Using Wavelet Transform Self-Similarity for Effective Multiple Description Video Coding

Roya Choupani
Faculty of Electrical Engineering,
Mathematics and Computer Science
Delft University of Technology
Delft, the Netherlands
e-mail: r.choupani@tudelft.nl

Stephan Wong
Faculty of Electrical Engineering,
Mathematics and Computer Science
Delft University of Technology
Delft, the Netherlands
e-mail: J.S.S.M.Wong@tudelft.nl

Mehmet Tolun
Department of Electrical Engineering
Aksaray University
Aksaray, Turkey
e-mail: mehmet.tolun@aksaray.edu.tr

***Abstract*—Video streaming over unreliable networks requires preventive measures to avoid quality deterioration in the presence of packet losses. However, these measures result in redundancy in the transmitted data which is utilized to estimate the missing packets lost in the delivered portions. In this paper, we have used the self-similarity property if the discrete wavelet transform (DWT) to minimize the redundancy and improve the fidelity of the delivered video streams in presence of data loss. Our proposed method decomposes the video into multiple descriptions after applying the DWT. The descriptions are organized in such a way that when one of them is lost during transmission, it is estimated using the delivered portions by means of self-similarity between the DWT coefficients. In our experiments, we compare video reconstruction in the presence of data loss in one or two descriptions. Based on the experimental results, we have ascertained that our estimation method for missing coefficients by means of self-similarity is able to improve the video quality by 2.14dB and 7.26dB in case of one description and two descriptions, respectively. Moreover, our proposed method outperforms the state-of-the-art Forward Error Correction (FEC) method in case of higher bit-rates.**

***Index Terms*—Multiple Description Coding, Video Transmission Error, Discrete Wavelet Transform, Self-Similarity.**

I. Introduction

Multiple Description Coding (MDC) methods are utilized for improving the error robustness of data transmission over unreliable networks. MDC methods provide error robustness by decomposing a certain video into various descriptions and transmitting each description over preferably an independent network channel [18]. The descriptions should be encoded in such a way that each stream is decodable independently [16]. Moreover, each delivered description should improve the quality of the reconstructed video regardless of the location of the delivered description data in the original video [9]. This decomposition should be optimized in such a way that the quality of the reconstructed video is maximized in case of a loss of one or more descriptions, and simultaneously maintaining the optimal coding efficiency by minimizing the redundancy in descriptions. The possibility of reconstructing the video (although in lower quality) when some of the descriptions are lost is provided by including redundant data in descriptions. Besides, the coding efficiency is reduced due to the elimination of the correlation between data during the decomposition. Minimizing the redundancy and improving the coding efficiency on the other hand, deteriorates the quality of video and results in distortions when some of the description are not delivered. Hence, a tradeoff between the encoder performance in terms of bit-rate and the imposed distortion is sought by adjusting coding parameters according to the channel conditions. In the present work we address the problem of minimizing the inaccuracy of the reconstructed video in presence of data loss or corruption. Our approach to the problem is based on utilizing the correlation present in order to estimate/iterpolate the missing data. Our proposed method utilizes the self-similarity feature of the Discrete Wavelet Transform (DWT) to estimate the missing data. We have presented a review of related works, the details of our proposed method, and its experimental evaluation in the following sections.

II. Related Work

A significant number of video coding methods using MDC schemes have been reported in literature [4][15][3][1]. A comprehensive overview paper on MDC methods is presented in [16]. Improving the robustness of MDC methods against packet loss through data redundancy [1] or selective protection of descriptions [21][10] reduces the bit-rate performance of the encoder [20]. In [13] the authors propose an algorithm to control the mismatch between the prediction loops at the encoder and decoder in multiple description (MD) video coders with motion-compensated predictions. They consider three different cases; one in which both descriptions are received and other two when either of individual descriptions is received. In [1] the authors propose to generate multiple scalable descriptions from a single SVC bit-stream by mapping scalability layers of different frames to different descriptions. Their scheme is intended for P2P streaming over multiple multicast trees and features several encoding parameters, such as base layer rate of descriptions and overall redundancy. They aim to optimize the mean rate-distortion performance of each description received over a packet loss network, range of extraction points of the SVC stream, and overall redundancy of their MDC scheme.
DWT-based MDC methods are also utilized together with SVC [11][2][6][5][12]. 3D DWT with MCTF is used in MDC methods where they either directly perform MCTF on the input video sequence before the spatial transform, or in the

978-1-4673-7218-3/15/$31.00 ©2015 IEEE

wavelet subband domain which is often referred to as in-band MCTF. Decomposing the DWT coefficients into independent descriptions are generally based on the spatial oriented trees introduced in [14]. In [15] the decomposition is carried out by dividing the DWT coefficients at each level into blocks of equal sizes, and obtaining the descriptions by distributing the blocks among them. However, to create balanced descriptions, the authors encode each block in both low and high distortion rates. Each description then contains low distortion coded versions of some of these blocks, and high distortion versions of the rest. The redundancy added in this way makes the method robust against the packet losses where they replace the missing low distortion blocks of the lost description with their high distortion version from the delivered description.

In [7], the authors propose a method which uses the scalability features of 3D DWT through the application of a t+2D wavelet transform [19] to each GOP. Subsequently, the authors divided the wavelet coefficients into three descriptions by utilizing a modified zigzag scanning methodology. Finally, based on the required quality and the date rate of each network channel, the descriptions were scaled by optimizing their threshold values. All missing data are replaced with zeros before reconstruction.

III. Self-Similar Descriptions

The presented method for combining MDC with SVC in wavelet domain has the robustness in terms of the network errors, and flexibility in terms of the bandwidth usage changes. In case of error when one or some descriptions are lost, the video is reconstructed by estimating the lost coefficients using the delivered coefficients before applying an inverse DWT transform. The estimation of the lost coefficients is performed by utilizing the self-similarity of data after applying DWT transform which is explained as follows. After applying DWT, most of the coefficients in the high frequency bands have very small absolute values. These small values are replaced by zeros after the quantization step [8], [17]. The discrete wavelet transform however, has the extra characteristic of self similarity. If we consider a multi-layer decomposition of an image using DWT, where the lower levels correspond to higher frequencies and higher levels correspond to lower frequencies, we can easily observe a decrease of energy when moving from a higher level to a lower level. Furthermore, if coefficients at a low level contain small energy, their corresponding coefficients at the same spatial orientation at a higher level will also contain low energy. This similarity between the coefficients at similar spatial locations of a multi-layer wavelet decomposition is called self similarity characteristic. This characteristic is a property of natural images since the object boundaries in these images are not completely sharp and are reflected at different frequency levels. The self-similarity characteristic of the DWT can be exploited to interpolate the missing data providing better bit error rate in video streams.

A. Organizing DWT Coefficient in Self-Similar Descriptions

In the method presented here, the wavelet coefficients as depicted in Figure 1, are decomposed into three descriptions. The wavelet transform is repeated twice and the low frequency part of the coefficients is repeated redundantly in each description. The wavelet coefficient content of the descriptions are as given in Table I. The labels LLLL, LLHL, HL, LLHH, HH, LLLH, and LH refer to the group of wavelet transform coefficients as depicted in Figure 1. The low frequency coefficients (LLLL) are repeated in each description hence always a minimum level of fidelity in the reconstructed video is guaranteed. The reconstruction in presence of error or loss of a description is carried out by estimating the missing coefficients with the corresponding coefficients in other sub-bands. For each description, we have computed a parameter termed as similarity coefficient (ζ) which indicates the average ratio of the low frequency sub-band coefficients to their corresponding high frequency coefficient. Besides, a new scheme for structuring the descriptions is proposed. The new scheme provides the facility of utilizing the self-similarity characteristic of DWT for estimating the coefficients of the missing description. In our proposed method, we define three coefficient groups as:

TABLE I
The coefficients included in each description.

Description Number	Coefficients Included
Description 1	LLLL, LLLH, LH
Description 2	LLLL, LLHH, HH
Description 3	LLLL, LLHL, HL

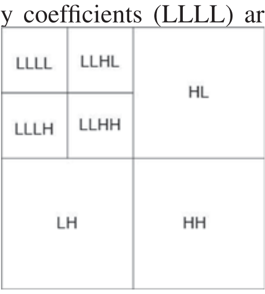

Fig. 1. 2D Wavelet transform coefficients.

Group 1	LLLL, LLLH, LH
Group 2	LLLL, LLHH, HH
Group 3	LLLL, LLHL, HL

The main idea in the proposed method is that when some of the coefficients in a coefficient group are lost, they can be estimated by means of the existing self-similarity. However, if we decompose the coefficients in a way that each coefficient group is transmitted in one description, in case of a loss or corruption in the description, the whole coefficient group is lost. Therefore, estimating the values of the lost coefficients by means of self-similarity will not be possible. However, if each description contains coefficients from different coefficient groups, then in case of a description loss, the coefficients can be estimated from the delivered description. The new organization of the coefficient groups in the descriptions is presented in Table II.

TABLE II
THE COEFFICIENTS INCLUDED IN EACH RE-ORGANIZED DESCRIPTION.

Description Number	Coefficients Included
Description 1	LLLL, LLLH, HH
Description 2	LLLL, LLHH, HL
Description 3	LLLL, LLHL, LH

The self-similarity between LLHL and HL for instance can be utilized to estimate the coefficients when one of these groups is lost. In case that HL is not available, LLHL can be up-sampled for an estimation, and when LLHL is lost, HL is down-sampled to obtain an approximation for LLHL. A similar method is used for (LLLH, LH) and (LLHH, HH) coefficient groups. As mentioned above, the proposed method ensures that the groups of self-similar coefficients are always transmitted in distinct descriptions In this way, by assuming that only one description is lost, the reconstructor will receive one coefficient group completely while, the remaining two coefficient groups are received partially. The partial coefficient groups can be completed by either up-sampling or down-sampling the available coefficients.

B. *Reconstructing the Video*

The frame reconstruction is presence of data loss in one or two descriptions is explained below, and the trivial case of no packet loss is not explained. We have observed that the self-similarity in a description is directly proportioned to the frequency content of the macro-block. This means that although there exists a strong correlation between the DWT coefficients at a sub-band, the ratio of the low frequency coefficients to the high frequency coefficients varies with the content of the block. Hence, we define a self-similarity index for each macro-block which is computed for the coefficients at (LH, LLLH) group as given in Equation 1. A similar method is used for computing the similarity index in other sub-bands.

$$\xi = \frac{1}{m^2}\left\|\frac{[\uparrow LLLH]_{ij}}{[LH]_{ij}}\right\|, \quad [LH]_{ij} \neq 0 \tag{1}$$

where $[LL]_{ij}$ indicates the matrix element at ij position, and the division of LLLH by LH is an element-wise division. Besides, m^2 is the number of non-zero coefficients in $[LH]$, $\uparrow$ is used to represent up-sampling operation, and ξ is the self-similarity index. $\|A\|$ is the matrix norm as defined in Equation 2.

$$\|A\| = \sum_i \sum_j |A_{ij}| \tag{2}$$

Self-similarity index for each description is encoded and transmitted with the current description and its following description in a circular manner. This means that the similarity index of (LLLH, LH) coefficient group is transmitted in descriptions 1 and 2, the similarity index of (LLHH, HH) coefficient group is transmitted in descriptions 2 and 3, and finally the similarity index of (LLHL, HL) coefficient group is transmitted in descriptions 3 and 1. The similarity index values are rounded to nearest integer and an upper limit of 8 has been considered for their values (similarity index values greater than 8 are considered as 8).

1) Case 1: One Description is Lost: Assuming description 2 is lost the decoder should estimate the low frequency coefficients at LLHH and high frequency coefficients at LH. The similarity index of (LLHH, HH) coefficient group is included in description 3 as well. Therefore, down-sampling coefficients at HH and multiplying them by their corresponding similarity index provides the estimation of the missing coefficients. Similarly, the missing coefficients at LH are estimated by up-sampling LLLH coefficients and multiplying them by the similarity index included in description 1.

2) Case 2: Two Descriptions are Lost: Assuming description 2 and 3 are lost the decoder should estimate the low frequency coefficients at LLHH and LLHL, and high frequency coefficients at LH and HH. The similarity indices and coefficients LLLH and HL transmitted in description 1 are utilized to estimate LLHH and LH. As a result, description 2 is estimated from the coefficients and similarity indices delivered with description 1. However, description 3 is cannot be estimated by the proposed method and its coefficients are replaced with zeros.

IV. EXPERIMENTAL RESULTS

The proposed method is experimentally verified using several video sequences. In order to verify the performance of our method, we considered two cases of packet losses as below:

- Only one description is lost. In this case the information in the delivered descriptions is utilized for reconstructing the video.
- Two descriptions are lost. Since part of the coefficients belonging to the adjacent description is in the delivered description, our proposed method is able to reconstruct one of the lost descriptions.

In both experimental cases mentioned above we repeated the experiments by changing the missing description. The GOP length has been fixed to 32 frames. The DWT transform is applied twice as depicted in Figure 1.

To emphasize the important impact of the self-similarity feature of the DWT in reconstruction in presence of data loss, we have compared the reconstructed video when the missing description is estimated using self-similarity feature, and the same video when the coefficients at the missing description are replaced with zeros. Figures 2 and 3 provide the comparative results for one description loss in low and high bit rates respectively which are averaged over the blocks of each frame. The experimental results provided in Figures 2 and 3 indicate that the self-similarity based estimation of the missing data is more effective in higher bit-rates which can be related to the fact that in low bit-rates the higher frequency coefficients are mostly zeros. Despite the fact that the results are not much different in low bit-rates, in high bit-rates we can see an average improvement of 2.14(dB) in terms of PSNR values in Figure 3.

Our next experiment is evaluation of the method when one description is lost. In [21] the authors combine layered coding

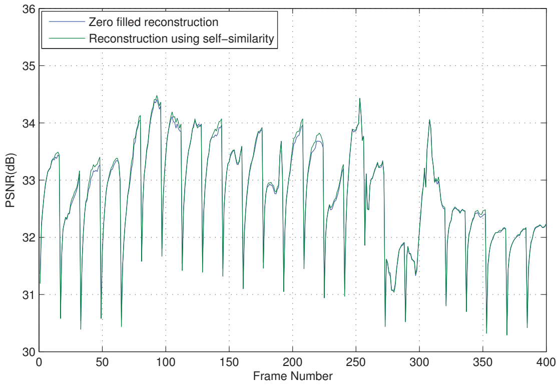

Fig. 2. PSNR values of the reconstructed frames by replacing missing coefficients with zero, and estimating using self-similarity in low bit-rate.

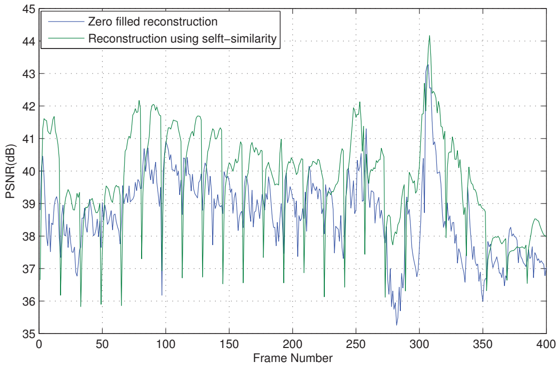

Fig. 3. PSNR values of the reconstructed frames by replacing missing coefficients with zero, and estimating using self-similarity in high bit-rate.

MDC methods for error-resilient video transmission over unreliable channels. They used unequal loss protection to provide the base layer with the highest level of channel error protection through the use of Forward Error Correction (FEC) coding. In order to cope with the network congestion which is main cause of packet losses, they have considered erasure codes for data protection. The FEC code creates redundancy in the transmitted video which makes the method proposed in [21] similar to our proposed method as our proposed method repeats the low frequency coefficients in all descriptions. The authors in [21] divide a bitstream into two portions where the first portion (b_1) consists of the base layer and is further divided into sub-bitstreams. The second portion (b_2) includes the enhancement layers and is also divided into sub-bitstreams. A description is created by including x sub-bitstreams from b_1 and y sub-bitstreams from b_2. As it is assumed the descriptions are transmitted over channels with different probability of data loss, they are protected against packet losses in an unbalanced way through FEC. Besides, in [1] the authors propose a SVC method which decomposes the video into multiple descriptions. Their method combines video segments coded at high and low rates and transmits the high rate segments from one stream together with the low rate segments of the other streams in each description. The low rate coefficients are used for reconstructing the missing description(s) in a lower quality. The authors also propose a Multiple-Objective Optimization (MOO) framework for selection of the best encoding configuration to achieve the best tradeoff between redundancy and reliability. Since their proposed method includes redundancy in each description to attain a better quality level in presence of packet losses, we have considered their method as a benchmark to compare the performance of our MD video coder.

Figure 4 depicts the comparative performance of the proposed method and the methods proposed in [1] and [21] . We have assumed that only one description is lost and later on reconstructed by using the redundancy available in other descriptions. When the self-similarity feature of the DWT is utilized, our proposed method outperforms the other two methods in high bit-rates. One important feature of the method proposed in [1] is that the redundancy is proportional to the intended total bit rate while in our proposed method, the low frequency part of the coefficients are repeated in all descriptions almost independently from the bit rate.

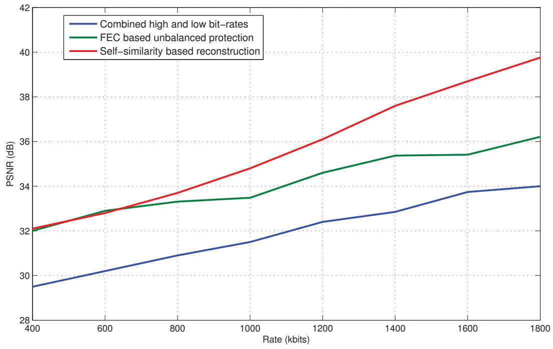

Fig. 4. PSNR values of the reconstructed frames by the proposed method, the combined high-low rate coding method [1], and FEC based unbalanced protection [21] when one description is lost.

In our next experiment we have assumed that two out of three descriptions are lost during transmission. For fair performance analysis and comparison, we have modified the proposed methods given in [21] and [1] in order to include three descriptions. Figure 5 depicts the result of the reconstruction versus average PSNR value. The performance of the proposed method is considerably better in presence of high packet losses such as the case depicted in Figure 5. This result indicates that the proposed method is suitable for transmitting high rate videos over unreliable networks. The experimental results indicate that the proposed method outperforms the traditional video coding methods in presence of frame losses. When two descriptions are delivered, the average PSNR values with and without using self-similarity index for reconstruction are 35.69(dB) and 33.55(dB) respectively. The average PSNR values when only one description is delivered are 34.38(dB) and 27.12(dB) for reconstruction with and without using self-similarity index respectively. The redundancy imposed by repeating the low frequency coefficients of the DWT can be minimized by increasing the number of times the DWT is applied to frame

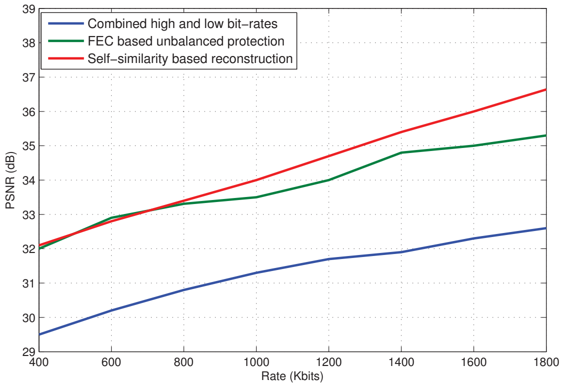

Fig. 5. PSNR values of the reconstructed frames by the proposed method, the combined high-low rate coding method [1], and FEC based unbalanced protection [21] when two descriptions are lost.

data. Moreover, a better performance of the proposed method in case of higher bit-rates indicates that the proposed method is more suitable for streaming over unreliable networks of high bandwidths.

V. Conclusions

A new DWT based video coding method for transmitting video over unreliable networks is proposed. The frame blocks are decomposed into three descriptions after applying the DWT transform. The proposed method improves the performance of the existing MDC methods by utilizing the self-similarity feature of the DWT. The experimental results indicate that the proposed method outperforms the existing methods when the video bit rate and the packet loss rate are high. The redundancy added by repeating the low frequency data in each description can be minimized by increasing the number of times that the DWT applied. However, the number of DWT levels should be optimized with the number of self-similarity index values which are transmitted in descriptions. Besides, the optimization can be performed by considering the available bandwidth of the underlying network.

References

[1] T. B. Abanoz and A. M. Tekalp. SVC-based scalable multiple description video coding and optimization of encoding configuration. *Signal Processing: Image Communication*, 24:691–701, 2009.

[2] N. Adami, A. Signoroni, and R. Leonardi. State-of-the-art and trends in scalable video compression with wavelet-based approaches. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(9):1238–1255, 2007.

[3] E. Akyol, A. M. Tekalp, and M. R. Civanlar. A flexible multiple description coding framework for adaptive peer-to-peer video streaming. *IEEE Journal of Selected Topics in Signal Processing*, 1:231–245, 2007.

[4] M.R. Ardestani, A. A. Beheshti Shirazi, and M. R. Hashemi. Low-complexity unbalanced multiple description coding based on balanced clusters for adaptive peer-to-peer video streaming. *Signal Processing: Image Communication*, 26:143–161, 2011.

[5] M. Biswas, M. R. Frater, and J. F. Arnold. Multiple description wavelet video coding employing a new tree structure. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(10):1361–1368, 2008.

[6] S. Cho and W. A. Pearlman. A full-featured, error-resilient, scalable wavelet video codec based on the set partitioning in hierarchical trees (SPIHT) algorithm. *IEEE Transactions on Circuits and Systems for Video Technology*, 12(3):157–171, 2002.

[7] R. Choupani, S. Wong, and M. Tolun. Scalable video transmission over unreliable networks using multiple description wavelet coding. *The 7th International Conference on Digital Content, Multimedia Technology and its Application (IDCTA2011)*, pages 5–10, 2011.

[8] M. L. Comer, K. Shen, and E. J. Delp. Rate-scalable video coding using a zerotree wavelet approach. *Proceedings of the Ninth Image and Multidimensional Digital Signal Processing Workshop*, pages 162–163, 1996.

[9] V. K. Goyal. Multiple description coding: Compression meets the network. *IEEE Signal Processing Magazine*, 18:74–94, 2001.

[10] F. A. Lopez-Fuentes. P2P video streaming combining SVC and MDC. *International Journal of Applied Mathematics and Computer Science*, 21(2):295–306, 2011.

[11] A. Mavlankar and E. Steinbach. Multiple description video coding using motion-compensated lifted 3d wavelet decomposition. *IEEE International Conference on Acoustics, Speech, and Signal Processing,(ICASSP '05)*, 2:65–68, 2005.

[12] N. Mehrseresht and D. Taubman. A flexible structure for fully scalable motion-compensated 3-d dwt with emphasis on the impact of spatial scalability. *IEEE Transaction on Image Processing*, 15(3):740–753, 2006.

[13] A.R. Reibman, H. Jafarkhani, Y. Wang, M.T. Orchard, and R. Puri. Multiple-description video coding using motion-compensated temporal prediction. *IEEE Transaction on Circuits and Systems for Video Technology*, 12:193–204, 2002.

[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*, 6(3):243–250, 1996.

[15] T. Tillo, M. Grangetto, and G. Olmo. Multiple description image coding based on lagrangian rate allocation. *IEEE Transaction on Image Processing*, 16(3):673–683, 2007.

[16] R. Venkataramani, G. Kramer, and V.K. Goyal. Multiple description coding with many channels. *IEEE Transaction on Information Theory*, 49:2106–2114, 2003.

[17] Q. Wang and M. Ghanbari. Scalable coding of very high resolution video using the virtual zerotree. *IEEE Transactions on Circuits and Systems for Video Technology*, 7(5):719–727, 1997.

[18] Y. Wang, A. R. Reibman, and L. Shunan. Multiple description coding for video delivery. *Proceedings of IEEE*, 93:57–70, 2005.

[19] M. Weeks and M.A. Bayoumi. Three-dimensional discrete wavelet transform architectures. *IEEE Transactions on Signal Processing*, 50(8):2050–2063, 2002.

[20] M. Wien, H. Schwarz, and T. Oelbaum. Performance analysis of SVC. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(9):1194–1203, 2007.

[21] W. Xiang, C. Zhu, C. K. Siew, Y. Xu, and M. Liu. Forward error correction-based 2-d layered multiple description coding for error-resilient H.264 SVC video transmission. *IEEE Transaction on Circuits and Systems for Video Technology*, 19(12):1730–1738, 2009.