is performed before demultiplexing) changes the statistics of the captured noise. Indeed, demosaicking (i.e., averaging of n neighboring chromatic values) asymptotically reduces the noise. Now if for all \mathbf{u} we have either

$$\|\mathbf{a}^{i,u} \odot \mathbf{d}^i(\mathbf{x})\|_\infty \geq 1 \quad (5)$$

i.e. the corresponding pixels are sampled in two different micro-images in the raw data, or

$$\|\mathbf{a}^{i,u}\|_\infty > n/2 \quad (6)$$

i.e. the baseline between the views is bigger than half of the demosaicking window's width (e.g., $n = 4$ for bilinear demosaicking of raw data), then these corresponding pixels are obtained from two different neighborhoods on the raw data, and therefore their noise values remain statistically independent after the demosaicking step. The Central Limit theorem therefore indicates that averaging the corresponding values (here with disparity-guided view recovery) further reduces the noise to normally distributed around 0, iff many of such pixels are used. For the few other pixels where none of the above conditions are satisfied, however, the corresponding pixels come from the same raw data neighborhood, and therefore the corresponding pixel noise values are not statistically independent. In those cases, the noise is reduced by averaging the values, but can not be guaranteed to converge to 0.

III. EXPERIMENTAL RESULTS

In this section, we perform both quantitative and qualitative assessments of our view recovery framework. We use the method of [9] for obtaining the disparities. We then demosaic the raw data and demultiplex according to the method of [9]. We next proceed with our view recovery method first in a Monte Carlo setup to assess the noise reduction performed by our method.

Next, we compare our results with the relevant state-of-art methods in view recovery of type 1.0 plenoptic cameras, i.e., the image processing pipelines presented in [10], [11] and [6]. The comparison is performed both employing an objective non-reference image quality metric called “metric Q” [12], as well as subjective inspection of the recovered views of the scenes. The code and data of [10] are available on-line. The authors of [11] kindly agreed to provide us with their code, and the authors of [6] kindly agreed to run their code on our data, to allow comparisons.

The computational costs of the proposed method is negligible compared to the costs of estimating disparities. In particular, on a five-core Intel CPU exploiting multi-threading with OpenMP [13], the disparity estimation in [9] takes 90 seconds in average per view of Lytro images. In return, the presented denoising and up-sampling method takes a few milliseconds per view on the same machine.

We encourage the reader to inspect the results on the pdf version of this paper, rather than a hard copy.

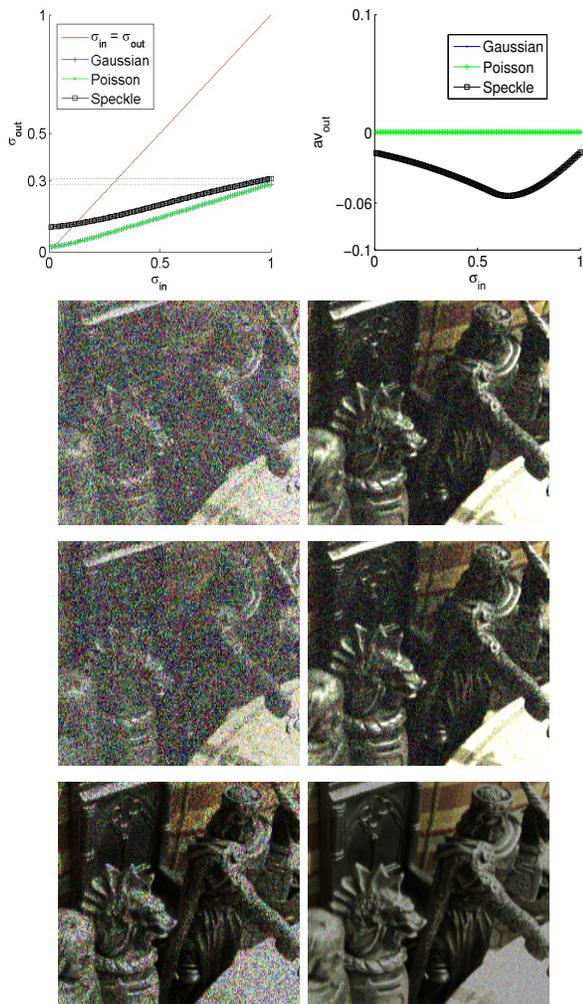


Fig. 2. **Monte Carlo analysis of the proposed view recovery system.** The first row shows the standard deviation σ (Left) and the average (Right) of the distance between the denoised output and the original light field (the distance being calculated as the difference between the denoised and original signal for the additive noise, and as the ratio of denoised to original signal for the multiplicative noise). The results of additive Gaussian, additive Poisson and multiplicative Speckle noise are plotted in blue, green and black, respectively. In both plots the results of additive Gaussian and additive Poisson noise are very similar to each other. The next two rows show the noisy input of the lightfield 4 in Fig. 3, with maximum noise $\sigma_{in} = 1$, and the recovered view, respectively. Additive Gaussian, additive Poisson and multiplicative Speckle noise are shown from left to right.

A. Monte Carlo analysis of noise reduction performance

To quantify the noise reduction obtained by the proposed method, we performed a set of Monte Carlo studies using the demultiplexing and disparity estimation proposed in [9]. After estimating the disparities according to [9], three types of noise were applied to the raw data of the light fields. The noisy raw data was then demosaicked and demultiplexed. We then proceeded to denoise, re-grid and up-sample the views as proposed in Section II. The results of our denoising method were compared to the originally demultiplexed light fields without any applied noise. In particular, we used the estimated

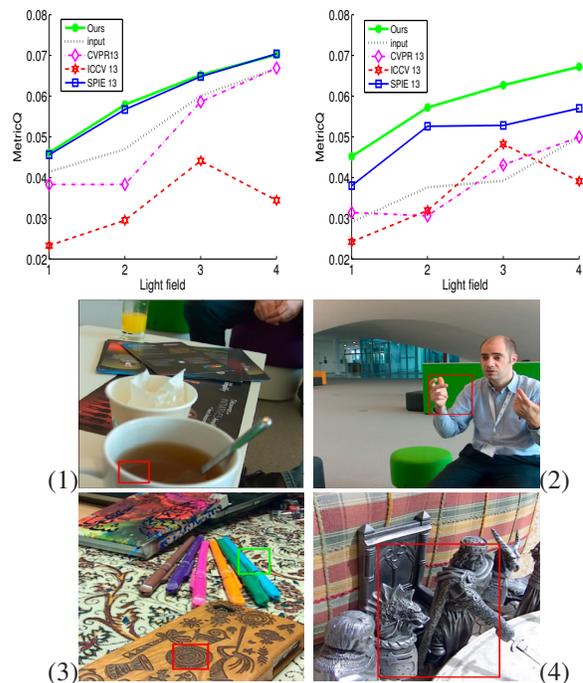


Fig. 3. **Quantitative assessment with metric Q .** First row: metric Q plotted for 4 different light fields, for a central view (Left) and a peripheral view (Right). Caption labels stand for *CVPR13* [10], *ICCV13* [11] and *SPIE13* [6]. Next two rows provide thumbnails of the four light fields with marked regions used in the other figures of this paper.

disparities obtained from the originally captured raw data to separate the performance of the disparity estimation of [9] from our proposed view recovery.

Noise categories were chosen as additive Gaussian noise, additive Poisson noise, and multiplicative Speckle noise. For each category, the standard deviation σ of the noise samples was gradually changed between 0 and 1. The noise was applied to normalized light field data (*i.e.*, RGB image values in the range of $[0, 1]$). Fifty noise samples per pixel were simulated for each σ . In particular, for every σ , the additive Gaussian noise was centered at 0 and the additive Poisson noise at σ^2 . The multiplicative speckle noise was generated to have uniform distribution between $1 - \sqrt{12}\frac{\sigma}{2}$ and $1 + \sqrt{12}\frac{\sigma}{2}$.

Fig. 2 shows some statistics on the distance between the denoised output and the original signal as a function of applied noise's σ . This distance is designed to assess the remaining noise in the denoised output. It is calculated as the difference between the denoised and original signal for additive noise, and as the ratio of denoised to original signal for the multiplicative noise. Indeed, the proposed method reduces the applied noise in all cases, having zero average in both cases of additive Gaussian and Poisson noise. Note that the expected average of the input Poisson noise grows theoretically as the noise variance. Nevertheless, the average of the remaining noise in our the views remains close to 0. The effect of re-gridding is the most evident in the case of multiplicative (Speckle) noise, where the average is close to

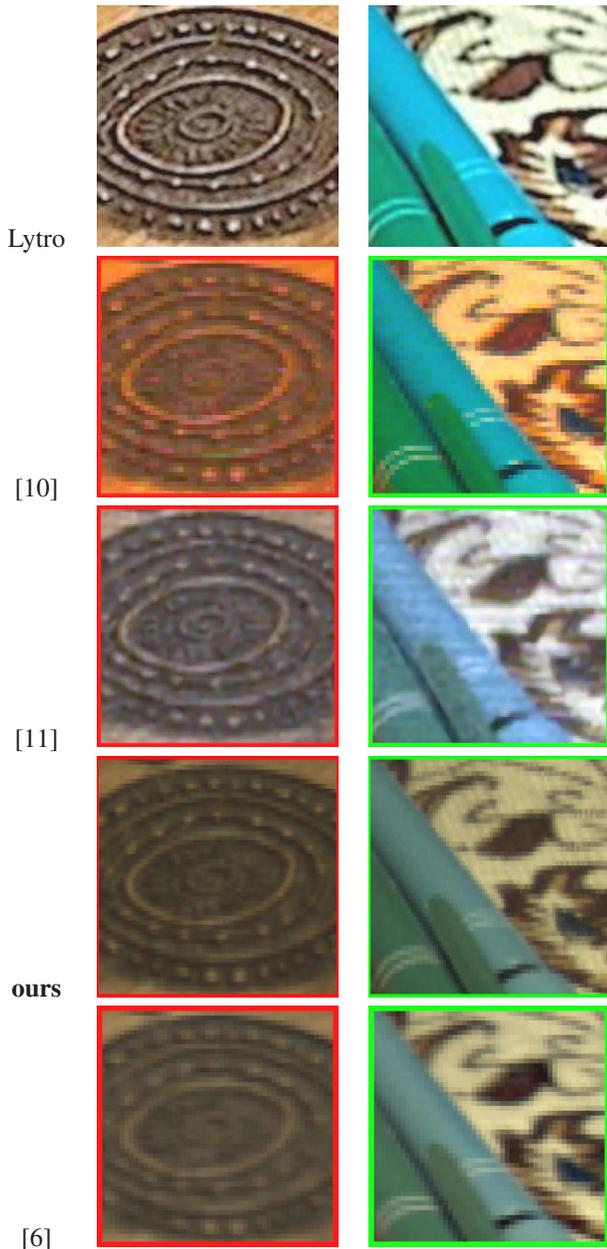


Fig. 4. **Qualitative comparisons on recovering central views:** Close-ups on the central views recovered by different methods at two different depths from the third lightfield in Fig. 3. Note the additional sharpening on strong edges in Lytro images, the loss of high frequency and noise presence for [10], the noise magnification for [11] with images twice of our results in resolution, and the loss of high frequency in [6]. In particular, none of the state of art methods preserve the texture of the tablecloth. Difference in the color appearance of the results come from different color correction methods implemented by each toolbox.

0, but does not necessarily converge to 0. Finally, in presence of low magnitudes of the applied noise, the level of the noise in the original data, the effect of re-gridding on the views after re-gridding/denoising, and/or possibly the in-accuracies of the estimated disparities are revealed.

B. Non-reference assessment with metric Q

Due to the absence of ground truth for the captured light fields, we calculate metric Q [12] on the output views of each toolbox, to quantify the quality of the views recovered by each method. In particular, we include the view rectification step in [10], and the dictionary learning super-resolution step in [11] to assess the full pipeline of each work. Metric Q is based on the singular value decomposition of local gradients of the image. In particular, the image is divided into non-overlapping blocks of $B \times B$ pixels. For block b , the horizontal and vertical intensity gradients are stacked in a $B^2 \times 2$ matrix. Denoting by z_1^b and z_2^b the singular values of this matrix, the thresholded coherence R^b of the block b is defined as:

$$R^b = \begin{cases} 1 & \text{if } \frac{|z_1^b - z_2^b|}{z_1^b + z_2^b} \geq \vartheta, \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

To consider only anisotropic blocks (*i.e.*, the blocks for which $R^b = 1$), metric Q is defined as

$$Q = \sum_b R^b z_1^b \frac{|z_1^b - z_2^b|}{z_1^b + z_2^b}. \quad (8)$$

It is shown in [12] that, although the value of metric Q strongly depends on the content of the scene, it is conveniently correlated with the level of noise, the sharpness and the intensity contrast of the image. Notably, a bigger value of this metric indicates a higher quality of the image, that is lower noise, higher sharpness and better contrast. The two parameters of this metric are the size B of the blocks and the threshold ϑ on the coherence between the singular values. We set these parameters similar to the ones proposed by the authors, *i.e.*, the block size $B = 8$ and the threshold in Eq. 7 $\vartheta = 0.02$.

Fig. 3 shows plots of the metric Q values estimated for four different light fields on the central and peripheral views. As can be seen, our method and the denoising of [6] out-perform the others in the central views, and our technique out-performs all of the others, including [6] on the peripheral views according to this non-reference image quality metric.

C. Subjective comparisons

We show in Figs. 4 and 5 close ups of the recovered central and peripheral views for the different light-fields and the different methods from Fig. 3. Note that subjective comparison should be conducted while taking into account the different spatial resolutions delivered by the different methods: The sub-aperture views of [10] are 328×328 pixels, Lytro provides all-in-focus images (the combination of all views) of size 1080×1080 pixels, our recovery method gives images of size 655×655 pixels, and the size of the sub-aperture views of [11] depends on the estimated microlens rotation (978×976 pixels for our light field camera).

In practice, our method does visually better than [6] (denoising in frequency domain with a hyperfan filter) at preserving the high frequency content and reducing the ringing artifacts in the peripheral views (see Fig. 5). It also out-performs [10]

(with their view rectification) and [11] (with their dictionary learning step and images twice of ours in resolution) in noise reduction in the homogeneous areas. Note the over exposure of the bright areas in the Lytro image, which is caused by the integration of the light over the corresponding μ -images. Besides, none of the state-of-art methods preserve the texture of the tablecloth in Fig. 4 as ours does.

IV. CONCLUSION

In this paper, we address the physical mis-alignments of the optical elements that result in non-integer spatial and angular sampling of the signal in a plenoptic camera. Particularly, having the pixel correspondences obtained from disparity estimation, multiple samples of the signal are exploited to spatially re-grid the signal on the views and to reduce the contaminating noise. Objective assessment of the algorithm (through a Monte Carlo study and a non-reference image quality metric) shows the improved quality of our results w.r.t. the state of art in terms of noise reduction, image sharpness and intensity contrast.

It should be noted that the sub-pixel misalignment of the μ -image center corresponds to both spatial *and* angular coordinates of the demultiplexed pixels. In this work, we only exploit the spatial aspect of the sub-pixel accuracy of the estimated μ -image centers. A more rigorous handling of both spatial and angular aspects implied by the sub-pixel center estimation is left to future work.

Also note that the proposed recovery method can be seen as a variation of the Non-Local means (NL-means) denoising [14] applied on light field data, where the blocks are found through the correspondences provided by the estimated disparities. While NL-means performs denoising within a single image, our approach exploits multiple instances of the same scene locality to perform denoising. The latter reduces the possible introduction of intra-image correlation between pixels.

Finally, when light field imaging is used only to perform scene refocussing [4], [15], our re-gridding, denoising and up-sampling method can be adapted to proceed directly on the targeted refocusing plane.

REFERENCES

- [1] T. E. Bishop and P. Favaro, "The light field camera: Extended depth of field, aliasing, and superresolution," *(IEEE TPAMI)*, vol. 34(5), pp. 972–986, 2012.
- [2] S. Wanner and B. Goldluecke, "Variational light field analysis for disparity estimation and super-resolution," *(IEEE TPAMI)*, 2014.
- [3] A. Lumsdaine and T. Georgiev, "The focused plenoptic camera," in *(IEEE ICCP)*, 2009.
- [4] C.-K. Liang and R. Ramamoorthi, "A light transport framework for lenslet light field cameras," *ACM Transactions on Graphics (TOG)*, 2015.
- [5] A. Jarabo, B. Masia, A. Bousseau, F. Pellacini, and D. Gutierrez, "How do people edit light fields?" vol. 33, no. 4, 2014.
- [6] D. G. Dansereau, D. L. Bongiorno, O. Pizarro, and S. B. Williams, "Light field image denoising using a linear 4d frequency-hyperfan all-in-focus filter," in *SPIE, Computational Imaging XI*, vol. 8657, 2013, pp. 86 570P–86 570P–14.
- [7] K. Mitra and A. Veeraraghavan, "Light field denoising, light field superresolution and stereo camera based refocussing using a GMM light field patch prior," in *(IEEE CVPRW)*, Jun. 2012, pp. 22–28.

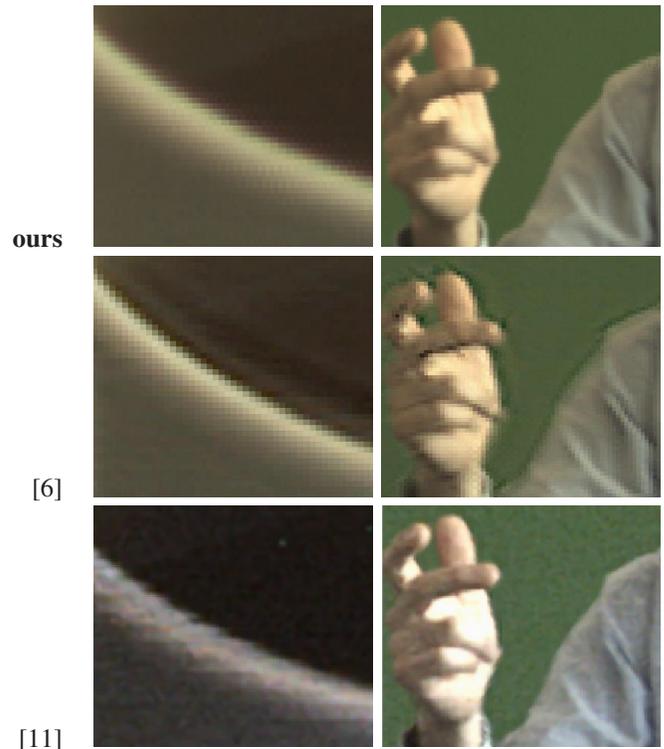


Fig. 5. **Qualitative comparisons on recovering peripheral views:** Close-ups on the peripheral views recovered by different methods from the two first lightfields in Fig. 3. The three rows show results obtained by our approach, [6] and [11], respectively. Note the artifacts of [6] around edges and the magnification of noise on homogeneous areas for [11].

- [8] Z. Li, S. Vaddadi, H. Jin, and S. K. Nayar, "Multiple view image denoising," *CVPR*, pp. 1542–1549, 2009.
- [9] N. Sabater, M. Seifi, V. Drazic, G. Sandri, and P. Perez, "Accurate disparity estimation for plenoptic images," 2014.
- [10] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Decoding, calibration and rectification for lenselet-based plenoptic cameras," in *(IEEE 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 *(IEEE ICCV)*, 2013.
- [12] X. Zhu and P. Milanfar, "Automatic parameter selection for denoising algorithms using a no-reference measure of image content," *(IEEE TIP)*, vol. 19, no. 12, pp. 3116–3132, Dec. 2010.
- [13] L. Dagum and R. Menon, "openmp: an industry standard api for shared-memory programming," *Computational Science and Engineering, IEEE*, vol. 5, no. 1, pp. 46–55, 1998.
- [14] A. Buades, Y. Lou, J. M. Morel, and Z. Tang, "Multi image noise estimation and denoising," *HAL documents*, vol. 1, 2010.
- [15] Z. Yu, J. Yu, A. Lumsdaine, and T. Georgiev, "an analysis of color demosaicing in plenoptic cameras," in *Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2012, pp. 901–908.