IMPROVING RESIDUAL CODING OF WASP LIGHT FIELD CODEC

Pekka Astola and Ioan Tabus

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

ABSTRACT

We study the residual coding by JPEG 2000 in a recently proposed light field compression method, WaSP, with the goal of optimizing the performance with respect to a weighted PSNR criterion, which is the objective performance criterion utilized in the JPEG Pleno Light Field standardization. In WaSP the residual view is encoded by using JPEG 2000. In here we show that instead of encoding by JPEG 2000 the RGB residual view (using the rate-distortion optimization performed by JPEG 2000 for the RGB image), one can obtain better results by performing first an RGB to YCbCr color transformation and reallocating the given bitrate to each color component optimally with respect to the weighted PSNR criterion. Then JPEG 2000 is called separately for each YCbCr component with the selected bitrates. We experiment with the proposed scheme inside the WaSP codec, and show that the weighted PSNR results are improved by nearly as much as 1 dB when compared to the default RGB encoding by JPEG 2000.

Index Terms— Light field compression, residual coding, image coding, rate-distortion optimization.

1. INTRODUCTION

Recording and rendering light fields involve a huge amount of data and hence the necessity of dedicated compression methods for light field data. The current standardization work for JPEG Pleno Light Field [1] is centered around the compression of an $N \times M$ array of subaperture images, acquired with a high density camera array, or with a plenoptic camera. The call for proposals for the new standard [2] has defined a weighted average of the sum of the individual PSNRs of the views as one objective performance criterion, where PSNR of the color components are weighted with fixed numbers, based on perceptual quality considerations. Although the final evaluation of the proposed compression methods in the standardization process will be done through subjective testing, the intermediate development stages are based on objective criteria. The weighted PSNR and SSIM are the recommended criteria within the JPEG Pleno Light Field project [2] [3]. The good correlation between the proposed objective criteria and the subjective evaluation results was shown in [4], which evaluated four light field compression methods.

In the light field compression literature there is a rich body of publications transforming the initial light field array of views into a pseudo-temporal sequence of views and compressing the sequence using efficient video coding methods such as HEVC or multi-view HEVC [5] [6]. The rate allocation mechanism built-in in the underlying video compression tools are implicitly used in most proposals, which leads to a good performance of the overall compression scheme. One step further, specific mechanisms for light field images were recently proposed for optimizing the bit rate allocation for each view (pseudo- frame) in the underlying HEVC codec [7] [8].

In the 80th meeting, the JPEG Pleno work group adopted our light field codec WaSP [9] as the initial verification software of the upcoming standard. In the 81st meeting an improved version of WaSP was selected to continue as the latest version of the verification software. The codec represents an unification of most useful features of the earlier codecs presented in [10] and [11], which were compressing first several *reference* angular views, and then encoding conditionally on these references all other views called *intermediate views*.

In the light field codec WaSP the intermediate views are reconstructed by first warping several neighbor reference views to the location of the current side view, then merging the warped versions to obtained a predicted version of the current side view, and then by encoding the residuals between the original side view image and its predicted version. We call residual view the image of the above residuals, and we encode this residual view by JPEG 2000 [9].

WaSP added a new mechanism of "inter-view" encoding of the reference views, by using a hierarchical scheme for warping-prediction and residual encoding of the reference views, which was shown to provide a performance on par with the encoding by HEVC the pseudo-temporal sequence of views. However, WaSP uses JPEG 2000 [12] as the underlying codec for each reference view and for each residual view, which is justified by the backward compatibility requirement in the call for proposals [2].

As a new development, a different option of coding will

be added for the reference views in the next verification model of JPEG Pleno Light Field project, based on 4D DCT transforms, [13]. Still, the residual view coding by JPEG 2000 remains a main feature of the verification model. In this paper we study the ways to increase the weighted PSNR performance of the overall WaSP scheme, by improving the bit allocation for the individual color components in JPEG 2000 encoding of the views. The results are evaluated under a weighted PSNR criterion, as defined in the JPEG Pleno Light Field call for proposals.

The underlying codec for each reference view is considered in WaSP to be JPEG 2000. JPEG 2000 has a very precise capability of allocating a given rate to any color component of an image. Consequently, the bit allocation problem is formulated as follows. Given a bitrate R allocated for one view, find the optimal split of bitrate into R_1 and $2R_2$, where R_1 is used for the luminance component and R_2 is used for each chrominance component. So the main algorithmic improvement to the baseline version of WaSP is to find a suitable split curve, $R_1(R)$ for each encoded reference or residual view. We note that we are allocating for each of $\mathcal{C}_b$ and $\mathcal{C}_r$ components the same number of bits, since in general their rate-distortion properties are very similar, but the same technique can be used for allocating separately the bitrates for $\mathcal{C}_b$ and $\mathcal{C}_r$, if one accepts a slight increase of complexity.

In the first part of the paper we generate ideally optimal $R_1(R)$ curves (optimal under the weighted PSNR criterion) for the models extracted from experimental data of PSNR-Rate obtained by JPEG compression of each color component at a given view. We then perform a similar study when the optimality criterion is the weighted MSE criterion (which is the default criterion when JPEG 2000 optimizes its bitrate when encoding jointly an RGB image). We compare the two ideally optimal curves and notice that the optimal split for high bitrates is very different for the two optimality criteria (weighted PSNR and weighted MSE). That leads to a simple practical algorithm for obtaining an improved weighted PSNR performance, by testing experimentally at each view several split ratios.

In the Sections II and III we intend to precisely model the fidelity-rate curves obtained with JPEG 2000 for each individual color component and obtain a very dense set of points (PSNR-Rate), close to the ones obtained by the internal optimization of JPEG 2000 codec, without the need of extensively running the JPEG 2000 codec. We show in Subsection 2.3 that a high polynomial model for the dependency (log MSE-log rate) is better suited than the linear model (log MSE-log rate) used for example in [8].

We apply this technique for independently coding of the reference views, and then also for the coding of prediction residuals of dependent (intermediate) views. We notice different patterns of the $R_1(R)$ curves, different in the two cases (weighted PSNR and weighted MSE).

We present in Section II the modeling problem and then we show in Section III the method for obtaining the optimal rate allocation for the models found in Section II and exemplify the obtained rate allocation for two cases: first, encoding of reference images, independently by JPEG 2000, and second, encoding the residual images (obtained after warping and prediction) by JPEG 2000. In Section IV we give the experimental algorithm used for improving the split in WaSP, as a fast bit allocation strategy, which picks from several allocation rules the winner according to the JPEG Pleno weighted PSNR criterion. We show in Section V experimental results obtained over four light field datasets. We finish with conclusions in Section 6.

2. MODELLING PSNR-VS.-RATE

We present a modeling tool for the fidelity-rate curve obtained with JPEG 2000 for each individual color component, in the $\mathcal{Y}$, $\mathcal{C}_b$, and $\mathcal{C}_r$ space, and we will use it for analyzing the independent encoding of an image and then also for encoding an image residual. The need to analyze the two cases separately arises because the JPEG 2000 compressor was intended for encoding of original photographic images, for which its rate-distortion performance was optimized, while in the second case, of encoding the residual of a current image (obtained after subtracting the optimally merged warped predictions from the current image), the compression of the residual image is expected to be more difficult, mainly due to the loss of smoothness for the residual image.

2.1. The modeling problem

In this section we exemplify the modeling problem with PSNR-rate or distortion-rate plots obtained over the angular views of the lenslet dataset "Bikes" [14], each view having dimensions $n_r \times n_c$. The color RGB to YCbCr transformation is chosen here to be performed according to ITU-R BT.709-6 recommendation.

We denote $\mathcal{Y}$, $\mathcal{C}_r$, and $\mathcal{C}_b$ the color component images. We define a set of "test" bitrates $R_i, i = 1, \ldots, n_{br}$, spread logarithmically over the range of interest $[0.01; 5]$ bits per pixel (hence the total budget for each color component is given by $n_r n_c R_i$).

For defining the performance criteria we denote

$$MSE_Y(R_i) = \frac{1}{n_r n_c} \sum_{i_r=1}^{n_r} \sum_{i_c=1}^{n_c} (\mathcal{Y}_{i_r,i_c} - \tilde{\mathcal{Y}}_{i_r,i_c}(R_i))^2,$$

where $\tilde{\mathcal{Y}}$ is obtained by compressing the signal component image $\mathcal{Y}$ with JPEG 2000 using $n_r n_c R_i$ bits, and similarly we define $MSE_{C_r}(R_i)$ and $MSE_{C_b}(R_i)$.

The componentwise PSNRs are defined by

$$PSNR_Y(R_i) \quad = \quad 10 \log_{10} \frac{2^{2M}}{MSE_Y(R_i)}, \qquad (1)$$

where M is the bitdepth of the initial image (in our experiments all images have $M = 10$ bits per pixel), and in a similar manner we define $PSNR_{C_r}(R_i)$ and $PSNR_{C_b}(R_i)$.

By taking as initial image data the central subaperture image (SAI) of the lenslet dataset "Bikes", the collected plots $PSNR_Y(R_i)$, $PSNR_{C_b}(R_i)$, and $PSNR_{C_r}(R_i)$ are shown in Fig. 1. One can notice that one effect of the multiplication by the color matrix of the original RGB components (that were initially rather similar in terms of PSNR-vs.-Rate) is that the chroma components $\mathcal{C}_r$, and $\mathcal{C}_b$ have a much higher PSNR than the $\mathcal{Y}$ component, at a given R_i. In the following we will use only two models: first $PSNR_Y(R_i)$, for the component $\mathcal{Y}$, and second $PSNR_C(\tilde{R}_i)$, accounting for the two chroma components, $PSNR_C(\tilde{R}_i) = (PSNR_{C_r}(R_i) + PSNR_{C_b}(R_i))$, where $\tilde{R}_i = 2R_i$.

2.2. A polynomial model for PSNR

After testing several alternatives, we found a suitable model in the form of a polynomial in the log-bitrate. Let denote by $PSNR(R_i)$ the criterion of interest, e.g., $PSNR(R_i) = PSNR_Y(R_i)$. Starting from the experimental data available, $\{PSNR(R_1), \ldots, PSNR(R_{n_{br}})\}$, we estimate a model of the following form

$$PSNR(R) \quad = \quad \sum_{i=0}^{n_P} A_i \cdot (\log R)^i,$$

where n_P is the degree of the polynomial, taken $n_P = 6$ for a reasonably good fit, and log denotes the natural logarithm. We obtain by least squares the estimates of the parameters $\hat{A}_0, \ldots, \hat{A}_{n_P}$ from the given data, and obtain the particular criterion $PSNR(R) = \sum_{i=0}^{n_P} \hat{A}_i (\log R)^i$.

With the estimated model we can generate an approximation of $PSNR(R)$ for a much finer grid of bitrates in the initial range. We illustrate in Fig. 1 the model points, represented using the continuous line, showing a very precise match to data. This model, polynomial in the logarithm of the rate, is much more precise than a polynomial model on the rate (the later one matches well the data at lowest rates but starts to deviate significantly from the data for the large rates). Since the two chroma components have very similar curves, we choose to treat them in a single term, cumulating $PSNR_C(\tilde{R}_i) = PSNR_{C_r}(R_i) + PSNR_{C_r}(R_i)$, which requires $\tilde{R}_i = 2R_i$ bits.

For the rate-fidelity type of criterion optimization we need the derivatives of the criterion PSNR with respect to the rate, which can be found easily from the model as:

$$\frac{\partial PSNR}{\partial R} = \sum_{i=1}^{n_P} i \hat{A}_i (\log R)^{i-1} \frac{1}{R}. \tag{2}$$

2.3. The corresponding polynomial model for MSE

Since the PSNR model showed a very good match to experimental data, we extract from it a model for the MSE. It is simple to consider the PSNR-MSE relationship to get

$$\begin{aligned} \log MSE(R) \quad &= \quad \frac{20M \log 2 - PSNR \log 10}{10} \\ &= \quad \sum_{i=0}^{n_P} C_i (\log R)^i, \end{aligned}$$

with the coefficients $\{C_i\}$ resulting in a straightforward manner from $\{A_i\}$. So this model is a "log-log polynomial" model, representing the logarithm of MSE as a polynomial of the logarithm of the rate. We should note that such a polynomial model of order $n_P = 1$ has the equivalent form $MSE(R) = \alpha R^\beta$, which is the model class used in [8]. However the linear matching in the plots of Fig. 7 (represented in log-log coordinates) will not provide a good approximation, while the $n_P = 6$ is seen there to provide very close matching.

The derivative of the MSE model with respect to rate is

$$\begin{aligned} &\frac{d}{dR} MSE(R) = \frac{d}{dR} exp(\sum_{i=0}^{n_P} C_i (\log R)^i) \\ &= exp(\sum_{i=0}^{n_P} C_i (\log R)^i) \sum_{i=1}^{n_P} i C_i (\log R)^{i-1} \frac{1}{R} \end{aligned} \tag{3}$$

3. OPTIMIZING THE ALLOCATION OF COMPONENT BITRATE FOR PSNR-RATE AND MSE-RATE

3.1. The optimization problem for PSNR optimization

We want to find for a given bitrate R the optimal split between the luminance and chroma components, so that the weighted PSNR is maximized

$$\begin{aligned} &\max \{\alpha PSNR_Y(R_1) + \beta (PSNR_{C_b}(R_2) + PSNR_{C_r}(R_2))\} \\ &\text{Subject to} \quad R_1 + 2R_2 = R \end{aligned}$$

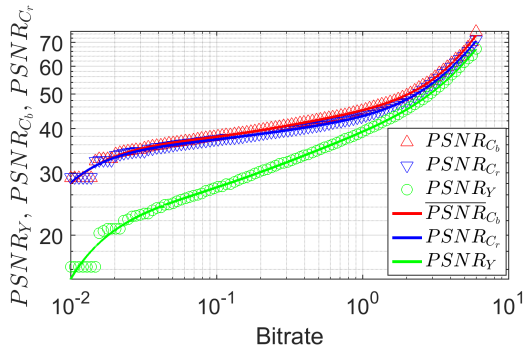
where the scalars α and β are given, e.g., in the current JPEG Pleno call for proposals the values were set to $\alpha = \frac{6}{8}$ and $\beta = \frac{1}{8}$ [3]. Such a criterion was shown to correlate well the objective criterion values to the MOS obtained in subjective testing for several light field coding schemes [4].

Denoting by $\{\hat{A}_i^1\}$ the model parameters for luminance, i.e., $J_1(R) = PSNR_Y(R) = \sum_{i=0}^{n_P} \hat{A}_i^1 (\log R)^i$, and by $\{\hat{A}_i^2\}$ the model parameters for the combined chroma criterion, i.e., $PSNR_C(\tilde{R}) = \sum_{i=0}^{n_P} \hat{A}_i^2 (\log \tilde{R})^i$, the weighted criterion becomes

$$\begin{aligned} &\max \alpha PSNR_Y(R_1) + \beta PSNR_C(\tilde{R}_2) \\ &\text{Subject to} \quad R_1 + \tilde{R}_2 = R. \end{aligned}$$

Using the Lagrange multiplier method, the extended criterion becomes

$$\max \alpha PSNR_Y(R_1) + \beta PSNR_C(\tilde{R}_2) + \lambda (R - R_1 - \tilde{R}_2)$$



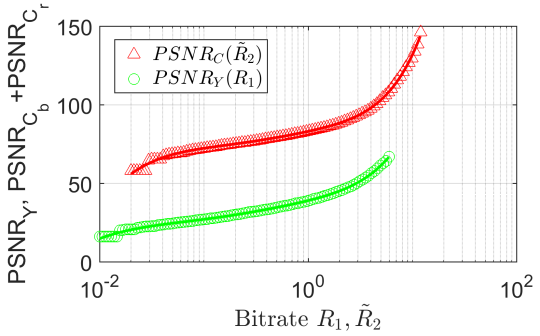


Fig. 1. (Left) The PSNR vs. Rate for the $\mathcal{Y}$, $\mathcal{C}_b$, and $\mathcal{C}_r$ components of the central angular view of dataset "Bikes" [14]. The markers are showing the experimental values obtained with JPEG 2000 and the full lines are showing the output of the corresponding log-polynomial models. (Right) The corresponding models $PSNR_Y(R_i)$ and $PSNR_C(R_i)$ (since $PSNR_{C_b}(R_i)$ and $PSNR_{C_r}(R_i)$ are very similar, they are grouped together as $PSNR_C(\tilde{R}_i) = (PSNR_{C_r}(R_i) + PSNR_{C_b}(R_i))$).

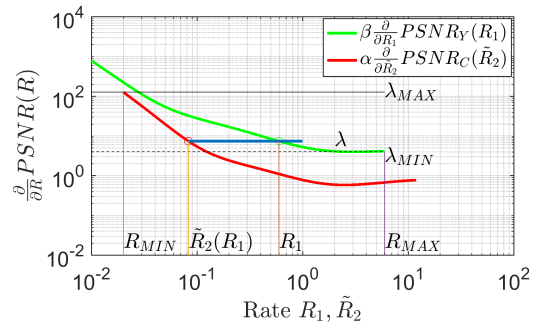


Fig. 2. The derivatives of PSNR vs. Rate for the $\mathcal{Y}$, and combined $\mathcal{C}$ components at the central angular view of dataset "Bikes". For optimal allocation, the correspondence $\tilde{R}_2(R_1)$ is found by tracing the horizontal line λ at all values $\lambda \in [\lambda_{MIN}; \lambda_{MAX}]$ and taking the crossing with the second curve as R_1 and the crossing with the first curve as $\tilde{R}_2(R_1)$.

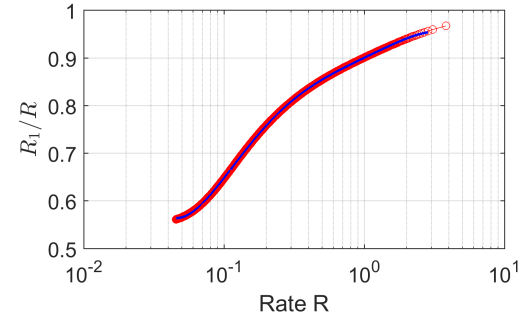


Fig. 3. The optimal allocation R_1/R versus R, where the total available rate is $R = R_1 + 2R_2(R_1) = R_1 + \tilde{R}_2(R_1)$, with the corresponding R_1 and $\tilde{R}_2$ found from the plots of Fig. 2. The same shape, monotonically increasing, was obtained by the optimization process for all views encoded by WaSP as reference views in the analyzed light fields in Section 5.

with the necessary optimality conditions

$$\alpha \frac{\partial PSNR_Y(R_1)}{\partial R_1} = \lambda$$
$$\beta \frac{\partial PSNR_C(\tilde{R}_2)}{\partial \tilde{R}_2} = \lambda.$$

For the polynomial log-bitrate model

$$\begin{aligned} \alpha \sum_{i=1}^{n_P} i \hat{A}_i^1 (\log R_1)^{i-1} \frac{1}{R_1} &= \lambda, \\ \beta \sum_{i=1}^{n_P} i \hat{A}_i^2 (\log \tilde{R}_2)^{i-1} \frac{1}{\tilde{R}_2} &= \lambda. \end{aligned} \quad (4)$$

Fig. 2 shows the two weighted curves

$$\alpha \sum_{i=1}^{n_P} i \hat{A}_i^1 (\log R_1)^{i-1} \frac{1}{R_1},$$
$$\beta \sum_{i=1}^{n_P} i \hat{A}_i^2 (\log \tilde{R}_2)^{i-1} \frac{1}{\tilde{R}_2},$$

and the way the parametrization by λ in (4) leads to obtaining the optimal $\tilde{R}_2(R_1)$. Since the total available rate is $R = R_1 + \tilde{R}_2(R_1)$, we obtain the optimal dependency $R(R_1) = \frac{R_1}{R_1 + \tilde{R}_2(R_1)}$. This relationship is show in the more convenient representation R_1/R versus R in Fig. 3. It shows what is the optimal ratio between luminance bitrate and available bitrate, depending on the available bitrate for that view. This quantity is the relevant one in the partitioning an available rate of R bits between the components $\mathcal{Y}$, $\mathcal{C}_r$, and $\mathcal{C}_b$, with R_1 bits to $\mathcal{Y}$, and $\tilde{R}_2/2$ bits to each of $\mathcal{C}_r$, and $\mathcal{C}_b$. In practice, one can obtain the curve $R_1(R)$ by collecting statistics from the following faster process: at a given rate R, one can run the JPEG 2000 coding with R_1 for luminance and $(R - R_1)/2$ for each chrominance component, for a range of values of R_1,

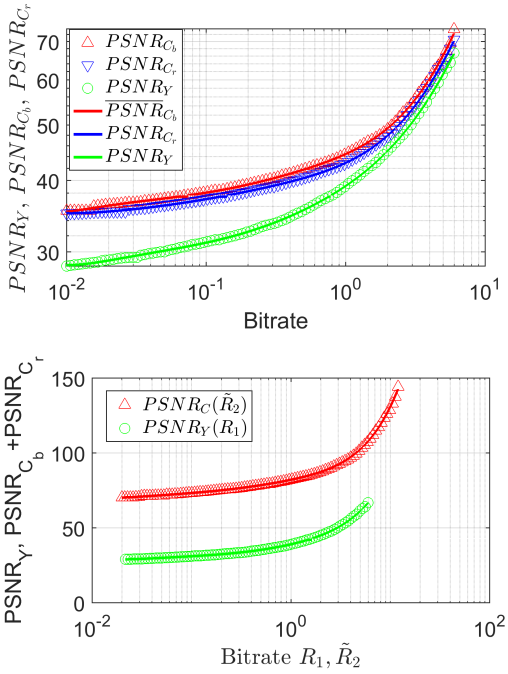


Fig. 4. (Top) The PSNR vs. Rate for the $\mathcal{Y}$, $\mathcal{C}_b$, and $\mathcal{C}_r$ components at one side view of the dataset "Bikes". The side view is reconstructed starting from the predicted version obtained by warping the central view plus the encoded residual component images, obtained at the shown rates. The lines have the same significance as in Fig. 1.

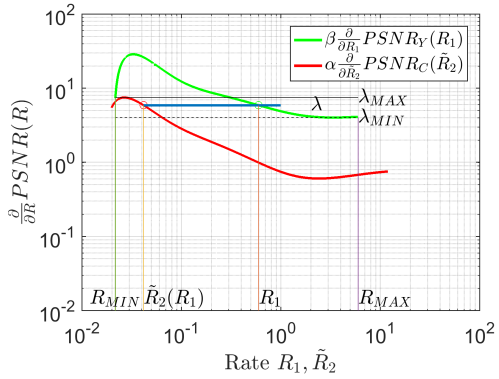


Fig. 5. The derivatives of PSNR vs. Rate for the $\mathcal{Y}$, and combined $\mathcal{C}$ components at a side view reconstructed from a warped version plus encoded residuals, with the shown bitrates. The correspondence $\tilde{R}_2(R_1)$ is found similarly to the procedure shown in Fig. 2.

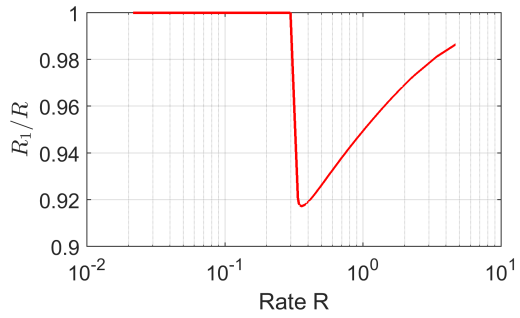


Fig. 6. The optimal allocation R_1/R versus R, where the total available rate is $R = R_1 + 2R_2(R_1) = R_1 + \tilde{R}_2(R_1)$, with the corresponding R_1 and $\tilde{R}_2$ found from the plots of Fig. 5. Similar shapes were observed when encoding side views conditional on reference views for the dataset "Bikes" .

say $R_1 \in \{0.6, 0.7, \ldots, 0.9\}$, and observe for which value of R_1, say $\hat{R}$, one obtains the largest weighted PSNR. Repeating the process for several R values one will obtain points $\hat{R}_1(R)$ approximately sitting on the curve from Fig. 3. Since the curve in Fig. 3 is monotonically smooth, one will need only few points, e.g., $R_1 \in \{0.6, 0.7, \ldots, 0.9\}$, to obtain a good approximation of Fig. 3. In all cases of reference views that we studied with the modelling-optimality approach from this section we found the same monotonically shape as in Fig. 3, so that the practical approach may provide the needed characteristic $R_1(R)$ only using a few running of the JPEG 2000.

3.2. The optimization problem for reconstruction by prediction plus residuals

We study now the case of optimally encoding the residuals using the weighted PSNR criterion. For exemplifying we use one of the side views from dataset Bikes, for a given prediction image obtained by warping the central view to the location of this side view [9].

The resulting PSNR values for the color components of the reconstructed image are shown in Fig. 4, as functions of the bitrates used for each color component. The derivatives shown in Fig. 5 are now having a different relationship than the ones observed in Fig. 2. Especially important is the fact that for a large range of values of the parameter λ, namely those above the line marked λ_{MAX} in Fig. 4, there is a single intersection with the derivative curve, and therefore bitrate should be allocated only to the corresponding color component, in this instance the luminance component. The allocation of the bitrate for R_1 is shown for this case in Fig. 6, where at the left range of the total rate R the whole bitrate is allocated to luminance, while the chrominance components are skipped from encoding. This happens at the low bitrates, where it pays off to improve only the luminance of the predicted view, leaving the predicted chrominance components untouched. After the bitrate R becomes larger, it becomes optimal to improve also the chrominance components, since nonzero bitrates will be allocated to their residuals. As in the discussion for Fig. 3, one can obtain ex-

perimentally the dependency from Fig. 6, by performing a few JPEG encoding with a small number of R and R_1 values, since the shape is very regular, being composed of a constant segment at $R_1/R = 1$ and then a monotonically increasing smooth curve. The quick decreasing transition between the two curves is reflecting the transition between the two phases of the characteristics from Fig. 5, and it is optimal only for a very limited range of rates R, so one could neglect it, and replace it with the $R_1 R = 1$, without dramatic loss.

3.3. The optimization problem for weighted MSE criterion

We now address the problem of optimal allocation when the criterion is a weighted MSE. We note that this is the case in the optimization done in JPEG 2000 [15], so this is the optimal allocation performed implicitely in the baseline WaSP, which we want to improve.

We now formulate the similar problem, of finding for a given bitrate R the optimal split between the rates for luminance and chroma components, so that the weighted MSE is minimized

$$\min \{\gamma MSE_Y(R_1) + \delta(MSE_{C_b}(R_2) + MSE_{C_r}(R_2))\}$$
$$\text{Subject to} \quad R_1 + 2R_2 = R$$

where the scalars γ and δ are given. Denoting $MSE_1(R) = MSE_Y(R)$ and $MSE_2(\tilde{R}) = MSE_{C_b}(R) + MSE_{C_r}(R)$, with $\tilde{R}_2 = 2R_2$, it is easy to derive similarly to the PSNR problem that the necessary optimum conditions will be now:

$$\gamma \frac{\partial MSE_1(R_1)}{\partial R_1} = \lambda$$
$$\delta \frac{\partial MSE_2(\tilde{R}_2)}{\partial \tilde{R}_2} = \lambda \tag{5}$$

and finding of the optimal solution will follow the same procedure, now using the conditions (5). The necessary condition of the optimality for the pair of rates R_1 and $\tilde{R}_2$ is

$$\frac{\partial MSE_1(R_1)}{\partial R_1} = \frac{\delta}{\gamma} \frac{\partial MSE_2(\tilde{R}_2)}{\partial \tilde{R}_2}. \tag{6}$$

It is interesting to see what is the connection between the optimal weighted PSNR and optimal MSE solutions. Since the criteria are connected as:

$$\begin{aligned} \frac{\partial PSNR_1}{\partial R_1} &= \frac{-\partial 10 \log_{10}(MSE_1)}{\partial R_1} \\ &= \frac{-10}{\log 10} \frac{\partial \log(MSE_1)}{\partial R_1} \\ &= \frac{-10}{MSE_1 \log 10} \frac{\partial MSE_1}{\partial R_1} \end{aligned} \tag{7}$$

when we require at a certain λ that

$$\frac{\partial PSNR_1(R_1)}{\partial R_1} = \frac{\beta}{\alpha} \frac{\partial PSNR_2(\tilde{R}_2)}{\partial \tilde{R}_2},$$

that is equivalent of requiring that

$$\frac{\partial MSE_1(R_1)}{\partial R_1} = \frac{\beta}{\alpha} \frac{MSE_1(R_1)}{MSE_2(\tilde{R}_2)} \frac{\partial MSE_2(\tilde{R}_2)}{\partial \tilde{R}_2}.$$

Comparing to equation (6), we see that the weighted PSNR solution is equivalent to a weighted MSE solution, where the weights $\frac{\delta}{\gamma}$ are replaced with varying weights $\frac{\beta}{\alpha} \frac{MSE_1(R_1)}{MSE_2(\tilde{R}_2)}$.

We exemplify in Figs. 7-9 the same elements as in Figs. 1-3, this time for the optimality criterion being the weighted combination of MSE. We take the MSE weights in the way they are usually set, with equal weights for all components, $\gamma = 1$ and $\delta = 1$. One can see that the optimal allocation of the bitrate R to component Y is different than the one shown in Fig. 3, especially at the high bitrates. So, when the criterion of interest is weighted PSNR, one should utilize the type of allocation from Fig. 3, instead the one from Fig. 9.

4. A PRACTICAL SPLIT ALLOCATION ALGORITHM

In order to obtain a gain in the overall performance of WaSP, without increasing very much the computational complexity, we choose to add before the residual encoding step the operations described in the following algorithm, motivated by the findings in the previous two sections. The implementation of JPEG 2000 used here is the free implementation by Kakadu.

1. Given: one angular view A (which is a $n_r \times n_c$ RGB image) from the light field array, a prediction $\hat{A}$ of A based on the already encoded reference views, and the current rate R allocated in WaSP for encoding the residual view $A^{res} = A - \hat{A}$.

2. Transform the RGB image A^{res} into the three components $\mathcal{Y}$, $\mathcal{C}_b$, and $\mathcal{C}_r$.

3. For $\rho \in \mathcal{R}$, set $R_1 = \rho R$: encode and decode the three components by JPEG 2000 with $\mathcal{Y}$ using R_1 bits, $\mathcal{C}_b$ using $(R - R_1)/2$ bits, and $\mathcal{C}_r$ using $(R - R_1)/2$ bits. Transform the decoded components back in RGB space, denoting $\hat{A}^{res}$ the RGB decoded image. Obtain the reconstructed image as $A^{rec} = \hat{A}^{res} + \hat{A}$. Compute the weighted $PSNR_{YUV}(R_1)$ criterion between the image A^{rec} and A.

4. Pick the R_1^* for which the maximum $PSNR_{YUV}(R_1^*)$ is obtained.

5. Encode also with JPEG 2000 the RGB image using the default configuration (with parameter "-noweights"), and evaluate the result as $PSNR_{YUV}^{default}(R)$.

6. Pick the better method by comparing $PSNR_{YUV}^{default}(R)$, and $PSNR_{YUV}(R_1^*)$, and use it for finally encoding the residual image.

We note that the set of rates $\mathcal{R}$ may contain only a few points since we expect a smooth behavior of the performance with respect to R_1, as shown in the previous sections. Also, the set $\mathcal{R}$ might contain the important point $\rho = R_1/R = 1$, which is needed for covering the optimal behavior shown in

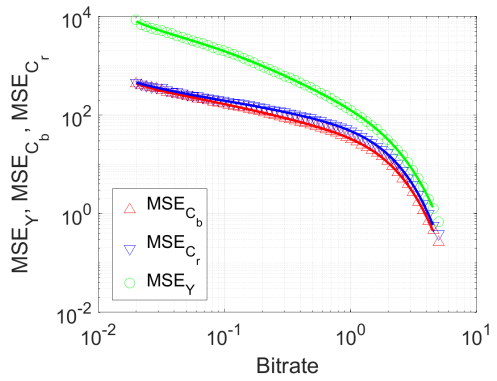


Fig. 7. The MSE vs. Rate for the $\mathcal{Y}$, $\mathcal{C}_b$, and $\mathcal{C}_r$ components. The markers are showing the experimental values obtained with JPEG 2000 and the full lines are showing the output of the corresponding log-log-polynomial models of order $n_P = 6$.

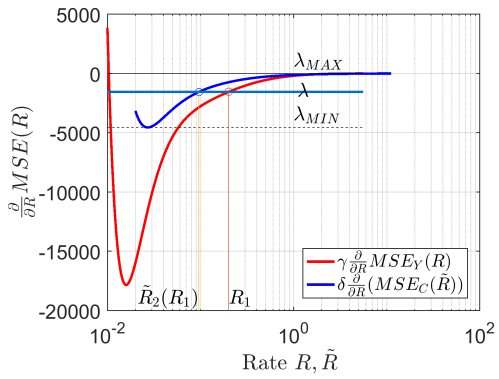


Fig. 8. The derivatives of MSE vs. Rate for the $\mathcal{Y}$ and combined $\mathcal{C}_b$, and $\mathcal{C}_r$ components, weighted by $\gamma = 1$ and $\delta = 1$, respectively. The correspondence $\tilde{R}_2(R_1)$ is found by tracing the horizontal line λ at all values $\lambda \in [\lambda_{MIN}; \lambda_{MAX}]$ and taking the crossing with the second curve as R_1 and the crossing with the first curve as $\tilde{R}_2(R_1)$.

Fig. 6. We perform in Step 6 the evaluation of the default RGB encoding by JPEG 2000 for the reason that the RGB version might obtain better results, because of exploiting additional correlations between the components during the inner operations of JPEG 2000, which the individual calls for each component in Step 3 are not exploiting. The results are shown in the next section.

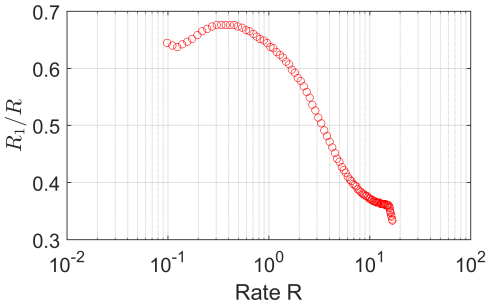


Fig. 9. The optimal allocation $R_1(R) = \frac{R_1}{R_1+2R_2(R_1)}$ found from the plots of Fig. 8.

5. EXPERIMENTAL RESULTS

We investigated the effect of having a quasi-optimal allocation of bitrate for color components in the JPEG Pleno Light Field verification model. We have used two possible allocations: the first one is to call JPEG 2000 for the combined RGB image, where the JPEG 2000 encoder performs internally the rate allocation to color components and also utilizes the possible correlations between the components. The second method calls JPEG 2000 individually for the Y, Cb, Cr components, and searches for the best point out of ten possible ratios R_1/R spaced equidistantly between 0.85 and 1.00. We experimented with a selection of datasets from [14].

In Fig. 10 we show in red the $PSNR_{YUV}$ results for the JPEG 2000 default rate allocation, and in blue the results for the presented YCbCr rate allocation. The improvements by the presented method can be seen especially well at the higher bit rates where the gain is almost 1 dB on the weighted PSNR metric.

Fig. 11 illustrates the result on the SSIM metric. The improvements can be seen at all bit rates, and as with the PSNR we can observe an increase in the gain at the higher rates.

6. CONCLUSIONS

In this paper we present a rate allocation method for improving the residual coding performance in WaSP light field codec. We show that the codec's performance can be improved by splitting the residual rate unevenly between the luminance, and the chrominance components. We report gains of nearly 1dB on the weighted PSNR metric used by the JPEG Pleno light field work group.

7. REFERENCES

[1] 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.

[2] ISO/IEC JTC 1/SC29/WG1 JPEG, "JPEG Pleno Call

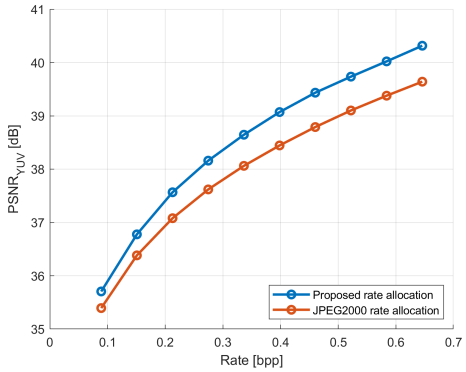


Fig. 10. The average $PSNR_{YUV}$ obtained over four light field datasets: "Bikes", "Danger de Mort", "Stone Pillars Outside", and "Fountain and Vincent 2" [14].

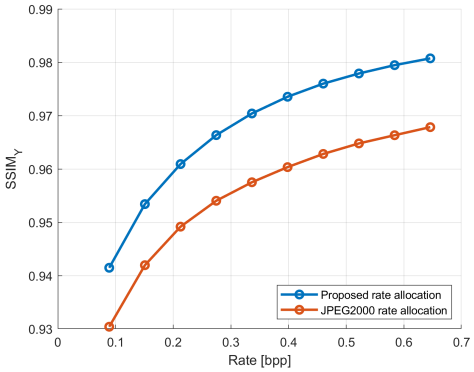


Fig. 11. The average $SSIM_Y$ obtained over four light field datasets: "Bikes", "Danger de Mort", "Stone Pillars Outside", and "Fountain and Vincent 2" [14].

for Proposals on Light Field Coding," Doc. N74014, Geneva, Switzerland, January 2017.

[3] ISO/IEC JTC 1/SC29/WG1 JPEG, "JPEG Pleno Light Field Coding Common Test Conditions," in ISO/IEC JTC 1/SC29/WG1 JPEG, Doc. N80027, July 2018.

[4] I. Viola and T. Ebrahimi, "Quality Assessment of Compression Solutions for ICIP 2017 Grand Challenge on Light Field Image Coding", 2018 International Conference on Multimedia and Expo Workshops, San Diego, California, USA, July 23-27, 2018.

[5] S. Zhao and Z. Chen, "Light Field Image Coding via Linear Approximation Prior", in 2017 IEEE International Conference on Image Processing (ICIP), Sept 2017, pp. 4562–4566.

[6] W. Ahmad, R. Olsson and M. Sjöström, "Interpreting Plenoptic Images as Multi-view Sequences for Improved Compression", in 2017 IEEE International Conference on Image Processing (ICIP), Sept 2017, pp. 4557–4561.

[7] L. Li, Z. Li, B. Li, D. Liu and H. Li, "Pseudo-Sequence-Based 2-D Hierarchical Coding Structure for Light-Field Image Compression," in IEEE Journal of Selected Topics in Signal Processing, vol. 11, no. 7, pp. 1107-1119, Oct. 2017.

[8] B. Guo, Y. Han and J. Wen, "Convex Optimization Based Bit Allocation for Light Field Compression Under Weighting and Consistency Constraints", 2018 Data Compression Conference, pp 107-116, March 2018.

[9] P. Astola and I. Tabus, "WaSP: Hierarchical Warping, Merging, and Sparse Prediction for Lightfield Image Compression", EUVIP, Tampere, November 2018. (also on-line at http://www.cs.tut.fi/∼tabus/PID5611181.pdf)

[10] I. Tabus, P. Helin, P. Astola, "Lossy Compression of Lenslet Images from Plenoptic Cameras Combining Sparse Predictive Coding and JPEG 2000", 2017 International Conference on Image Processing, Beijing, September 2017.

[11] P. Astola and I. Tabus, "Light Field Compression of HDCA Images Combining Linear Prediction and JPEG 2000", EUSIPCO, Rome, September 2018.

[12] C. Christopoulos, A. Skodras and T. Ebrahimi, "The JPEG2000 still image coding system: an overview", in IEEE Transactions on Consumer Electronics, vol. 46, no. 4, pp. 1103-1127, Nov 2000.

[13] M. B. de Carvalho, M. P. Pereira, G. Alves, E. A. B. da Silva, C. L. Pagliari, F. Pereira, V. Testoni , "A 4D DCT-Based Lenslet Light Field Codec," 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, pp. 435-439, 2018.

[14] M. Řeřábek and T. Ebrahimi, "New Light Field Image Dataset", 8th International Conference on Quality of Multimedia Experience (QoMEX), Lisbon, Portugal, June 2016.

[15] D. Taubman, "High Performance Scalable Image Compression with EBCOT," in IEEE Transactions on Image Processing, vol. 9, no. 7, pp. 1158-1170, July 2000.