

## CHROMA CODING IN DISTRIBUTED VIDEO CODING

Vijay Kumar Kodavalla<sup>1</sup> and P. G. Krishna Mohan<sup>2</sup>

<sup>1</sup>Semiconductor and Systems Division, Wipro Technologies, Bangalore, India,  
E-mail: vijay.kodavalla@wipro.com.

<sup>2</sup>Dept of Electronics & Comm Engineering., JNTU College of Engineering, Hyderabad, India,  
E-mail: pgkmohan@yahoo.com.

### ABSTRACT

Distributed Video Coding (DVC) is a video coding method for emerging wireless video surveillance networks, wireless video sensor networks and wireless mobile video applications, which is not yet been standardized. DVC is relatively new video coding paradigm, which is not to compete but to complement the popular predictive coding standards such as H.26x, MPEG, VC1 and DivX etc., for the emerging applications. Certain wireless video applications would need low complex encoder, even at the expense of relatively more complex decoder; which is in contrast to the predictive coding standards. Various DVC Architectures developed so far addressed only luma component coding and its Rate Distortion (RD) performance evaluation. In this paper color components (Chroma) coding method of DVC are proposed and results are presented.

**Keywords:** Chroma components, side information, H.264 intra, RD performance, wyner-ziv.

### 1. INTRODUCTION

Today's predictive coding standards such as H.264, MPEG2/4, VC1 and DivX etc. are suitable for application like broadcasting. In these predictive coding methods, encoder is typically 5-10 times more complex than the decoder by design. In applications like broadcasting, there is one encoder and many decoders and hence decoder has to be less complex. But there are many other emerging mono-view and multi-view applications and services, where reverse encoder/decoder complexity is desired compared to that of predictive coding standards. A next generation video coding method known as Distributed Video Coding (DVC) is suitable for the emerging applications and services. DVC is designed to have less complex encoder, even if the decoder is more complex. DVC is still in research and not yet been standardized. There are still many challenges with DVC such as non availability of complete coding method including chroma coding, necessity of feedback channel, lower RD performance, interoperability due to no standardization and flicker. The DVC gaps and challenges are detailed in Authors another paper [4]. In the literature various DVC architectures proposed [6, 7] dealt with only luma component coding and its RD performance comparison with predictive intra coding standard. The main objective of this paper is to provide the details of DVC chroma component coding method along with implemented 'C' code results and RD performance benchmarking. The Section 2 details proposed DVC codec Architecture and implementation details, especially for chroma components. The Section 3 presents the results with 'C' code implementation, followed by conclusion in Section 4.

### 2. DVC CODEC ARCHITECTURE AND IMPLEMENTATION DETAILS

The proposed DVC Codec Architecture (as shown in Figure 1) and implementation details are included in Authors another paper [3]. The encoding process of DVC is intentionally made simple by design. As a first step, consecutive frames in the incoming video sequence are split into various groups based on the cumulative motion of all the frames in each group crossing a pre-defined threshold. And the number of frames in each such group is called a GOP (Group of Pictures).

In a GOP, first frame is called key frame and remaining frames are called Wyner-Ziv (WZ) frames. The key frames are encoded by using H.264 main profile intra coder. The WZ frames undergo block based transform, i.e. DCT is applied on each  $4 \times 4$  block. The DCT coefficients of entire WZ frame are grouped together, forming DCT coefficient bands. After the transform coding, each coefficient band is uniform scalar quantized with pre-defined levels. Bit-plane ordering is performed on the quantized bins. Each ordered bit-plane is encoded separately by using Low Density Parity Check Accumulator (LDPCA) encoder. LDPCA encoder computes a set of parity bits representing the accumulated syndrome of the encoded bit-planes. An 8-bit Cyclic Redundancy Check (CRC) sum is also sent to decoder for each bit plane to ensure correct decoding operation. The parity bits are stored in a buffer in the encoder and progressively transmitted to the decoder, which iteratively requests more bits during the decoding operation through the feedback channel.

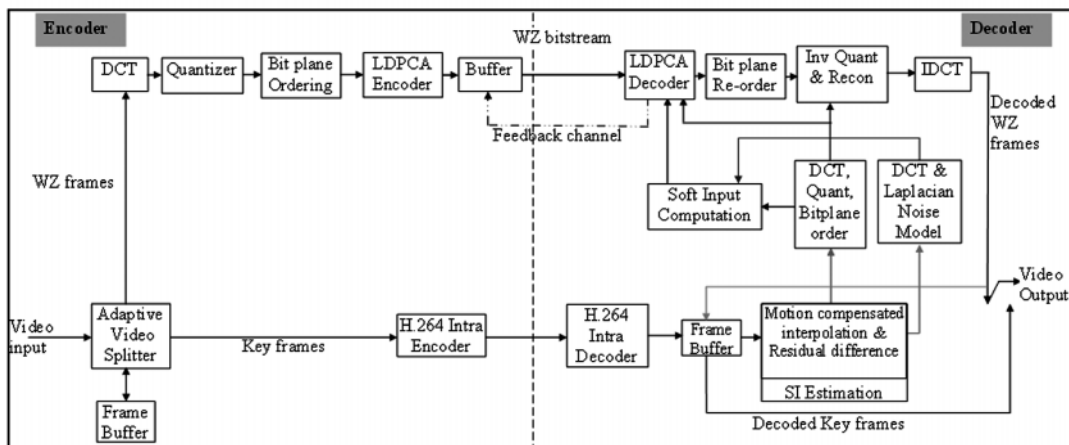


Figure 1: DVC Codec Architectural Block Diagram [3].

The DVC decoding process is relatively more complex, to keep the encoder less complex by design. The key frames are decoded by using H.264 main profile intra decoder. The decoded key frames are used for reconstruction of side information (SI) at the decoder, which is an estimation of the WZ frame available only at the encoder. A motion compensated interpolation between the two closest reference frames is performed, for SI generation. The difference between the WZ frame and the corresponding SI can be correlation noise in virtual channel. An online Laplacian model is used to obtain a good approximation of the residual (WZ-SI). The DCT transform is applied on the SI and an estimate of the coefficients of the WZ frame are thus obtained. From these DCT coefficients, soft input values for the information bits are computed, by taking into account the statistical modeling of the virtual noise. The conditional probability obtained for each DCT coefficient is converted into conditional bit probabilities by considering the previously decoded bit-planes and the value of side information. These soft input values are fed to the LDPCA decoder. The decoder success or failure is verified by an 8-bit CRC sum received from encoder for current bit plane. If the decoding fails, i.e. if the received parity bits are not sufficient to guarantee successful decoding with a low bit error rate, more parity bits can be requested using the feedback channel. This process is iterated until successful decoding. After the successful decoding of all bit planes, inverse bit plane ordering is performed. Inverse quantization and reconstruction is performed on the decoded bins. Next, inverse DCT (IDCT) is performed and each WZ frame is restored to pixel domain. Finally, decoded WZ frames and key frames are interleaved as per GOP to get the decoded video sequence.

The encoding and decoding procedure described shall be applied on all three luma and chroma ( $C_b$  and  $C_r$ ) components separately for example in  $YC_bC_r$  4:2:0 video format. The luma component encoding and decoding method is explained in Authors another paper [3]. In this paper Authors proposed chroma coding method for achieving lower encoder complexity without compromising much on RD performance and lower decoding delay.

## 2.1. DVC Encoder

As already explained, the key frames are H.264 intra coded. All three luma and chroma components are intra encoded as per H.264 standard. And as per Figure 1, the WZ frames are transformed (DCT), quantized and LDPCA encoded. The luma component block size considered for DCT is  $4 \times 4$ , and hence block size considered for chroma component DCT is  $2 \times 2$ , due to 4:2:0 format. Once the DCT operation has been performed on all the  $2 \times 2$   $C_b$  and  $C_r$  samples of image, the DCT coefficients of  $C_b$  and  $C_r$  are separately grouped together according to the standard Zig-Zag scan order within the  $2 \times 2$  DCT coefficient blocks. After performing the Zig-Zag scan order,  $C_b$  and  $C_r$  coefficients are organized into four bands each. First band containing low frequency information is often called the DC band and the remaining bands are called AC bands which contains the high frequency information.

To encode WZ frames, each  $C_b$  and  $C_r$  DCT bands are quantized separately using predefined number of levels, depending on the target quality. DCT coefficients representing lower spatial frequencies (DC bands) are quantized using uniform scalar quantizer with low step sizes, i.e. with higher number of levels. AC bands are quantized using dead zone quantizer with doubled zero interval to reduce the blocking artifacts. The dynamic range of each AC band is calculated instead of using fixed value, to have quantization step size adjusted to the dynamic range of each band. The dynamic data range is calculated separately for each AC band to be quantized, and transmitted to the decoder along with the encoded bit stream. Depending on the target quality and data rates, eight different  $2 \times 2$  chroma quantization matrices are used, which are derived from that of  $4 \times 4$  luma. The eight luma quantization matrices ( $Q_1$  to  $Q_8$ ) used are listed in Table 1, where  $Q_1$  and  $Q_8$  correspond to lowest and highest quality levels respectively. In the table entries, '0' means that no WZ parity bits are transmitted to the decoder for the corresponding bands. The decoder will replace the

DCT bands by the corresponding side information DCT coefficients bands determined at the decoder, when no WZ parity bits are sent for a given band. Three types of chroma quantization matrices of eight each are considered for evaluation, which are listed in Table 2 and 3. The chroma quantization matrices type-1 and 2 given in Table 2 and 3 are derived from sub-sampling luma quantization matrices shown in Table 1. Though type-1 chroma quantization matrices are expected to result in higher PSNR, the data rate required will be higher too. Due to higher chroma quantization matrix element values in type-1, the number of bit-planes for LDPCA encoding will be high. More number of LDPCA encoded bit-planes would mean more parity bits for error correction to decoder, thereby higher bit rate. The type-3 chroma quantization matrices with all '0' elements is preferred as data rate will be lowest if decoded chroma components can meet PSNR. All '0' elements chroma quantization matrices would mean, no parity bits will be sent from encoder to decoder for chroma components of WZ frames. In this case, the chroma components side information (SI) generated at the decoder is to be used as decoded WZ frames. Hence the data rate with this quantization matrices type will be the lowest. After quantizing the  $C_b$  and  $C_r$  DCT coefficient bands, the quantized symbols are converted into bit-stream. The quantized symbol bits of the same significance (Ex: MSB) are grouped together forming the corresponding bit-plane, which are separately encoded by LDPCA encoder.

**Table 1**  
Luma Quantization Matrices [3]

16 8 0 0	32 8 0 0	32 8 4 0	32 16 8 4
8 0 0 0	8 0 0 0	8 4 0 0	16 8 4 0
0 0 0 0	0 0 0 0	4 0 0 0	8 4 0 0
0 0 0 0	0 0 0 0	0 0 0 0	4 0 0 0
$Q_1$	$Q_2$	$Q_3$	$Q_4$
32 16 8 4	64 16 8 8	64 32 16 8	128 64 32 16
16 8 4 4	16 8 8 4	32 16 8 4	64 32 16 8
8 4 4 0	8 8 4 4	16 8 4 4	32 16 8 4
4 4 0 0	8 4 4 0	8 4 4 0	16 8 4 0
$Q_5$	$Q_6$	$Q_7$	$Q_8$

**Table 2**  
Type-1 Chroma Quantization Matrices

16 0	32 0	32 4	32 8
0 0	0 0	4 0	8 0
$Q_1$	$Q_2$	$Q_3$	$Q_4$
32 8	64 8	64 16	128 32
8 4	8 4	16 4	32 8
$Q_5$	$Q_6$	$Q_7$	$Q_8$

**Table 3**  
Type-2 Chroma Quantization Matrices

0 0	0 0	4 0	8 0
0 0	0 0	0 0	0 0
$Q_1$	$Q_2$	$Q_3$	$Q_4$
8 4	8 4	16 4	32 8
4 0	4 0	4 0	8 0
$Q_5$	$Q_6$	$Q_7$	$Q_8$

**2.2. DVC Decoder**

As already explained, key frames are H.264 intra decoded. All three luma and chroma components are intra decoded as per H.264 standard. And all the three luma and chroma components of WZ frames are decoded using side information (SI) and WZ parity bits from the encoder. In case of type-1 and 2 chroma quantization matrices, WZ chroma components decoding are based on both chroma components side information and WZ chroma parity bits from encoder. Where as in case of type-3 chroma quantization matrices, WZ chroma components decoding depends on only chroma components side information and there are no WZ chroma parity bits sent from the encoder. Hence in case of type-3 chroma quantization matrices, chroma components PSNR directly depend on how good the generated SI is at the decoder.

Motion compensated frame interpolation method is used for SI estimation at the decoder, either by using key frames or previously decoded WZ frames ( $X_B$  and  $X_F$ ) based on GOP. Motion compensated interpolation is a motion estimation technique based on only information in the reference frames, and doesn't have any information from the frame being predicted. The motion vectors are computed for luma (Y) component using forward motion estimation and bidirectional motion estimation in conjunction with spatial motion smoothing [8], as shown in Figure 2. Full search based forward motion estimation is performed with full-pel step size using  $16 \times 16$  blocks, with search range as  $\pm 32$ . And the bidirectional motion estimation is performed with half-pel and hierarchical block sizes (first  $16 \times 16$  and then followed by  $8 \times 8$ ). Next, weighted vector median filters based spatial motion smoothing algorithm is used to make the final motion vector field smoother. There is no separate motion estimation performed for chroma components for SI estimation. The luma motion vectors are used for chroma components as well. However each luma  $8 \times 8$  block corresponds to a  $4 \times 4 C_b$  block and  $4 \times 4 C_r$  block in  $YC_bC_r$  4:2:0 format. This half-pel motion vector values for each  $4 \times 4 C_b$  and  $4 \times 4 C_r$  block are obtained by dividing corresponding  $8 \times 8$  luma block half-pel motion vector value by two. However as  $8 \times 8$  block luma motion vectors themselves are in half-pel values, the division may lead to quarter-pel motion vector values for chroma components. These chroma quarter-pel motion vector

values are rounded to half-pel values. Once motion vectors are obtained, the interpolated frame can be obtained by using bidirectional motion compensation. And bidirectional motion compensation will be performed separately for all three luma and chroma components using computed motion vectors.

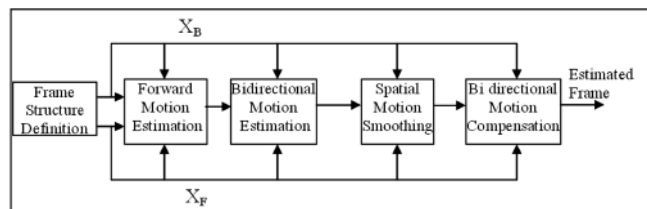


Figure 2: Side Information (SI) Estimation [3].

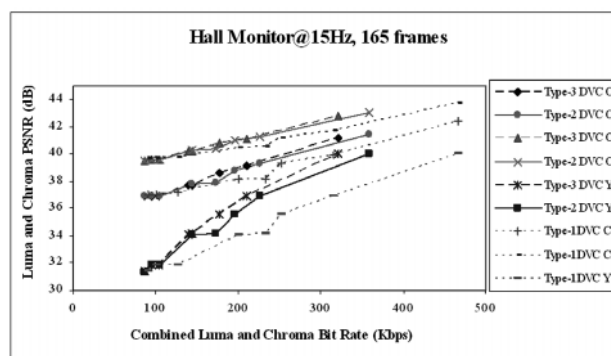
In DVC, decoding efficiency of WZ frame critically depends on the capability to model the statistical dependency between the WZ information at the encoder and the side information computed at the decoder. A Laplacian distribution, which has good tradeoff between model accuracy and complexity, is used to model the correlation noise, i.e. the error distribution between corresponding DCT bands of SI and WZ frames. The method used to estimate the Laplacian distribution parameter of chroma components is same as that of luma components. After laplacian distribution model, the next step is soft input computation for LDPCA decoding. As a part of soft input computation, conditional bit probabilities are calculated by considering the previously decoded bit planes and the value of side information (SI). The probability calculations of DC and AC bands (derived from  $2 \times 2$  DCT block) of chroma components are same as that of luma component. Similarly, there is no other difference between handling chroma and luma components for other steps including LDPCA decoder, inverse quantization (with chroma quantization matrices) and reconstruction and inverse DCT.

### 3. DVC 'C' MODEL IMPLEMENTATION RESULTS AND DISCUSSION

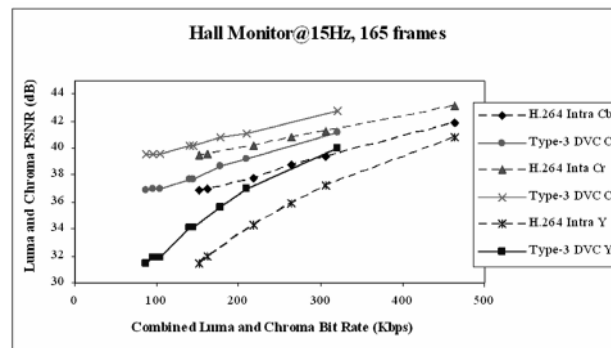
The DVC encoder and decoder described are completely implemented in 'C'. The implemented codec has been evaluated with four standard test sequences namely hall monitor, coast guard, foreman and soccer sequences with QCIF resolution and 15 Hz vertical frequency. The chosen test video sequences are representative of various levels of motion activity. The Hall monitor video surveillance sequence has low to medium amount of motion activity. And Coast guard sequence has medium to high amount of motion activity, whereas Foreman sequence has very high amount of motion activity. And Soccer sequence has significant motion activity. The H.264 coder profile in key frame path is main profile, which can encode 4:2:0 sequences. DVC luma and chroma quantization matrices given in Table 1, 2 and 3 are used for evaluation, apart from all '0' element chroma quantization matrices.

### 3.1. Hall Monitor Sequence Results

The Figure 3(a) shows RD performance of hall monitor luma and chroma components, with various quantization matrix types. From this figure, it can be concluded that RD performance with type-3 chroma quantization parameters outperformed other types. And Figure 3(b) shows comparison of DVC RD performance (with type-3 chroma quantization parameters) with that of H.264 intra. The PSNR achieved with DVC is better by upto 2.2dB, 1.6dB and 1.4dB for luma,  $C_b$  and  $C_r$  components respectively for a given combined luma and chroma bit rate; compared to that of H.264 intra. Similarly, DVC achieved similar or better PSNR for luma and chroma components with combined luma and chroma bit rate lesser by upto 100Kbps, compared to that of H.264 intra.



(a)



(b)

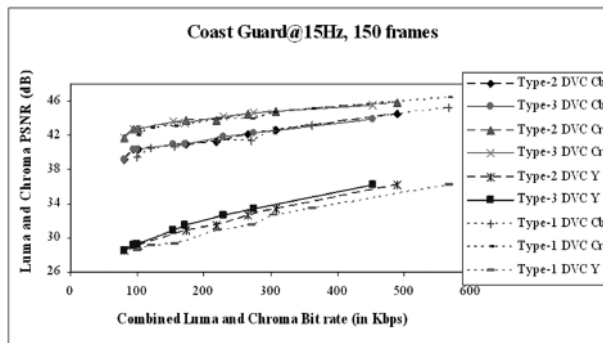
Figure 3: Hall Monitor (a) Type-1, 2 and 3 RD Performance Comparison, (b) Type-3 RD Performance Comparison with H.264 Intra.

### 3.2. Coast Guard Sequence Results

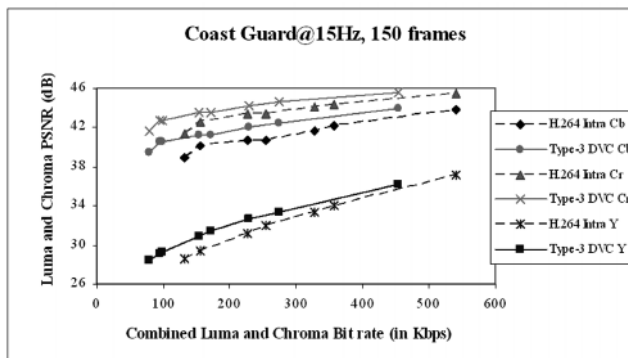
The Figure 4(a) shows RD performance of Coast guard luma and chroma components, with various quantization matrices types. From this figure, it can be concluded that RD performance of luma with type-3 chroma quantization parameters outperformed other types. Whereas, RD performance of chroma components with all the three types of chroma quantization matrices almost matched. Hence it can be concluded that even with Coast guard sequence, type-3 quantization matrices worked well. Also it should be noted that for Hall monitor sequence, even chroma components RD performance with type-3

chroma quantization parameters is better than other types, whereas it is not the case with Coast guard sequence. This can be attributed to relatively high motion activity in Coast guard sequence compared to that of Hall monitor sequence. With high motion, SI quality at the decoder reduces and thereby chroma components PSNR dropped. And Figure 4 (b) shows comparison of DVC RD performance (with type-3 chroma quantization parameters) with that of H.264 intra. The PSNR achieved with DVC is better by upto 1.5dB each for luma,  $C_b$  and  $C_r$  components for a given combined luma and chroma bit rate; compared to that of H.264 intra. Similarly, DVC achieved similar or better PSNR for luma and chroma components with combined luma and chroma bit rate lesser by upto 80Kbps, compared to that of H.264 intra.

quantization parameters) with that of H.264 intra. The PSNR achieved with DVC is better by upto 0.8 dB, 0.4dB and 0.4dB for luma,  $C_b$  and  $C_r$  components respectively for a given combined luma and chroma bit rate; compared to that of H.264 intra. Similarly, DVC achieved similar or better PSNR for luma and chroma components with combined luma and chroma bit rate lesser by upto 20Kbps, compared to that of H.264 intra. However for higher quantization parameters (low compression), H.264 intra RD performance is better than that of DVC. This is because for high quantization parameters, DVC decoder parity bit request rate through feedback channel increases to improve SI quality. At some point the request rate increases so high that H.264 intra scores better from there on. In summary, type-3 chroma quantization matrices are suitable for Foreman sequence also.



(a)

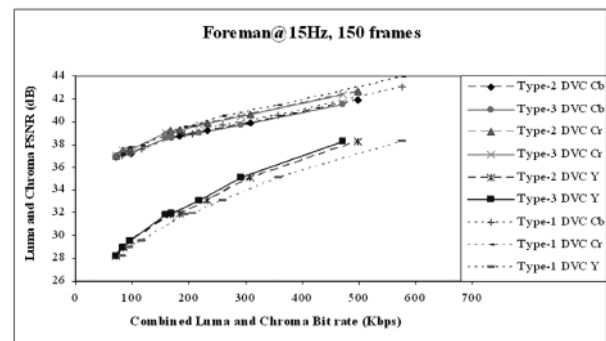


(b)

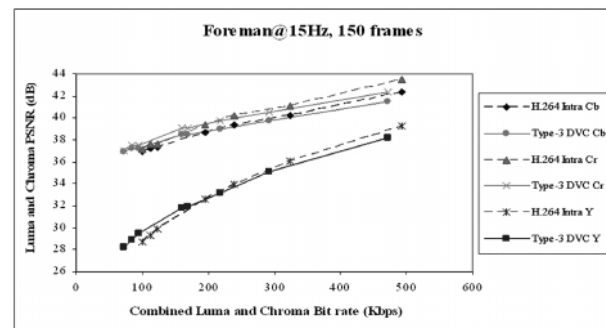
Figure 4: Coast Guard (a) Type-1, 2 and 3 RD Performance Comparison, (b) Type-3 RD Performance Comparison with H.264 Intra.

### 3.3. Foreman Sequence Results

The Figure 5(a) shows RD performance of Foreman luma and chroma components, with various quantization matrices types. From this figure, it can be concluded that RD performance of luma with type-3 chroma quantization parameters outperformed other types. Whereas, chroma components RD performance with type-3 quantization matrices are slightly lower than that of with type-1 and 2. This can be again attributed to relatively high motion activity in Foreman sequence compared to that of Hall monitor or Coast guard sequences. And Figure 5(b) shows comparison of DVC RD performance (with type-3 chroma



(a)



(b)

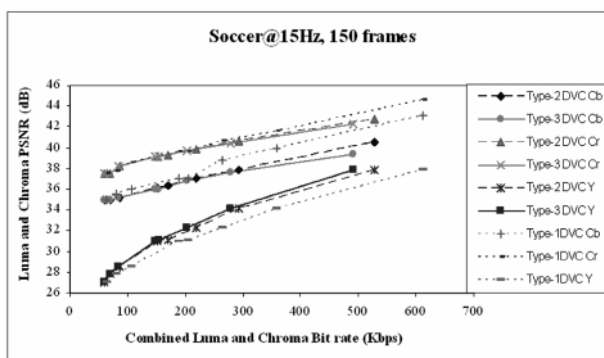
Figure 5: Foreman (a) Type-1, 2 and 3 RD Performance Comparison, (b) Type-3 RD Performance Comparison with H.264 Intra.

### 3.4. Soccer Sequence Results

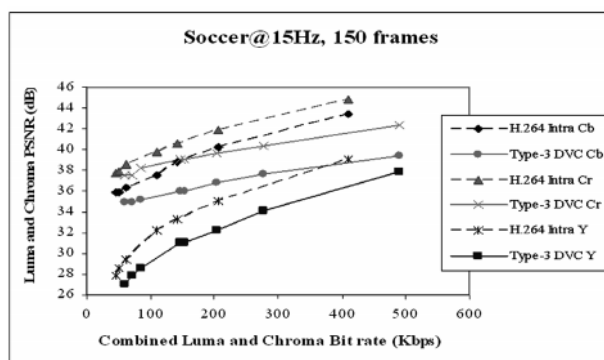
The Figure 6(a) shows RD performance of Soccer sequence luma and chroma components, with various quantization matrices types. From this figure, it can be concluded that RD performance of luma with type-3 chroma quantization parameters outperformed other types. Whereas, chroma components RD performance with type-3 quantization matrices are lower than that of with type-1 and 2. This can be again attributed to relatively very high motion activity in Foreman sequence compared to that of Hall monitor or Coast guard or Foreman sequences. And Figure 6(b) shows comparison of DVC RD performance

(with type-3 chroma quantization parameters) with that of H.264 intra. It can be seen that H.264 intra RD performance achieved for all the three luma and chroma components are better than that of DVC, which is in contrast with other video sequences. This is due to significantly high motion of video sequence and there-by SI quality is not good enough.

The chroma components RD performance achieved is inline with that of luma component, based on motion activity of video sequence. It has been proved that DVC RD performance of all three luma and chroma components (with type-3 chroma quantization matrices) are better than that of H.264 intra for low to very high motion activity video sequences. However for significantly high motion activity video sequences, DVC RD performance is lesser than that of H.264 intra. Hence proposed DVC method works well upto very high motion activity sequences.



(a)



(b)

**Figure 6: Soccer (a) Type-1, 2 and 3 RD Performance Comparison, (b) Type-3 RD Performance Comparison with H.264 Intra.**

#### 4. CONCLUSION

The DVC Architecture, implementation details and results of luma and chroma components are presented in this paper. It has been proved that luma and chroma components RD performance of DVC is better than that of H.264 intra for video sequences upto medium motion activity. For high motion sequences, luma and chroma components RD performance of DVC and H.264 intra are

comparable. Whereas, DVC RD performance of very high motion sequences are lower than that of H.264 intra. Also it has been shown that chroma components RD performance is inline with that of luma component, with all '0' element chroma quantization matrices (type-3). Hence it can be concluded that type-3 chroma quantization matrices can be deployed in DVC for chroma components without any considerable loss in PSNR. That means it is not necessary to encode and send parity bits of chroma components corresponding to WZ frames. At the decoder these chroma components can be reconstructed from chroma components of decoded key frames, using motion compensated frame interpolation method of SI estimation. Hence the objectives of DVC for luma and chroma components are met with proposed method.

#### REFERENCES

- [1] Vijay Kumar Kodavalla, Prasad Bhatt, Udaya Kamath and Vanisha Joseph, "Transition from Smart TV to Smart Viewing - Architectural Tradeoffs to Reach There!", ARM TechCon 2011, Santa Clara, California, USA, October 2011.
- [2] Vijay Kumar Kodavalla and P.G. Krishna Mohan, "Qualitative Distributed Video Coding Paradigm", *International Journal on Graphics Vision and Image Processing (GVIP)*, **11**, Issue 3, pp. 23-30, June 2011.
- [3] Vijay Kumar Kodavalla and P.G. Krishna Mohan, "Distributed Video Coding: Codec Architecture and Implementation", *Signal & Image Processing: An International Journal (SIPIJ)*, **2**, No. 1, pp. 151-163, March 2011.
- [4] Vijay Kumar Kodavalla and P.G. Krishna Mohan, "Distributed Video Coding (DVC): Challenges in Implementation and Practical Usage", IP-SOC 2010, Grenoble, France, November 2010.
- [5] Vijay Kumar Kodavalla and P.G. Krishna Mohan, "Distributed Video Coding: Adaptive Video Splitter", IP-SOC 2010, Grenoble, France, November 2010.
- [6] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov and M. Ouaret, "The Discover Codec: Architecture, Techniques and Evaluation, Picture Coding Symposium (PCS)", Lisboa, Portugal, November 2007.
- [7] R. Puri, A. Majumdar and K. Ramchandran, "PRISM: A Video Coding Paradigm with Motion Estimation at the Decoder", *IEEE Transactions on Image Processing*, **16**, No. 10, pp. 2436-2448, October 2007.
- [8] J. Ascenso, C. Brites and F. Pereira, "Improving Frame Interpolation with Spatial Motion Smoothing for Pixel Domain Distributed Video Coding", *5<sup>th</sup> EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services*, Smolenice, Slovak Republic, July 2005.
- [9] D. Varodayan, A. Aaron and B. Girod, "Rate-Adaptive Codes for Distributed Source Coding", *EURASIP Signal Processing Journal, Special Section on Distributed Source Coding*, **86**, No. 11, November 2006.
- [10] J. Ascenso, C. Brites and F. Pereira, "Content Adaptive Wyner - Ziv Video Coding Driven by Motion Activity", *IEEE International Conference on Image Processing*, Atlanta, USA, October 2006.