# LIGHT FIELD COMPRESSION USING DISPARITY-COMPENSATED LIFTING

*Bernd Girod, Chuo-Ling Chang, Prashant Ramanathan, and Xiaoqing Zhu*

Stanford University
Information Systems Laboratory, Department of Electrical Engineering
350 Serra Mall, Stanford, CA 94305, USA

*Invited Paper*

**ABSTRACT**

We propose a novel approach for light field compression that incorporates disparity compensation into 4-D wavelet coding using disparity-compensated lifting. With this approach, we obtain the benefits of wavelet coding, including compression efficiency and scalability in all dimensions. Additionally, our proposed approach solves the irreversibility limitations of previous wavelet coding approaches. Experimental results show that the compression efficiency of the proposed technique outperforms current state-of-the-art wavelet coding techniques by a wide margin.

## 1. INTRODUCTION

Image-based rendering has become an important graphics technique, especially for interactive photo-realistic applications. A *light field* [1][2] is a data set for image-based rendering. It captures the outgoing radiance from a particular scene or object, at all points in 3-D space and in all directions. In practice, the light field can be represented as a 2-D set of 2-D camera views. A novel view from an arbitrary position and direction can be generated by appropriately combining image pixels from the acquired views.

The uncompressed size for a large photo-realistic light field can easily exceed tens of Gigabytes. Compression is therefore an important component of any light field system, and accordingly, there have been several different light field compression algorithms that have been described in the literature to date. These compression algorithms all try to exploit the intra-view and inter-view coherence in the data set; intra-view refers to the relationship among pixels within the same view, and inter-view refers to the relationship between pixels of views captured from different viewpoints. An embedded, scalable representation of a light field is also desirable.

An early light field compression algorithm based on vector quantization (VQ) [1] exploits both intra-view and inter-view coherence, with fast decoding capability as an important feature. To provide scalability, the discrete wavelet transform (DWT) has also been proposed for light field compression [3][4].

Several papers further exploit the inter-view coherence by using ideas akin to motion compensation in video compression, referred to as *disparity compensation*. Disparity-compensated prediction can be performed in a block-wise manner to improve the efficiency of predictive coding [5][6]. Disparity compensation can be further refined by incorporating some form of geometry information as described in [7].

In addition to predictive coding, the geometry model is also helpful for transform coding by aligning the views before applying the inter-view transform. In [8], a geometry model is used to warp the various views to a set of aligned view-dependent texture-maps, which is then encoded using a 4-D Haar wavelet coder.

The alignment of views is essential to fully exploit the inter-view coherence for transform coding. However, parts of the views may experience contraction, an irreversible many-to-one mapping, resulting in a loss in resolution. Together with the interpolation involved, the resampling process is generally irreversible. Therefore, when converting the resampled data set back to the acquired camera views, there is a loss in image quality [9]. The degraded reconstruction quality of the acquired views can affect the quality of the rendered views.

We propose a wavelet coding scheme that does not suffer from the irreversibility problems of previous transform-based approaches. Our scheme, *disparity-compensated lifting*, incorporates disparity compensation into the DWT using lifting. In video coding, the analogous technique of motion-compensated lifting [10][11][12] has recently been proposed that incorporates unrestricted motion compensation into 3-D subband coding in an reversible fashion.

The remainder of the paper is organized as follows. In Section 2, we discuss disparity-compensated lifting. The proposed light field compression system is described in Section 3. The experimental results and comparison with existing techniques are shown in Section 4.

## 2. DISPARITY-COMPENSATED LIFTING

Lifting is a procedure that can be used to implement discrete wavelet transforms [13]. Suppose that, in the context of light field compression, we have a set of $N$ views, $x[n]$, $n = 0, \cdots, N-1$. Assuming $N$ is even, we split up this set into two sets of $\frac{N}{2}$ views: an even set $x_0[k]$, $k = 0, \cdots, \frac{N}{2} - 1$, and an odd set $x_1[k]$, $k = 0, \cdots, \frac{N}{2} - 1$. Wavelet analysis can be factorized into one or more lifting steps, each consisting of a prediction and an update filter. The lifting structure transforms $x_0[k]$ and $x_1[k]$ into $y_0[k]$ and $y_1[k]$, the low-pass and the high-pass subbands resulting from the DWT of $x[n]$ respectively.

For reconstruction, as long as the filters used in wavelet synthesis are identical to those in wavelet analysis, the reversibility of the transform is ensured. We can use any kind of filters in lifting, including non-linear or data-adaptive filters, while still preserving the reversibility. In the context of light field compression, we in-

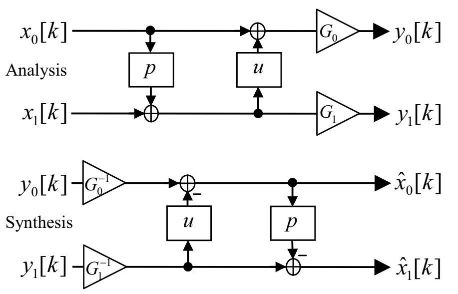


**Fig. 1**. Wavelet analysis and synthesis using one lifting step: $p$ is the prediction filter and $u$ is the update filter. $G_0$ and $G_1$ are the scaling factors to normalize the transform.

corporate disparity compensation into the prediction and update filters, using the disparity value derived from an approximate geometry model.

Let $v_0[k]$ and $v_1[k]$ denote the view-point, i.e., the viewing position and direction, of $x_0[k]$ and $x_1[k]$ respectively. Let $w_{01}^{(k)}$ be the function that warps its input, either an even view $x_0[k]$ or a low-pass subband image $y_0[k]$, from view-point $v_0[k]$ to $v_1[k]$ using the disparity information. Similarly, $w_{10}^{(k)}$ warps its input, either $x_1[k]$ or $y_1[k]$, from $v_1[k]$ to $v_0[k]$. As an example, $w_{01}^{(k)}(x_0[k])$ denotes the warped view (with view-point $v_1[k]$), derived from the given view $x_0[k]$ (with view-point $v_0[k]$).

The disparity-compensated lifting approach uses the warping functions, $w_{01}^{(k)}$ and $w_{10}^{(k)}$, as the first stage of the prediction and update filters, respectively. For the Haar wavelet, disparity-compensated lifting can be described by the following equations:

$$y_1[k] = x_1[k] - w_{01}^{(k)}(x_0[k]) \tag{1a}$$

$$y_0[k] = x_0[k] + \tfrac{1}{2}w_{10}^{(k)}(y_1[k]) \tag{1b}$$

$$= (x_0[k] - \tfrac{1}{2}w_{10}^{(k)}(w_{01}^{(k)}(x_0[k]))) + \tfrac{1}{2}w_{10}^{(k)}(x_1[k])$$

$$\hat{x_0}[k] = y_0[k] - \tfrac{1}{2}w_{10}^{(k)}(y_1[k]) = x_0[k] \tag{1c}$$

$$\hat{x_1}[k] = y_1[k] + w_{01}^{(k)}(x_0[k]) = x_1[k] \tag{1d}$$

Note that $y_1[k]$ needs to be computed prior to $y_0[k]$ in the lifting structure. We first generate a warped view $w_{01}^{(k)}(x_0[k])$ from $x_0[k]$ to predict $x_1[k]$. The resulting disparity-compensated prediction residual, $y_1[k]$, corresponds to the high-pass subband of the Haar wavelet. This high-pass subband is then warped and added to $x_0[k]$ in order to generate $y_0[k]$, which is approximately the disparity-compensated average of $x_0[k]$ and $x_1[k]$ that corresponds to the low-pass subband.

For this Haar wavelet example, only one lifting step is needed as shown in Figure 1. The prediction filter, $p$, and the update filter, $u$, consist of disparity compensation followed by a scaling of $-1$ and $\frac{1}{2}$, respectively. The additional scaling factors $G_0$ and $G_1$ needed to normalize the transform (Figure 1) are omitted in (1).

Note that unlike the texture-map approach which needs an explicit geometry model [8], the lifting structure can use other methods to provide the disparity information, such as block-matching. In addition, the lifting structure is not limited to the Haar wavelet. Any discrete wavelet transform can be factorized into lifting steps.

## 3. SYSTEM DESCRIPTION

In this paper, we consider the case in which the light field data set is generated by acquiring a 2-D array of views. A wavelet transform is performed on all 4 dimensions of the light field data set, through a 2-D inter-view transform and a 2-D intra-view transform. A modified version of the *Set Partitioning in Hierarchical Trees* (SPIHT) algorithm [14] is used to encode the 4-D transform coefficients into a compressed bitstream.

In applications where the target is an object as opposed to more complex scenery, shape adaptation can be incorporated into the proposed system so as to encode only the foreground object and discard the background.

In the following subsections, each component of the system, including the inter-view transform, intra-view transform, coefficient coding, and shape adaptation will be described. In addition, the scalability provided by the proposed system will be examined.

### 3.1. Inter-View Transform

The 2-D inter-view transform is carried out by applying 1-D transforms horizontally and vertically across the 2-D array of views, using disparity-compensated lifting as described in Section 2.

Various wavelet kernels can be implemented using lifting. In this work, the Haar wavelet and the biorthogonal Cohen-Daubechies-Feauveau $5/3$ wavelet are adopted due to their simplicity and effectiveness. To increase coding speed, a truncated version of the wavelet kernels can also be used, in which case the low-pass subband images are replaced by the even views.

If the number of view points is sufficiently large, a multi-level transform can be performed. The inter-view transform can be applied again, on the low-pass subband images $y_0[k]$, with $v_0[k]$ as the corresponding view-points. This procedure can be repeated several times depending on the density and total number of the view points.

### 3.2. Intra-View Transform

After the inter-view transform, the resulting subband images still exhibit coherence among neighboring pixels, especially for the low-pass subbands. To further exploit the remaining coherence, a multi-level 2-D DWT is applied to each of the subband images. The biorthogonal Cohen-Daubechies-Feauveau $9/7$ wavelet, popular for image compression, is chosen for the intra-view transform.

### 3.3. Coefficient Coding

To encode the 4-D DWT coefficients, the SPIHT algorithm is applied two-dimensionally for each of the subband images. It is further modified to work in a block-wise manner, similar to [15], for exploitation of local statistics as well as better random access support. Note that although the coefficients are coded together within each block, the intra-view wavelet transform is performed on the entire image.

Given a certain bit-rate or reconstruction quality target, bitstreams for all blocks need to be truncated and assembled in a rate-distortion optimal way. Candidate truncation points are limited to the end of each coding pass in the SPIHT algorithm, so as to reduce the overhead for indicating the truncation position. Reconstruction distortion is estimated in the coefficient domain, assuming additive contributions from different subbands and ignoring the effect of disparity compensation. Finally, the optimal bit allocation is decided using the Lagrangian multiplier technique.

### 3.4. Shape Adaptation

If the light field target is an object and only the foreground pixels contained in the object are of interest, shape adaptation can be incorporated to discard the background using Shape-Adaptive DWT (SA-DWT) [16]. 2-D object shape in each view, obtained from segmentation, is encoded and transmitted to the decoder as side information. In the proposed system, the approximate shape derived from the geometry model is used to predict the exact shape in order to encode it more efficiently.

### 3.5. Scalability

Different reconstruction qualities can be obtained from a single encoding process by assembling the bitstreams using different truncation points. This provides reconstruction-quality scalability. View-point scalability is supported by the inter-view wavelet transform. Specifically, we can decode only the low-pass subband images, if necessary. Moreover, the intra-view wavelet transform provides image-resolution scalability. Depending on the applications, the views in the light field can be decompressed up to the full resolution, or only a fraction of it, from a single compressed bitstream.

## 4. EXPERIMENTAL RESULTS

Experimental results are shown for two data sets, *Garfield* and *Penguin*, both having a hemispherical view arrangement, where there are 8 latitudes each containing 32 views, each with a resolution of 288x384 pixels.

With shape adaptation, only the object pixels are encoded. Hence, the conventional *bit-per-pixel* (bpp) measurement is modified to *bit-per-object-pixel* (bpop). Similarly, the *PSNR* measurement is modified to be computed by averaging over only the object pixels in all of the views. For each data set, a geometry model with 2048 triangles is reconstructed from the views. The geometry model is encoded at a bit-rate of 0.0090 and 0.0114 bpop, and the shape information is encoded at 0.0461 and 0.0602 bpop, for *Garfield* and *Penguin* respectively.

For the two data sets we used, experiments show that there is not much benefit by applying a multi-level inter-view transform over the single-level one, possibly due to the fact that the viewpoints in these data sets are rather sparse. Therefore, only one level of inter-view transform is used in the following results. The 4-level intra-view transform with the biorthogonal 9/7 wavelet was chosen throughout the experiments as it gives the best empirical performance.

### 4.1. Inter-View Wavelet Kernels

The compression performance of four different inter-view wavelet kernels, Haar, 5/3, truncated-Haar, and truncated-5/3, are compared together with the intra-coding scheme with no inter-view transform. Only the luminence component is coded, and the rate-PSNR curves for *Garfield* are shown in Figure 2. Results for *Penguin* are similar.

At the same bit-rate, the 5/3 wavelet performs about 1-1.5 dB better than the Haar wavelet. The truncated kernels perform slightly worse than their non-truncated counterparts for lack of low-pass filtering in the wavelet analysis.

Compared to the intra-coding scheme, the inter-view transform with the 5/3 wavelet provides 3-4 dB gain in terms of PSNR at the same bit-rate, or equivalently a bit-rate reduction of 50% for the same reconstruction quality.

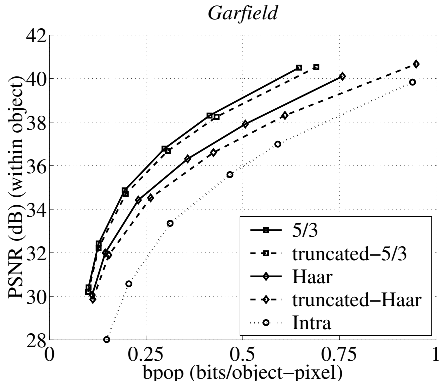


**Fig. 2**. Rate-PSNR curves for different inter-view wavelet kernels

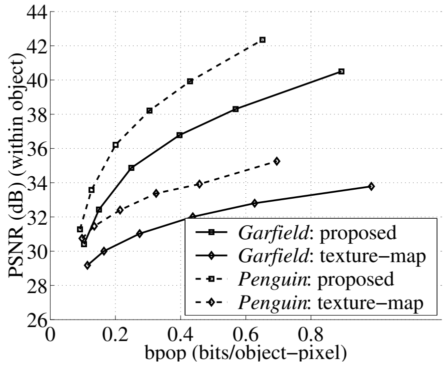


**Fig. 3**. Comparison between the proposed coder and the texture-map coder: For both data sets, the proposed coder gives a gain of up to 6 dB over the texture-map coder when coding at the same bit-rate, or equivalently, a bit-rate reduction about 70% at the same reconstruction quality.

### 4.2. Comparison with existing techniques

We compare the proposed coder, using 5/3 wavelet, with the texture-map approach described in [8]. The experiments are performed on color images using the (Y, Cb, Cr) color representation, with the chrominance components sub-sampled by a factor of 2 in each image dimension. The rate-PSNR curves are shown in Figure 3. The bit-rates include all three color components whereas the PSNR values are for luminance only.

Note that only the proposed coder needs the shape information, whereas the texture-map coder implicitly uses the approximate shape derived from the geometry. Consequently the PSNR and bit-rate values for the texture-map coder is obtained by only considering the object pixels within the approximate shape.

Compared to the texture-map coder, the proposed coder performs consistently better, except for the very-low-bit-rate region, where the extra overhead for the shape information is no longer negligible. With increasing bit-rate, the performance gap grows, since reconstruction quality of the texture-map coder is limited by the irreversible resampling process involved, whereas the proposed coder can achieve perfect reconstruction. The proposed coder gains more than 6 dB in PSNR for high bit-rates over the texture-map approach. Alternatively, there is a reduction of 70% in bit-rate for the same reconstruction quality.

Furthermore, as shown in Figure 4, the object shape is distorted for the texture-map coder, whereas the proposed coder retains the original object shape.

**Fig. 4**. Reconstructed views: *Garfield*: (left) proposed at 0.615 bpop (right) texture-map at 0.627 bpop, *Penguin*: (left) proposed at 0.490 bpop (right) texture-map at 0.458 bpop

## 5. CONCLUSIONS

Our light field compression system based on disparity-compensated lifting avoids the irreversibility problems of previous transform-based techniques. Our system also provides scalability in image resolution, view-point and reconstruction quality. Experimental results from two light field data sets show that the compression efficiency of the proposed approach outperforms current state-of-the-art wavelet coding techniques. We obtain a gain of up to 6 dB in overall reconstruction quality at the same bit-rate, or, equivalently, a bit-rate reduction of up to 70% at the same reconstruction quality.

## 6. REFERENCES


[1] M. Levoy and P. Hanrahan, "Light field rendering," in *Computer Graphics (Proceedings SIGGRAPH 96)*, August 1996, pp. 31–42.

[2] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, "The lumigraph," in *Computer Graphics (Proceedings SIGGRAPH 96)*, August 1996, pp. 43–54.

[3] I. Ihm, S. Park, and R. K. Lee, "Rendering of spherical light fields," in *Proc. Pacific Graphics 1997*, 1997, pp. 57–68.

[4] M. Magnor, A. Endmann, and B. Girod, "Progressive compression and rendering of light fields," in *Proceedings Vision, Modelling and Visualization 2000*, November 2000, pp. 199–203.

[5] X. Tong and R. M. Gray, "Coding of multi-view images for immersive viewing," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing ICASSP 2000*, Istanbul, Turkey, June 2000, vol. 4, pp. 1879–1882.

[6] M. Magnor and B. Girod, "Data compression for light field rendering," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 3, pp. 338–343, April 2000.

[7] M. Magnor, P. Eisert, and B. Girod, "Model-aided coding of multi-viewpoint image data," in *Proceedings of the IEEE International Conference on Image Processing ICIP-2000*, Vancouver, Canada, September 2000, vol. 2, pp. 919–922.

[8] M. Magnor and B. Girod, "Model-based coding of multi-viewpoint imagery," in *Proceedings SPIE Visual Communications and Image Processing VCIP-2000*, Perth, Australia, June 2000, vol. 1, pp. 14–22.

[9] M. Magnor and B. Girod, "Two approaches to incorporate approximate geometry into multi-view image coding," in *Proceedings of the IEEE International Conference on Image Processing ICIP-2000*, Vancouver, Canada, September 2000, vol. 2, pp. 5–8.

[10] A. Secker and D. Taubman, "Motion-compensated highly scalable video compression using an adaptive 3D wavelet transform based on lifting," in *Proceedings of the IEEE International Conference on Image Processing ICIP-2001*, Thessaloniki, Greece, October 2001.

[11] B. P.-P. and V. Bottreau, "Three dimensional lifting schemes for motion compensated video compression," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP-2001*, Salt Lake City, UT, USA, May 2001, pp. 1793–1796.

[12] L. Luo, J. Li, S. Li, et al., "Motion compensated lifting wavelet and its application in video coding," in *Proceedings of the IEEE International Conference on Multimedia and Expo 2001*, Tokyo, Japan, August 2001, pp. 481–484.

[13] W. Sweldens, "The lifting scheme: A construction of second generation wavelets," *SIAM Journal on Mathematical Analysis*, vol. 29, no. 2, pp. 511–546, 1998.

[14] A. Said and W. A. Pearlman, "A new fast and efficient image codec based on Set Partitioning in Hierarchical Trees," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, pp. 243–250, June 1996.

[15] K. K. Lin and R. M. Gray, "Video residual coding using SPIHT and dependent optimization," Snowbird, UT, USA, March 2001, pp. 113–122.

[16] S. Li and W. Li, "Shape-adaptive discrete wavelet transforms for arbitrarily shaped visual object coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, pp. 725–743, August 2000.