Modeling the Pairwise Disparities in High Density Camera Arrays

Ioan Tabus and Pekka Astola
Department of Signal Processing
Tampere University of Technology
Tampere,Finland
ioan.tabus@tut.fi and pekka.astola@tut.fi

***Abstract*—We discuss in this paper models for the disparity information needed when pairwise warping the angular views in a light field data set formed of N views. In one scenario of light field data compression, first a set of M reference views is encoded and then each of the remaining views is predicted by warping several reference views using disparity information. The necessary disparity information in this case may be as high as M(N-1) pairwise view disparity maps, estimated and transmitted independently for each pair (reference, target). We propose an estimation model which can be used in a flexible way for any selected configuration of references and predicted views. We study the estimation of the global model from the matching information provided by a pairwise matching program. The model may be defined in several ways, by considering the vertical and horizontal matches at various views and by allowing different model parameters for the regions from a segmentation of the scene. The regions based model is shown to perform better than a single region model. The performance of the model in synthesizing the unseen color views at specified locations in the views array is presented for several configurations of the estimation and prediction sets.**

I. Introduction

The light field images acquired by dense camera arrays have become recently available having high image resolution and high number of views in the array. In the standardization project JPEG Pleno light field [1] [2] it is of interest how the disparity information extracted from a light field dataset could be encoded in the most efficient way and then be utilized for synthesizing some views based on reference views. One of the problems in compression is how to utilize the most precise disparity models, available for pairs of views, so that the prediction by disparity based warping will produce high quality images.

Here we propose to extract a single overall model, expressed in terms of the depth for the scene at each reference view, out of which all the pairwise needed disparities can be computed. Even for a carefully prepared dataset the overall model will produce however only an approximate reproduction of the pairwise disparity estimated from real data, due to optical and geometrical imperfections when acquiring and processing the multiview images.

For now we address the simple question of how to optimally reconstruct the pairwise disparities obtained by a state-of-the-art optical flow estimation method. We present an algorithm which operates only on the input matches. It avoids the expensive color warping operations to produce the color distortion measure. The usefulness of the results will largely depend on the precision of the optical flow routine. In here we trust completely the matches offered by the program presented in efficient coarse-to-fine patch match (CPM) estimation method [7], since our optimization criterion will seek to reproduce these matches by our proposed criterion. As better flow estimation programs will become available, the quality of the input data will also improve.

The estimation of light field disparity was studied for long time, see for example [3] [4] and reference therein for existing public literature. Also, especially for the HDCA data considered here, there is prior work in [5] [6] proposed in the standardization literature, which used CPM pairwise matching information in a heuristically motivated estimation algorithm.

II. Pairwise matching views in light field data

We consider a light field composed of $N = K \times L$ angular views, $\{\mathbf{A}_{k,\ell}, k = 1,\dots,K; \ell = 1,\dots,L\}$, obtained by a camera taking N pictures of the scene, with the camera positions at coordinates $(Y(k,\ell), X(k,\ell))$ in a rectangular grid, resulting in pictures which are highly redundant.

We are choosing an angular view $\mathbf{A}_{k_0,\ell_0}$ as a reference and any other view, say $\mathbf{A}_{k,\ell}$, is taken as a target. We obtain a list of matches between the two views, by using the CPM method [7]. The goal is to encode the target $\mathbf{A}_{k,\ell}$ based on the reference $\mathbf{A}_{k_0,\ell_0}$ by using warping. Each angular view, $\mathbf{A}_{k,\ell}$, is a $n_r \times n_c$ RGB color image having at pixel location (i,j) the vector of color $\mathbf{A}_{k,\ell}(i,j)$, with three components indexed by $c \in \{1,2,3\}$, e.g., $\mathbf{A}_{k,\ell,1}$ is the value of the red component. In our experiments we use the dataset described in [8], which is used for core experiments in JPEG Pleno light field. Each angular view is a 4K image having $n_c = 3840$ and $n_r = 2160$. The full HD case can be obtained by cropping the central part of each 4K image to obtain a HD sub-image, $n'_c = 1920$ and $n'_r = 1080$, as we did in the color view synthesizing experiment.

The list of matches has entries of the form (i,j,i',j'), for the pair of views $\mathbf{A}_{k_0,\ell_0}$ $\mathbf{A}_{k,\ell}$, where the RGB color at $\mathbf{A}_{k_0,\ell_0}(i,j)$ was found to correspond to the RGB color at $\mathbf{A}_{k,\ell}(i',j')$. We refer to (i,j) as a scene point being anchored at the reference and (i',j') as being the same scene point anchored at the target and denote $\Psi_{(k_0,\ell_0),(k,\ell)}$ the set of

ISBN 978-90-827970-1-5

pixels (i,j) for which matches are found. The entries in the list define disparities at the locations $(i,j) \in \Psi_{(k_0,\ell_0),(k,\ell)}$ as follows: row displacement $D^r_{k_0,\ell_0,k,\ell}(i,j) = i' - i$ and column displacements $D^c_{k_0,\ell_0,k,\ell}(i,j) = j' - j$.

The matching information provides evidence, the closest to data, about the matches that happen between the pair of views $(\mathbf{A}_{k_0,\ell_0}, \mathbf{A}_{k,\ell})$. Hence one would be tempted to use the list of matches, i.e., the vertical $D^r_{k_0,\ell_0,k,\ell}(i,j)$ and horizontal $D^c_{k_0,\ell_0,k,\ell}(i,j)$ disparity images, to perform warping for predicting $\mathbf{A}_{k,\ell}$ based on $\mathbf{A}_{k_0,\ell_0}$, obtaining the warped image $\mathbf{W}_{k,\ell}$:

$$\mathbf{W}_{k,\ell}(i + D^r_{k_0,\ell_0,k,\ell}(i,j), j + D^c_{k_0,\ell_0,k,\ell}(i,j)) = \mathbf{A}_{k_0,\ell_0}(i,j).$$

This will presumably offer a very good PSNR of the predicted $\mathbf{A}_{k,\ell}$, but will require encoding for each side view the pair of disparity images $D^r_{k_0,\ell_0,k,\ell}(i,j), D^c_{k_0,\ell_0,k,\ell}(i,j)$, hence the associate bitrate will be very large for a separate encoding of the disparities, at each side view. The remedy is to encode jointly the disparities using the high redundancy that exists between them.

Due to occlusions, the warped image $\mathbf{W}_{k,\ell}$ will not be defined in the whole $(n_r \times n_c)$ grid, and hence several other references may be used to get a complete prediction of $\mathbf{A}_{k,\ell}$.

A. Correspondence between vertical and horizontal disparities

If the camera optical axis had an ideal translation from the camera center $(Y_{k_0,\ell_0}, X_{k_0,\ell_0})$ to $(Y_{k,\ell}, X_{k,\ell})$, then the pixel displacements will be given by

$$\begin{aligned} i' - i &= \frac{(Y_{k,\ell} - Y_{k_0,\ell_0})f}{z(i,j)} \\ j' - j &= \frac{(X_{k,\ell} - X_{k_0,\ell_0})f}{z(i,j)} \end{aligned} \quad (1)$$

where $z(i,j)$ is the depth of the pixel with location (i,j) in $\mathbf{A}_{k_0,\ell_0}$ and f is a focal parameter (which will not be needed, since it will not appear as an explicit parameter in our model). Hence the ratios $(i'-i)/(j'-j)$ should be constant for all entries in the list for the pair of views $\mathbf{A}_{k_0,\ell_0}$ $\mathbf{A}_{k,\ell}$,

$$\rho(i,j) = \frac{(i'-i)}{(j'-j)} = \frac{(Y_{k,\ell} - Y_{k_0,\ell_0})}{(X_{k,\ell} - X_{k_0,\ell_0})} = \rho^0_{(k_0,\ell_0),(k,\ell)}. \quad (2)$$

The image acquisition scenario in HDCA is that $Y_{k,\ell} - Y_{k_0,\ell_0} = (k-k_0)\delta_Y$ and $X_{k,\ell} - X_{k_0,\ell_0} = (\ell - \ell_0)\delta_X$, with δ_Y and δ_X specific to the experiment, which results in

$$\frac{(Y_{k,\ell} - Y_{k_0,\ell_0})}{(X_{k,\ell} - X_{k_0,\ell_0})} = \frac{(k-k_0)\delta_Y}{(\ell-\ell_0)\delta_X} = \frac{(k-k_0)}{(\ell-\ell_0)}\mu, \quad (3)$$

where $\mu = \delta_Y/\delta_X$.

However, the exact positioning of the robot arm that carries the camera at a desired position is not possible, and hence the ratio $\frac{(Y_{k,\ell} - Y_{k_0,\ell_0})}{(X_{k,\ell} - X_{k_0,\ell_0})}$ should be estimated experimentally, by (robustly) averaging $\rho(i,j) = \frac{(i'-i)}{(j'-j)}$ form (2) over all $(i,j) \in \Psi_{(k_0,\ell_0),(k,\ell)}$. We collect the distribution of $\rho(i,j)$ for all $(i,j) \in \Psi_{(k_0,\ell_0),(k,\ell)}$ and estimate the location parameter as the median value, denoted as $\hat{\rho}^M_{(k_0,\ell_0),(k,\ell)}$. A robust estimate of the scale is taken as the median absolute deviation (MAD) defined as the median of the absolute differences $|\rho(i,j) - \hat{\rho}^M_{(k_0,\ell_0),(k,\ell)}|$ over all $(i,j) \in \Psi_{(k_0,\ell_0),(k,\ell)}$. Assuming a normal distribution corrupted by outliers, we estimate the normal scale as $\sigma^M = 1.4826 MAD(\{\rho(i,j)\})$. The ideal constant value $\mu = \delta_Y/\delta_X$ is hence estimated separately for each angular view as $\hat{\mu}^M_{(k_0,\ell_0),(k,\ell)} = \hat{\rho}^M_{(k_0,\ell_0),(k,\ell)} \frac{(\ell-\ell_0)}{(k-k_0)}$. The experimental setup for the HDCA data that we use has $\mu = 1.5$.

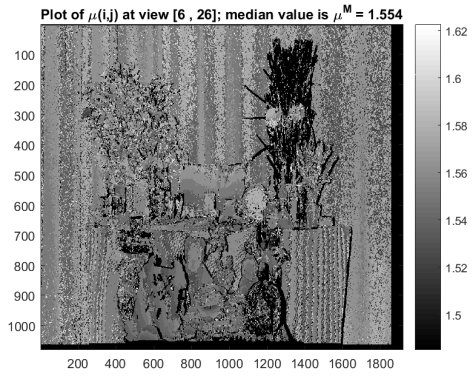


Fig. 1. Plot of the conversion factor between horizontal disparities and vertical disparities at the view $(2,26)$. The ideal ratio would be $\mu^{ideal} = 60mm/40mm = 1.5$. The horizontal and vertical disparities are obtained by estimating the optical flow between the center view and the view $(2,26)$ with the efficient coarse-to-fine patch match (CPM) estimation method [7].

When combining the information from the list of matches, one should take into account the variability of $\rho(i,j) = \frac{(i'-i)}{(j'-j)}$, between angular views, and inside each angular view.

To obtain a first picture of how much the experimental data deviates from the ideal scenario (equidistant sampling and parallel translation), we take the center view as a reference, $k_0 = 10, \ell_0 = 50$, and we illustrate in Figure 1 the estimated location and the estimated normal scale for one angular view. In order to provide more robust estimates, we excluded from each list of matches those matches for which $\frac{(i'-i)}{(j'-j)} \notin (\hat{\rho}^M - 3\sigma^M, \hat{\rho}^M + 3\sigma^M)$.

III. Formulating the maximum likelihood problem

We consider one region Ω in the image, and the disparities anchored at (k_0,ℓ_0) obtained by matching between view (k_0,ℓ_0) and (k_1,ℓ_1) for this region, $D^c_{k_1,\ell_1}(i,j)$ (in this section we omit the first two subindices, k_0,ℓ_0 for D^c) . We denote $z(i,j)$ the ideal depth of the scene point represented as the pixel (i,j), anchored at view (k_0,ℓ_0), and for simplicity of notations, denote $\psi(i,j) = f/z(i,j)$ where f is the camera focal parameter and $\psi(i,j)$ is called for short reciprocal-depth (we have an additional f factor in the definition of reciprocal-depth, compared to [6]). We consider the disparity model, agreeing to (1),

$$\begin{aligned} D^c_{k_1,\ell_1}(i,j) &= (X_{k_1,\ell_1} - X_{k_0,\ell_0}) f \frac{1}{z(i,j)} + e(i,j) \\ &= (X_{k_1,\ell_1} - X_{k_0,\ell_0})\psi(i,j) + e(i,j) \end{aligned}$$

ISBN 978-90-827970-1-5 © EURASIP 2018

and collect all equations for the pixels belonging to the non-occluded part of region Ω in the vector form:

$$\mathbf{x}_1 = \mathbf{r}_1 C_1^x + \mathbf{e}_1, \tag{4}$$

where $C_1^x = (X_{k,\ell_1} - X_{k_0,\ell_0})$; we arranged the pixels $(i_\tau, j_\tau), \tau = 1, \ldots, n_1$ from the two-dimensional region Ω_1 by scanning columnwise the region Ω_1, resulting in the vectors $\mathbf{x}_1$ and $\mathbf{r}_1$ having the elements with index τ as $\mathbf{x}_1(\tau) = D^c_{k_1,\ell_1}(i_\tau, j_\tau)$ and $\mathbf{r}_1(\tau) = f/z(i_\tau, j_\tau) = \psi(i_\tau, j_\tau)$, respectively.

Proceeding in a similar way for matching to the views with indices $(k_2, \ell_2), \ldots (k_n, \ell_n)$ we get the model

$$\begin{cases} \mathbf{x}_1 &= \mathbf{r}_1 C_1^x + \mathbf{e}_1 \\ \vdots & \\ \mathbf{x}_n &= \mathbf{r}_n C_n^x + \mathbf{e}_n. \end{cases} \tag{5}$$

Each equation refers to a different subset of Ω, due to different occlusions in different views; however, in general there is significant overlap between the matched pixels sets $\Omega_1, \ldots, \Omega_n$, which makes the equations in (5) to be highly interconnected.

The system of equations is used to estimate both the reciprocal-depth values (the set of elements $\Phi = \{\psi(i_\tau, j_\tau)\}$ appearing in the vectors $\mathbf{r}_1, \ldots, \mathbf{r}_n$) and the constants $\mathcal{C}^x = \{C_1^x, \ldots, C_n^x\}$. Assuming Gaussian distribution $N(0, \sigma_m^2)$ for elements of the error vectors $\mathbf{e}_m$, and independence for all errors, the negative log-likelihood function is a function of $\Phi, \mathcal{C}^x$ and $\mathcal{V}^x = \{\sigma_1^2, \ldots, \sigma_n^2\}$

$$\mathcal{J}^1_{\Phi, \mathcal{C}^x, \mathcal{V}^x} = \sum_{m=1}^{n} \sum_{q=1}^{n_m} \left(\frac{e_m(q)^2}{2\sigma_m^2} + \frac{1}{2} \log \sigma_m^2 \right) \tag{6}$$

One can easily show that the variances from the set $\mathcal{V}^x$ at optimality should satisfy

$$\begin{aligned} \hat{\sigma}_m^2 &= \frac{\sum_{q=1}^{n_m} \hat{e}_m(q)^2}{n_m} \\ &= \frac{\sum_{(i_\tau, j_\tau) \in \Omega_m} \left(D^c_{k_m,\ell_m}(i_\tau, j_\tau) - C_m^x \psi(i_\tau, j_\tau) \right)^2}{n_m} \end{aligned} \tag{7}$$

The criterion to be minimized with respect to $\Phi, \mathcal{C}^x$ remains

$$\mathcal{J}^2_{\Phi, \mathcal{C}^x} = \sum_{m=1}^{n} \left(\frac{1}{2} \log \frac{\sum_{q=1}^{n_m} (x_m(q) - \hat{r}_m(q) \hat{C}_m^x)^2}{n_m} \right). \tag{8}$$

The minimization of (8) with respect to the parameters $\{\psi(i_\tau, j_\tau)\}$ and $C_1^x, \ldots, C_n^x$ will be done alternately, by first re-estimating the set of elements $\{\psi(i_\tau, j_\tau)\}$ appearing in in $\mathbf{r}_1, \ldots, \mathbf{r}_n$ considering the current estimates of $C_1^x, \ldots, C_n^x$, and then reversing the role of the current and re-estimated parameters. We start with the stage of initial values for the parameters $C_1^x, \ldots, C_n^x$, which are known from the experiment setting. For given $C_1^x, \ldots, C_n^x$, the estimation of the element $\psi(i_\tau, j_\tau)$ involves all equations where $(i_\tau, j_\tau) \in \Omega_m$, leading to

$$\psi(i_\tau, j_\tau) = \frac{\sum_{m | (i_\tau, j_\tau) \in \Omega_m} \frac{D^c_{k_m,\ell_m}(i_\tau, j_\tau) C_m^x}{n_m \sigma_m^2}}{\sum_{m | (i_\tau, j_\tau) \in \Omega_m} \frac{(C_m^x)^2}{n_m \sigma_m^2}}. \tag{9}$$

The weighing by the noise variances is not possible at the first iteration, where we take all these variances equal and then they cancel from (9). Starting from the second iteration of (9), the current estimates of noise variances computed by (7) are used.

The iteration for finding new estimates of $C_1^x, \ldots, C_n^x$ uses the current estimates for reciprocal-depths and the current variances from (7):

$$\hat{C}_m^x = \frac{\sum_{(i_\tau, j_\tau) \in \Omega_m} D^c_{k_m,\ell_m}(i_\tau, j_\tau) \psi(i_\tau, j_\tau)}{\sum_{(i_\tau, j_\tau) \in \Omega_m} (\psi(i_\tau, j_\tau))^2} \tag{10}$$

Completely analogously we can treat the model for vertical matches

$$\begin{aligned} D^r_{k_1,\ell_1}(i,j) &= (Y_{k_1,\ell_1} - Y_{k_0,\ell_0}) f \frac{1}{z(i,j)} + \varepsilon(i,j) \\ &= (Y_{k_1,\ell_1} - Y_{k_0,\ell_0}) \psi(i,j) + \varepsilon(i,j) \end{aligned}$$

and obtain the vector equation for one region Ω_m

$$\mathbf{y}_m = \mathbf{r}_1 C_m^y + \varepsilon_m \tag{11}$$

and get a system of equations

$$\begin{cases} \mathbf{y}_1 &= \mathbf{r}_1 C_1^y + \varepsilon_1 \\ \vdots & \\ \mathbf{y}_n &= \mathbf{r}_n C_n^y + \varepsilon_n \end{cases} \tag{12}$$

similar to (5), by considering several views, $(k_1, \ell_1), \ldots, (k_n, \ell_n)$.

In fact the unknowns $\mathbf{r}_1, \ldots, \mathbf{r}_n$, $C_1^x, \ldots, C_n^x$, and $C_1^y, \ldots, C_n^y$ of the systems (5) and (12) can be solved together, by merging the systems (5) and 12) and solving with the same alternate approach. We present results with this joint version in this paper.

IV. Experimental results

The light field data used in the experiments is the set S_2 from the HDCA data [8], from which we keep the subarray 11×33 from the vertical locations $0:2:21$ and horizontal locations $2:3:98$ (according to the labeling of view files). The central view is shown in Figure 2a. We consider five reference views $\Gamma_{ref} = \{(1,1), (11,1), (1,33), (11,33), (6,16)\}$.

The horizontal and vertical disparities are obtained by estimating the optical flow between each reference view and the rest of 362 views, with the efficient coarse-to-fine patch match (CPM) estimation method [7].

We have estimated the reciprocal depthmaps at each reference view, using our proposed re-estimation algorithm. The PSNR of reconstructed disparity versus the iteration stage of the re-estimation algorithm is shown in Figure 3 in red. The PSNR corresponds to the MSE obtained by practically summing over all 362 views σ_m^2 given in (7). Convergence is fast, practically after two iterations the estimated quantities don't change significantly. The resulted reciprocal depthmap for the reference view $(6, 17)$ is shown in Figure 2 b.

ISBN 978-90-827970-1-5 © EURASIP 2018

A. Comparing the re-estimation for the whole image versus running re-estimation separately over each region

In order to improve the performance of the re-estimation procedure, we show that we can run the procedure separately over the regions of a partition. We consider the simple case of partition based on the depth values, as shown in Figure 2c. The partition into regions is based on the estimates of reciprocal depth obtained previously for the whole scene.

Again the convergence of the routine over each region is very fast, the PSNR for the whole image reconstruction being shown in blue in Figure 3 over each iteration (cumulating the results of the three different runs). The performance of reconstructing the initial pairwise matching data has improved. In a compression application this will involve only additional cost for encoding not only one set of estimated camera coordinates, $(\mathcal{C}^x, \mathcal{C}^y)$, but three of them, $\{(\mathcal{C}^x(\Omega^p), \mathcal{C}^y(\Omega^p))\}$. In Figure 4 we show the locations of $(\mathcal{C}^x(\Omega^p), \mathcal{C}^y(\Omega^p))$ for various regions and various views. The differences between a single region and specific regions are extremely small, so the cost of differentially encoding $(\mathcal{C}^x(\Omega^p), \mathcal{C}^y(\Omega^p))$ is very small.

B. Synthesizing color views not available at the encoder using re-estimated disparities

For the array of (11×33) views we consider the following scenario: We have available at the encoder a subset Γ_{design} of views, which can be used for the estimation of the reciprocal depth map Φ. The encoder will transmit the estimated map $\hat{\Phi}$ to the decoder and also the estimated centers of the views from the subset Γ_{design}. The encoder will also transmit the color views for the reference views set Γ_{ref}, which in our case is formed of the corner and center of the array, i.e., $\Gamma_{ref} = \{(1,1), (11,1), (1,33), (11,33), (6,17)\}$. The decoder needs to display the views on a freeview display, and will need to decode the color lightfield views from Γ_{ref}. Based on the decoded reference views, the decoder will then synthesize and render the views at specified camera positions, $\Gamma_{predict}$, in our case positions in the (11×33) array, where we know the ground truth. We consider the case when $\Gamma_{predict}$ does not contain any of the views used for design, Γ_{design}, so that the estimated map $\hat{\Phi}$ could not gather directly informations about disparities relevant to this set.

We consider three configurations of Γ_{design}: in Figure 5 are shown the configurations SUBSET and BORDER, and the last configuration is ORACLE, which would have access at the encoder on all the (11×33) views.

We use our re-estimation procedure at each of these configurations, for each of the reference views, obtaining $\{\hat{\Phi}_i^{ORACLE}, i \in \Gamma_{ref}\}$, $\{\hat{\Phi}_i^{SUBSET}, i \in \Gamma_{ref}\}$ and $\{\hat{\Phi}_i^{BORDER}, i \in \Gamma_{ref}\}$. Each estimated map is quantized into 511 levels and is losslessly encoded. For each configuration, the given target bitrate is used for encoding the depthmap, and the rest of the bits are used for encoding the five RGB references. Each view (k, ℓ) from $\Gamma_{predict}$ is synthesized using the five references, by warping first the closest reference to the view (k, ℓ), resulting in a warped image $\mathcal{W}$, where not all pixels are defined, due to occlusions. Then the second closest reference is warped to view with position (k, ℓ), but this time only the resulting locations that were occluded $\mathcal{W}$ are filled in. The process continues in the same way with the rest of the references.

The $PSNR_{YUV}$ values, [1], for the synthesized locations are computed and displayed in pseudocolor in Figure 6, for all the configurations, for all views from $\Gamma_{predict}$, and for five bitrates. One can see first that the views closer to the central view have a reasonable reconstruction, obtaining $PSNR_{YUV}$ as high as 39 dB, but as the location of the view goes further from the closest reference view, the $PSNR_{YUV}$ drops below 30 dB. One can note the almost radial distribution of the PSNR in the view array, according to the distance between each view and its closest neighbor, which is the array center $(6, 17)$ for $\Gamma_{predict}$.

The ORACLE configuration performs better than the others, as expected, showing that the reciprocal depth designed from all available views has a better quality and performs better in warping. However, for the design performed with least number of used views, which is the configuration BORDER, the estimated reciprocal depth map still performs well, loosing only 1dB when compared to the performance of ORACLE configuration.

V. Conclusions

We have introduced a re-estimation procedure for the reciprocal depth, given pairwise match estimates for pairs of views. The routine can be applied in a flexible way, and it was shown that for a simple partition into three regions the PSNR of reconstruction of matches is better than for a single region. With larger partitions the PSNR performance may still improve, at the cost of having to encode additional sets of coordinates parameters, one set for each region.

References

[1] ISO/IEC JTC 1/SC29/WG1 JPEG, "JPEG Pleno Call for Proposals on Light Field Coding," Doc. N74014, Geneva, Switzerland, January 2017.

[2] T. Ebrahimi, S. Foessel, F. Pereira and P. Schelkens, JPEG Pleno: Toward an Efficient Representation of Visual Reality, in *IEEE MultiMedia*, vol. 23, no. 4, pp. 14-20, Oct.-Dec. 2016.

[3] T. C. Wang, A. A. Efros, and R. Ramamoorthi, "Occlusion-aware depth estimation using light-field cameras," in *ICCV*, pp. 3487–3495, Dec 2015.

[4] M. W. Tao, Sunil Hadap, J. Malik, and R. Ramamoorthi, "Depth from combining defocus and correspondence using light-field cameras," in *ICCV 2013*, Washington, DC, USA, pp. 673–680, 2013.

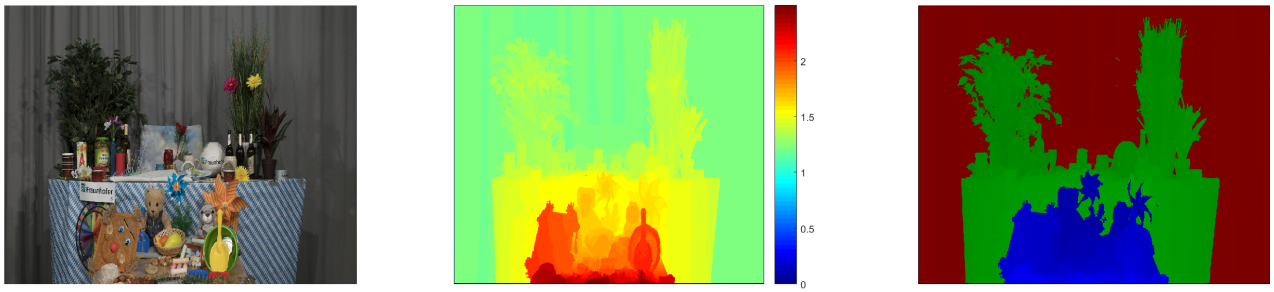
[5] A. Naman, R. Mathew, D. Ruefenacht, D. Taubman, "UNSW Depth Reciprocal Fields for the HDCA Dataset", ISO/IEC JTC 1/SC29/WG1 JPEG, Doc. N78000, Dec 2017.

[6] D. Ruefenacht, A. Naman, R. Mathew, D. Taubman, "Inter-View Compression Framework with Base Anchored Modeling and Inference", ISO/IEC JTC 1/SC29/WG1 JPEG, Doc. N78051, Jan 2018.

[7] Y. Hu, R. Song and Y. Li, Efficient Coarse-to-Fine Patch Match for Large Displacement Optical Flow, in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 5704-5712.

[8] M. Ziegler, R. op het Veld, J. Keinert and F. Zilly, "Acquisition system for dense lightfield of large scenes," in 2017 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), Copenhagen, Denmark, 2017, pp. 1-4.

ISBN 978-90-827970-1-5



(a) (b) (c)

Fig. 2. (a) The central view of the lightfield S_2; (b) Estimated reciprocal depthmap anchored at the central view, obtained by running the re-estimation algorithm for the whole scene Ω; (c) Partition of the scene $\mathcal{P} = \{\Omega^1, \Omega^2, \Omega^3\}$ into three regions $\Omega^p = \{(i,j)|\psi(i,j) \in \mathcal{I}_p\}$ with $\mathcal{I}_1 = (0; 1.27]$, $\mathcal{I}_2 = (1.27; 1.6]$, $\mathcal{I}_3 = (1.6; 2.5]$.

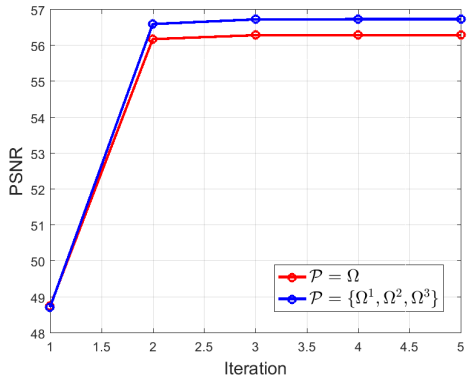


Fig. 3. PSNR when reconstructing the pairwise matching points, versus the iteration step in the re-estimation algorithm. In the case $\mathcal{P} = \Omega$ the algorithm is run on the whole image, considered as a single large region Ω. In case $\mathcal{P} = \{\Omega^1, \Omega^2, \Omega^3\}$ the re-estimation algorithm is run three times, once for each region Ω^p, obtaining the reciprocal depthmap for that region, plus a set of camera coordinate positions $(\mathcal{C}^x(\Omega^p), \mathcal{C}^y(\Omega^p))$ for each region. The central view is partioned into three regions as shown Figure 2c). It can be seen that the algorithm practically converges in two steps in both cases.

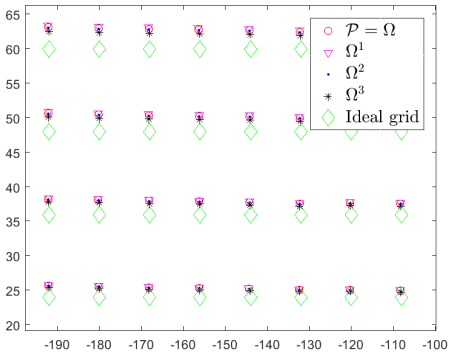
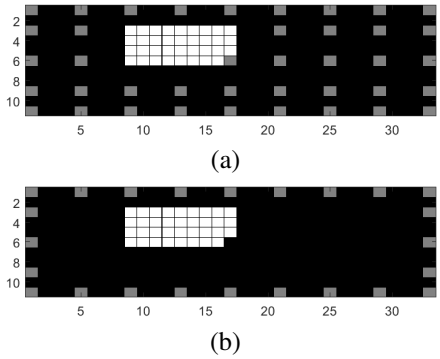


Fig. 4. Estimated positions of the ideal cameras $(\mathcal{C}^x(\Omega^p), \mathcal{C}^y(\Omega^p))$, when appying the re-estimation algorithm for the whole scene as a single Ω region (red circles) and when applying the re-estimation separately for the three regions forming the partition $\mathcal{P} = \{\Omega^1, \Omega^2, \Omega^3\}$, as shown in Figure 3. Also shown are the ideal positions, in the regular grid (green diamonds). For better viewing only the positions for a small subset of views are shown.



(a) (b)

Fig. 5. The array of (11×33) views: in gray, the views used for reciprocal depth estimation; in white: the views for which the color image has to be synthesized by warping the references. We assume five references: (1,1);(11,1);(1,33);(11,33);(6,17) (a) The SUBSAMPLE configuration for estimating Φ; (b) the BORDER configuration for estimating Φ. The ORACLE configuration uses all (11×33) views for reciprocal depth estimation, Φ.

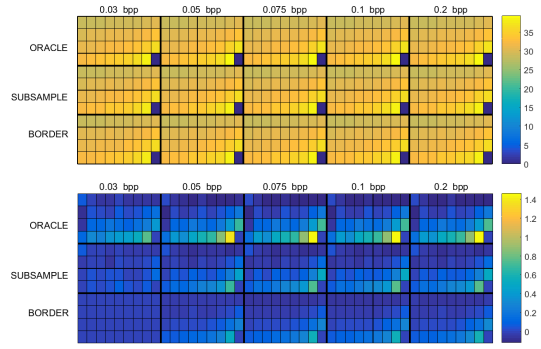


Fig. 6. The $PSNR_{YUV}$ performance over the color synthesized views, under different experimental configurations and different bitrates. The location of the (4×9) represented blocks of synthesized views is shown in Figure 5. The view $(6, 17)$ is a reference and is not synthesized, so no $PSNR_{YUV}$ value is given for it. (Top) Here the $PSNR_{YUV}$ values are represented in pseudocolor in each of the 5 bitrates and 3 configurations; (Bottom) To increase the legibility of (a), here the improvements are shown in pseudocolor, where improvements are with respect to the worst performance (which is at 0.03 bpp, for the BORDER configuration). As much as 0.2 dB are gained in SUBSAMPLE configurations, and 0.8 dB in ORACLE configuration, between corresponding bitrates. .

ISBN 978-90-827970-1-5 © EURASIP 2018