

HMLFC: Hierarchical Motion-Compensated Light Field Compression for Interactive Rendering

S. Pratapa¹ and D. Manocha²

¹University of North Carolina, Chapel Hill.

²University of Maryland, College Park.

Abstract

We present a new motion-compensated hierarchical compression scheme (HMLFC) for encoding light field images (LFI) that is suitable for interactive rendering. Our method combines two different approaches, motion compensation schemes and hierarchical compression methods, to exploit redundancies in LFI. The motion compensation schemes capture the redundancies in local regions of the LFI efficiently (local coherence) and the hierarchical schemes capture the redundancies present across the entire LFI (global coherence). Our hybrid approach combines the two schemes effectively capturing both local as well as global coherence to improve the overall compression rate. We compute a tree from LFI using a hierarchical scheme and use phase shifted motion compensation techniques at each level of the hierarchy. Our representation provides random access to the pixel values of the light field, which makes it suitable for interactive rendering applications using a small run-time memory footprint. Our approach is GPU friendly and allows parallel decoding of LF pixel values. We highlight the performance on the two-plane parameterized light fields and obtain a compression ratio of 30–800× with a PSNR of 40–45 dB. Overall, we observe a ~2–5× improvement in compression rates using HMLFC over prior light field compression schemes that provide random access capability. In practice, our algorithm can render new views of resolution 512 × 512 on an NVIDIA GTX-980 at ~200 fps.

CCS Concepts

• **Computing methodologies** → **Image-based rendering; Image compression; Graphics processors; Graphics file formats; Virtual reality;**

1. Introduction

Virtual reality (VR) is being increasingly used for immersive multimedia experiences and telepresence applications. To achieve a high degree of presence in VR, we need to generate high-fidelity renderings of real world scenes at interactive rates. Photo-realistic renderings increase the sense of immersion in real world scenes and provide artistic, life-like experiences in VR[†]. The plenoptic function (7D) describes the total flow of light through all the points in space [AB91]. Light Fields (LF) are a low-dimensional (4D or 5D) function of the plenoptic function that capture the radiance of the light rays over a specific region of space. Yu [Yu17] outlines the emergence of light fields and lists the advantages of using LF technology to generate high-quality content for VR applications.

Levoy & Hanrahan [LH96] and Gortler et al. [GGSC96] describe a 4D parameterized LF and practical approaches for capturing and rendering static scenes using 2D image samples. To generate photo-

realistic renderings from different viewpoints, such image-based-rendering (IBR) techniques need large amounts of data to be captured, which is a major issue for interactive applications. The number of image samples required for a good quality rendering using LF is generally in the order of tens of thousands [CTCS00]. The data sizes of the sampled LF vary from hundreds of MB [LH96] to hundreds of GB [LS00, LPC*00] depending on the scene complexity, sampling rate, and sampling resolution. For 360° panoramic light fields [OEED18] the LF data-sizes are close to 4–6 GBs. Therefore, compressing the LF is necessary for storing, transmitting, and interactive rendering.

The LF-based rendering algorithms involve retrieving pixel values from LFI and interpolating the pixel values to compute a new view. The pixels required for computing the new view may be located in different regions of different LFI. Therefore, for real-time LF rendering, the relevant portions of uncompressed LFI should be present in the local memory. To render a new view, only a portion of data is required from the entire LFI. As a result, we do not need the entire uncompressed LFI in memory, as it may result in memory bottlenecks. During rendering pixel data is continuously fetched from memory, and in real-time systems, with limited bandwidth

[†] MIT Technology Review: VR is still a novelty, but Googles light-field technology could make it serious art. <https://goo.gl/F79udn>

(mobile and untethered AR/VR), can cause a significant performance bottleneck [Fen03]. Random access compression schemes help in mitigating the memory and bandwidth bottlenecks. Random access compression schemes of LFI have two main properties: (1) selective decoding of only the required data; (2) allowing fast hardware decompression. To enable interactive LF rendering applications, it is necessary to develop LFI compression schemes that maintain the properties of random access compression schemes as well as provide good compression rates.

Prior LFI compression schemes can be broadly categorized into hierarchical schemes and motion compensated schemes. Hierarchical compression approaches for LFI compute a tree (parent, child dependencies) from the LFI using image transformations and create levels of hierarchy [PS01, PM19]. Motion compensation methods capture redundancies in nearby LFI by using a pair of motion vectors or disparity values [ZL00, OEED18]. The hierarchical schemes and motion compensation schemes exhibit different characteristics in terms of capturing the redundancies across the LFI.

Main Results: We present a new motion compensated hierarchical compression scheme (HMLFC) for encoding LFI for interactive rendering. Ours is a hybrid method that combines two different random access compression approaches to maximize the redundancies captured across the LFI. The first class of methods is motion compensation schemes in which the redundancies present in the small regions of the LFI are efficiently captured using extensive search based techniques. The other class of methods is hierarchical compression approaches in which image manipulation and transformation techniques are applied to the entire LFI to capture redundancies across the LFI in a global manner. We use a hierarchical light field compression approach to capture the redundancies in a global fashion and then apply phase shifted motion compensation to various levels of the hierarchy. We apply motion compensation to all the levels of the hierarchy by selecting a set of reference frames at each level, creating a new motion-compensated hierarchy. The tree structure computed in the underlying hierarchical scheme is maintained after applying motion compensation at each level. After motion compensation, the amount of data in the levels of the hierarchy is reduced by a significant factor leading to a higher compression rate. We also present a simple and fast scheme to decompress the light fields and use them for interactive rendering on commodity GPUs. The main contributions of our approach include:

1. A novel compression approach combining two different schemes (motion-compensation and hierarchical schemes) for LFI compression to achieve better compression performance (Section- 3);
2. New phase shifted motion-compensation technique suitable for the properties of the images computed in the hierarchy (Section- 4);
3. A hybrid compression scheme (HMLFC) that provides many benefits including random access, progressive decoding, and parallel decompression on commodity hardware (Section- 4).

Our compression algorithm, HMLFC, provides a $2\text{--}5\times$ improvement in compression rate for similar compression quality compared to prior hierarchical schemes as well as motion compensation schemes that provide random access capability (Section- 5). The decompression memory overhead and decompression time overhead due to our hybrid combination is minimal. We can render new

views at a resolution of 512×512 using an NVIDIA GTX-980 at ~ 200 fps (Section- 5).

2. Prior Work

In this section, we give a brief overview of prior work on light field rendering and compression algorithms.

2.1. Light Field Rendering

The plenoptic function describes the flow of light in space. Adelson & Bergen [AB91] use the term plenoptic function (7D) to describe the light intensity (L) at any point (x, y, z) and orientation (θ, ϕ) in free space at any given time (t), and over a range wavelengths (λ) in the visible spectrum: $L = P(x, y, z, \theta, \phi, t, \lambda)$. Levoy & Hanrahan [LH96] and Gortler et al. [GGSC96] describe a low-dimensional (4-D) form of the plenoptic function, called *light field* as a set of outgoing light rays from a static object or scene. Levoy & Hanrahan [LH96] use two parallel planes described by (u, v) and (s, t) , a two-plane parameterization, to describe the 4-D function. Several other low-dimensional parameterizations such as the spherical [IPL97, OEED18] or the unstructured [DLD12] have been proposed to describe and capture the plenoptic function. In all the parameterizations, the light rays are captured by densely sampling 2-D camera images from multiple viewpoints around the scene. New views from arbitrary positions in space are generated by interpolating the captured light rays (pixel values). High sampling rates (orders of tens of thousands) are required to achieve photo-realistic reconstructions, which need huge amounts of data to store the captured images [CTCS00].

2.2. Light Field Compression

A large amount of image data is needed for LF rendering and it creates a bottleneck for interactive applications. LF compression schemes are used to transmit and store LFI for rendering. JPEG Pleno [EFPS16] has been launched by the JPEG standards committee with the goal of establishing standards for the broader adaptability of 4D LF applications. Several compression schemes have been proposed to handle the image data problem in LF rendering. We categorize the existing schemes into two types: *hierarchical compression schemes*, which apply image transformations and image manipulations (wavelet, image warping, and arithmetic manipulations) to the original LFI and build hierarchical structures that exploit redundancies; *motion compensated compression schemes*, which use standard techniques similar to MPEG video compression (motion vectors) or disparity compensation to capture redundancies. In addition to these two categories, Levoy & Hanrahan [LH96] use vector quantization (VQ) to compress the LFI using a 4D dictionary. The compression rates attained using VQ are around 10:1 to 20:1. A survey of compression schemes for LFI is presented in Viola et al. [VRE17]. More recently, deep learning methods to synthesize new light field views (dense set) from a sparsely sampled set of images are presented [ZLX*14, SRASC14]. Using [ZLX*14, SRASC14], a dense sampling of light field images can be generated from a sparse set, enabling high-quality reconstruction. Sparse sampling of light fields can be viewed as a form of compression because the total number of views required

for good reconstruction quality is reduced, decreasing the total size of the data captured.

2.2.1. Motion-Compensated Compression Schemes:

We further categorize motion compensated schemes into two sub-categories: *high efficiency schemes*, which provide very high compression ratios without random access properties and *random access schemes*, which enable interactive rendering by allowing random access to the pixel values of the LFI.

High-efficiency schemes: The primary approaches in LFI compression adopt methods similar to image and video compression methods. These approaches apply techniques such as motion-vector compensation, domain transform (DCT, wavelet), and image-warping to exploit redundancies among the LFI. In the case of two-plane 4D parameterized LF, the light rays are sampled using uniform camera motion between adjacent samples. Using this observation, Girod et al. [GCRZ03]; Jagmohan et al. [JSA03]; and Magnor & Girod [MG00] describe methods that use a single disparity value instead of a pair of motion-vector values to encode the LFI predictively. The compression ratios achieved are close to 100:1 to 200:1. Moreover, these methods do not provide random access capabilities. Very high compression efficiency schemes that provide compression ratios of 100:1 to 1000:1 have been proposed [CHC18, LWL*16, PA16]. In Liu et al. [LWL*16], the grid of LFI is first processed to arrange them in sequential order to get an optimal pseudo-temporal ordering that maximizes when compressed using HEVC encoding []. Chen et al. [CHC18] process the LFI using predictive and image-warping methods from which from a small set of key-views are selected and the rest of the LFI are predicted using the key-views. After the pre-processing, the images are temporally ordered using the method in Liu et al. [LWL*16] and then compressed using HEVC. Techniques that use additional information about scene geometry and characteristics in addition to image-warping techniques are presented in Chang et al. [CZRG06]. Image homography is used to warp the LFI onto a fixed set of reference images to find redundancies in Kundu [Kun12], yielding compression ratios of 10:1 to 50:1. Although some of the above methods provide large compression ratios, they fail to address the problems of random access and heavy memory consumption for interactive rendering. These methods are efficient for transmitting and streaming LF data over the internet but require the entire LF to be decoded for rendering.

Random Access schemes: The method in Zhang & Li [ZL00] is the first approach that uses motion compensation and provides random access capabilities for LFI compression. They describe a multi-reference, frame-based motion compensation approach that provides compression ratios of 80:1. Overbeck et al. [OEED18] present a scheme for compressing 360° panoramic light fields captured using their LF capturing system. They achieve compression ratios of 40:1 to 200:1 on the complex panoramic LFI datasets.

2.2.2. Hierarchical Compression Schemes:

Peter & Straßer [PS01] present a 4D wavelet hierarchical scheme for compressing LF that provides random access. This method

uses 4D Haar wavelets to transform the LFI into wavelet coefficients and organizes the coefficients into a tree structure. They attain compression rates of 20:1 to 40:1 and their method makes assumptions about the scene captured in the light field. Pratapa & Manocha [PM19] present a hierarchical compression scheme that is based on computing representative and residual views at each level of the hierarchy to exploit redundancies across the LFI. The top-level images of the hierarchical tree capture the redundant common details among the LFI and the other levels of the tree store the low-level, high-frequency details of the LFI. Their method obtains compression rates of 20:1 to 200:1, provides random access to the compressed stream, enables progressive decompression, and supports fast hardware decoding. Magnor & Girod [MG99] describe a hierarchical predictive-based encoding scheme using disparity maps. In Magnor & Girod [MG99], an explicit hierarchical tree is not constructed, but a hierarchical relationship among the LFI is established by iteratively dividing the LFI into sub-quadrants. This method provides compression rates of around 400:1, but it does not provide random access capability for interactive rendering.

3. Motion Compensation & Hierarchical Compression: Coherence

In this section, we present high-level descriptions of motion compensation and hierarchical compression methods for LFI that provide random access. We discuss the advantages and limitations of each of these approaches and motivate the design of our hybrid compression scheme. In the following discussion, we assume 4D two-plane parameterization, though our approach can be extended to other LF parameterizations.

For a set of light field images captured using two-plane parameterization, redundancies are present across all the captured LFI. The amount of coherence between two captured light field images varies based on the distance between the actual image capture points in the space. Adjacent light field images exhibit higher coherence, while far off images exhibit lesser coherence. We refer to the coherencies that are commonly present across the entire LFI as *global coherencies*. We refer to the coherencies present across the adjacent images of the LFI as *local coherencies*. An example of LFI highlighting local and global coherencies is highlighted in the suppl. material[‡] (Section 3).

3.1. Motion Compensation Methods

We use the following terminology to give an overview of prior motion compensation schemes used for LF compression.

Reference Images: The set of images selected from the original LFI that are encoded independently, using standard compression schemes and used as the reference set for encoding the rest of the images in the LFI in the motion compensation schemes.

Predictive Images: The set of images from the original LFI that are

[‡] Supplementary material is available at: <https://bit.ly/2KphqEY>

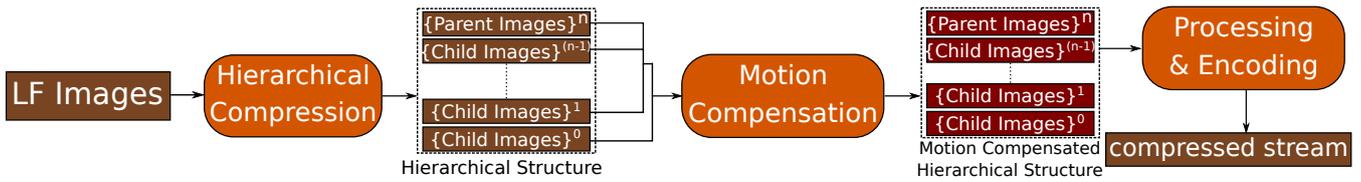


Figure 1: Overview of our HMLFC compression pipeline: The compression pipeline consists of different stages. In the first stage, a hierarchical compression scheme is applied to LFIs to compute levels of new, transformed parent and child images. In the next step, all the levels of the computed hierarchy are processed using a motion compensation scheme to compute a new motion-compensated hierarchy. In the final stage, the motion compensated hierarchy is further processed and encoded to generate the compressed bit stream.

associated with a reference image and are encoded using motion compensation techniques.

At a high level, motion compensation schemes start by selecting a subset of frames from the LFI as *reference images*. The process of selecting the *reference images* varies depending on the exact compression scheme, as discussed in Section 2. Once the *reference images* are selected, the rest of the LFI are marked as *predictive images*. Each of the *predictive images* is associated with a *reference image* and encoded using motion compensation. The predictive images are divided into non-overlapping rectangular blocks, and each block of pixels is predicted (computed as difference) from the reference image using a pair of motion vectors. Motion compensation schemes typically use an exhaustive search over a large region in the reference image to minimize the residual difference for each block in the predictive images. Therefore, they compute the redundancies between a given reference image and the associated predictive images very efficiently using an exhaustive search.

The LFI exhibit a large amount of coherence across all the LFI captured. Motion compensation schemes capture *local coherencies* effectively using exhaustive search. Although the motion compensation schemes exploit *local coherencies* (reference and predictive image sets) of the LFI efficiently using an exhaustive search, they fail to capture the redundancies present across the entire LFI in a global fashion (e.g., coherencies across grids in Fig 2).

3.2. Hierarchical Methods

We use the following terms to present an overview of the hierarchical approach:

Parent Images & Child Images: The new sets of transformed images computed using image manipulations and transformations from the original LFI in hierarchical compression schemes. The new sets of transformed images are parent image sets and children image sets forming hierarchical relationships between the sets, creating a hierarchy.

Hierarchical schemes use image manipulation and transformation techniques on the LFI to compute a new set of images capturing the redundancies across the entire LFI. The new set of transformed images is partitioned into two subsets, parent images and child images, creating a hierarchy. The image manipulation and transformations (wavelet transforms, image warping, image filtering, and arithmetic manipulations) used to compute the new set of

images and the exact parent-child relationships depend on the particular compression scheme. Typically, the parent images capture the common redundant details across the LFI and the children contain image specific low-level details of the LFI. The parent subset is further processed recursively to compute the next level of the hierarchy.

The primary advantage of hierarchical methods is that they capture the *global coherencies* across distant images of the LFI, which the motion compensation schemes fail to capture. Due to the lack of an exhaustive search for redundancies, the global redundancies that are encapsulated in the parent images are limited by the image transformation and manipulation techniques employed in the compression scheme. Figure 2 (b) shows a high-level overview of a hierarchical LFI compression scheme.

3.3. Challenges in the Hybrid Approach

Our goal is to develop a hybrid scheme that captures the benefits of motion compensation schemes in terms of local coherency and hierarchical schemes in terms of global coherency. We design a hybrid approach that is based on applying an additional layer of motion compensation to a hierarchical representation. To obtain good compression rates and provide random access capabilities, we need to address these issues:

1. Once the redundancies across the LFI are captured globally, the properties of the parent and children images computed using a hierarchical scheme differ significantly from the properties of typical images or original LFI. To effectively capture the local coherency, we need new motion compensation schemes to account for the properties of the transformed images in the hierarchy for achieving further compression.
2. The resulting motion compensation scheme should conserve all the properties and benefits of the underlying hierarchical scheme, including the hierarchy structure and random access capability.
3. The overhead of the additional costs of decompression after an additional layer of motion compensation should be minimal.

4. Our Method: HMLFC

In this section, we describe our novel hybrid compression algorithm that captures local and global coherency and addresses the challenges highlighted above.

4.1. Overview

To tackle the limitations of the *motion compensation methods* and *hierarchical methods*, we combine both approaches to capture redundancies in both *global* and *local* fashion, resulting in better compression rates. In other words, we first apply a hierarchical approach to the LFI gathering all the *global coherencies*, then apply an additional layer of compression to the images at each level to capture remaining redundancies using motion compensated search.

4.2. Exploiting Local and Global Coherence

Our approach (HMLFC) uses a hierarchical motion compensation scheme to capture the redundancies present across the entire set of LFI in a global fashion (*global coherence*). Next, we treat the parent images and each of the children subsets at all levels computed from a hierarchical scheme as separate subsets of images. Our goal is to apply motion compensation methods to each of the subsets independently and design a new scheme that exploits the properties of these subsets. The application of motion compensation on the children and parent subsets further exploits the local redundancies (*local coherence*) efficiently by using the exhaustive search that the hierarchical methods fail to exploit.

The decoding properties (random access, progressive decoding, and hardware decoding) of our hybrid approach depend mainly on the underlying hierarchical scheme used for computing the hierarchy and the motion compensation scheme. In the following section we present the overview of the underlying hierarchical scheme and the details of the compression and decompression of our hybrid approach.

4.3. RLFC: Hierarchical Compression Scheme

We choose the RLFC hierarchical compression algorithm described in Pratapa & Manocha [PM19] because it provides random access to the compressed data and allows hardware decompression. In addition to RLFC, we use a novel motion compensation scheme on the levels of the hierarchy. We maintain the random access property of RLFC after this motion compensation step.

In RLFC, the LFI are clustered based on the spatial locations of the samples. For each of the clusters, a new image referred to as the *representative key view (RKV)* is computed by filtering all the image samples in the clusters. The RKV images encapsulate common details among all the images in a given cluster, and the set of RKVs from all the clusters forms a new level (parent images) in the hierarchy. After computing the new level of RKVs, the differences between the RKV and the images in the corresponding cluster are computed. The difference images are referred to as *sparse residual views (SRV)* and the new set of SRVs are the child images in the hierarchy. The SRVs are high-frequency images that contain the specific low-level details of the images that are not captured in the RKVs. This process is recursively implemented on the new RKVs until the tree height reaches a user-set level. We refer the readers to Pratapa & Manocha [PM19] for exact details of the hierarchy and tree structure computed in RLFC.

Notation: We use the following notation for explaining the approaches: R_i denotes the i^{th} reference image in the motion compensation methods; P_i^k denotes the k^{th} predictive image associated

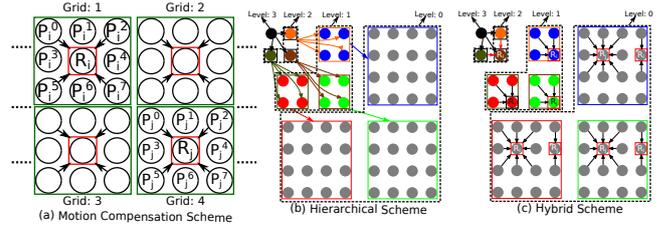


Figure 2: An example overview of our hybrid compression scheme. (a) *Motion Compensation:* The local coherencies are effectively captured in small regions (grids marked in green) of the LFI using an exhaustive search. A set of images from LFI is selected as the reference images, highlighted in red (R_i, R_j). The rest of the images are marked as predictive images and are compressed from the reference images using motion compensation. (b) The coherencies present across the entire LF are captured using image manipulation and transformations applied to the LFI in a global fashion. The levels of the hierarchy are marked and arrows indicate parent-child relationships in the hierarchy. (c) *Our Hybrid Approach:* Once the global coherencies are encapsulated using the hierarchical scheme, the additional redundancies at each level of the hierarchy are captured as local coherencies by applying motion compensation at every level. At each level of the hierarchy, reference images (R_i, R_j) are indicated using a bounding box. The predictive images associated with the reference images are indicated with arrows pointing towards the reference images.

with R_i in the motion compensated methods; (x, y) denotes the motion vector pair in the motion compensation schemes; $B_{P_i^k}$ denotes a block of pixels in the k^{th} predictive image; $B_{R_i}^{xy}$ denotes a block of pixels for motion vectors (x, y) in the reference image R_i ; Δ represents the prediction residual error computed between the reference block and the predictive block.

4.4. Phase-shifted Motion Compensation

As shown in Fig. 3 (left), the SRV images exhibit significant local coherency at each level of the hierarchy. The SRV images are computed as the difference between RKV images at a given level and images in the level below. Due to the difference computation, the pixels in the SRVs have both negative and positive intensity values. The negative and positive pixel intensity values correspond to the inversions of pixel intensity values across the SRV image signals at a given level. We refer to these inversions as *phase-shifts* in the SRV image signals. We present a new phase-shifted motion compensation to capture *local coherencies* in the levels of the hierarchy. These phase shifts in the SRV image signals need to be accounted while applying motion compensation to the SRV images. More details about the phase shifts that occur in the SRV images are presented in the suppl. material, Sec-1.

For a selected SRV reference image R_i at any given level, let $\{P_i^0, P_i^1, \dots, P_i^n\}$ denote the set of predictive SRV images associated with R_i . Each block ($B_{P_i^k}$) in the predictive P_i^k is motion compensated by searching over a large search window W in the reference SRV image R_i using a pair of motion vectors (x, y) . We include the phase shifts in our motion prediction scheme by computing two

residual errors for each block (number of pixels in a block: N) as follows:

$$\Delta_-^{xy} = \sum_{l=1}^N |B_{P_i^k}(l) - B_{R_i}^{xy}(l)|, \quad (1)$$

$$\Delta_+^{xy} = \sum_{l=1}^N |B_{P_i^k}(l) + B_{R_i}^{xy}(l)|. \quad (2)$$

For a given pair of motion vectors (x, y) , we compute a subtractive prediction residual error Δ_-^{xy} and an additive prediction residual error Δ_+^{xy} to include possible phase shifts between the reference image and the predictive images in a given region.

$$\Delta_- = \min_{(x^-, y^-)} (\Delta_-^{xy}) \quad \forall x, y \in [-W, W],$$

$$\Delta_+ = \min_{(x^+, y^+)} (\Delta_+^{xy}) \quad \forall x, y \in [-W, W].$$

The minimum subtractive prediction residual error Δ_- and the minimum additive prediction residual error Δ_+ are computed for each block ($B_{P_i^k}$). (x^-, y^-) and (x^+, y^+) are the motion vectors corresponding to the minimum prediction residuals. The final motion prediction residual error Δ and the corresponding motion vectors are computed as follows:

$$\Delta = \min(\Delta_-, \Delta_+),$$

$$(x, y) = \begin{cases} (x^-, y^-) & \text{if } \Delta = \Delta_-, \\ (x^+, y^+) & \text{if } \Delta = \Delta_+. \end{cases}$$

Next, we perform a replacement step in which the original pixel values in the block $B_{P_i^k}$ in the predictive SRV image (P_i^k) are replaced with predictive residuals of the block. The replacement step modifies the SRV images in the original RLFC tree and computes a new HMLFC tree, but the tree structure and the hierarchy remain exactly the same as the original RLFC tree. Figure 3 shows the predictive residual SRV images after applying our phase inclusive motion compensation. The predictive residual SRV images computed after motion compensation are much sparser than the original SRV images. Therefore, the predictive residual SRV images can be compressed more significantly without quality loss, resulting in better compression rates. A zoomed-in (16X) visual comparison between an original SRV image and the corresponding predictive residual SRV image after motion compensation is shown in suppl. material, Sec-5. A quantitative comparison between the sparsity of the predictive residual SRV images and corresponding original SRV images is shown in Table 3.

4.5. Compressing HMLFC tree

YCoCg [MSS08] color space is used in our implementation to decorrelate the RGB color channels and the chroma channels are sub-sampled. The dynamic range (number of bits to store pixels) of the pixel values is adjusted accordingly to avoid any loss of information due to transformations. The hierarchy computation and motion compensation are performed separately on all the channels. After the motion compensation, the top-level RKV images of the hierarchy are similar to the standard images (RGB) and are compressed using JPEG2000 in the lossless mode.



Figure 3: (left) An SRV image cluster from the RLFC hierarchy from level: 0. It is evident that SRV images are visually similar to each other and exhibit a lot of coherency between them. We exploit the redundancy by applying motion compensation to achieve further compression. (right) The SRV image cluster is shown at left after applying motion compensation. The reference image is highlighted in red, and the rest of the images are residual difference SRV images after motion compensation. Compared to the original SRV images, the residual difference SRV images are sparser leading to significantly better compression rates.

The compression rate and compression quality of the scheme are controlled by encoding parameters set as user-input to the compression method. The main encoding parameters in the RLFC scheme are *tree height*, *block size*, and *block thresholds*. In addition to the three encoding parameters another important encoding parameter is *search window size* used in the phase-shifted motion compensation to perform the exhaustive search. The SRV images at all levels of the hierarchy are divided into *block size* non-overlapping rectangular blocks and motion compensation is applied to all the blocks as described in the previous section.

After applying the motion compensation to the hierarchy and computing the HMLFC tree, SRV images (both residual difference SRVs and reference SRVs) are thresholded to discard insignificant data based on the two *block thresholds* set as encoding parameters. For each block in the SRV images, an energy value is computed by summing up the absolute values of the pixels in the block. If the energy is less than the user set threshold the block is marked as insignificant and not stored in the final compression. Due to motion compensation, any losses introduced in the reference SRVs due to thresholding gets propagated to the associated predictive SRVs. To avoid that we use two independent *block thresholds* for thresholding the motion compensated residual SRVs and reference SRVs.

4.6. Bounded Integer Sequence Encoding

In the construction of the hierarchy, we need to perform lossless integer computations, and the final pixel values in all the images of the hierarchy are integer values. BISE [NLP*12] presents an efficient way of encoding a sequence of integer values within a fixed range $[0, N - 1]$ and allows for fast random access decoding in constant time with minimal hardware. The straightforward solution for allowing fast hardware random access to the sequence of integers as bit strings of their corresponding binary representations. However, this solution is only optimal when N is a power of two because it uses $\log_2 N$ bits to store the integer values equivalent to the

information present in each integer value. Besides the simple case when N is a power of two BISE provides an efficient encoding that is close to the information theoretic bounds for other ranges of N . The significant blocks in the SRV images after thresholding in the HMLFC tree are encoded using BISE. The resulting formulation is easily supported by the hardware and provides lossless computations.

4.7. Compressed Stream Structure

We further process and compress the HMLFC tree and the additional motion vector values computed from the motion compensation step. The HMLFC tree is linearized using breadth-first search (BFS) traversal indexing all the SRVs in the traversal order starting from the top of the tree. The BISE compressed blocks of the SRV images in the tree are arranged in the same BFS linearized order and appended to the compressed stream. To maintain fast random access property, we extend the application of BISE to encode the motion vector values. Compressing motion vectors using BISE also preserves the fast hardware decompressible property of our stream.

Algorithm 1 Decompress light field image block

Input:

LFI compressed stream: **CompLFI**
 Image index: **ImgIdx**
 Block index: **BlkIdx**

Output:

Pixel values: **PixVals**

// Load the stream into memory and separate

Initialization:

$RKVStream \leftarrow ReadRKVStream(CompLFI)$
 $RKV^n \leftarrow DecompressJPEG2000(RKVStream)$
 $BlockOffsets \leftarrow ReadBlockOffsets(CompLFI)$
 $SRVStream \leftarrow ReadSRVStream(CompLFI)$
 // Decode the reference SRV images
 $RefSRVImages \leftarrow DecodeRefSRVImages(RefSRVIdx)$
 $SparseRefSRV \leftarrow SparseMatrixRep(RefSRVImages)$

function DECOMPRESSLFI_BLOCK(ImgIdx, BlkIdx)

// Get the start location of BlkIdx in bitstream
 $StartOffset \leftarrow BlockOffsets[BlkIdx]$
 // Read the top level filtered values from RKV
 $RKVBlock \leftarrow ReadBlock(RKV^n, BlkIdx)$
 // Compute the location of the parent blocks in stream
 $ParentIdx \leftarrow GetParentIndices(ImgIdx)$
 // Read the BISE encoded stream of required blocks
 $OrderedBlockBISE \leftarrow ReadBlocks(SRVStream,)$
 // Decode the BISE blocks
 $Blocks \leftarrow BISEDecode(OrderedBlockBISE)$
 // Motion Compensate the decoded blocks
 $ReCompBlocks \leftarrow ReMotionComp(Block, ImgIdx)$
 // Combine the residual values with filtered pixel values
 $PixVals \leftarrow CombineBlocks(RKVBlock, ReCompBlocks)$

4.8. Decompression

4.8.1. Decoding Procedure:

Decoding a block of pixel values from a particular location in the LFI consists of two main steps: 1. Decoding the blocks from the HMLFC hierarchy using tree traversal; 2. Applying motion re-compensation for the motion compensated blocks. At the start of the decoding procedure, the top-level RKV images are decoded and stored in memory. Next, the reference SRV images at all levels of the hierarchy are decoded. The reference SRV images are necessary for performing the motion re-compensation step. Both the top-level RKV images and the reference SRV images are decoded and stored prior to decompression. To decode a particular block of pixels from the HMLFC tree, we traverse the tree starting at the top and collect the BISE compressed streams required for decoding the block. For a given block location, we find the locations of all parent-child images in the hierarchy that belong to the block using the tree structure constructed during compression. The BISE compressed streams gathered from all the levels are decoded to compute the blocks of the residual pixel values.

After the residual pixels are decoded from the hierarchy, motion re-compensation is required to get the original SRV image pixels. Using the block location and the tree structure, we can determine whether a decoded residual block belongs to a predictive SRV image. If the block belongs to the predictive SRV image, the corresponding pixels from the reference SRV image are read from the memory and a motion re-compensation step is performed. Finally, all the residual pixel values are aggregated with the top-level RKV pixel values to get the required block of pixel values. Algorithm-1 shows the high-level pseudo-code of the decompression scheme.

4.8.2. Decompression Memory Overhead:

The pixels from the reference SRV images are necessary during decoding to perform the additional motion re-compensation step. To avoid the additional time required for decoding the reference SRV pixels, at the start of the decompression, the reference SRV images at all the levels are decoded and loaded into the memory. The SRV images in the hierarchy are highly sparse in terms of the pixel distribution present in the images (Fig. 3). We use a sparse matrix representation [NR12] to store the decompressed reference images in the memory while rendering. The sparse matrix representation reduces the additional run-time memory overhead required for the motion re-compensation step to decode a given block of pixels. Although there is a minor time overhead in reading the pixels from sparse matrices, the overhead is much smaller than the time required for decoding the reference SRV pixels for motion re-compensation. The size of the additional memory overhead depends on two factors: One of them is a user-set encoding parameter such as the number of levels in the hierarchy and number of reference images in each level. The second factor is the sparsity of the SRV images which depends on details of the scene captured in the light field images.

4.8.3. Decompression Time Overhead:

The additional layer of motion-compensated step on top of the hierarchy requires decompression and results in additional decompression overhead. This includes tree traversal decoding operations

LF Dataset (Resolution): Size (MB)	Compression rate (bpp)	PSNR (dB)
Amethyst ($16 \times 16 \times 768 \times 1024$): 576	0.045	40.7
Bracelet ($16 \times 16 \times 1024 \times 640$): 480	0.143	40.1
Bunny ($16 \times 16 \times 1024 \times 1024$): 768	0.027	41
Jelly Beans ($16 \times 16 \times 1024 \times 512$): 384	0.029	40.5
Lego Knights ($16 \times 16 \times 1024 \times 1024$): 768	0.157	41
Lego Gallantry ($16 \times 16 \times 640 \times 1024$): 480	0.155	40.1
Tarot Cards ($16 \times 16 \times 1024 \times 1024$): 768	0.68	40.3

Table 1: The compression computed using our HMLFC algorithm and the quality for several LF datasets from the Stanford light field archive. All the image samples are 24-bit color RGB images. For a similar PSNR quality, the compression rate varies for each LF depending on the details of the scene recorded in the LF.

needed to retrieve a block of pixels. Furthermore, the HMLFC algorithm performs three additional basic operations: 1. Bit manipulations required to decode the corresponding motion vectors; 2. Loading the bytes of data (pixels) from the reference image in memory into the registers; 3. Performing arithmetic operations to compute the motion re-compensated block. In terms of these additional operations required to decode a block of pixels, only the memory load operations are slightly more expensive. In our parallel GPU decoding implementation and experiments (Sec. 5), we noticed this overhead to be minimal.

4.8.4. Random Access for Interactive Rendering:

Random access to the pixel values in the HMLFC tree is guaranteed by the tree traversal decoding operations described (Section 4.8.1). To decode a required pixel value, a block of pixel values corresponding to the required pixel value is decoded. Following that, only the motion vectors corresponding to the predictive residual blocks are retrieved from the BISE compressed motion vector stream. Only a part of the compressed stream is decoded to retrieve required blocks of pixels and the corresponding motion vector values, while the rest of the compressed stream remains intact. Our method also supports parallel decompression of different pixel values from the compressed stream enabling fast GPU decoding. To retrieve a single pixel value of LFI using our decoding, a block of pixel values are decompressed. As a block of pixels are decoded, our method benefits any LF rendering scheme by providing fast access to neighboring pixels for interpolation to compute new views. A set of new views computed for different camera positions and for given LF geometry are shown in suppl. material Sec-8.

4.9. Compression Analysis

We identify two primary properties of the LFI that affect the final compression rate of our method, and we briefly discuss their relationship with the encoding parameters used in our approach. 1. Distance between the captured light field image samples; 2. Details of the scene captured in the light field.

As the distance between the light field samples increases the disparity for a real-world scene point in the pixel space of the adjacent light images also increases. The RKVs are computed as weighted filtering (pixel-wise) of the close light field images; as the disparity gets higher, the correlation between the same pixels decreases. As a consequence, the redundancies captured in the RKVs decrease

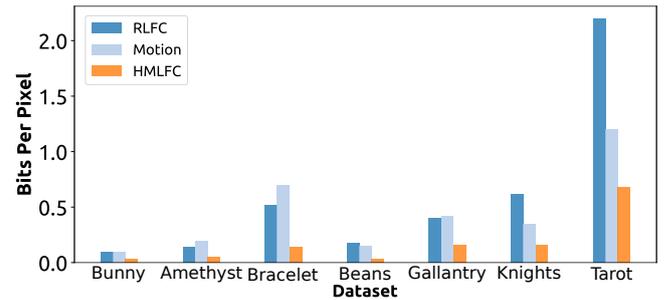


Figure 4: We compare HMLFC with RLFC and motion compensation schemes in terms of compression rates (bpp) for several datasets. The datasets are compressed to have similar compression qualities for each of the methods. Overall HMLFC improves the compression rate by a factor of $\sim 2-5\times$, compared to prior schemes.

leading to a decrease in the sparsity of the SRV images and more additional redundancies across the SRV images in a given level of the hierarchy. As a result, for a fixed search window size, as the sampling distance between LFI increases, the resulting bit rate increases. For a scene with extensive details, the sampled light field images contain a lot of high-frequency components. In this case, even for a small capture distance between light field images due to the vast regions of high-frequency components, the resulting SRV images have low sparse regions with large intensity values. For a given block threshold, as the complexity of the scene increases, the resulting compression rate also increases as the number of significant blocks in the SRV images increases. However, the redundancies in the high-frequency components of the SRV images in a level can be captured using a motion compensated search.

5. Evaluation & Performance Analysis

We present the results from the evaluation of our hybrid approach and analyze its performance on the Stanford LF archives [LH96, WJV*05]. We use peak-signal-to-noise-ratio (PSNR) [OSS*12] for quality comparison (suppl. material, Sec-5) and bits per pixel (bpp) to present the compression rates. We present a comparison of our hybrid method with RLFC [PM19] and a motion compensation scheme that enables random access in terms of compression rates and compression quality. The motion compensation scheme is implemented based on Zhang & Li [ZL00].

In Table 1, the compression rates and PSNR values are shown for different datasets from the Stanford LF archive. For a similar PSNR quality, the compression rates vary from 0.029 bpp to 0.68 bpp due to the variation of the details captured in the scenes of the datasets. The encoding parameters are varied across the datasets to achieve a similar compression quality. The threshold values used for compressing the datasets are provided in the suppl. material, Sec-9. The top-level RKV image is included as a compressed JPEG2000 in all our evaluations.

Figure 4 shows the comparison of compression rates for several datasets for similar compression quality (variation in PSNR quality 0.5 – 1.5 dB) for different methods. In some datasets (Bunny, Amethyst, Jelly Beans) RLFC provides similar or better compression

sion than motion compensation. In other datasets (Lego Knights, Tarot Cards) with complex and high-frequency details, we notice that the motion compensation scheme provides better compression rates. We notice that by combining both approaches, HMLFC achieves better compression in both cases. HMLFC improves the compression rate by $\sim 2 - 5\times$ compared to RLFC for datasets where RLFC provides better compression rates. For other datasets with complex and high-frequency details, HMLFC improves the compression by $\sim 3 - 5\times$ compared to RLFC and improves the compression by $\sim 2 - 3\times$ over motion compensation schemes.

We analyze the rate-distortion properties of HMLFC by varying the following encoding parameters: *block size*, *block threshold*, and *search window size*. Table 2 shows the variation of the compression rate and resulting quality with a change in the *block size*. Increasing the *block size* with a fixed *block threshold* causes a decrease in the thresholding errors, which results in an increase of PSNR and bpp. Figure 5 shows the effect of varying the window size on the compression rates and compression quality. As the search window size increases, the predictive blocks find better matching blocks in the reference images resulting in a sparser predictive residual. The increase in sparsity of the predictive residuals leads to a reduction in the compression rate. Better matching blocks in the reference images lead to better compression quality and an increase in the PSNR. The coherency between the predictive blocks and reference images is limited to only a certain local region and is diminished beyond a certain search window size. As the window size gets larger than a certain range, we notice that the compression rate and compression quality become saturated. If the spatial distance between the sampled light field images is large, we notice a large benefit in terms of compression (Tarot, Bracelet) as the search window size increases. The results of varying the search window size agree with the compression analysis presented in Section-4.8. The effect of varying the threshold on the sparsity of the SRV images is included in the suppl. material, Sec-10.

As estimated in the compression analysis in Section - 4.8 and, as presented in Table 1 as the details of the contents captured in the scene (example images of the dataset are shown in suppl. material, Sec-2) increase we notice an increase in the bit rate. In the new Stanford LF archive, the Tarot Cards scene is captured with two different sampling distances (small and large) between the light field images. The compression rate on the dataset with the larger sampling distance is 0.68 bpp for a PSNR of 40.3 dB; on the dataset with smaller sampling distance it is 0.47 bpp for a PSNR of 40.2 dB.

The variations of compression quality with compression rate for both HMLFC and RLFC are shown in the Figure 6. The rate-distortion for both methods is computed by varying the block threshold. We notice that for different ranges of PSNR, HMLFC achieves better compression compared to RLFC. Visual quality comparison between RLFC and HMLFC is shown in Figure 7. The encoding time using our current single-threaded implementation required for compressing varies from 30–90 minutes depending on the input size of the LFI and resolution of the LFI. More compression evaluations of HMLFC (in comparison with RLFC) on datasets Heidelberg LF benchmark [HJKG16] in suppl. material,

LF Dataset	Metric	Block Size: 2	Block Size: 4	Block Size: 8
Amethyst	PSNR	38.7	43.69	48.35
	bpp	0.0592	0.106	0.707
Bunny	PSNR	40.52	43.35	47.52
	bpp	0.0173	0.0411	0.548
Bracelet	PSNR	37.03	44.15	48.79
	bpp	0.033	0.35	1.108
Knight	PSNR	38.27	43.046	47.89
	bpp	0.096	0.243	1.15
Beans	PSNR	37.3	44.53	49
	bpp	0.0067	0.052	0.377
Gallantry	PSNR	37.22	42.54	46.95
	bpp	0.0235	0.235	1.08
Tarot	PSNR	36.98	42.14	46.79
	bpp	0.272	0.831	2.6

Table 2: The effect of varying the block size on the compression rate and quality is highlighted. Increasing the block size for a fixed block threshold reduces the total number of thresholding errors, resulting in an increase of bit rate and PSNR. The block threshold is set to 75, the search window size is set to 16, and the tree height is set to 3.

Sec-6. Novel views not present in the original LFI computed for new camera viewpoints are presented in suppl. material, Sec-8.

Decompression Analysis: We have implemented a GPU LF rendering (more details in suppl. material, Sec-7) using a basic ray-tracing method to test the implementation of our decompression scheme on an NVIDIA GTX-980. We tested the decompression scheme on the Lego Knights dataset compressed using the following encoding parameters: *block size 4*, *tree height 3*, *search window size 16*. Our method takes 3 – 8 milliseconds to generate frames at resolution 512×512 , depending on the number of blocks decoded per frame. The resulting average frame rate for rendering new views is ~ 200 fps. The average frame rate to render new views at resolution 1024×1024 is close to ~ 110 fps. Although HMLFC involves few additional steps in decoding a block of pixels, it achieves similar frame rates as RLFC using our current implementation. We speculate that the decompression of RLFC is bottle-necked on the number of memory operations required to perform the decoding of a block. The inclusion of a few additional memory operations to perform the extra step of motion re-compensation for decoding HMLFC is negligible on the overall rendering performance. The decompression memory overhead to store the sparse matrix representation of the reference images while rendering is ~ 800 KB using the Lego Knights dataset.

To quantitatively test the run-times of HMLFC with RLFC, similar to the GPU decoder we have implemented a single threaded CPU-based decoder and LF renderer for generating new views. We compute an average of all the times required for decoding all the blocks of pixel values while generating over several hundreds of different new views. The average time to decompress a block of pixels using RLFC are: 1. Y-Channel: 2.61 microseconds; 2. Co-Channel: 1.62 microseconds; 3. Cg-Channel: 1.42 mi-

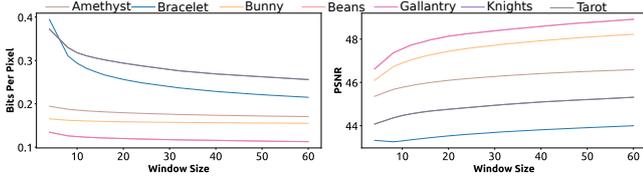


Figure 5: We highlight the variation in the compression rates and compression quality of HMLFC with the change in the window size. (left) The increase in the search window size leads to better matching blocks resulting in smaller prediction residual errors and better compression rates. (right) The prediction residual errors are reduced with an increase in the search window size and the resulting compression quality increases. The block size is set to 4 and tree height is set to 3. The block threshold is varied across different datasets to keep the PSNR within a certain range.

croseconds. The average times required to decompress a block of pixel values using HMLFC are: 1. Y-Channel: 3.32 microseconds; 2. Co-Channel: 2.86 microseconds; 3. Cg-Channel: 2.21 microseconds. The effect of including additional memory reads and motion re-compensation steps required for decoding HMLFC can be seen in the resulting average times. As shown, the additional time required by HMLFC for decoding on a single thread on a CPU is not significantly high.

6. Conclusions, Limitations & Future Work

Conclusions: We present a novel hybrid compression scheme that combines two prior compression methods, *hierarchical schemes* and *motion compensation schemes*, to encode LFI. Our approach captures the local and global coherencies in the LFI and improves the compression rate by a factor of $\sim 2 - 5\times$ without any significant loss in the compression quality. Our scheme provides random access capability and can be used for interactive rendering on current GPUs. We have highlighted its benefits on standard benchmarks and observe compression rates of $30 - 800\times$ with a PSNR of $40 - 45$ dB.

Limitations: Our approach has some limitations. The primary limitation of the hybrid approach is in designing a suitable motion compensation scheme for the transformed images in the hierarchy. Without proper motion compensation suitable for the underlying hierarchy the benefits, from the hybrid combination might be limited. Another limitation of our method as pointed out in the results (Sec - 5) is that the compression rate is dependent on the distance between the light field images in the light field samples. In the case of light fields captured with a sparse sampling rate, the performance of our compression scheme is reduced. The current GPU decoder for our compression scheme is not optimized in terms of the memory operations required for decoding.

Future Work: In the current implementation, we use per-pel motion compensation, i.e., a search for a matching block is performed at a pixel level. Using sub-pixel motion compensation, i.e., sub-pixel level motion compensation to search for a matching block using bi-linear interpolation methods could provide better compression rates. We have implemented the hybrid approach for one spe-

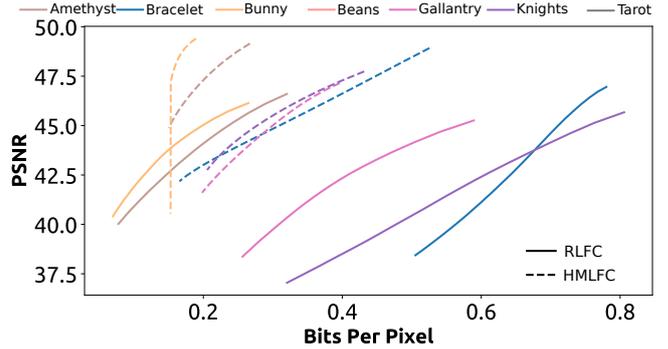


Figure 6: The variation of compression quality with bit rate is highlighted for HMLFC and RLFC. HMLFC provides better compression rates for all the datasets over a range of PSNR values. The block size is set to 4, tree height is set to 3, and the search window size is set to 16.

LF Dataset	RLFC (%SRV sparsity)	HMLFC (%SRV sparsity)
Amethyst	20.4 %	59.7 %
Bracelet	17.2 %	58.2 %
Bunny	18.2 %	56.2 %
Beans	19.3 %	75.4 %
Knights	10.0 %	45.1 %
Gallantry	14.3 %	47.9 %
Tarot	2.92 %	18.8 %

Table 3: The average sparsity of the predictive SRV images before motion compensation (RLFC) and after motion compensation (HMLFC) for the LF datasets. The sparsity is measured as the ratio of the total number of zero-valued pixels to the total number of pixels in the SRV images. The sparsity is increased by a factor for all datasets after applying motion compensation, which leads to more compression.

cific hierarchical scheme (RLFC), and we would like to extend and test our hybrid approach for other hierarchical compression schemes that allow random access (e.g., Peter & Straßer [PS01]). Adding depth information to the light field pixel data improves the rendering quality by a significant factor and provides more parallax. Extending our approach to compress depth information alongside image data is also a good direction for future work. Also, using motion compensation vectors for parallax correction to reduce artifacts during LF rendering may be possible. Our current implementation focuses on 4D two-plane parameterization of the light fields; in the future, we would like to extend our compression approach to more complex parameterizations such as spherical [IPL97], panoramic [OEED18], and unstructured LF [DLD12]. Our current implementation for encoding LFI is single threaded and slow. The encoding speed can be improved by a factor with a multithread and parallelized implementation on the CPU or the GPU. Our current GPU decoder implementation can be optimized to execute efficient parallel memory accesses and requests while rendering further improving the performance of our approach. Exploring the application of the current ideas to dynamic and animated light field videos is a great direction for future work.

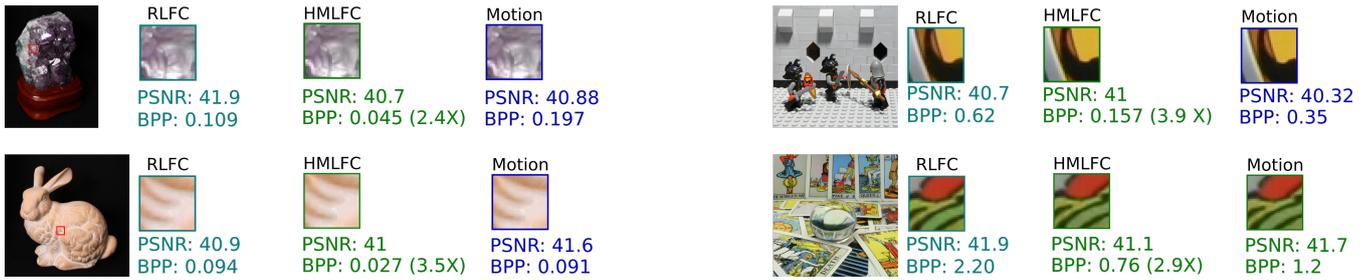


Figure 7: We show a zoomed in comparison between decoded images from the different schemes. A small region of size 32×32 marked in red box is selected and scaled upto 512×512 for visual quality comparison. The PSNR and bpp values for each of the methods are mentioned in the figure. For same PSNR values We find no additional visual degradation in HMLFC compared to RLFC and motion compensation. The factor of improvement of HMLFC over RLFC is highlighted in the bracket.

7. Acknowledgements

This research is supported in part by Intel. We would like to thank Tom Olson, Alex Chalfin, and Andreas Selvik for their inputs and valuable discussions during this research.

References

- [AB91] ADELSON E. H., BERGEN J. R.: The plenoptic function and the elements of early vision. In *Computational Models of Visual Processing* (1991), MIT Press, pp. 3–20. 1, 2
- [CHC18] CHEN J., HOU J., CHAU L.-P.: Light field compression with disparity-guided sparse coding based on structural key views. *IEEE Transactions on Image Processing* 27, 1 (2018), 314–324. 3
- [CTCS00] CHAI J.-X., TONG X., CHAN S.-C., SHUM H.-Y.: Plenoptic sampling. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques* (2000), ACM Press/Addison-Wesley Publishing Co., pp. 307–318. 1, 2
- [CZRG06] CHANG C.-L., ZHU X., RAMANATHAN P., GIROD B.: Light field compression using disparity-compensated lifting and shape adaptation. *IEEE transactions on image processing* 15, 4 (2006), 793–806. 3
- [DLD12] DAVIS A., LEVOY M., DURAND F.: Unstructured light fields. In *Computer Graphics Forum* (2012), vol. 31, Wiley Online Library, pp. 305–314. 2, 10
- [EFPS16] EBRAHIMI T., FOESSEL S., PEREIRA F., SCHELKENS P.: JPEG Pleno: Toward an efficient representation of visual reality. *IEEE MultiMedia* 23, 4 (Oct 2016), 14–20. doi:10.1109/MMUL.2016.64.2
- [Fen03] FENNEY S.: Texture compression using low-frequency signal modulation. In *Proceedings of the ACM SIGGRAPH/EUROGRAPHICS conference on Graphics hardware* (2003), HWWS '03, Eurographics Association, pp. 84–91. URL: <http://dl.acm.org/citation.cfm?id=844174.844187.2>
- [GCRZ03] GIROD B., CHANG C.-L., RAMANATHAN P., ZHU X.: Light field compression using disparity-compensated lifting. In *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on* (2003), vol. 4, IEEE, pp. IV–760. 3
- [GGSC96] GORTLER S. J., GRZESZCZUK R., SZELISKI R., COHEN M. F.: The lumigraph. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques* (New York, NY, USA, 1996), SIGGRAPH '96, ACM, pp. 43–54. URL: <http://doi.acm.org/10.1145/237170.237200>, doi:10.1145/237170.237200. 1, 2
- [HJKG16] HONAUER K., JOHANNSEN O., KONDERMANN D., GOLDLUECKE B.: A dataset and evaluation methodology for depth estimation on 4d light fields. In *Asian Conference on Computer Vision* (2016), Springer, pp. 19–34. 9
- [IPL97] IHM I., PARK S., LEE R. K.: Rendering of spherical light fields. In *Computer Graphics and Applications, 1997. Proceedings., The Fifth Pacific Conference on* (1997), IEEE, pp. 59–68. 2, 10
- [JSA03] JAGMOHAN A., SEHGAL A., AHUJA N.: Compression of light-field rendered images using coset codes. In *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on* (2003), vol. 1, IEEE, pp. 830–834. 3
- [Kun12] KUNDU S.: Light field compression using homography and 2d warping. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on* (2012), IEEE, pp. 1349–1352. 3
- [LH96] LEVOY M., HANRAHAN P.: Light field rendering. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques* (New York, NY, USA, 1996), SIGGRAPH '96, ACM, pp. 31–42. URL: <http://doi.acm.org/10.1145/237170.237199>, doi:10.1145/237170.237199. 1, 2, 8
- [LPC*00] LEVOY M., PULLI K., CURLESS B., RUSINKIEWICZ S., KOLLER D., PEREIRA L., GINZTON M., ANDERSON S., DAVIS J., GINSBERG J., ET AL.: The digital michelangelo project: 3d scanning of large statues. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques* (2000), ACM Press/Addison-Wesley Publishing Co., pp. 131–144. 1
- [LS00] LIN Z., SHUM H.-Y.: On the number of samples needed in light field rendering with constant-depth assumption. In *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on* (2000), vol. 1, IEEE, pp. 588–595. 1
- [LWL*16] LIU D., WANG L., LI L., XIONG Z., WU F., ZENG W.: Pseudo-sequence-based light field image compression. In *Multimedia & Expo Workshops (ICMEW), 2016 IEEE International Conference on* (2016), IEEE, pp. 1–4. 3
- [MG99] MAGNOR M., GIROD B.: Hierarchical coding of light fields with disparity maps. In *Proceedings 1999 International Conference on Image Processing (Cat. 99CH36348)* (Oct 1999), vol. 3, pp. 334–338 vol.3. doi:10.1109/ICIP.1999.817130. 3
- [MG00] MAGNOR M., GIROD B.: Data compression for light-field rendering. *IEEE Transactions on Circuits and Systems for Video Technology* 10, 3 (2000), 338–343. 3
- [MSS08] MALVAR H. S., SULLIVAN G. J., SRINIVASAN S.: Lifting-based reversible color transformations for image compression. In *SPIE Applications of Digital Image Processing* (August 2008), International Society for Optical Engineering. URL: <http://research.microsoft.com/apps/pubs/default.aspx?id=102040.6>
- [NLP*12] NYSTAD J., LASSEN A., POMIANOWSKI A., ELLIS S., OLSON T.: Adaptive scalable texture compression. In *Proceedings of the Fourth ACM SIGGRAPH/Eurographics conference on High-Performance Graphics* (2012), Eurographics Association, pp. 105–114. 6
- [NR12] NEELIMA B., RAGHAVENDRA P. S.: Effective sparse matrix

- representation for the GPU architectures. *International Journal of Computer Science, Engineering and Applications* 2, 2 (2012), 151. 7
- [OEED18] OVERBECK R. S., ERICKSON D., EVANGELAKOS D., DEBEVEC P.: Welcome to light fields. In *ACM SIGGRAPH 2018 Virtual, Augmented, and Mixed Reality* (New York, NY, USA, 2018), SIGGRAPH '18, ACM, pp. 32:1–32:1. URL: <http://doi.acm.org/10.1145/3226552.3226557>, doi:10.1145/3226552.3226557. 1, 2, 3, 10
- [OSS*12] OHM J. R., SULLIVAN G. J., SCHWARZ H., TAN T. K., WIEGAND T.: Comparison of the coding efficiency of video coding standards; including high efficiency video coding (hevc). *IEEE Transactions on Circuits and Systems for Video Technology* 22, 12 (Dec 2012), 1669–1684. doi:10.1109/TCSVT.2012.2221192. 8
- [PA16] PERRA C., ASSUNCAO P.: High efficiency coding of light field images based on tiling and pseudo-temporal data arrangement. In *Multimedia & Expo Workshops (ICMEW), 2016 IEEE International Conference on* (2016), IEEE, pp. 1–4. 3
- [PM19] PRATAPA S., MANOCHA D.: Rlfc: Random access light field compression using key views and bounded integer sequence encoding. In *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games* (New York, NY, USA, 2019), I3D '19, ACM, pp. 9:1–9:10. URL: <http://doi.acm.org/10.1145/3306131.3317018>, doi:10.1145/3306131.3317018. 2, 3, 5, 8
- [PS01] PETER I., STRASSER W.: The wavelet stream: Interactive multi resolution light field rendering. In *Rendering Techniques 2001*. Springer, 2001, pp. 127–138. 2, 3, 10
- [SRASC14] SHARIF RAZAVIAN A., AZIZPOUR H., SULLIVAN J., CARLSSON S.: Cnn features off-the-shelf: An astounding baseline for recognition. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops* (June 2014). 2
- [VŘE17] VIOLA I., ŘEŘÁBEK M., EBRAHIMI T.: Comparison and evaluation of light field image coding approaches. *IEEE Journal of Selected Topics in Signal Processing* 11, 7 (Oct 2017), 1092–1106. doi:10.1109/JSTSP.2017.2740167. 2
- [WJV*05] WILBURN B., JOSHI N., VAISH V., TALVALA E.-V., ANTUNEZ E., BARTH A., ADAMS A., HOROWITZ M., LEVOY M.: High performance imaging using large camera arrays. In *ACM Transactions on Graphics (TOG)* (2005), vol. 24, ACM, pp. 765–776. 8
- [Yu17] YU J.: A light-field journey to virtual reality. *IEEE MultiMedia* 24, 2 (Apr 2017), 104–112. doi:10.1109/MMUL.2017.24. 1
- [ZL00] ZHANG C., LI J.: Compression of lumigraph with multiple reference frame (mrf) prediction and just-in-time rendering. In *Data Compression Conference, 2000. Proceedings. DCC 2000* (2000), IEEE, pp. 253–262. 2, 3, 8
- [ZLX*14] ZHOU B., LAPEDRIZA A., XIAO J., TORRALBA A., OLIVA A.: Learning deep features for scene recognition using places database. In *Advances in Neural Information Processing Systems 27*, Ghahramani Z., Welling M., Cortes C., Lawrence N. D., Weinberger K. Q., (Eds.). Curran Associates, Inc., 2014, pp. 487–495. URL: <http://papers.nips.cc/paper/5349-learning-deep-features-for-scene-recognition-using-places-database.pdf>. 2