

Repositório ISCTE-IUL

Deposited in *Repositório ISCTE-IUL*:

2019-05-27

Deposited version:

Post-print

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Conti, C., Nunes, P. & Soares, L. D. (2018). Light field image coding with jointly estimated self-similarity bi-prediction. *Signal Processing: Image Communication*. 60, 144-159

Further information on publisher's website:

[10.1016/j.image.2017.10.006](https://doi.org/10.1016/j.image.2017.10.006)

Publisher's copyright statement:

This is the peer reviewed version of the following article: Conti, C., Nunes, P. & Soares, L. D. (2018). Light field image coding with jointly estimated self-similarity bi-prediction. *Signal Processing: Image Communication*. 60, 144-159, which has been published in final form at <https://dx.doi.org/10.1016/j.image.2017.10.006>. This article may be used for non-commercial purposes in accordance with the Publisher's Terms and Conditions for self-archiving.

Use policy

Creative Commons CC BY 4.0

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in the Repository
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Light Field Image Coding with Jointly Estimated Self-Similarity Bi-Prediction

Caroline Conti*, Paulo Nunes, and Luís Ducla Soares

Instituto Universitário de Lisboa (ISCTE-IUL), Instituto de Telecomunicações, Lisbon, Portugal

Abstract

This paper proposes an efficient light field image coding (LFC) solution based on High Efficiency Video Coding (HEVC) and a novel Bi-prediction Self-Similarity (Bi-SS) estimation and compensation approach to efficiently explore the inherent non-local spatial correlation of this type of content, where two predictor blocks are jointly estimated from the same search window by using a locally optimal rate constrained algorithm. Moreover, a theoretical analysis of the proposed Bi-SS prediction is also presented, which shows that other non-local spatial prediction schemes proposed in literature are suboptimal in terms of Rate-Distortion (RD) performance and, for this reason, can be considered as restricted cases of the jointly estimated Bi-SS solution proposed here. These theoretical insights are shown to be consistent with the presented experimental results, and demonstrate that the proposed LFC scheme is able to outperform the benchmark solutions with significant gains with respect to HEVC (with up to 61.1 % of bit savings) and other state-of-the-art LFC solutions in the literature (with up to 16.9 % of bit savings).

Keywords: Light Field, Plenoptic, Holoscopic, Integral Imaging, Image Coding, HEVC, Self-Similarity, Bi-Prediction

1. Introduction

Light Field (LF) imaging based on a single-tier camera equipped with a Microlens Array (MLA) – also known as holoscopic, plenoptic, and integral imaging – derives from the fundamentals of light field/radiance sampling [1], where not only the spatial information about the Three Dimensional (3D) scene is represented, but also the angular viewing direction, i.e., the “whole observable” scene.

Recently, LF imaging has become a prospective imaging approach for providing richer content capture, visualization, and manipulation, being applicable in many different areas of research, e.g., 3D television [2,3], biometric recognition [4], and medical imaging [5]. Among the advantages of employing an LF imaging system is the enabling of new degrees of freedom in terms of content production and manipulation, thus supporting functionalities not straightforwardly available in conventional imaging systems, namely, post-production refocusing, changing depth-of-field, and changing viewing perspective.

However, deploying LF image and video applications with its appealing functionalities will require the use of efficient coding schemes to deal with the large amount of data involved in such types of systems. In this context, novel initiatives on LF image and video coding standardization are also emerging. Notably, the Joint Photographic Experts Group (JPEG) committee has recently started the JPEG Pleno standardization initiative [6] that addresses representation and coding of emerging new imaging modalities. In addition, the Moving Picture Experts Group (MPEG) group has recently started a new work item on coded representations for immersive media (MPEG-I) [7].

1.1. Related Work

Previous Light Field Coding (LFC) schemes available in the literature can be categorized in three main approaches: i) based on transform coding [8,9], ii) based on view extraction [10–17], and iii) based on non-local spatial prediction [18–22]. Generally, all coding schemes try to take advantage of the particular planar intensity distribution of the LF image. Notably, as a result of the used optical system, the raw LF image corresponds to a 2D array of Micro-Images (MIs), where both light intensity and direction information are recorded.

* Corresponding author. Tel.: +351 218418164.

E-mail address: caroline.conti@lx.it.pt.

1.1.1. LFC Based on Transform Coding

Most of the early proposed LFC schemes adopted the transform-based approach by using a Discrete Cosine Transform (DCT) or a Discrete Wavelet Transform (DWT). In [8], a 3D DCT was applied to a stack of MIs to exploit the existing correlation between adjacent MIs, as well as the redundancy within each MI. In [9], the LF content was separated into various viewpoint images by extracting one pixel with the same position from each MI and a 3D DWT was then applied to a stack of them. Afterwards, the lower frequency coefficients were transformed using a Two Dimensional (2D) DWT followed by arithmetic encoding, while the remaining high frequency coefficients were simply quantized and arithmetic encoded. Recently, it has been concluded in the literature that HEVC Main Still Image Profile [23] presents significant compression performance improvements in comparison to previous transform-based still image coding technologies [24,25] – such as JPEG (DCT-based) and JPEG 2000 (DWT-based) standards. Moreover, similar conclusions have been also reached for LF image coding, in [26,27], where HEVC presented significantly better performance than JPEG and JPEG 2000.

1.1.2. LFC Based on View Extraction

Alternatively, other schemes proposed to extract a set of views from the LF data for coding. In [10–15], MIs or Viewpoint Images (VIs) were extracted from the LF content in order to represent the LF data as a set of views and to use inter-view prediction for achieving compression. In [10,11], these views were then encoded as multiview content using Multiview Video Coding (MVC) [28]. Differently, in [12–15], the views were encoded as a Pseudo Video Sequence (PVS) using a 2D video coding standard, such as H.264/AVC [28], in [12], or High Efficiency Video Coding (HEVC) [23] in [13–15]. Although conceptually different (in terms of coding architecture), both multiview- and PVS-based coding approaches have the same basic purpose of proposing an efficient prediction configuration for better exploiting the correlations between the views. For this, different scanning patterns for ordering the views, as well as different prediction structures have been proposed. In [29], it was shown that the PVS-based coding solution outperformed a transform-based solution (similar to the LF coding solution proposed in [9]) with significant gains, notably, at lower bit rates. In addition, an alternative to the multiview representation based on these low resolution MIs/VIs was proposed in [16,17] using super-resolved rendered views. In this case, the scalable coding architecture proposed in [16] was used, which supported backward compatibility to legacy 2D and 3D multiview displays in the lower layer while the highest layer supports the entire LF content. In [17], the associated disparity information was also encoded and transmitted in the lower layers along with the set of views.

1.1.3. LFC Based on Non-Local Spatial Prediction

Schemes based on the non-local spatial predictive approach rely on a non-local prediction techniques that exploit the existing redundancy between MIs in a (spatial) neighborhood to encode the entire raw LF image, being usually integrated (but not necessarily so) on a standard 2D image codec. The idea of exploiting non-local spatial redundancy has been firstly proposed for 2D image and video compression in order to further enhance the performance of H.264/AVC intra prediction [30]. Notably, the intra macroblock compensation technique was proposed in [30] to extend the usage of motion compensated prediction for intra-coded frames.

In the context of LF content coding, previous work of the authors [18,19] showed that further improvements are still possible for LF images with respect to the state-of-the-art for 2D image coding using the HEVC Main Still Picture profile [24,25,31] by using the concept of Self-Similarity (SS) compensated prediction. Similar to the intra macroblock compensation [30], the SS estimation process uses a block-based matching over the previously coded and reconstructed area of the current picture (referred to as SS reference [18]), to find the ‘best’ predictor for the current block. As a result, the chosen block becomes the candidate predictor and the relative position between the two blocks is signaled by an SS vector. In [19], a novel vector prediction scheme was also proposed to take advantage of the particular characteristics of the SS prediction data and thus increase coding efficiency. Subsequently, in [20], a scheme to extend the SS prediction concept by using HEVC inter B frame bi-prediction was proposed for LF image coding. However, in this case, to guarantee that the two prediction signals came from two different MIs, the search area was proposed to be separated into two non-overlapping parts [20] to perform the prediction estimation as in conventional HEVC bi-prediction. Although not targeting LF image coding, another prediction scheme similar to the SS compensated prediction, known as Intra Block Copy (IntraBC) [32], has been recently proposed in the literature in the context of Screen Content Coding (SCC) [32]. In this case [32], the prediction estimation is performed considering only integer pixel accuracy and the search window is expanded to the entire CB row or column, or to the entire previously coded area of the picture by using a hash-based search [32].

Furthermore, instead of using a block-based matching approach, an alternative prediction scheme based on locally linear embedding was proposed in [21], where a set of nearest neighbor patches were estimated from the same search area and linearly combined to predict the current block. More recently, in [22], a prediction scheme based on Gaussian Process

Regression (GPR) was also proposed for LF image coding. In this case, two separate search areas are adopted for finding a set of nearest neighbor patches and the prediction is modeled as a non-linear (Gaussian) process for estimating the predictor block.

1.2. Motivations and Contributions

Motivated by the authors' results in [18,19,21], this paper proposes an improved LF image coding solution based on HEVC and a novel Bi-predicted Self-Similarity (Bi-SS) estimation approach using the generic concept of superimposed prediction [33], which allows bi-prediction using samples from the same search area. Therefore, instead of dividing the search area into two non-overlapping parts to derive each predictor block from different MIs (as in [20]), these predictor blocks can be located in the same MI and in overlapped pixel positions. Moreover, instead of simply combining the two (independent) best uni-predicted candidate predictor blocks for bi-prediction (as in [20]), the locally optimal rate-constrained algorithm [34] is used for jointly estimating these two predictor blocks.

In addition to this, a theoretical analysis of the proposed Bi-SS prediction is also presented, which shows that other non-local spatial prediction schemes – such as the IntraBC [32], the preceding uni-predicted SS solution in [19], and the bi-prediction proposed in [20] – are suboptimal in terms of Rate-Distortion (RD) performance and, for this reason, can be considered as restricted cases of the jointly estimated Bi-SS solution proposed here. Furthermore, studies about the influence of the MI cross-correlation and the weighting factors used for bi-prediction on the RD efficiency of the Bi-SS prediction are also presented to experimentally validate the theoretical assumptions used for LF image coding.

Experimental results show that the proposed LFC solution using the jointly estimated Bi-SS prediction – from now on referred to as LFC Bi-SS solution – is able to outperform with significant coding gains various state-of-the-art LFC solutions based on different non-local spatial predictions.

1.3. Paper Outline

The remainder of the paper is organized as follows: Section 2 describes the proposed LFC Bi-SS solution architecture; Section 3 proposes the jointly estimated Bi-SS prediction and presents the theoretical and experimental analyses of its prediction efficiency improvement for LF image coding; Section 4 presents the test conditions and experimental results; and, finally, Section 5 concludes the paper.

2. LFC Bi-SS Solution Architecture

The proposed LFC Bi-SS solution is not tuned for any particular optical acquisition setup since it does not require any explicit knowledge about it (e.g., microlens size, focal length, and distance of the microlenses to the image sensor). Notice that, although these parameters may be provided by camera makers, many of them are highly dependent on the manufacturing process, being different even from camera to camera of the same model (e.g., each microlens may vary slightly in shape, size, and relative layout position). This means that the LF content from each specific LF camera is also different, and compression tools that use this kind of information need to be robust to these variations. For this reason, using compression tools that are less dependent on a very precise calibration pre-process may be advantageous for supporting LF visualization without increasing the processing complexity.

In this sense, this LFC Bi-SS solution may be mainly advantageous for applications in which the LF content is consumed by the end user in a format similar to the captured format, as for example, in the case in which the captured LF content is visualized in an LF display that also makes use of an MLA in its optical system, or is consumed by using a proprietary LF rendering algorithm that makes use of the same (raw) 2D format. In this scenario, the captured LF image can be encoded with the proposed LFC Bi-SS solution and may be then transmitted to a receiver over heterogeneous networks. Alternatively, this LFC solution may be also advantageous for improving the storage compression efficiency. The proposed solution is also advantageous in terms of the encoder/decoder computational complexity and necessary memory, which are no larger than that of HEVC inter B frame coding. Detailed information about HEVC computational complexity can be found in [19].

Fig. 1 presents the architecture of the proposed LFC Bi-SS solution, which is based on HEVC and comprises both additional and modified modules to efficiently handle LF content. Basically, the proposed codec introduces an additional type of prediction – the Bi-SS – and the encoder will choose the best, among Bi-SS and HEVC intra prediction, based on a conventional Rate-Distortion Optimization (RDO) process.

More specifically, enhancing the HEVC coding architecture with Bi-SS compensated prediction requires adaptations at the following stages of the coding process (as explained in the following): i) SS estimation; ii) prediction modes and block partitioning; iii) Bi-SS compensation; iv) Bi-SS vector prediction; and v) reference picture management.

2.1. Bi-SS Estimation

The Bi-SS estimation (depicted in Fig. 1) is used to exploit the cross-correlation existing in an MI neighborhood (see Fig. 2a) by estimating the prediction block with the highest similarity (according to appropriate criteria) to the current block in the previously coded and reconstructed area of the current picture itself (the SS reference, as seen in Fig. 2b). Hence, the relative position between the current and the ‘best’ candidate block is signaled by an SS vector, \mathbf{v}_0 , (see Fig. 2b). Similarly to the conventional HEVC inter P frame prediction, the best SS vector, \mathbf{v}_0^{best} , for the SS prediction can be found by minimizing the Lagrangian cost function in (1) [35],

$$J_{uni-ss} = \min_{\mathbf{v}_0} \|I(\mathbf{x}) - \tilde{I}(\mathbf{x} - \mathbf{v}_0)\|_1 + \lambda R(\mathbf{v}_0) \quad (1)$$

where $I(\mathbf{x})$ is a matrix variable representing the current block at position $\mathbf{x} = (x, y)$ in the LF image; $\tilde{I}(\mathbf{x} - \mathbf{v}_0)$ represents a candidate block in the SS reference, \tilde{I} , with $\mathbf{x} - \mathbf{v}_0 \in \mathbf{W}$ (see Fig. 2b); $R(\mathbf{v}_0)$ corresponds to an estimated number of bits for encoding the SS vector \mathbf{v}_0 (i.e., the estimated number of bits necessary to encode the motion vector difference between \mathbf{v}_0 and its predictor, selected as shown in Section 2.4); and λ is the Lagrangian multiplier. In addition, to keep the complexity low, the l_1 -norm (or Sum of Absolute Differences (SAD)), $\|\cdot\|_1$, is used and a limited causal search window \mathbf{W} is adopted. However, it is worth noting that the search area shall be larger than the MI size to be able to exploit the inherent MI cross-correlations. Finally, the SS predictor block, $\tilde{I}(\mathbf{x})$, is derived as $\tilde{I}(\mathbf{x} - \mathbf{v}_0^{best})$. As done in HEVC reference software version 14.0 [36], when SAD is used as the distortion measure, λ is given by $\sqrt{\lambda_{intra}}$, where λ_{intra} is the Lagrangian multiplier computed for prediction mode selection in intra-coded frames.

Notice that the SS estimation process in (1) only considers a single compensated signal for prediction of the current block, as previously proposed by the authors in [19], and for this reason will be hereinafter referred to as Uni-predicted Self-Similarity (Uni-SS) estimation. Moreover, further improvements to this estimation process will be proposed in Section 3 for allowing jointly estimated Bi-SS prediction.

2.2. Prediction Modes and Block Partitioning

The Bi-SS prediction is evaluated for all Coding Block (CB) sizes (i.e., from 64×64 down to 8×8) in the conventional RDO process to choose the best prediction mode. For this, the proposed Bi-SS method defines the following two additional prediction modes:

- **Bi-SS mode** – In this case, Bi-SS estimation (see Fig. 1) is used to find a prediction for encoding the current block. As in HEVC inter coding, the eight partition patterns, i.e., $M \times M$, $M \times (M/2)$, $(M/2) \times M$, $(M/2) \times (M/2)$, $M \times (M/4)$, $M \times (3M/4)$, $(M/4) \times M$, and $(3M/4) \times M$ [37], are allowed to define a flexible way to partition the CB for the Bi-SS estimation process in Fig. 2b.

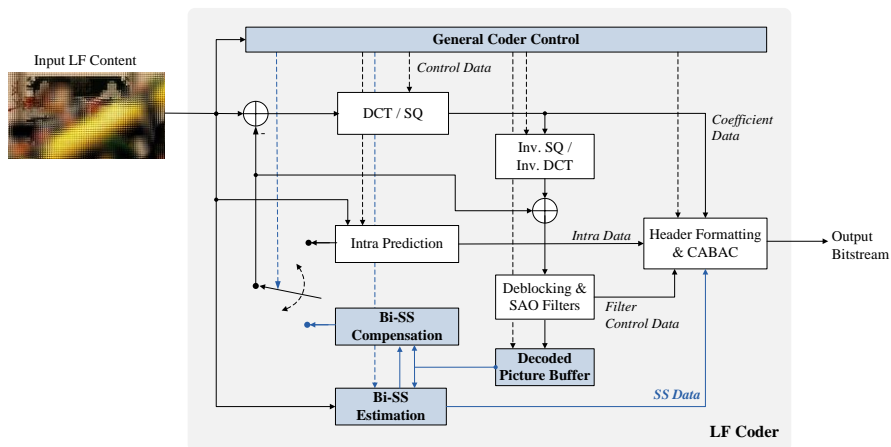


Fig. 1 Coding architecture of the proposed LFC Bi-SS solution based on HEVC (the novel and modified blocks are highlighted in blue). The HEVC blocks for motion estimation and compensation are here omitted since temporal prediction is not used in the proposed light field image coding solution.

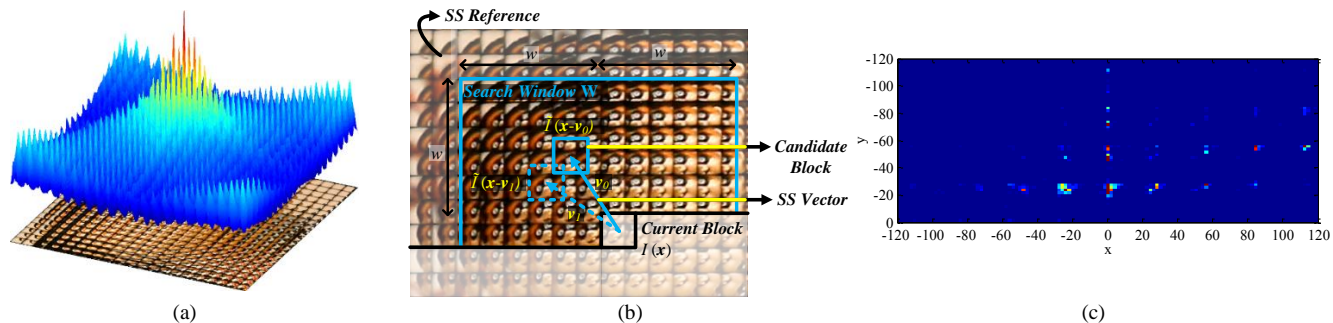


Fig. 2 SS prediction: (a) inherent MI cross-correlation in a light field image neighborhood; (b) Bi-SS estimation process (example of a second candidate block and SS vector for bi-prediction is shown in dashed blue line); and (c) Heat map showing the SS vectors distribution when coding an LF image

- **SS-skip mode** – The SS-skip is employed only for the $M \times M$ partition pattern, and, in this case, the SS vector is directly derived from the Bi-SS vectors prediction presented Section 2.4, and no further information is transmitted.

2.3. Bi-SS Compensation

In this process (shown in Fig. 1), the inverse quantized and inverse transformed causal residual block is added to the prediction block to obtain the reconstructed block. The SS reference is then updated (for each CB) by including this reconstructed block as it will be available at the decoder side.

2.4. Bi-SS Vectors Prediction

In conventional video coding solutions, since neighboring motion vectors are likely to be correlated, they are usually predictively coded based on motion vectors of neighboring blocks. Regarding HEVC [23], this predictive coding of motion vectors was improved, relatively to previous video coding standards, by introducing a tool called Advanced Motion Vector Prediction (AMVP). Furthermore, an HEVC technique called merge mode is used to derive all motion data of a block (i.e., motion vectors and indices of the used reference pictures) from the neighboring blocks, replacing the direct and skip modes of the previous H.264/AVC standard. In these methods, a vector candidate (or merge candidate) list is built by selecting vectors from CBs in the spatial and temporal (co-located) neighborhood [37]. From these spatio-temporal candidates, the encoder selects the best predictor vector in an RDO sense, and transmits only the index of the chosen candidate in this list.

In addition to this, a set of new SS candidate vectors proposed in [19], referred to as MI-based vector prediction (MIVP) candidate vectors, is also included into AMVP and merge candidate lists to further improve the RD performance. As explained in [19], up to three MI-based candidate vectors (i.e., left, above, and above-left) are computed to force the candidate vectors to be distributed according to the structure of MIs. A detailed description of the MIVP candidate vectors selection is given in [19].

Regarding the proposed Bi-SS mode, if bi-prediction is used, two predictor vectors are derived from the AMVP method (one for each estimated SS vector) and the difference between the two SS estimated vectors and the corresponding predictor vectors are transmitted along with the indices of the chosen candidates in the list. In this case, the AMVP candidate list is constructed with the following candidates (for intra-coded frames):

- **Spatial AMVP vector candidates** – Up to two spatial vector candidates are derived from a set of five spatial neighboring CBs that were previously coded with Bi-SS mode. The position of these neighboring CBs are defined in HEVC standard [23].
- **MI-based vector candidate** – When less than two spatial vector candidates are available, one MI-based vector candidate is derived from the set of left, above, and above-left defined in MIVP.
- **Zero vector candidates** – Afterwards, zero vectors are added to fill, when necessary, the AMVP candidate list with up to two final candidates (as in HEVC [23]).

In the case of the SS-skip mode, bi-predicted merge candidates may be derived from the following merge candidate list:

- **Spatial merge vector candidates** – Up to four spatial candidates from the set of five [23] neighboring CBs that were coded with Bi-SS mode are included.
- **MI-based vector candidates** – The maximum size of the merge candidate list is signaled in the slice header syntax (being equal to five as defined by default in HEVC standard [23]). After selecting the spatial candidates, up to three MIVP merge candidates are included into the merge candidate list until the maximum number of candidates is reached.

- **Additional merge candidates** – Furthermore, if the merge candidate list is still not fully populated, bi-predicted candidates can also be derived by combining two existing candidates from different reference picture lists. When the list is still not full, zero motion candidates are included to complete the list.

An analysis of the influence of this MIVP scheme in the RD performance achieved by the Bi-SS prediction is presented in Section 3.5.

2.5. Reference Picture Management

To allow the Bi-SS estimation and Bi-SS compensation in intra-coded frames of HEVC, the reference lists construction and signaling need to be altered so as to include the SS reference. This process is similar to the temporal lists construction on HEVC inter-coded frames, and is managed by the general coder control block in Fig. 1 by using the concept of Reference Picture Set (RPS), which is signaled for each slice [37]. For this, the SS reference is made available at the decoded picture buffer (see Fig. 1) and marked to be used as a reference.

3. Proposed Jointly Estimated Bi-SS Compensated Prediction

To further improve the performance of the proposed Uni-SS LFC solution [19], a novel jointly estimated Bi-SS estimation and compensation scheme, which is based on the generic concept of superimposed prediction [33], is here proposed to replace the aforementioned Uni-SS estimation and compensation processes.

To motivate the adoption of this Bi-SS prediction in the LFC solution presented in Section 2, a theoretical analysis is firstly presented, which shows that other non-local spatial prediction schemes – such as the IntraBC [32], the preceding Uni-SS prediction presented in Section 2.1, and the solution for bi-prediction proposed in [20] – can be considered as restricted cases of the Bi-SS solution proposed here. Then, the Bi-SS candidate predictor estimation is proposed in Section 3.2 and the theoretical assumptions for its improved RD efficiency are experimentally analyzed in Sections 3.3 and 3.4, respectively, in terms the MI cross-correlation in LF images and in terms of the weighting factor used in the bi-prediction. Finally, the influence of the MIVP in the RD performance achieved for Bi-SS prediction is also analyzed in Section 3.5.

3.1. Theoretical Bi-SS Performance Analysis

The RD performance improvement due to the adoption of the jointly estimated Bi-SS prediction (presented in more detail in the following section) is based on three main hypotheses, which will be analyzed in this section:

- 1) With a large enough search window, \mathbf{W} , (see Fig. 2b), it is possible to find two predictor blocks that properly represent the current block, $I(\mathbf{x})$, i.e., with low residual signal.
- 2) By combining two good predictor blocks, it is possible to further minimize the residual signal of the SS compensated prediction, compared to only using the uni-predicted SS candidate (as in the Uni-SS prediction [19] and in the IntraBC scheme [32]).
- 3) Jointly estimating the predictor blocks leads to better RD performance than deriving them independently (as in reference software for HEVC inter B frame coding [36] and in the bi-prediction solution proposed in [20]).

For this analysis, the performance of the proposed Bi-SS prediction is here modeled by the uncertainty [35] (or inaccuracy [33,38]) in the SS compensated prediction signal.

Regarding the first abovementioned hypothesis, it is valid due to the following facts:

- Given the small baseline between adjacent microlenses in the acquisition process, a significant cross-correlation exists between neighboring MIs, as shown by the autocorrelation function in Fig. 2a. It can be seen (Fig. 2a) that the autocorrelation function presents a regular structure of spikes and the constant distance between these regular spikes corresponds to the MI spacing in the array [19]. Since these highly-correlated samples are distributed along the MIs, it is likely that similarly good predictor blocks will be also distributed accordingly.
- It was shown in [19] that, when using the SS compensated prediction scheme for exploiting the inherent MI cross-correlation, the distribution of the chosen SS vectors is also related to the size and arrangement of the MIs in the MLA. This can be illustrated by the heat map in Fig. 2c, where brighter areas correspond to more frequent SS vector amplitudes. Hence, since these most frequent best uni-predicted SS vectors are distributed in all directions according to the MI arrangement, it is possible to consider that the second-best SS vector, which can also represent the current block properly, is likely to be found according to this distribution in a different direction.

Regarding the second and third hypotheses, the residual signal for the Bi-SS compensated prediction is given by (2),

$$e(\mathbf{x}) = I(\mathbf{x}) - \sum_{p=0}^1 h_p(\mathbf{x}) \cdot \tilde{I}(\mathbf{x} - \mathbf{v}_p) \quad (2)$$

where $\mathbf{x} - \mathbf{v}_p \in \mathbf{W}$, and h_p is the weight for each of these predictor blocks. For instance, for the bi-predicted SS candidate predictor proposed in Section 3.2, $h_p = 1/2, \forall p \in \{0,1\}$. However, as discussed in [35], the residual signal given by (2) can be actually generalized as in (3) for N predictor blocks.

$$e(\mathbf{x}) = I(\mathbf{x}) - \sum_{p=0}^{N-1} h_p(\mathbf{x}) \cdot \tilde{I}(\mathbf{x} - \mathbf{v}_p) \quad (3)$$

The general case represented by (3) can also incorporate other types of candidate predictors reflecting the very flexible set of inter coding tools of HEVC [23]. In this case, $\mathbf{h} = (h_0, h_1, \dots, h_{N-1})$ corresponds to a weight vector that is able to, for example, incorporate [35]: i) the filtering used to generate the quarter-pixel interpolated signal in the SS estimation; and ii) the deblocking filter that can be applied in the SS reference. Each $\tilde{I}_p(\mathbf{x}) = \tilde{I}(\mathbf{x} - \mathbf{v}_p)$ term can be interpreted as each of the multiple compensated signals available for prediction of the current block. Hence, the uncertainty in a given Bi-SS compensated prediction can be modeled, as in [35], by an a posteriori probability density function, $h_p(\mathbf{x})$, conditioned on the encoded data. Therefore, since the expected value (the second term on the right-hand side of (3)) is the estimator that minimizes the mean-square error in the prediction of a random variable [35], it is possible to say that the residual signal in (3) can be minimized and, consequently, the accuracy of the prediction can be improved by using a larger set of multiple compensated signals and an optimized weight vector [35].

In addition, another possibility is to analyze the performance of the Bi-SS compensated prediction by modeling the inaccuracy of each used displacement vector \mathbf{v}_p , as in [33,38]. For this, Fig. 2b shows that although the pixel correlation in the raw LF image is not as smooth as in conventional 2D images, each MI itself has some degree of inter-pixel redundancy as in common 2D images (see Fig. 2b). Thus, it is possible to consider that samples inside each MI follow the same correlation model as samples in a 2D image (i.e., an isotropic exponentially decaying autocorrelation function). This assumption is reasonable at least for blocks $I(\mathbf{x})$ smaller than the MI resolution, and has been also adopted in [20,39] for LF images. With this assumption, the accuracy of the SS compensation can be measured by the displacement error variance [38], and the same signal model used in [33,38] can also be considered for the SS compensated prediction signal. In this case [33,38], \mathbf{h} denotes a row vector of impulse responses $(h_0, h_1, \dots, h_{N-1})$ of a 2D prediction filter [38], and the residual signal is given by (4),

$$e(\mathbf{x}) = I(\mathbf{x}) - \mathbf{h}(\mathbf{x}) * \tilde{\mathbf{I}}(\mathbf{x}) \quad (4)$$

where the second term on the right-hand side of the equation denotes a 2D convolution of the prediction filter \mathbf{h} with a column vector of N multiple compensated signals $\tilde{\mathbf{I}} = (\tilde{I}_0, \tilde{I}_1, \dots, \tilde{I}_{N-1})^T$. In this model, both I and each component of $\tilde{\mathbf{I}}$ are assumed to be wide sense stationary random processes with an additive Gaussian noise signal. Hence, these noisy signals may comprise all signal components of the SS compensated prediction that cannot be described by the translational displacement model [38].

Based on the abovementioned approximations, the conclusions from [33,38] also hold for validating the second and third hypotheses. Notably, with high rate assumptions:

- Concerning the second hypothesis, the optimal filter \mathbf{h} (i.e., that minimizes the mean square error) can be interpreted as a low pass filter that removes high frequency components from $\tilde{\mathbf{I}}$ that are too noisy or that change too rapidly [38]. From the theoretical analysis in [38], it was concluded that increasing the number of equally good predictor blocks always led to bitrate savings compared to a more limited set of predictor blocks, even if the simple average filter is used, instead of considering an optimal filter \mathbf{h} in (4). Therefore, this suggests that increasing from one predictor block (in the previously proposed Uni-SS prediction [19]) to two predictor blocks (in the Bi-SS prediction proposed here) minimizes the residual signal in the SS compensated prediction.
- Concerning the third hypothesis, an extended analysis was performed in [33] for the case where the multiple compensated samples of $\tilde{\mathbf{I}}$ are jointly estimated. In this case, the displacement error of all components of $\tilde{\mathbf{I}}$ are assumed to be correlated, instead of being independent as assumed in [38]. Moreover, this analysis considered the simple average filtering case as for the proposed Bi-SS. Therefore, it was shown that a combination of two jointly estimated predictor blocks is more efficient than two independent predictor blocks [33]. Furthermore, it was concluded that, for jointly estimated predictors, the major portion of the gain is already achievable by only two predictor blocks [33]. This suggests that further RD gains can be achieved by jointly estimating the two predictor blocks for bi-prediction compared to

simply combining two best uni-predicted candidates (as in the HEVC reference software inter B frame coding [23] and in bi-predicted solution proposed in [20]).

Therefore, without loss of generality, the IntraBC scheme in [32], and the previously proposed LFC solutions in [19,20] can be seen as restricted cases of the Bi-SS solution being proposed here. In these cases [19,20,32], restrictions are imposed in the number of predictor blocks, the number of allowed partition patterns and sizes, and in the bi-prediction estimation process for each predictor block that is independently employed in different areas of the SS reference.

As it will be seen in Section 4, the theoretical insights from this section for the proposed Bi-SS solution are supported by the experimental results for LF images. Moreover, to extend the conclusions from this analysis, Section 3.4 also analyzes the RD performance for LF image coding when using different sets of weighting coefficients for Bi-SS prediction.

3.2. Bi-SS Candidate Predictors Estimation

Motivated by the theoretical analysis presented above, the proposed Bi-SS prediction is here presented, which is based on the generic concept of superimposed prediction [33].

More specifically, there is only a single reference picture available in the Bi-SS compensated prediction, i.e., the SS reference [18], and only two possible candidate predictors (instead of the three candidates of HEVC [23] that are used in [20]) are derived to predict the current block, namely: i) the Uni-SS candidate, and ii) the Bi-SS candidate.

The Uni-SS candidate predictor corresponds to the previously proposed SS prediction [19] (see Section 2.1), in which the predictor block is found by minimizing the Lagrangian cost function in (1).

The proposed Bi-SS candidate predictor differs from the HEVC reference software inter B frame bi-prediction, as well as from the bi-predicted solution in [20], for two main reasons:

- 1) The two predictor blocks in the Bi-SS solution are derived from the same reference picture (the SS reference) and are estimated in the same search window, \mathbf{W} (see Fig. 2b). Consequently, they can be located in the same MI (different to [20]) and in overlapped pixel positions (as illustrated by the dashed blue lines in Fig. 2b).
- 2) To further improve the prediction efficiency, these two predictor blocks are jointly estimated in the complete search window (as explained below), instead of combining two best uni-predicted candidates (as in [20]).

For jointly estimating the two predictor blocks, the locally optimal rate-constrained algorithm proposed in [34] (see Fig. 3) is used. This algorithm avoids searching through all possible combinations of two candidate predictor blocks $\tilde{I}(\mathbf{x} - \mathbf{v}_0)$ and $\tilde{I}(\mathbf{x} - \mathbf{v}_1)$ inside \mathbf{W} . For this, in each algorithm iteration, k , an optimal SS candidate vector $\mathbf{v}_q^{(k+1)}$ (with index $q \in \{0,1\}$) is found by minimizing the Lagrangian cost function conditioned to the optimal SS candidate vector found in the previous iteration $\mathbf{v}_p^{(k)}$ (with $p \in \{0,1\}$). Therefore, the algorithm is focused on finding an optimized vector \mathbf{v}_1 conditioned to a known vector \mathbf{v}_0 in even iterations, and vice versa in odd iterations. For instance, in the first iteration, $k = 0$, $p = 0$, $q = 1$, and the optimal $\mathbf{v}_1^{(1)}$ is found by fixing $\mathbf{v}_0^{(0)} = \mathbf{v}_0^{best}$ (the best uni-predicted SS candidate vector). Similarly, in the second iteration, $k = 1$, $p = 1$, $q = 0$, and the optimal $\mathbf{v}_0^{(2)}$ is found by fixing $\mathbf{v}_1^{(1)}$ (which was found in the previous iteration). Similarly to (1), the Lagrangian cost function show in Fig. 3 is used to find the optimal SS vector in each iteration, where λ is computed as $\sqrt{\lambda_{Intra}}$ [36] and $R(\mathbf{v}_q^{(k+1)}) + R(\mathbf{v}_p^{(k)})$ corresponds to the estimated number of bits for encoding the SS vectors \mathbf{v}_0 and \mathbf{v}_1 given in each

$$\begin{aligned} & \mathbf{Initialization:} \quad k = 0, \quad h_0 = h_1 = 1/2, \quad \mathbf{v}_0^{(k)} = \mathbf{v}_0^{best}, \quad J_{Bi-SS} = J_{MAX} \\ & \mathbf{do} \\ & \quad J_{Bi-SS}^{best} = J \\ & \quad p = (k) \bmod(2), \quad q = 1 - p \\ & \quad J_{Bi-SS} = \|\|I(\mathbf{x}) - [h_q \cdot \tilde{I}(\mathbf{x} - \mathbf{v}_q^{(k+1)}) + h_p \cdot \tilde{I}(\mathbf{x} - \mathbf{v}_p^{(k)})]\|\|_1 + \lambda[R(\mathbf{v}_q^{(k+1)}) + R(\mathbf{v}_p^{(k)})] \\ & \quad \mathbf{v}_q^{(k+1)} = \arg \min_{\mathbf{v}_q^{(k+1)}} J_{Bi-SS} \\ & \quad k = k + 1 \\ & \mathbf{while} \quad k \neq K \text{ or } J_{Bi-SS} < J_{Bi-SS}^{best} \end{aligned}$$

Fig. 3 Algorithm for jointly estimating the two predictor blocks $\tilde{I}(\mathbf{x} - \mathbf{v}_0)$ and $\tilde{I}(\mathbf{x} - \mathbf{v}_1)$ for the proposed Bi-SS candidate predictor. The index q defines which of the two vectors (\mathbf{v}_0 or \mathbf{v}_1) will be optimized in a particular iteration k , while the index p defines the vector that will be kept fixed.

Table 1 Influence of MI cross-correlation in mode selection statistics and RD performance for high bitrate

Image	Prediction Modes Statistics			SS Prediction Statistics		CBS Statistics				LFC Bi-SS vs. LFC Uni-SS	
	Intra	SS	SS-skip	Uni-Prediction	Bi-Prediction	64×64	32×32	16×16	8×8	BD-PSNR	BD-BR
<i>Plane and Toy (frame 123)</i>	57.2 %	42.3 %	0.5 %	60.4 %	39.6 %	0.0 %	0.9 %	9.3 %	89.8 %	0.3 dB	-4.5 %
<i>Plane and Toy (frame 23)</i>	67.8 %	30.9 %	1.3 %	54.2 %	45.8 %	0.0 %	3.0 %	11.4 %	85.6 %	0.4 dB	-6.1 %
<i>Demichelis Spark</i>	53.3 %	39.9 %	6.9 %	48.8 %	51.2%	0.1%	3.8 %	18.2 %	77.9 %	0.5 dB	-14.9 %
<i>Laura</i>	36.3 %	62.9 %	0.7 %	44.9 %	55.1 %	0.0 %	1.1 %	20.7 %	78.3 %	0.5 dB	-8.6 %
<i>Jeff</i>	39.6 %	58.6 %	1.7 %	44.5 %	55.5 %	0.0 %	4.5 %	27.6 %	67.9 %	0.5 dB	-10.4 %
<i>Seagull</i>	27.3 %	66.3 %	6.4 %	47.3 %	52.7 %	0.0 %	3.2 %	29.8 %	67.0 %	0.7 dB	-14.4 %
<i>Zhengyun1</i>	58.3 %	40.7 %	1.0 %	47.5 %	52.5 %	0.0 %	3.3 %	20.8 %	76.0 %	0.4 dB	-9.9 %
<i>Flowers</i>	72.7 %	26.5 %	0.8 %	45.3 %	54.7 %	0.8 %	15.3 %	15.8 %	68.1 %	0.1 dB	-3.1 %
<i>Vespa</i>	29.9 %	65.6 %	4.5 %	26.7 %	73.3 %	0.4 %	7.9 %	27.7 %	64.0 %	0.5 dB	-18.9 %
<i>Ankylosaurus_&_Diplodocus_1</i>	14.7 %	84.6 %	0.7 %	10.8 %	89.2 %	0.2 %	3.5 %	30.0 %	66.4 %	0.7 dB	-44.1 %
<i>Fountain_&_Vincent_2</i>	40.9 %	57.4 %	1.7 %	16.3 %	83.7 %	0.6 %	12.3 %	23.1 %	64.0 %	0.7 dB	-16.6 %
<i>Color_Chart_1</i>	20.3 %	75.4 %	4.3 %	16.2 %	83.8 %	0.1 %	2.5 %	28.4 %	69.0 %	0.4 dB	-22.5 %
<i>ISO_Chart_12</i>	23.5 %	73.7 %	2.8 %	21.3 %	78.7 %	0.1 %	5.1 %	28.3 %	66.5 %	0.6 dB	-22.7 %

Table 2 Influence of MI cross-correlation in mode selection statistics and RD performance for low bitrate

Image	Prediction Modes Statistics			SS Prediction Statistics		CBS Statistics				LFC Bi-SS vs. LFC Uni-SS	
	Intra	SS	SS-skip	Uni-Prediction	Bi-Prediction	64×64	32×32	16×16	8×8	BD-PSNR	BD-BR
<i>Plane and Toy (frame 123)</i>	30.5 %	46.2 %	23.3 %	87.1 %	12.9 %	0.1 %	3.9 %	22.3 %	73.7 %	0.3 dB	-4.9 %
<i>Plane and Toy (frame 23)</i>	43.7 %	30.5 %	25.8 %	75.0 %	25.0 %	0.6 %	14.9 %	32.4 %	52.1 %	0.4 dB	-6.5 %
<i>Demichelis Spark</i>	18.2 %	46.1 %	35.7 %	61.1 %	38.9 %	8.5 %	40.0 %	33.0 %	18.5 %	0.7 dB	-17.3 %
<i>Laura</i>	22.5 %	45.1 %	32.4 %	74.1 %	25.9 %	0.1 %	11.0 %	36.0 %	52.9 %	0.5 dB	-9.5 %
<i>Jeff</i>	21.0 %	38.7 %	40.3 %	76.8 %	23.2 %	1.1 %	18.5 %	40.5 %	39.9 %	0.4 dB	-8.8 %
<i>Seagull</i>	15.9 %	36.5 %	47.6 %	81.0 %	19.0 %	0.7 %	20.5 %	39.6 %	39.2 %	0.6 dB	-13.8 %
<i>Zhengyun1</i>	22.0 %	34.8 %	43.3 %	78.4 %	21.6 %	0.9 %	19.7 %	42.7 %	36.6 %	0.4 dB	-9.1 %
<i>Flowers</i>	75.7 %	18.7 %	5.5 %	62.5 %	37.5 %	12.3 %	43.9 %	17.0 %	26.8 %	0.1 dB	-3.4 %
<i>Vespa</i>	47.0 %	29.4 %	23.6 %	41.1 %	58.9 %	31.6 %	40.8 %	15.1 %	12.5 %	0.5 dB	-20.6 %
<i>Ankylosaurus_&_Diplodocus_1</i>	45.7 %	16.6 %	37.8 %	59.8 %	40.2 %	51.5 %	30.5 %	10.8 %	7.1 %	0.5 dB	-33.6 %
<i>Fountain_&_Vincent_2</i>	40.7 %	40.4 %	18.9 %	40.7 %	59.3 %	17.8 %	39.9 %	18.5 %	23.7 %	0.6 dB	-16.4 %
<i>Color_Chart_1</i>	26.4 %	33.6 %	40.0 %	26.1 %	73.9 %	40.7 %	33.9 %	13.6 %	11.8 %	0.3 dB	-15.0 %
<i>ISO_Chart_12</i>	46.9 %	31.5 %	21.6 %	40.2 %	59.8 %	25.2 %	45.3 %	21.2 %	8.3 %	0.6 dB	-20.8 %

iteration, i.e., the estimated number of bits necessary to encode the motion vector difference between the SS vectors and their predictor vectors and for signaling the vector predictor using AMVP.

The maximum number of iterations, K , defines a tradeoff between complexity and RD performance and can be adjusted according to the system constraints. In this work, $K = 2$ and, consequently, the corresponding complexity is similar to that of HEVC inter B-frame with one active reference in each reference picture list.

Finally, the best prediction between Uni-SS and Bi-SS candidates is also chosen in terms of conventional RDO [35] by comparing the associated Lagrangian costs J_{Uni-SS}^{best} and J_{Bi-SS}^{best} , respectively found in (1) and Fig. 3.

3.3. Bi-SS Prediction Analysis for Different MI Cross-Correlation

This section aims at analyzing how close the theoretical conclusions drawn in [33,38] and the approximations considered in Section 3.1 are to the experimental results when using the proposed Bi-SS prediction for LF image coding. More specifically, this analysis focuses on the influence of the MI cross-correlation, inherently present in LF images, in the performance of the Bi-SS prediction proposed in Section 3.2. For this, Tables 1 and 2 summarize some statistics of relevant results when encoding different LF test images (see Section 4), such as: percentages of prediction mode usage, SS bi-prediction usage, and coding block size (CBS) usage. These statistics are shown for higher bitrates in Table 1 (corresponding to quantization parameter value 22) and for lower bitrates in Table 2 (corresponding to quantization parameter value 42). In addition, some RD results comparing Uni-SS and Bi-SS prediction are presented using the Bjøntegaard delta (BD) [42] metrics, i.e., in terms of the luma PSNR of the LF image and the corresponding bitrate (BR) in terms of bits per pixel (bpp) for four different Quantization Parameter (QP) values. In this case, the sets of QP values $\{22, 27, 32, 37\}$ and $\{27, 32, 37, 42\}$ were considered for analyzing the performance, respectively, for high and low bitrates.

For better analyzing these results, two different situations where the MI cross-correlation is differently distributed in a neighborhood are considered (corresponding to the highlighted values in Tables 1 and 2). For the first situation, a second frame

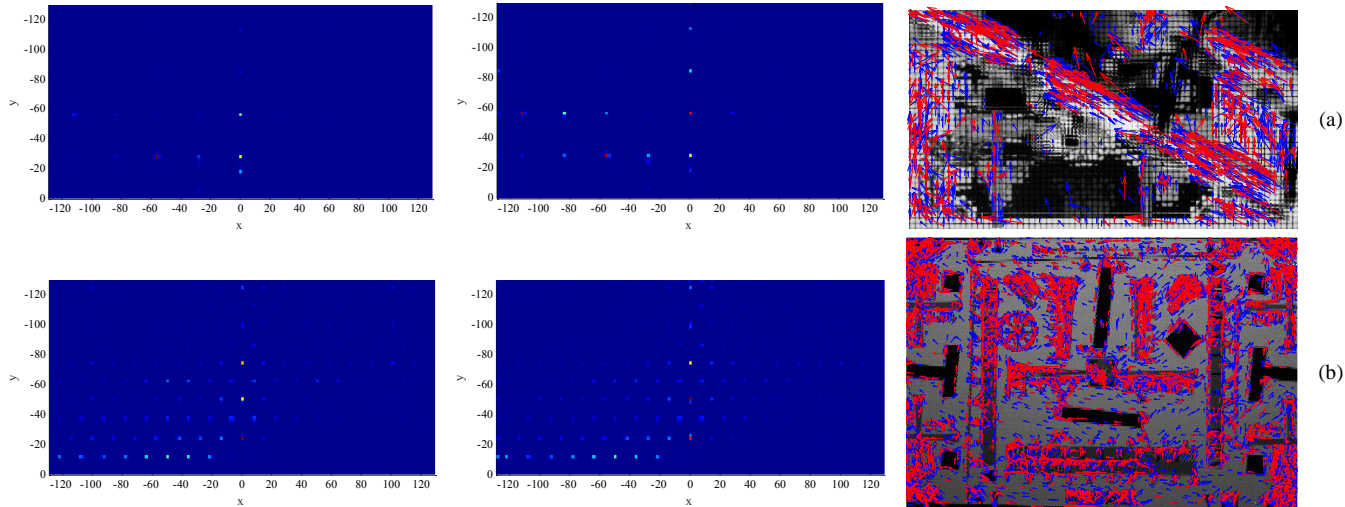


Fig. 4 Bi-predicted SS vector distribution (from left to right): heat map of SS vector distribution for first candidate predictor, heat map of SS vector distribution for second candidate predictor, distribution of first (in blue) and second (in red) SS vectors along the encoded raw image. Results are illustrated for: (a) *Plane and Toy* (frame 23); and (b) *ISO_Chart_12*.

of *Plane and Toy* sequence (frame 23) is used to exemplify the case where the MI cross-correlation varies for the same camera parameters due to the different distance of the main object relatively to the camera [40]. In this case, *Plane and Toy* (frame 23) presents a more rapid decrease in the MI cross-correlation in a neighborhood when compared to *Plane and Toy* (frame 123). For the second situation, two raw LF images, *Jeff* and *ISO_Chart_12*, with considerably larger and smaller MI resolutions, respectively, are used to illustrate the case in which the MI cross-correlation varies due to a change in the camera parameters. In this case, the aperture of the microlens (which usually corresponds to the size of the MI) limits its field of view [41], and consequently, the smaller the MI is, the less correlated it will be with MIs in its neighborhood (i.e., less overlapped areas of the scene will be captured by neighbor MIs). For completeness, Tables 1 and 2 also show statistics for all LF test images.

The results in Table 1 illustrate the influence of the MI cross-correlation in the usage and performance of the proposed Bi-SS approach for higher bitrates. Comparing the results for *Plane and Toy* (frame 123) and *Plane and Toy* (frame 23), it can be seen that the percentage of SS bi-prediction is larger in the case where the MI cross-correlation decreases rapidly (frame 23). Moreover, the bi-prediction is also able to achieve larger bit savings in this case (frame 23) when compared to the LFC Uni-SS solution, where only uni-prediction is used. Somehow similar to the abovementioned conclusion, the usage percentage of bi-prediction tends to be also considerably larger for the case of *ISO_Chart_12* where the MI cross-correlation is smaller due to changes in the camera parameters (compared to *Jeff*).

In addition, Fig. 4 illustrates the Bi-SS vectors distribution for the LF images *Plane and Toy* (frame 23) and *ISO_Chart_12*. It can be seen that the Bi-SS vectors are also distributed according to MI structure of each tested LF image (as was also concluded for the uni-predicted SS vectors [19]). This is evident for *ISO_Chart_12* (in Fig. 4b) where the MIs in the raw LF image are distributed according to a hexagonal grid, and, consequently, the SS vectors distribution (Fig. 4b) follow the same structure. This fact supports the assumptions made in Section 3.1 that more than one good predictor is likely to be found in all directions, distributed according to the MI size and arrangement in the array. Moreover, analyzing the distribution of these SS vectors along the raw LF image (rightmost images in Fig. 4), it can be seen that, although the amplitude of the SS vectors respects the MI structure, the direction tends to be dictated by the object boundaries and textures.

It is also worthwhile to notice (Table 1) that, for all tested LF images considered in Section 4, most of the coding blocks tend to be partitioned down to 8×8 blocks, being smaller than the MI resolution. These results support the assumptions used in the theoretical analysis from Section 3.1, which considered that samples inside each MI follow the same correlation model as samples in a 2D image.

To complete this analysis, Table 2 illustrates the statistical results also for lower bitrates, so as to analyze the performance of the Bi-SS prediction when the SS reference degrades. Comparing the results in Tables 1 and 2 for all test images, it can be observed that using higher QP values results in increasing percentages of usage of the SS uni-prediction, as well as SS-skip modes. This is due to the fact that the Lagrangian multiplier, in (1) and Fig. 3, increases with increasing QP values [35] and, for this reason, the possible quality improvements of using the bi-prediction do not justify the higher number of bits needed for transmitting the SS vectors when minimizing the Lagrangian cost.

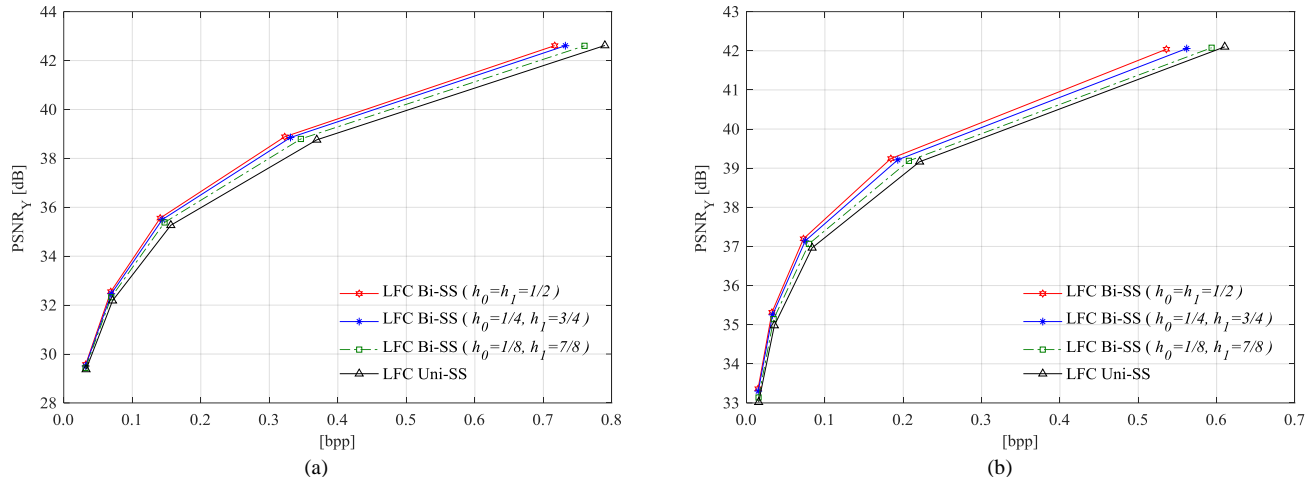


Fig. 5 RD performance of the proposed LFC Bi-SS solution (compared to the LFC Uni-SS solution) with different sets of weighting coefficients (QP values 22, 27, 32, 37, and 42). The results are show for two LF test images: (a) *Seagull*, and (b) *Vespa*

3.4. Bi-SS Prediction Analysis for Different Weighting Coefficients

This section aims at analyzing the RD performance of the proposed LFC Bi-SS solution when different sets of weighting coefficients, h_0 and h_1 (see Fig. 3), are used for Bi-SS prediction.

For this, the HEVC weighted prediction signaling [37] is used. Basically, the usage of explicit weighted prediction in HEVC is activated by a flag in the Picture Parameter Set (PPS), and different integer weighting factors, w_p , and offset values, o_p , can be assigned for prediction in each slice [37]. The resulting predictor block $\hat{I}(x)$ for weighted bi-prediction can be then derived by [37]:

$$\hat{I}(x) = \left\lfloor \frac{w_0 \cdot \tilde{I}(x - v_0) + w_1 \cdot \tilde{I}(x - v_1) + (o_0 + o_1 + 1) \cdot 2^{LWD}}{2 \cdot 2^{LWD}} \right\rfloor \quad (5)$$

where LWD is a log weight denominator rounding factor [37] used to normalize the integer weighting factors and the sub-sample interpolation filtering process [43]. Notice that determining suitable weighting factors and offsets is out of the scope of HEVC standard and they are directly derived from the bitstream at the decoder side.

In this case, the following three LFC Bi-SS solutions are here tested and compared:

- 1) *LFC Bi-SS* ($h_0 = h_1 = 1/2$) – This corresponds to the proposed Bi-SS prediction, where the average weighting coefficients are used. In (5), this corresponds to having $w_0 = w_1 = 1$, $LWD = 0$, and $o_0 = o_1 = 0$.
- 2) *LFC Bi-SS* ($h_0 = 1/4$, and $h_1 = 3/4$) – In this case, weighted Bi-SS prediction is used, where the average weighting coefficients in Fig. 3 are replaced by $h_0 = 1/4$ and $h_1 = 3/4$. In (5), this corresponds to having $w_0 = 1$, $w_1 = 3$, $LWD = 1$, and $o_0 = o_1 = 0$.
- 3) *LFC Bi-SS* ($h_0 = 1/8$, and $h_1 = 7/8$) – In this case, weighted Bi-SS prediction is used with weighting coefficients $h_0 = 1/8$ and $h_1 = 7/8$. In (5), this corresponds to having $w_0 = 1$, $w_1 = 7$, $LWD = 2$, and $o_0 = o_1 = 0$.

Due to space limits, Fig. 5 illustrates the results for only two LF images (i.e., *Seagull* and *Vespa*), but the conclusions are consistent for all LF test images considered in Section 4. From these results, it can be seen that the LFC Bi-SS solution using the average weighting coefficients always presents the best RD performance. Moreover, the less asymmetric the weighting coefficients are, the better the RD efficiency of the LFC Bi-SS solution is shown to be. In addition, in all cases, the weighted Bi-SS always outperform the LFC Uni-SS solution. For this reason, the average weighting coefficients are here considered to the proposed LFC Bi-SS solution.

Notice that it was not under the scope of this work to develop an optimized set of weighting coefficients for each CB (instead of fixing it for each slice). In fact, the theoretical analysis in [38] suggested that it is still possible to achieve further RD performance improvements by adaptively estimating these weighting coefficients; this will be considered in future work, as well as the experimental validation, for LF image coding, of the theoretical assumption made in [33] that suggests that the major portion of the gain is already achievable by only two predictor blocks.

Table 3. RD performance for LFC Bi-SS solution using MIVP (for QP values 27, 32, 37, and 42)

Image	Uni-SS vs. Uni-SS w/o MIVP (from [19])		Bi-SS vs. Bi-SS w/o MIVP	
	PSNR [dB]	BR [%]	PSNR [dB]	BR [%]
<i>Plane and Toy</i>	0.2	-2.8	0.2	-3.1
<i>Demichelis Spark</i>	0.3	-8.3	0.1	-3.5
<i>Laura</i>	0.1	-2.5	0.1	-2.3
<i>Jeff</i>	0.2	-4.7	0.1	-3.1
<i>Seagull</i>	0.3	-6.6	0.2	-5.5
<i>Zhengyun1</i>	0.2	-5.2	0.2	-4.7
Average	0.2	-5.0	0.2	-3.7

3.5. MIVP Efficiency Analysis for Bi-SS Prediction

To separately analyze the RD efficiency of using the MIVP vector prediction for Bi-SS prediction, Table 3 shows some preliminary results for comparing the performance of Bi-SS prediction with MIVP (referred to as Bi-SS w/ MIVP) and without MIVP (referred to as Bi-SS w/o MIVP). These results are then compared to the achieved RD improvements presented for the previously proposed LFC Uni-SS solution in [19] (i.e., Uni-SS solution with and without MIVP).

In these tests, the same test conditions adopted in [19] are here used. Notably, HEVC reference software version 14.0 [36] is used as the benchmark, as well as the base software for implementing the proposed codec. RD performance is evaluated here for six different LF images through the BD [42] metrics, i.e., in terms of the luma PSNR of the LF image and the corresponding bitrate (BR) in terms of bits per pixel (bpp) for four QP values (27, 32, 37, and 42). These results of the MIVP influence on RD performance for Uni-SS prediction can also be found in [19].

From these results, it is possible to see that the gains of including the MIVP candidate vectors for Bi-SS prediction are slightly lower than for Uni-SS prediction. However, the MIVP is still relevant for the LFC Bi-SS solution, leading to further bit savings of up to 5.5% (for LF test image *Seagull*).

4. Performance Evaluation

This section assesses the performance of the complete LFC Bi-SS codec. For this, the test conditions are firstly introduced and, then, the obtained results are presented and discussed.

4.1. Test Conditions

The test conditions to evaluate the performance of the proposed LFC Bi-SS solution can be summarized as follows.

4.1.1. Test Images

Twelve LF test images with different camera setups and scene characteristics are used (see Fig. 6 and Table 4) so as to achieve representative RD results. These are: *Plane and Toy* (frame number 123 of the sequence with same name) and *Demichelis Spark* [44] (first frame of a video sequence with same name); *Laura*, *Fredo*, *Seagull*, and *Zhengyun* [45]; and *Flowers*, *Vespa*, *Ankylosaurus_&_Diplodocus_1*, *Fountain_&_Vincent_2*, *Color_Chart_1*, and *ISO_Chart_12* [46].

The (raw) LF test images were converted to the Y’CbCr 4:2:0 color format before being encoded.

4.1.2. Coding Conditions

The experimental results that are presented in this section considered the following test conditions:

- 1) *Codec Software Implementation* – The reference software of HEVC version 14.0 [36] was used as the benchmark, as well as the base software for implementing the proposed codec.
- 2) *Search Range* – A search range value of 128 was adopted for all tested LF images (i.e., $w=128$ in Fig. 2b).
- 3) *Search Strategy* – The full search algorithm with the HEVC quarter-pixel accuracy was used since fast search algorithms proposed for 2D video coding and for SCC have shown to present significant drop in RD performance for LF image coding [20].
- 4) *Coding Configuration* – The results are presented using the Main Still Picture profile [23] and five QP values are considered, i.e.: 22, 27, 32, 37, and 42.

Table 4. Description of LF test images in Fig. 6

Image	Resolution	LF camera setup	MLA structure	MI Resolution
<i>Plane and Toy</i>	1920×1088	Focused LF camera (Plenoptic camera 2.0)	Rectangular grid of squared MIs	28×28*
<i>Demichelis Spark</i>	2880×1620	Focused LF camera (Plenoptic Camera 2.0)	Rectangular grid of circular MIs	38×38*
<i>Laura</i>	7240 ×5232	Focused LF camera (Plenoptic Camera 2.0)	Rectangular grid of squared MIs	74×74*
<i>Jeff</i>	7240 ×5232	Focused LF camera (Plenoptic Camera 2.0)	Rectangular grid of squared MIs	74×74*
<i>Seagull</i>	7240 ×5232	Focused LF camera (Plenoptic Camera 2.0)	Rectangular grid of squared MIs	74×74*
<i>Zhengyun1</i>	7240 ×5232	Focused LF camera (Plenoptic Camera 2.0)	Rectangular grid of squared MIs	74×74*
<i>Flowers</i>	7728×5368	Traditional LF camera (Plenoptic Camera 1.0)	Hexagonal grid of circular MIs	15×15*
<i>Vespa</i>	7728×5368	Traditional LF camera (Plenoptic Camera 1.0)	Hexagonal grid of circular MIs	15×15*
<i>Ankylosaurus_&_Diplodocus_1</i>	7728×5368	Traditional LF camera (Plenoptic Camera 1.0)	Hexagonal grid of circular MIs	15×15*
<i>Fountain_&_Vincent_2</i>	7728×5368	Traditional LF camera (Plenoptic Camera 1.0)	Hexagonal grid of circular MIs	15×15*
<i>Color_Chart_1</i>	7728×5368	Traditional LF camera (Plenoptic Camera 1.0)	Hexagonal grid of circular MIs	15×15*
<i>ISO_Chart_12</i>	7728×5368	Traditional LF camera (Plenoptic Camera 1.0)	Hexagonal grid of circular MIs	15×15*

*Integer approximation

5) *RD Evaluation* – The RD performance was evaluated in terms of three different objective quality metrics: i) Overall PSNR; ii) Rendering-dependent PSNR (PSNR_{5×5Views}); and iii) Rendering-dependent SSIM (SSIM_{5×5Views}). The overall PSNR is calculated by taking the luma PSNR of the raw LF image. Differently, the rendering-dependent metrics (PSNR_{5×5Views} and SSIM_{5×5Views}) are measured in terms of the average luma PSNR and SSIM calculated for a set of views rendered from the reconstructed LF content, similarly to the metrics proposed in [6]. To have a representative number of rendered views, a set of 5×5 views was rendered from equally distributed directional positions. For rendering the views from LF images captured using a focused LF camera setup (Table 4), the algorithm proposed in [47] and referred to as Basic Rendering algorithm was used. In this case, the plane of focus was chosen to represent the case where the main object of the scene is in focus. For LF images captured using the traditional LF camera setup (Table 4), 5×5 VIs were extracted. The rate is given in terms of bpp value, which is calculated with the number of bits needed for encoding the LF image divided by its corresponding number of pixels given in Table 4 (bpp).

4.1.3. Benchmark Solutions

In order to assess the RD performance for different local and non-local spatial prediction schemes, five HEVC-based coding solutions are compared against the proposed LFC Bi-SS. To guarantee a fair comparison between all of them, the same test conditions presented in Section 4.1.2 are also adopted. These five benchmark solutions are:



Fig. 6 Central rendered view from each test LF image: (a) *Plane and Toy*, (b) *Demichelis Spark*; (c) *Laura*, (d) *Jeff*. (e) *Seagull*, (f) *Zhengyun1*. (g) *Flowers*, (h) *Vespa*. (i) *Ankylosaurus_&_Diplodocus_1*, (j) *Fountain_&_Vincent_2*, (k) *Color_Chart_1*, and (l) *ISO_Chart_12*

- 1) *HEVC* – In this case, the original LF image is encoded with HEVC [36], using the Main Still Picture profile [23].
- 2) *HEVC SCC* – In this case, the original LF image is encoded using the HEVC SCC reference software version 1.0 [48], where IntraBC prediction [32] is used. As previously discussed in this paper, this solution is a restricted case of the proposed LFC Bi-SS since a reduced set of coding options is used, such as: i) only uni-prediction estimation is allowed considering only integer-pixel precision; ii) partition patterns are limited (i.e., $M \times M$, $M \times (M/2)$, $(M/2) \times M$, and $(M/2) \times (M/2)$); iii) CB sizes are also limited (CBs larger than 16×16 are skipped based on a threshold on the RD cost); and iv) only one dimensional (1D) vectors for 16×16 CBs are allowed. However, it is worth mentioning that some improvements for the IntraBC prediction have been continuously included. For instance, in the HEVC SCC reference software 1.0, the search window was expanded over the entire CB row or column (for 16×16 CBs), and over some positions in the entire picture by using a hash-based search (for 8×8 CBs).
- 3) *LFC Uni-SS* – In this case, the original LF image is encoded with the authors’ previous solution proposed in [19], where only the uni-predicted candidate is available for the SS estimation and compensation. This solution also uses the MIVP candidate vectors for improving the coding performance (as in the proposed LFC Bi-SS). A search area with $w=128$ (in Fig. 2b) is also adopted in this case.
- 4) *LFC Restricted-SS* – In this case, the original LF image is encoded with the author’s implementation of the solution proposed in [20], where bi-prediction is also allowed by simply using the HEVC inter B-frame prediction. For this, and as explained in [20], the SS reference search area is separated into two different parts, which are assumed to be two different reference pictures [20]. Therefore, as in HEVC inter B-frame prediction, three candidate predictors can be derived: the two best (uni-predicted) candidates from each of the two reference pictures, and a linear combination of them for bi-prediction. As discussed in Section 3, this solution can be seen as a restricted case of the Bi-SS prediction proposed here. It is also worthwhile to notice that, the solution presented in [20] (as well as the author’s implementation of this solution) does not include the MIVP candidate vectors that are used in both LFC Uni-SS and LFC Bi-SS solutions. A search area with $w=128$ (in Fig. 2b) is also adopted in this case.
- 5) *LFC GPR* – In this case, an HEVC-based LFC solution using the implicit GPR-based prediction method proposed in [22] is considered. Different from the SS prediction, the predictor block is here given as a linear combination of six nearest neighbor patches, which are implicitly found in two different search windows: i) horizontal search window (defined as the causal area in the same CB row); ii) the vertical search window (defined as the causal area in the same CB column). Afterwards, the set of weighting coefficients for combining these six patches is implicitly determined by solving a GPR and the same process to find the nearest neighbor patches and weighting coefficients is replicated at the decoder side to derive the prediction. For this comparison, the results presented in [22] for coding using this LFC GPR solution are compared to the results for the proposed LFC Bi-SS solution considering the same test conditions (described in [22]). Notably, the results are shown using the BD metrics in terms of the overall PSNR with respect to HEVC reference software [36] with “Intra, main” configuration and considering four different QP values: 22, 27, 32, and 37 (as in [22]).

It is important to notice that, other benchmark solutions have already been compared against the authors’ previous solution LFC Uni-SS in [19], (e.g., a multiview-based method, similar to the proposed in [12,13]). For a more complete performance comparison, please also refer to [19].

4.2. Overall RD Performance

Figs. 7 and 8 illustrate the RD performance of the proposed LFC Bi-SS using three different objective quality metrics, i.e., the overall PSNR, and the rendering-dependent metrics $PSNR_{5 \times 5 \text{Views}}$ and $SSIM_{5 \times 5 \text{Views}}$. Additionally, Tables 5 and 6 present the LFC Bi-SS RD performance against four benchmark solutions (i.e., HEVC, HEVC SCC, LFC Uni-SS, and LFC Restricted-SS) in terms of BD metrics, respectively, for higher and lower bpp values. It is worth noting that the comparison between the proposed LFC Bi-SS and the LFC GPR solutions will be separately performed in Section 4.4 since, in this case, the results presented in [22] for the LFC GPR solution is directly compared to the LFC Bi-SS solution RD results when using the same set of LF images as well as the same test coding conditions adopted in [22].

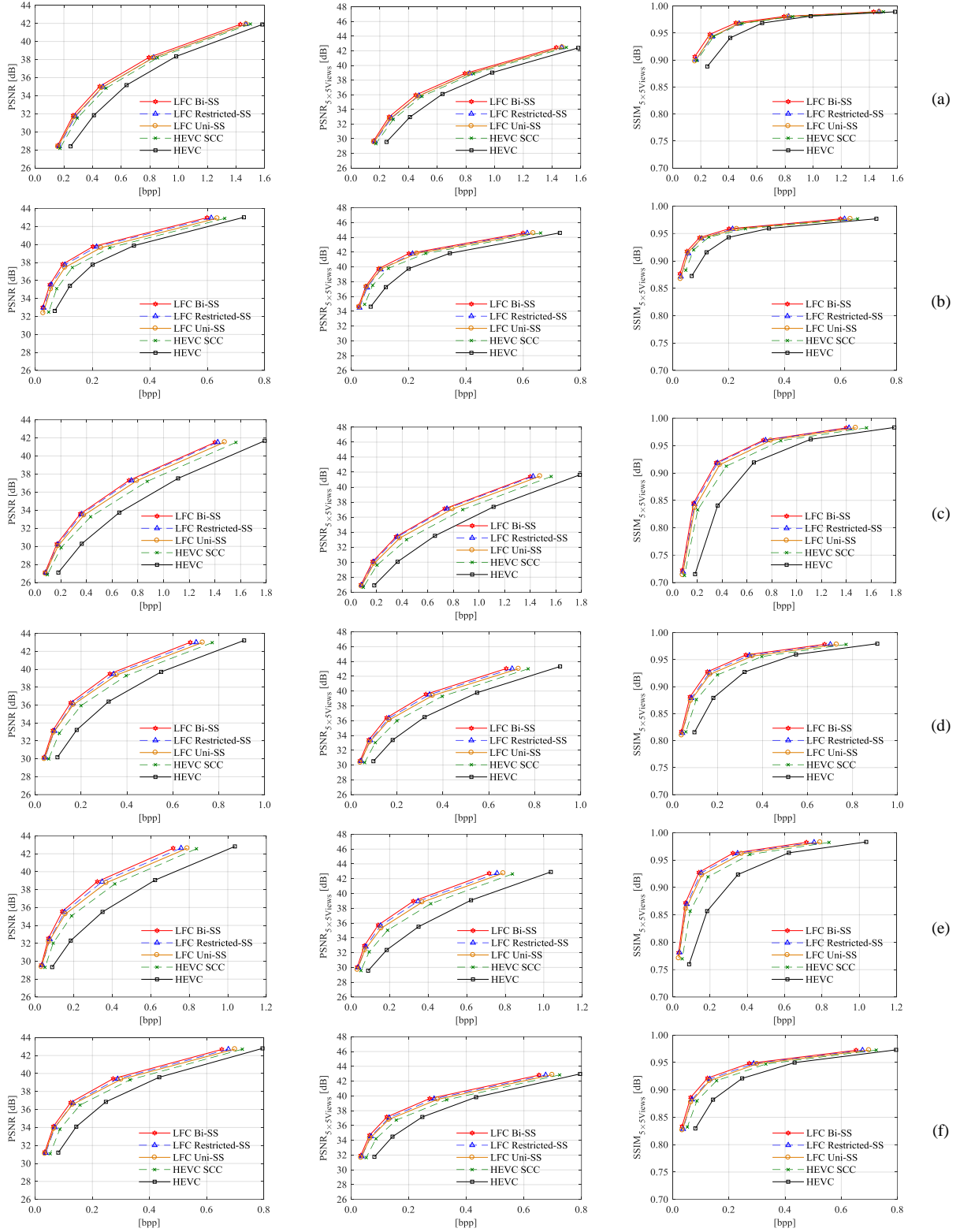


Fig. 7 RD performance (QP values 22, 27, 32, 37, and 42) for the LF test images: (a) *Plane and Toy*, (b) *Demichelis Spark*; (c) *Laura*, (d) *Jeff*, (e) *Seagull*, and (f) *Zhengyun1*

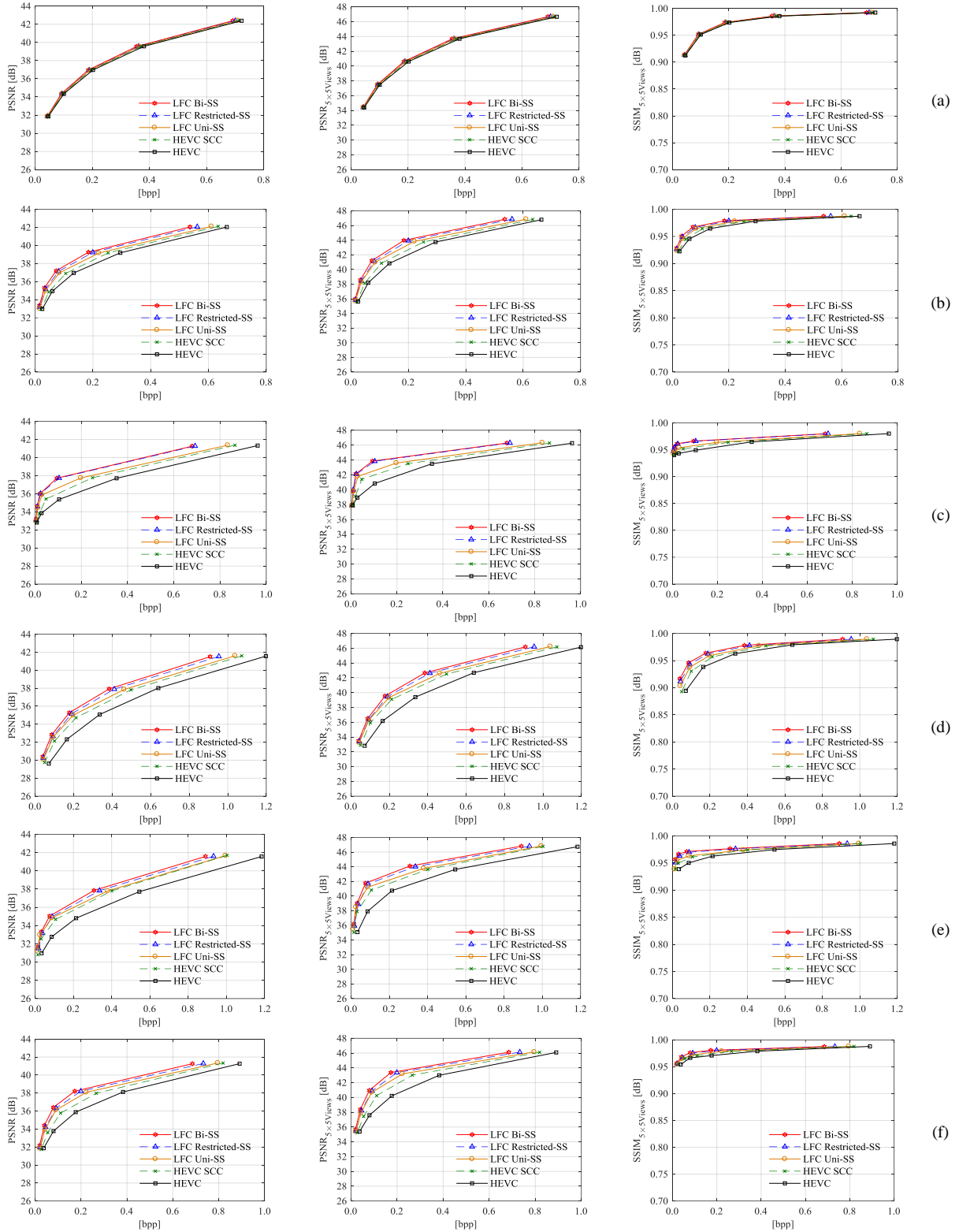


Fig. 8 RD performance (QP values 22, 27, 32, 37, and 42) for the LF test images: (a) *Flowers*, (b) *Vespa*, (c) *Ankylosaurus_&_Diplodocus_1*, (d) *Fountain_&_Vincent_2*, (e) *Color_Chart_1*, and (f) *ISO_Chart_12*

Table 5. LFC Bi-SS overall RD performance with respect to the benchmark solutions for each image in Fig. 6 (QP values 22, 27, 32, and 37)

LF Image	LFC Bi-SS vs. HEVC		LFC Bi-SS vs. HEVC SCC		LFC Bi-SS vs. LFC Uni-SS		LFC Bi-SS vs. LFC Restricted-SS	
	BD-PSNR [dB]	BD-BR [%]	BD-PSNR [dB]	BD-BR [%]	BD-PSNR [dB]	BD-BR [%]	BD-PSNR [dB]	BD-BR [%]
<i>Plane and Toy</i>	1.6	-22.8	0.6	-9.4	0.3	-4.5	0.3	-4.9
<i>Demichelis Spark</i>	1.7	-43.0	1.0	-27.4	0.5	-14.9	0.2	-6.8
<i>Laura</i>	2.8	-37.9	1.2	-19.4	0.5	-8.6	0.2	-3.0
<i>Jeff</i>	2.9	-43.7	1.3	-23.8	0.5	-10.4	0.3	-6.3
<i>Seagull</i>	3.6	-51.9	1.5	-28.3	0.7	-14.4	0.4	-8.3
<i>Zhengyun1</i>	2.7	-47.8	1.0	-21.8	0.4	-9.9	0.3	-6.7
<i>Flowers</i>	0.3	-7.1	0.2	-4.6	0.1	-3.1	0.1	-2.0
<i>Vespa</i>	1.4	-41.3	1.0	-31.3	0.5	-18.9	0.2	-9.4
<i>Ankylosaurus_&_Diplodocus_1</i>	2.3	-68.9	1.4	-56.1	0.7	-44.1	0.1	-6.8
<i>Fountain_&_Vincent_2</i>	2.2	-41.8	1.1	-25.0	0.7	-16.6	0.3	-7.7
<i>Color_Chart_1</i>	2.1	-51.1	0.8	-26.7	0.4	-22.5	0.3	-11.3
<i>ISO_Chart_12</i>	2.1	-54.1	1.2	-38.1	0.6	-22.7	0.3	-12.9
Average	2.1	-42.6	1.0	-26.0	0.5	-15.9	0.2	-7.2

Table 6. LFC Bi-SS overall RD performance with respect to the benchmark solutions for each image in Fig. 6 (QP values 27, 32, 37, and 42)

LF Image	LFC Bi-SS vs. HEVC		LFC Bi-SS vs. HEVC SCC		LFC Bi-SS vs. LFC Uni-SS		LFC Bi-SS vs. LFC Restricted-SS	
	BD-PSNR [dB]	BD-BR [%]	BD-PSNR [dB]	BD-BR [%]	BD-PSNR [dB]	BD-BR [%]	BD-PSNR [dB]	BD-BR [%]
<i>Plane and Toy</i>	2.4	-31.0	0.8	-12.3	0.3	-4.9	0.4	-6.4
<i>Demichelis Spark</i>	3.1	-56.0	1.6	-37.0	0.7	-17.3	0.3	-8.8
<i>Laura</i>	3.4	-48.0	1.2	-23.1	0.5	-9.5	0.2	-4.6
<i>Jeff</i>	3.6	-52.3	1.5	-28.6	0.4	-8.8	0.3	-7.0
<i>Seagull</i>	4.3	-61.1	1.8	-34.3	0.6	-13.8	0.5	-10.5
<i>Zhengyun1</i>	3.1	-51.7	1.4	-29.0	0.4	-9.1	0.4	-8.9
<i>Flowers</i>	0.3	-8.3	0.2	-5.6	0.1	-3.4	0.1	-2.4
<i>Vespa</i>	1.7	-50.3	1.2	-39.6	0.5	-20.6	0.3	-12.7
<i>Ankylosaurus_&_Diplodocus_1</i>	2.3	-82.4	1.7	-66.1	0.5	-33.6	0.2	-9.6
<i>Fountain_&_Vincent_2</i>	2.6	-51.1	1.2	-29.5	0.6	-16.4	0.4	-10.0
<i>Color_Chart_1</i>	2.4	-66.2	0.9	-36.5	0.3	-15.0	0.4	-16.9
<i>ISO_Chart_12</i>	2.5	-60.0	1.6	-42.6	0.6	-20.8	0.5	-15.8
Average	2.7	-51.5	1.3	-32.0	0.5	-14.4	0.3	-9.4

As can be observed in Figs. 7 and 8, independently of the adopted objective quality metrics, the results follow the same trend and the proposed LFC Bi-SS solution outperforms the other benchmark solutions in all cases. Moreover, as shown in Tables 5 and 6, significant gains can be achieved by the proposed LFC Bi-SS solution mainly for lower bpp values (Table 6), achieving in average 2.7 dB (or 51.5 % of bit savings) against HEVC, 1.3 dB (or 32.0 % of bit savings) against HEVC SCC, 0.5 dB (or 14.4 % of bit savings) against LFC Uni-SS, and 0.3 dB (or 9.4 % of bit savings) against LFC Restricted-SS. Additionally, as hypothesized in the theoretical analysis of Section 3.1, increasing the number of predictor blocks from the LFC Uni-SS solution to the LFC Bi-SS solution led to bit savings of up to 22.7 % (Table 5). Furthermore, comparing to the LFC Restricted-SS solution, it can be seen that by jointly estimating both candidate predictors, instead of devising them separately from different areas, further bit savings of up to 16.9 % (Table 6) can be achieved.

Moreover, comparing the achieved performance gains due to MIVP (Table 3) with the results against the LFC Uni-SS in Table 5, it can be seen that most of the LFC Bi-SS performance gains come from the usage of an improved bi-prediction scheme. However, it should be noticed that the MIVP is mainly advantageous for improving the performance against the LFC Restricted-SS (see Table 5), even when only Uni-SS prediction is allowed. For instance, for the LF image *Plane and Toy*, the LFC Uni-SS solution (using the MIVP) presents a slightly better performance than the LFC Restricted-SS solution (where the MIVP is not used).

Regarding the performance for different LF image characteristics, it was observed that the coding efficiency of the SS compensated prediction is not as good for close-up images. This can be seen by analyzing the results for LF image *Flowers* (close-up image) against *Fountain_&_Vincent_2* (see Fig. 6g and j, respectively), but it was also observed for other LF images of the dataset in [46]. In these close-up images, most of the coding blocks (about 75 % of them) are coded using HEVC intra prediction.

Table 7. LFC Bi-SS encoding complexity regarding the benchmark solutions with QP value set to 32 (for each image in Fig. 6)

Image	LFC Bi-SS Encoding Time [s]	LFC Bi-SS Encoding Time Ratio w.r.t:			
		HEVC	HEVC SCC	LFC Uni-SS	LFC Restricted-SS
<i>Plane and Toy</i>	1429.6	66.1	20.2	2.3	1.8
<i>Demichelis Spark</i>	2552.9	71.6	38.5	2.2	1.7
<i>Laura</i>	34479.6	84.8	1.9	2.1	1.9
<i>Jeff</i>	28360.0	91.7	10.4	2.0	1.7
<i>Seagull</i>	30665.7	86.8	2.9	1.9	1.7
<i>Zhengyun1</i>	30604.9	48.6	21.9	1.9	1.7
<i>Flowers</i>	39686.0	123.6	24.1	2.2	1.7
<i>Vespa</i>	23871.4	76.1	23.7	2.0	1.7
<i>Ankylosaurus_&_Diplodocus_1</i>	23289.8	78.0	16.0	2.0	1.7
<i>Fountain_&_Vincent_2</i>	32475.5	89.1	5.8	2.0	1.7
<i>Color_Chart_1</i>	32273.1	99.2	10.3	2.0	1.7
<i>ISO_Chart_12</i>	29985.8	93.5	9.7	1.8	1.6
Average		84.1	15.5	2.0	1.7

4.3. Computational Complexity

The significantly better performance of the LFC Bi-SS solution comes with the price of additional computational load compared to both LFC Uni-SS and LFC Restricted-SS.

To illustrate this fact, Tables 7 and 8 present, respectively, the encoding and decoding time of the LFC Bi-SS solution, and also the time ratio in terms of $Time_{Bi-SS}/Time_{bench.}$, with respect to HEVC, HEVC SCC, LFC Uni-SS, and LFC Restricted-SS benchmark solutions. For this, encoding/decoding times were obtained using a machine with an Intel Xeon E5-2620 v2 processor clocked at 2.10 GHz and using gcc 4.8.3 compiler.

Regarding the encoding complexity (Table 7), it can be seen that coding solutions that make use of a block-based matching algorithm (i.e., HEVC SCC, LFC Uni-SS, and LFC Restricted-SS and LFC Bi-SS) are generally much slower than the conventional HEVC intra coding (which is used as the HEVC benchmark solution with Still Picture Profile in this paper). For instance, LFC Bi-SS is 84.1 times slower than HEVC intra coding in average. The brute force RDO considering a larger set of available prediction modes is also a contributing factor for this increased encoding complexity in the LFC Bi-SS solution. Compared to HEVC SCC, the proposed LFC Bi-SS encoding is 15.5 slower (in average) mainly due to the brute force search that is performed to a larger set of CB partition patterns and sizes and in quarter-pixel positions, instead of the integer-pixel search that is performed in HEVC SCC using the faster hash-based search algorithm [32]. It is also important to notice that the encoding time using HEVC SCC may vary considerably for different LF images, mainly due to the different runtime complexity that can be achieved by the used hash-based search algorithm. This fact can be observed in Table 7 by comparing the significantly different encoding time ratios between *Laura* and *Flowers*. Additionally, it can be seen that the proposed LFC Bi-SS solution is 2.0 times slower than LFC Uni-SS and 1.7 times slower than LFC Restricted-SS (in average). In this case, this increase in encoding complexity is observed mainly due to the locally optimal rate constrained algorithm that is used for finding the two possible predictors Uni-SS and Bi-SS candidates. For instance, a search window with W (quarter) pixel positions will require $(K + 1) \times W$ SAD computations in the LFC Bi-SS solution compared to the W SAD computations needed for the LFC Uni-SS solution, where K is the maximum number of iterations allowed in the locally optimal rate constrained algorithm (in this work $K = 2$).

Nevertheless, note that a fast search approach can still be adopted for the proposed LFC Bi-SS solution, for instance, by taking advantage of the regular SS vector distribution depicted in Fig. 4. In this case, instead of performing the full search algorithm, the number of positions that are visited for the block-based matching algorithm can be significantly reduced by considering only the most probable positions shown in Fig. 4.

Regarding the decoding complexity (Table 8), it can be observed that the proposed LFC Bi-SS decoding complexity does not vary significantly with respect to the LFC Uni-SS and LFC Restricted-SS solutions. However, the proposed LFC Bi-SS is 12.9 and 15.3 times slower than HEVC and HEVC SCC, respectively, mainly due to the quarter-pixel interpolation filter that is used for the SS compensation and due to the larger set of coding possibilities (e.g., prediction modes, CB partition patterns) which require more time for parsing at decoder.

Table 8. LFC Bi-SS decoding complexity regarding the benchmark solutions with QP value set to 32 (for each image in Fig. 6)

Image	LFC Bi-SS Decoding Time [s]	LFC Bi-SS Decoding Time Ratio w.r.t:			
		HEVC	HEVC SCC	LFC Uni-SS	LFC Restricted-SS
<i>Plane and Toy</i>	2.1	20.9	21.1	1.0	1.2
<i>Demichelis Spark</i>	1.6	10.5	12.6	0.9	1.5
<i>Laura</i>	38.1	18.6	23.1	0.9	0.9
<i>Jeff</i>	26.4	17.3	21.4	0.9	0.9
<i>Seagull</i>	31.6	20.8	26.8	1.2	1.0
<i>Zhengyun1</i>	29.1	18.6	23.8	1.2	1.0
<i>Flowers</i>	12.2	9.9	9.6	1.1	0.9
<i>Vespa</i>	8.0	7.4	7.9	1.0	0.8
<i>Ankylosaurus_&_Diplodocus_1</i>	4.8	4.2	6.0	0.8	0.8
<i>Fountain_&_Vincent_2</i>	12.7	9.4	10.3	0.9	0.9
<i>Color_Chart_1</i>	10.3	9.3	10.8	0.7	0.9
<i>ISO_Chart_12</i>	9.3	8.2	10.2	0.9	0.9
Average		12.9	15.3	1.0	1.0

Table 9. RD performance and complexity comparison between LFC Bi-SS and GPR-based solution proposed in [22] considering QP values 22, 27, 32, and 37 (according to [22])

LF Image	LFC Bi-SS vs. HEVC				LFC GPR vs. HEVC (from [22])			
	BD-PSNR [dB]	BD-BR [%]	Encoding Time Ratio	Decoding Time Ratio	BD-PSNR [dB]	BD-BR [%]	Encoding Time Ratio	Decoding Time Ratio
<i>Bike</i>	2.6	-39.2	81.5	20.9	2.4	-27.3	19.8	74.1
<i>Fountain</i>	2.6	-37.6	88.8	10.5	2.8	-36.4	20.4	97.7
<i>Fredo</i>	3.3	-45.9	48.6	18.6	3.0	-41.0	19.9	91.2
<i>Jeff**</i>	2.9	-43.7	91.7	17.3	2.2	-31.4	19.0	87.1
<i>Laura</i>	2.8	-37.9	84.8	20.8	3.0	-38.1	20.0	99.1
<i>Seagull</i>	3.6	-51.9	86.8	18.6	2.9	-41.4	22.0	101.3
<i>Sergio</i>	3.5	-46.8	88.3	9.9	2.3	-36.9	20.3	88.1
<i>Zhengyun1</i>	2.7	-47.8	48.6	7.4	2.3	-36.2	20.5	91.2
Average	3.0	-43.9	77.4	15.5	2.6	-36.1	20.2	91.2

** referred to as *Photographer* in [22]

4.4. Comparison between Bi-SS and GPR-based LF image coding

To complete the LFC Bi-SS RD performance analysis, Table 9 compares the results presented in [22] for the LFC GPR benchmark solution with the LFC Bi-SS solution RD results when using the same set of LF images as well as the same test coding conditions adopted in [22]. Namely, the results are shown using the BD metrics in terms of the overall PSNR with respect to HEVC reference software [36] with “Intra, main” configuration and considering four different QP values: 22, 27, 32, and 37 (as in [22]).

From these results, it can be seen that the proposed LFC Bi-SS solution is able to significantly improve the overall RD performance compared to the LFC GPR, leading to further bit savings of 43.9% with respect to HEVC (in average) compared to the 36.1 % of bit savings achieved by the LFC GPR solution (also with respect to HEVC).

Additionally, Table 9 also compares the encoding and decoding time ratios (with respect to HEVC) observed for both LFC Bi-SS and LFC GPR solutions. It can be seen that the LFC GPR solution is able to significantly reduce the encoding complexity by reducing the training data as proposed in [22]. However, the implicit prediction used in the LFC GPR solution is still a bottleneck in the decoding time since the same process for finding the nearest neighbor patches and determining the weighting coefficients used in the encoder needs to be repeated at decoder side. For this reason, the proposed LFC Bi-SS is shown to be significantly faster than the LFC GPR at the decoder side.

4.5. Visual Quality Inspection

In addition to the presented BD results, a portion of a central view from each of three different LF test images was used for a visual inspection in Fig. 9. For rendering the views, the algorithm proposed in [47] and referred to as Basic Rendering algorithm was used. For the image in Fig. 9c, the Light Field Toolbox version 0.4 [49] was firstly used to transform from the hexagonal to a square grid of MIs. For all compared solutions, the quantization parameter (QP) [23] of the encoder was

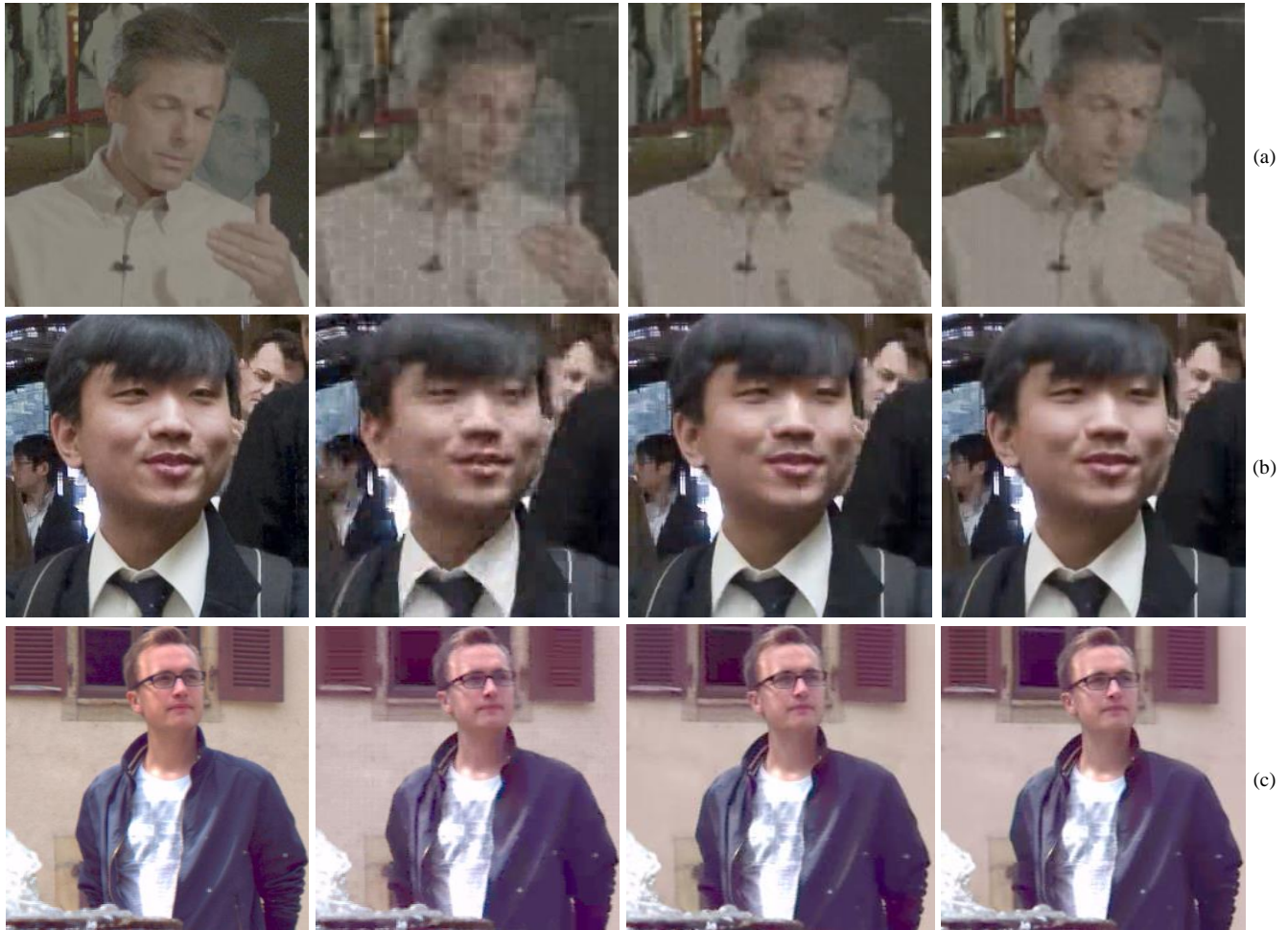


Fig. 9 Comparison of a portion from the central view rendered from (from left to right): original image; compressed image using HEVC; compressed image using LFC Restricted-SS; and compressed image using the proposed LFC Bi-SS solution. The results as shown for the bpp values: (a) 0.05 bpp for *Demichelis Spark*; (b) 0.06 bpp for *Zhengyun1*; and (c) 0.09 bpp for *Fountain_&_Vincent_2*

adjusted to lead to the same bpp value for all images in5 Fig. 9. Notice that, since there is still no consensus on the scientific community regarding subjective evaluation methodologies for LF content, these results are shown here as an illustrative qualitative analysis of the proposed coding solution.

From a visual inspection of the views rendered from the coded raw LF images in Fig. 9, it is possible to conclude that the proposed LFC Bi-SS presents considerably better visual quality than HEVC and improvements are also noticeable when compared to the LFC Restricted-SS solution (e.g., for *Demichelis Spark* and *Zhengyun1* LF images). Furthermore, it can be seen that the coding artifacts are less evident but still noticeable in the case of Fig. 9c.

5. Conclusions

This paper proposed an LF image coding solution based on self-similarity compensated bi-prediction (Bi-SS) where two predictor blocks can be jointly estimated from the same search window. As discussed in this paper, the proposed HEVC-based LFC coding architecture was shown to be advantageous in terms of the simplicity of the coding format, which is less dependent on a very precise LF camera calibration process, while keeping the encoder/decoder complexity and memory load comparable to HEVC inter coding. In addition, the proposed LFC Bi-SS led to significantly superior performance when compared to HEVC Main Still Picture Profile, presenting gains of up to 4.3 dB (or 61.1 % of bit savings). Furthermore, jointly estimating the two candidate blocks for Bi-SS prediction led to further RD improvements when compared to the case in which only one candidate block is estimated (with up to 44.1 % of bit savings with respect to the LFC Uni-SS solution), as well as compared to the case in which to the two candidate blocks are independently estimated (with up to 16.9 % of bit savings with respect to the LFC

Restricted-SS solution). Moreover, the proposed LFC Bi-SS solution is also able to achieve better overall RD performance when compared to the LFC GPR solution while presenting significantly smaller decoding complexity.

Acknowledgements

The authors acknowledge the support of FCT, under the project UID/EEA/50008/2013, and SFRH/BD/79480/2011 grant.

References

- [1] M. Levoy, P. Hanrahan, Light Field Rendering, in: Proc. 23rd Annu. Conf. Comput. Graph. Interact. Tech. - SIGGRAPH '96, New Orleans, LA, US, 1996: pp. 31–42. doi:10.1145/237170.237199.
- [2] A. Aggoun, E. Tsekles, M.R. Swash, D. Zarpalas, A. Dimou, P. Daras, P. Nunes, L.D. Soares, Immersive 3D Holoscopic Video System, *IEEE Multimed.* 20 (2013) 28–37. doi:10.1109/MMUL.2012.42.
- [3] J. Arai, Integral Three-Dimensional Television (FTV Seminar), ISO/IEC JTC1/SC29/WG11 M34199, Sapporo, Japan, 2014.
- [4] R. Raghavendra, K.B. Raja, C. Busch, Presentation Attack Detection for Face Recognition Using Light Field Camera, *IEEE Trans. Image Process.* 24 (2015) 1060–75. doi:10.1109/TIP.2015.2395951.
- [5] X. Xiao, B. Javidi, M. Martinez-Corral, A. Stern, Advances in Three-Dimensional Integral Imaging: Sensing, Display, and Applications [Invited], *Appl. Opt.* 52 (2013) 546–560.
- [6] JPEG Pleno Call for Proposals on Light Field Coding, ISO/IEC JTC 1/ SC29/WG1 N74014, Geneva, Switzerland, 2017.
- [7] K. Wegner, G. Lafruit, eds., Call for Immersive Visual Test Material, ISO/IEC JTC1/SC29/WG11 N16766, Hobart, Australia, 2017.
- [8] R. Zaharia, A. Aggoun, M. McCormick, Adaptive 3D-DCT Compression Algorithm for Continuous Parallax 3D Integral Imaging, *Signal Process. Image Commun.* 17 (2002) 231–242. doi:10.1016/S0923-5965(01)00020-0.
- [9] A. Aggoun, Compression of 3D Integral Images Using 3D Wavelet Transform, *J. Disp. Technol.* 7 (2011) 586–592. doi:10.1109/JDT.2011.2159359.
- [10] S. Shi, P. Gioia, G. Madec, Efficient Compression Method for Integral Images using Multi-View Video Coding, in: 2011 18th IEEE Int. Conf. Image Process., Brussels, Belgium, 2011: pp. 137–140.
- [11] J. Dick, H. Almeida, L.D. Soares, P. Nunes, 3D Holoscopic Video Coding Using MVC, in: 2011 IEEE EUROCON - Int. Conf. Comput. as a Tool, Lisbon, Portugal, 2011: pp. 1–4.
- [12] R. Olsson, Empirical Rate-Distortion Analysis of JPEG 2000 3D and H. 264/AVC Coded Integral Imaging Based 3D-Images, in: 2008 3DTV Conf. True Vis. - Capture, Transm. Disp. 3D Video, Istanbul, Turkey, 2008: pp. 113–116. doi:10.1109/3DTV.2008.4547821.
- [13] A. Vieira, H. Duarte, C. Perra, L. Tavora, P. Assuncao, Data Formats for High Efficiency Coding of Lytro-Illum Light Fields, in: 2015 Int. Conf. Image Process. Theory, Tools Appl., Orleans, France, 2015: pp. 494–497. doi:10.1109/IPTA.2015.7367195.
- [14] C. Perra, P. Assuncao, High Efficiency Coding of Light Field Images based on Tiling and Pseudo-Temporal Data Arrangement, in: 2016 IEEE Int. Conf. Multimed. Expo Work., Seattle, WA, US, 2016: pp. 1–4. doi:10.1109/ICMEW.2016.7574671.
- [15] D. Liu, L. Wang, L. Li, Zhiwei Xiong, Feng Wu, Wenjun Zeng, Pseudo-Sequence-Based Light Field Image Compression, in: 2016 IEEE Int. Conf. Multimed. Expo Work., Seattle, WA, US, 2016: pp. 1–4. doi:10.1109/ICMEW.2016.7574674.
- [16] C. Conti, P. Nunes, L.D. Soares, Inter-Layer Prediction Scheme for Scalable 3-D Holoscopic Video Coding, *IEEE Signal Process. Lett.* 20 (2013) 819–822. doi:10.1109/LSP.2013.2267234.
- [17] A. Dricot, J. Jung, M. Cagnazzo, B. Pesquet, F. Dufaux, Integral Images Compression Scheme Based On View Extraction, in: 2015 23rd Eur. Signal Process. Conf., Nice, France, 2015: pp. 101–105. doi:10.1109/EUSIPCO.2015.7362353.
- [18] C. Conti, P. Nunes, L.D. Soares, New HEVC Prediction Modes for 3D Holoscopic Video Coding, in: 2012 19th IEEE Int. Conf. Image Process., Orlando, FL, US, 2012: pp. 1325–1328. doi:10.1109/ICIP.2012.6467112.
- [19] C. Conti, L.D. Soares, P. Nunes, HEVC-Based 3D Holoscopic Video Coding using Self-Similarity Compensated Prediction, *Signal Process. Image Commun.* 42 (2016) 59–78. doi:10.1016/j.image.2016.01.008.
- [20] Y. Li, M. Sjöström, R. Olsson, U. Jennehag, Coding of Focused Plenoptic Contents by Displacement Intra Prediction, *IEEE Trans. Circuits Syst. Video Technol.* 26 (2016) 1308–1319. doi:10.1109/TCSVT.2015.2450333.
- [21] L.F.R. Lucas, C. Conti, P. Nunes, L.D. Soares, N.M.M. Rodrigues, C.L. Pagliari, E.A.B. da Silva, S.M.M. de Faria, Locally Linear Embedding-Based Prediction for 3D Holoscopic Image Coding using HEVC, in: 2014 Proc. 22nd Eur. Signal Process. Conf., Lisbon, Portugal, 2014: pp. 11–15.

- [22] D. Liu, P. An, R. Ma, C. Yang, L. Shen, 3D Holographic Image Coding Scheme Using HEVC with Gaussian Process Regression, *Signal Process. Image Commun.* 47 (2016) 438–451. doi:10.1016/j.image.2016.08.004.
- [23] G.J. Sullivan, J.-R. Ohm, W.-J. Han, T. Wiegand, Overview of the High Efficiency Video Coding (HEVC) Standard, *IEEE Trans. Circuits Syst. Video Technol.* 22 (2012) 1649–1668.
- [24] K. Ugur, J. Lainema, Updated Results on HEVC Still Picture Coding Performance, JCTVC-M0041, Incheon, South Korea, 2013.
- [25] P. Hanhart, M. Řeřábek, P. Korshunov, T. Ebrahimi, Subjective Evaluation of HEVC Intra Coding for Still Image Compression, JCTVC-L0380, Geneva, Switzerland, 2013.
- [26] Y. Li, Coding of Three-Dimensional Video Content, Ph.D Thesis, Mid Sweden University, 2015.
- [27] G. Alves, F. Pereira, E.A.B. da Silva, Light Field Imaging Coding: Performance Assessment Methodology and Standards Benchmarking, in: 2016 IEEE Int. Conf. Multimed. Expo Work., Seattle, WA, US, 2016: pp. 1–6. doi:10.1109/ICMEW.2016.7574774.
- [28] A. Vetro, T. Wiegand, G.J. Sullivan, Overview of the Stereo and Multiview Video Coding Extensions of the H.264/MPEG-4 AVC Standard, *Proc. IEEE.* 99 (2011) 626–642. doi:10.1109/JPROC.2010.2098830.
- [29] R. Olsson, Synthesis, Coding, and Evaluation of 3D Images Based on Integral Imaging, Ph.D Thesis, Mid Sweden University, 2008.
- [30] S.-L. Yu, C. Chrysafis, New Intra Prediction Using Intra-Macroblock Motion Compensation, JVT-C151, Fairfax, VA, US, 2002.
- [31] J. Lainema, F. Bossen, W. Han, J. Min, K. Ugur, Intra Coding of the HEVC Standard, *IEEE Trans. Circuits Syst. Video Technol.* PP (2012) 1. doi:10.1109/TCSVT.2012.2221525.
- [32] D.-K. Kwon, M. Budagavi, Fast Intra Block Copy (IntraBC) Search for HEVC Screen Content Coding, in: 2014 IEEE Int. Symp. Circuits Syst., Melbourne, Australia, 2014: pp. 9–12. doi:10.1109/ISCAS.2014.6865052.
- [33] M. Flierl, B. Girod, Video Coding with Superimposed Motion-Compensated Signals - Applications to H.264 and Beyond, 1st ed., Springer US, 2004. doi:10.1007/b105390.
- [34] M. Flierl, T. Wiegand, B. Girod, A Locally Optimal Design Algorithm for Block-Based Multi-Hypothesis Motion-Compensated Prediction, in: Proc. DCC '98 Data Compression Conf., Snowbird, UT, US, 1998: pp. 239–248. doi:10.1109/DCC.1998.672152.
- [35] G.J. Sullivan, T. Wiegand, Rate-Distortion Optimization for Video Compression, *IEEE Signal Process. Mag.* 15 (1998) 74–90. doi:10.1109/79.733497.
- [36] Il-Koo Kim, K. McCann, K. Sugimoto, B. Bross, W.-J. Han, G. Sullivan, High Efficiency Video Coding (HEVC) Test Model 14 (HM14) Encoder Description, JCTVC-P1002, San José, CA, US, 2014.
- [37] V. Sze, M. Budagavi, G.J. Sullivan, eds., High Efficiency Video Coding (HEVC): Algorithms and Architectures, Springer International Publishing, Cham, 2014. doi:10.1007/978-3-319-06895-4.
- [38] B. Girod, Efficiency Analysis of Multihypothesis Motion-Compensated Prediction for Video Coding., *IEEE Trans. Image Process.* 9 (2000) 173–83. doi:10.1109/83.821595.
- [39] P. Ramanathan, B. Girod, Rate-Distortion Analysis of Random Access for Compressed Light Fields, in: 2004 Int. Conf. Image Process. ICIP '04., Nanyang, Singapore, 2004: pp. 2463–2466. doi:10.1109/ICIP.2004.1421601.
- [40] C. Conti, L.D. Soares, P. Nunes, Influence of Self-Similarity on 3D Holographic Video Coding Performance, in: Proc. 18th Brazilian Symp. Multimed. Web - WebMedia '12, São Paulo, Brazil, 2012: pp. 131–134. doi:10.1145/2382636.2382666.
- [41] G. Baasantseren, J.-H. Park, K.-C. Kwon, N. Kim, Viewing Angle Enhanced Integral Imaging Display using Two Elemental Image Masks, *Opt. Express.* 17 (2009) 14405. doi:10.1364/OE.17.014405.
- [42] G. Bjøntegaard, Calculation of Average PSNR Differences between RD Curves, VCEG-M33, Austin, TX, US, 2001.
- [43] M. Wien, High Efficiency Video Coding, Springer Berlin Heidelberg, Berlin, Heidelberg, 2015. doi:10.1007/978-3-662-44276-0.
- [44] 3D Holographic Sequences (Download Link), (2013). <http://3dholographicsequences.4shared.com/> (accessed October 30, 2016).
- [45] T. Georgiev, Todor Georgiev Gallery of Light Field Data, (n.d.). <http://www.tgeorgiev.net/Gallery/> (accessed September 17, 2016).
- [46] M. Řeřábek, T. Ebrahimi, New Light Field Image Dataset, in: 8th Int. Conf. Qual. Multimed. Exp., Lisbon, Portugal, 2016.
- [47] T. Georgiev, A. Lumsdaine, Focused Plenoptic Camera and Rendering, *J. Electron. Imaging.* 19 (2010) 021106–021106. doi:10.1117/1.3442712.
- [48] R. Joshi, J. Xu, R. Cohen, S. Liu, Z. Ma, Y. Ye, Screen Content Coding Test Model 1 (SCM 1), JCTVC-Q1014, Valencia, Spain, 2014.
- [49] D.G.D.G. Dansereau, O. Pizarro, S.B.S.B. Williams, Decoding, Calibration and Rectification for Lenselet-Based Plenoptic Cameras, in: 2013 IEEE Conf. Comput. Vis. Pattern Recognit., Portland, OR, US, 2013: pp. 1027–1034. doi:10.1109/CVPR.2013.137.