

Bitrate Reduction Techniques for Stereoscopic Digital Cinema Distribution

Michael D. Smith^a and John Villasenor^b,

^aConsultant, Los Angeles, CA (miksmith@attglobal.net);

^bDept. of Electrical Engineering, University of California at Los Angeles (villa@icsl.ucla.edu);

ABSTRACT

We consider the performance of two candidate approaches to compressing stereoscopic digital cinema distribution images: decorrelation transforms and disparity compensation. We show that disparity compensation generally can provide superior performance when significant disparity exists, and furthermore, that the consideration of vertical displacement can be an important factor in maximizing this performance under certain conditions. For context, we also provide details about the current state of both 2D and stereoscopic digital cinema distribution as of the end of the year 2007.

Keywords: DCI, SMPTE, digital cinema, JPEG2000, stereoscopic, 3D, compression

1. INTRODUCTION

One of the most interesting open technical challenges related to digital cinema distribution concerns the handling of stereoscopic imagery. While it is intuitive that there is a strong correlation between the left and right images, there are subtle challenges associated with identifying ways to effectively capture that correlation given the resolution and processing attributes of digital cinema systems. Thus, while the topic of stereoscopic compression techniques is not new, its application in this context is of high current interest and involves unique challenges, and is thus the focus of a study subgroup recently formed under DC28.40 within the Society of Motion Picture and Television Engineers (SMPTE) DC28 Group efforts.

This paper examines the compression performance of two different frameworks that are receiving significant current attention in the SMPTE standardization efforts. The first uses multi-dimensional decorrelation transforms that can be implemented as a preprocessing step to the JPEG2000 Part-1 compression, or can be represented completely using the multi-component transform paradigm of JPEG2000 Part-2. The second framework uses depth maps and/or disparity estimation as a method of identifying correlations between the left and right images, and then exploits that correlation in the compression process. Both frameworks have been implemented in a software testbed environment and used to generate a set of results and comparisons in terms of objective image quality and in terms of the implementation challenges involved.

2. DIGITAL CINEMA DISTRIBUTION

The roll-out of Digital Cinema distribution is now underway. Digital Cinema playback servers are being placed in theaters across the USA and internationally. As of the end of the year 2007, there are approximately 4000 theaters in the world which have Digital Cinema playback capabilities. There are approximately 100,000 screens worldwide. The Society Of Motion Picture and Television Engineers (SMPTE) have almost completed standardizing the methods and formats for digital distribution of feature films. The SMTPE DCinema standards describe image and sound formats, properties, and interfaces. The JPEG2000 image compression standard was selected as the image compression format. SMPTE worked with the JPEG standard group to define two special profiles for the Digital Cinema application, called "Profile-3" and "Profile-4". The profiles have very specific requirements about coding-style, bitrate, resolution, and formatting of the JPEG2000 code-stream, more specific details can be found in [1].

Digital Cinema Distribution Master (DCDM) images are either 2K or 4K. A 4K DCDM has a maximum resolution of 4096x2160 and a 2K DCDM has a maximum resolution of 2048x1080. Most feature films are created in a "Scope" or

“Flat” aspect ratio, corresponding to 1.85:1 and 2.39:1 respectively. “Scope” DCDM images are 1998x1080 (2K) and 3998x2160 (4K) while “Flat” DCDM images are 2048x858 (2K) and 4096x1716 (4K).

For 24 frame/sec content, each JPEG2000 code-stream may not 1,302,083 bytes, which corresponds to 250 Mbits/sec. For 48 frame/sec content, each JPEG2000 code-stream may not exceed 651,041 bytes, which also corresponds to 250 Mbits/sec. Most feature films are mastered at 24 frames/sec. Digital cinema playback server products implement JPEG2000 decoding using either FPGAs or ASICs. A popular JPEG2000 ASIC is the Analog Devices ADV202 or ADV212. Most often, there are additional FPGA resources on the digital cinema servers that implement code-stream decryption, image formatting, forensic watermarking and link encryption.

3. STEREOSCOPIC DIGITAL CINEMA

Stereoscopic digital cinema uses a single projector for playback of both left and right eye images. Each eye is presented with a new image every 1/24 seconds. At any given time, the projector is only projecting one of the two images, which means that each left and right eye image only gets 1/48 seconds of display time in each 1/24 second period. Black is displayed for the other 1/48 seconds. The digital cinema projectors are usually configured to “double-flash” or “triple-flash”, which helps remove any perception of flickering.

Depth cues such as image size, perspective, occlusion, light and shade, texture and motion parallax are available in standard two-dimensional imaging to help a viewer’s brain determine the apparent three-dimensional location of an object in the image. Disparity is a depth cue that is only available a stereoscopic imaging. Disparity is the difference in location of the same object viewed by the left and right eye/camera. Figure 1 illustrates the perceived location of an object viewed in a stereoscopic theater and the corresponding disparity value.

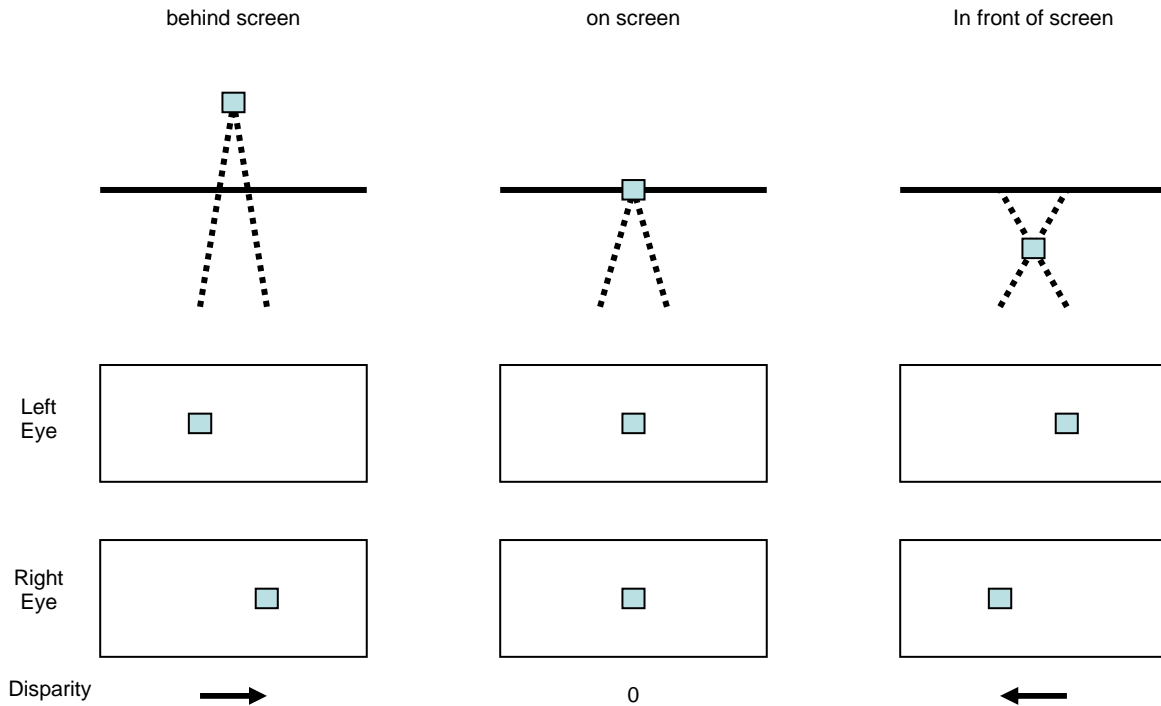


Fig. 1. Left-eye and Right-eye images showing perceived depth and corresponding horizontal disparity

As of the end of the year 2007, there are approximately 1000 theaters in the world which have Stereoscopic Digital Cinema playback capabilities. The images that are displayed on these systems are all compressed independently using

JPEG2000 Profile-3 with 250 Mbit/sec peak bitrate. An opportunity exists to improve the compression performance of these Stereoscopic Digital Cinema playback systems by compressing each stereo-pair together rather than independently. For easy implementation and adoption, the bitrate reduction technique used for stereoscopic digital cinema distribution should be similar to the existing “2D” playback method, which is simply the playback of independent JPEG2000 code-streams at a rate of 24 or 48 frames per second at a bitrate no greater than 250 Mb/sec. This means that JPEG2000 should be used for the core image coding functionality of the stereoscopic bitrate reduction technique.

4. BITRATE REDUCTION USING DECORRELATION-TRANSFORMS

The simplest bitrate reduction technique that has been suggested is the use of multi-dimensional decorrelation transforms. Two proposals have been made, one that can be described by (L, L-R) and the other by (L+R, L-R). The control that serves as a basis for comparison is always independent coding of left and right images, i.e. (L, R).

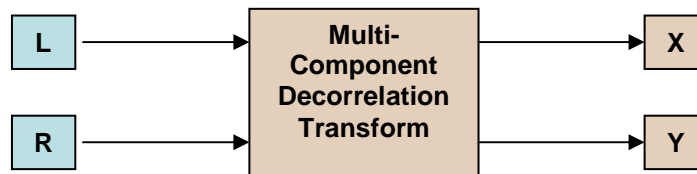


Fig. 2. Decorrelation Transform Framework

Each (left, right) pixel pair is transformed using the multi-component decorrelation transform, which can be expressed in matrix form as the following:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} L \\ R \end{bmatrix} = \begin{bmatrix} X \\ Y \end{bmatrix} \quad (1)$$

Where a, b, c and d are the coefficients of the multi-component decorrelation transform, L and R are the values of the left and right image pixels respectively, and X and Y are the two decorrelated pixel values.

When a = 1, b = 0, c = 0 and d = 1, the decorrelation matrix is an identity matrix, so each eye is coded independently. When a = 1, b = 0, c = 0.5 and d = -0.5, the left eye image is coded together with the difference between left and right eye images. When a = 0.5, b = 0.5, c = 0.5 and d = -0.5, the sum and difference between left and right eye images are coded. Other decorrelation matrix coefficients can be used as long as the resulting 2x2 matrix is invertible.

To optimize coding performance, the PCRD-opt step in the JPEG2000 encoding process should consider the matrix coefficients when computing the distortion-estimates for each coding-pass. The decorrelation transforms can either be implemented as a preprocessing step before/after JPEG2000 encoding/decoding, or they can be fully represented using the JPEG2000 Part-2 extensions.

The images in Figures 3 and 4 below show the results of the decorrelation techniques on stereoscopic image obtained from [3]. These images are from a video-based stereoscopic acquisition and have been processed by the decorrelation transforms with the three sets of matrix coefficients listed in the above text.

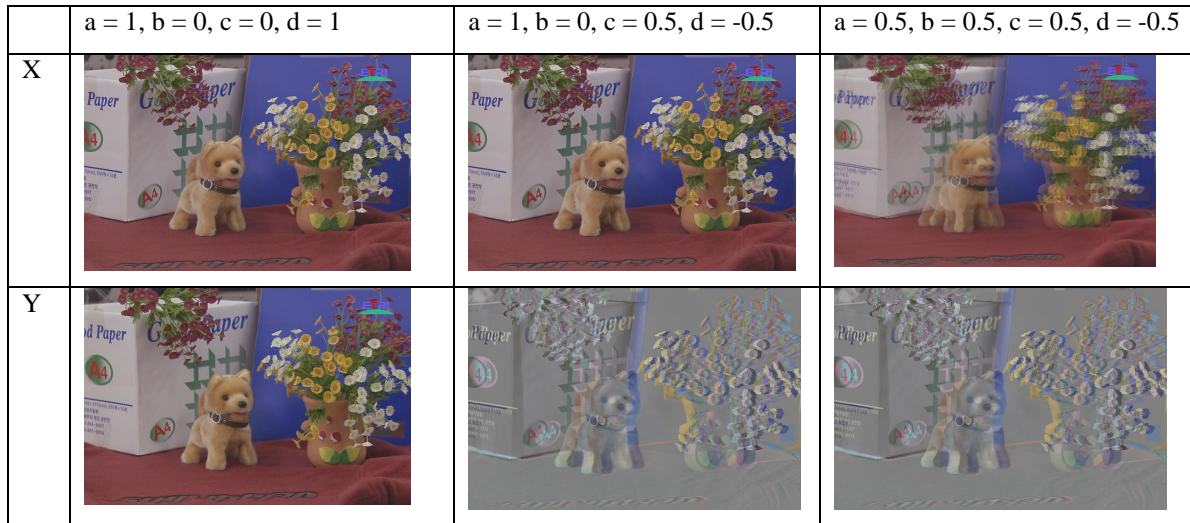


Fig. 3. Decorrelated Images – Puppy Sequence

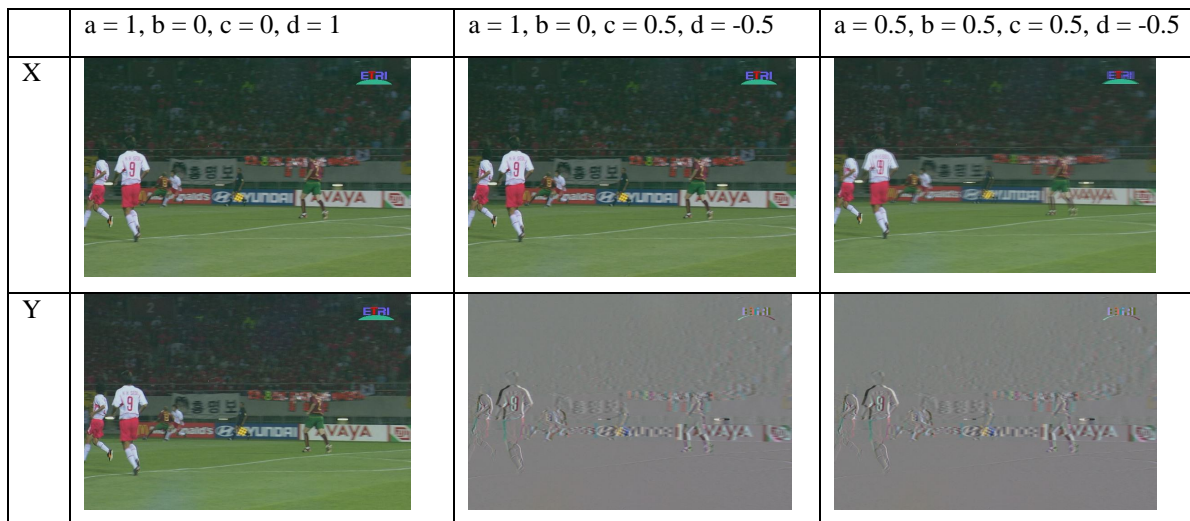


Fig. 4. Decorrelated Images – Soccer Sequence

5. BITRATE REDUCTION USING DISPARITY-COMPENSATION

A more sophisticated approach to stereoscopic bitrate reduction is to use disparity compensation between the left and right eye images. In this approach, the left eye image is coded independently. The decoded left eye image and the original right eye image are then compared in the disparity estimation step. The disparity estimates are used to predict the right eye image data from the decoded left eye image data. The difference between the predicted right eye image and original right eye image is computed. This residual image is then coded and transmitted. Figure 5 shows the processing steps involved in the disparity-compensation approach.

Disparity estimation is similar to motion estimation used in the MPEG and H.26x standards, with the difference that the system can be designed to exploit the fact that most of the displacement/disparity is in the horizontal direction. Horizontal disparity can range up to several hundred pixels for the high resolution 2K and 4K images used in digital cinema. Scharstein and Szeliski compare various methods of generating disparity estimates in [2].

For most computer animation content, the virtual cameras used to render the images are perfectly aligned in the vertical direction and thus the vertical disparity is zero. However, the assumption of zero disparity can fail to hold in film-based live action acquisition, since it is often impossible to maintain perfect camera alignment.

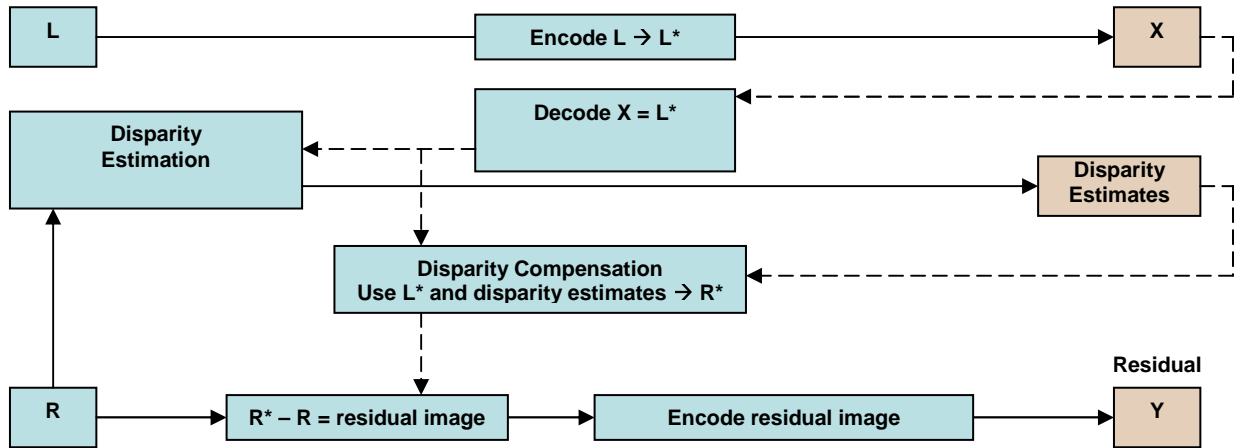


Fig. 5. Disparity-Compensation Framework

A depthmap may be provided with the DCDM images. The depth information in the depthmap image can be converted to horizontal disparity by an $\arctan()$ operation. A depthmap will not help with vertical disparity than can result when cameras are not perfectly aligned. The disparity estimation and compensation process has a lot of similarities with motion estimation/compensation in that the same tradeoffs between block size, integer vs. fraction pixel accuracy, and search range effect the quality of the predicted right image. Table 1 illustrates some of these tradeoffs when using a left and right image from a film-acquisition-based stereoscopic film.

Table 1 – Disparity Compensation performance under various block size and search range combinations

Block size (width, height)	Search Range (x,y)	PSNR Original R vs. Predicted R
(4,4)	(128x1)	25.40 dB
(4,4)	(128,16)	25.95 dB
(8,4)	(128,1)	24.94 dB
(8,4)	(128,16)	25.88 dB
(8,8)	(128,1)	24.35 dB
(8,8)	(128,16)	25.68 dB
(16,16)	(128,1)	22.85 dB
(16,16)	(128,16)	24.51 dB
(16,16)	(128,24)	24.58 dB

It is expected that allowing variable block sizes and performing in-loop deblocking filters may also improve the performance of the disparity estimation and compensation steps. Representing the disparity with more degrees of freedom such as affine or perspective distortion is also likely to improve prediction performance. The authors of [4] conclude that estimating and compensating for affine disparity results in a 3 dB improvement in prediction error in comparison with a purely translational disparity model.

6. EXPERIMENTS AND RESULTS

Both bitrate reduction frameworks have been implemented in a software testbed environment and were used to generate compression experiment results in terms of the objective image quality measure Peak Signal-to-Noise Ration (PSNR).

The decorrelation transform framework was implemented to allow flexible experimentation with decorrelation matrix coefficient values. As noted previously, the decorrelation transform step can be implemented as a preprocessing step before JPEG2000 Part-1 encoding, or it can be fully implemented by a JPEG2000 Part-2 encoder. For these experiments, we implemented it as a preprocessing step. Encoding performance was improved by incorporating distortion weights in the PCRD-opt step that account for the decorrelation matrix that is applied in preprocessing. Further details of Post-Compression Rate Distortion optimization (PCRD-opt) for JPEG2000 encoding can be found in [5] and [6].

The disparity-compensation framework was implemented to determine the feasibility and performance of a simple approach. To compute the disparity vectors, we implemented an integer-pixel block matching search. The block size and search range was made to be flexible to allow experimentation. Further work on the estimation and compensation algorithms is likely to improve performance as described in Section 5. For the results presented, we used 8x4 block size with a search range of (128, 16). Based on some initial experiments, we allocated 50% of the bytes to the left eye and 50% of the bytes to the residual image. We expect that a more sophisticated byte allocation strategy will lead to improved performance.

For the compression experiments, we compare the following 4 stereoscopic bitrate reduction techniques:

- 1) independent L and independent R image coding “1 0 0 1 : (L,R)”
- 2) independent L and dependent L-R image coding “1.0 0 0.5 -0.5 : (L, L-R)”
- 3) dependent L+R image and dependent L-R image coding “0.5 0.5 0.5 -0.5 : (L+R, L-R)”
- 4) independent L and disparity compensated R image coding “disparity-compensated”.

We used 3 different stereo-pair test images having different characteristics. The Puppy and Soccer2 image pairs are video-acquisition images and are publicly available [3]. The high-resolution film-based image is not publicly available. The Puppy image pair has a significant amount of disparity between left and right eye images. The Soccer2 image has only a modest amount of disparity. The film-based image was acquired with long telephoto lens and was enhanced in post-production to increase the disparity of various objects in the scene. The film-based image also contains a significant amount of film-grain.

Figures 6, 7 and 8 show the rate/distortion tradeoff for each experimental image compression scheme for the Puppy, Soccer2 and film-based images respectively. The disparity-compensated technique works well for the Puppy and Soccer2 images, resulting in approximately 1.0dB improvement over independent (L,R) coding across a wide range of bitrates. The (L+R, L-R) technique also works well for the Soccer2 image showing similar gains. The experimental coding techniques do not show any improvements for the high-resolution film-based image, due in part to the large amount of film grain. Our previous work shows that image prediction methods may not perform well on images with high amounts of film grain [9].

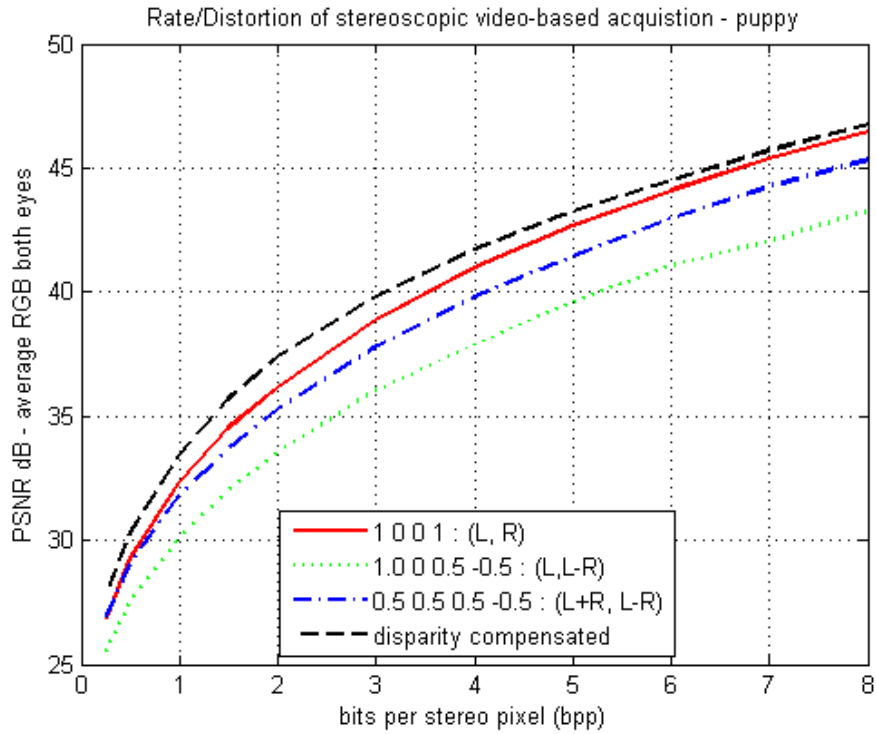


Fig. 6. Compression Performance of two frameworks – Puppy Image

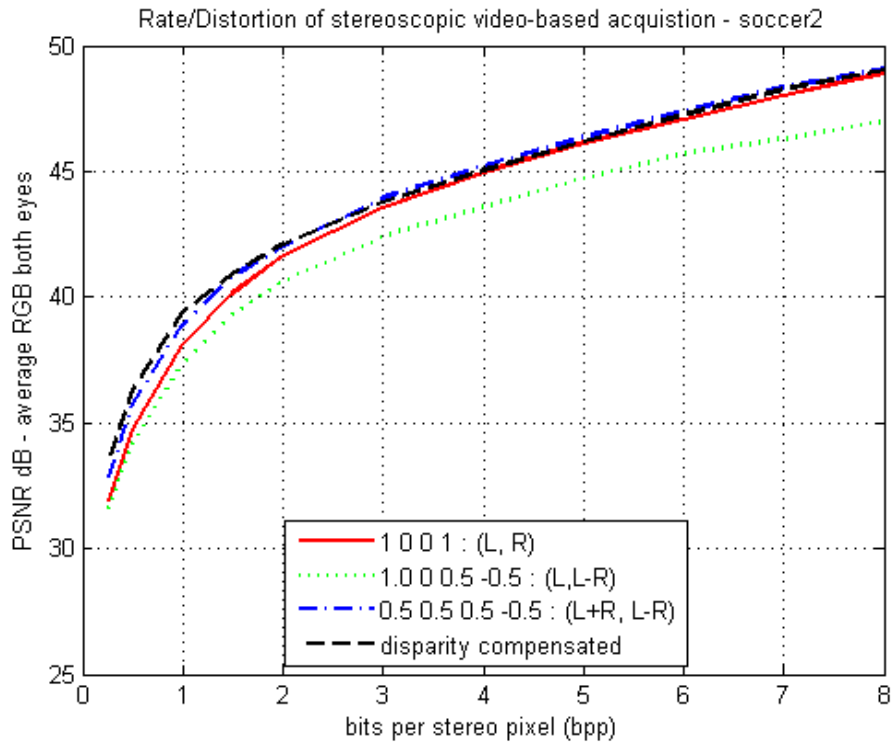


Fig. 7. Compression Performance of two frameworks – Soccer2 Image

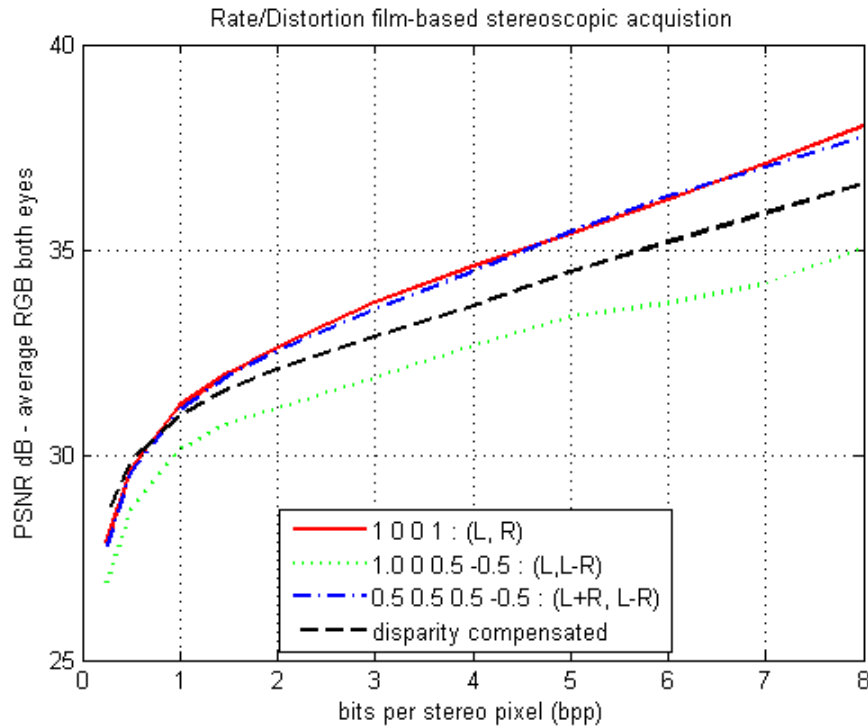


Fig. 8. Compression Performance of two frameworks – Soccer2 Image

7. CONCLUSIONS

We compared decorrelation-transform and disparity-compensation approaches to compressing stereoscopic digital cinema distribution images. Our results show that the disparity compensation method provides better performance when the images have a significant amount of disparity. We also showed that when using disparity compensation with larger block sizes, allowing vertical disparity to be compensated in addition to horizontal disparity can improve prediction results significantly (providing an improvement of about 2.5 dB PSNR in the images studied here). The decorrelation-transform approach shows some improvement when coding low noise low disparity images.

Based on these conclusions, it seems that the disparity-compensation approach is most promising for images with significant disparity and the decorrelation-transform approach is only appropriate when the images have little or no disparity. An adaptive approach that toggles between disparity-compensated, decorrelation-transforms and independent coding on a frame-by-frame basis may also be appropriate. Furthermore, we believe that our software testbed implementation of the disparity-compensation approach provides significant opportunities to introduce enhanced algorithmic approaches. Stereoscopic computer-based animation content is common in stereoscopic digital cinema distribution, but was not tested in these experiments.

REFERENCES

- ¹ ISO/IEC 15444-1:2004/FDAM 1 — Information Technology — JPEG2000 Image Coding System: Part 1—Core Coding System, Amendment 1: Profiles for Digital Cinema Applications.

- ² Daniel Scharstein and Richard Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms", Microsoft Research Technical Report MSR-TR-2001-81, November 2001, available <ftp://ftp.research.microsoft.com/pub/tr/tr-2001-81.pdf>
- ³ <http://www.3dtv-research.org/>
- ⁴ V.E.Seferidis and D.V.Papadimitriou, "Improved Disparity Estimation in Stereoscopic Television", IEEE Electronic Letters Vol.29, No.9, p.782-783 April 29 1993.
- ⁵ Taubman, David, "High Performance Scalable Image Compression with EBCOT," IEEE Transactions On Image Processing, pp. 1158-1170, Vol. 9, No. 7, July 2000.
- ⁶ Taubman, David S. and Michael W. Marcellin, JPEG2000 - Image Compression Fundamentals, Standards and Practice. Norwell: Kluwer Academic Publishers, 2002.
- ⁷ M. D. Smith and J. D. Villasenor, "Intra-frame JPEG-2000 vs. Inter-frame Compression Comparison: The benefits and trade-offs for very high quality, high resolution sequences", in Proc. SMPTE Technical Conf., Pasadena, CA, Oct. 2004, pp. 1-9.