MOTION-COMPENSATED WAVELET VIDEO CODING USING ADAPTIVE MODE SELECTION

Fan Zhai and Thrasyvoulos N. Pappas

Department of Electrical and Computer Engineering
Northwestern University, Evanston, IL 60208, USA
E-mail: {fzhai, pappas}@ece.northwestern.edu

ABSTRACT

A motion-compensated wavelet video coder is presented that uses adaptive mode selection (AMS) for each macroblock (MB). The block-based motion estimation is performed in the spatial domain, and an embedded zerotree wavelet coder (EZW) is employed to encode the residue frame. In contrast to other motion-compensated wavelet video coders, where all the MBs are forced to be in INTER mode, we construct the residue frame by combining the prediction residual of the INTER MBs with the coding residual of the INTRA and INTER_ENCODE MBs. Different from INTER MBs that are not coded, the INTRA and INTER_ENCODE MBs are encoded separately by a DCT coder. By adaptively selecting the quantizers of the INTRA and INTER_ENCODE coded MBs, our goal is to equalize the characteristics of the residue frame in order to improve the overall coding ef£ciency of the wavelet coder. The mode selection is based on the variance of the MB, the variance of the prediction error, and the variance of the neighboring MBs' residual. Simulations show that the proposed motion-compensated wavelet video coder achieves a gain of around 0.7-0.8dB PSNR over MPEG-2 TM5, and a comparable PSNR to other 2D motion-compensated wavelet-based video codecs. It also provides potential visual quality improvement.

1. INTRODUCTION

A hybrid motion-compensated video encoder usually consists of two major parts: the generation and compression of the motion vector £elds and compression of prediction error frames. A given macroblock (MB) can be intraframe coded, interframe coded using motion compensated prediction, or simply replicated from the previously decoded frame. These prediction modes are denoted as INTRA, INTER, and SKIP mode, respectively. Quantization and coding are performed differently for each MB according to its mode. The discrete cosine transform (DCT) is widely used in H.261, H.263, and MPEG standards to transform and encode the Intra frame and the prediction error because of its simplicity and ef£ciency [1, 2].

During the last decade, the discrete wavelet transform (DWT) has gained increased popularity in image coding due to the breakthrough work of Shapiro [3], Said and Pearlman [4], JPEG2000 [5], and others. JPEG2000, which makes use of wavelet and subband technologies, is a new standard for still image intended to overcome the shortcomings of the existing JPEG standard. The core compression is primarily based on Embedded Block Coding with Optimized Truncation (EBCOT) of the bitstream. Recently, there has also been active research applying the DWT to video coding [6–12]. The wavelet representation provides a multiresolution/multiscale expression of a signal with localization in both time and frequency. One of the advantages of DWT in both still image and video compression is that it is free of blocking artifacts. In addition, it offers the advantage of continuous data rate scalability, which is an important issue for video applications such as digital libraries, video database systems, and video streaming [9, 10]. For example, it is easier to do rate control for progressive source bitstreams, since they are highly ¤exible in adapting to time-varying channels.

Although 3D wavelet or subband video codecs have the inherent feature of full scalability [6–8], the coding ef£ciency is not competitive due to the inef£cient temporal £ltering in 3D technologies. For this reason, motion compensation is still the best known technique to remove temporal redundancy and improve coding ef£ciency in video sequences. One promising scheme is to combine motion compensation with the 3D wavelet or subband coding, such as the work in [13, 14]. However, the coding ef£ciency is still not satisfactory. In addition, 3D codecs usually require more computation and introduce longer

delay than 2D technologies. This work focuses on 2D wavelet with motion compensation. Pixel-based or dense motion £eld methods provide a fairly general description of motion, but are computationally expensive and involve a large amount of data [15]. Block-based motion estimation, on the other hand, requires fewer operations, and generates much less motion information. Another possibility is variable block size motion estimation. Compared to £xed-blocksize motion estimation, variable blocksize motion estimation is more general and can better adapt to motion discontinuities, though it requires a lot of computation [11, 12]. Thus, for both simplicity and ef£ciency, £xed blocksize motion estimation is still the most commonly used method in predictive coding.

In this work, we consider using £xed size block-based motion compensation in the spatial domain, and DWT on the resulting residue frame. In using block based coding (such as DCT), the prediction mode of each MB is usually decided based on the comparison of its variance and prediction error variance. In wavelet video coders, all MBs are usually set as INTER mode, and their prediction errors are grouped together to form a residue frame, which is then transformed by DWT and coded by one of the popular techniques such as EZW (embedded zerotree wavelet) coder in [3] and SPIHT (set partitioning in hierarchical trees) coder in [4]

One problem that arises here is that the residue frame may have large local variance due to motion compensation, which will make the use of DWT-based coders ineff£cient. Compared with natural images, the residue image typically contains important image characteristics such as sharp transitions, edges, or other singularities over different scales. One way to deal with this problem is to use nonlinear wavelets instead of classical linear wavelets whose representations are ill-suited for representing edge information [16, 17]. Another solution is to use adaptive subband structures to perform transformation, such as in [18]. Instead of tuning DWT to match the characteristics of residue images, we adjust the MB mode selection with the objective of smoothing the resulting residue frame, so that it matches the DWT's characteristics. In contrast to other motion-compensated wavelet video coders, where all the MBs are forced to be in INTER mode, we construct the residue frame by combining the prediction residual of the INTER MBs with the coding residual of the INTRA and INTER_ENCODE MBs. Different from INTER MBs that are not coded, the INTRA and INTER_ENCODE MBs are encoded separately by a DCT coder or by encoding only the DC coef£cient. By adaptively selecting the quantizers of the INTRA and INTER_ENCODE coded MBs, our goal is to equalize the characteristics of the residue frame in order to improve the overall coding ef£ciency of the wavelet coder. This paper presents a simple but ef£cient mode selection strategy, which enables mode switching for each MB based on the comparison result of its variance, the variance of the prediction error, and the variance of the neighboring MBs. This scheme adds little computational burden, while resulting in a considerably smoothed residue frame. This makes DWT work more ef£ciently, even if additional bits have to be used to encode the INTRA MBs. Experimental results demonstrate around 0.7-0.8dB PSNR gain over MPEG-2 TM5 at typical working bit rates, as well as improved visual quality. In addition, it has a comparable PSNR to other 2D motion-compensated wavelet-based video codecs and provides potential visual quality improvement.

The remainder of the paper is organized as follows. Section 2 explains why different MB modes are necessary for wavelet encoding and describes some other components used in the video coder. Section 3 describes the mode selection strategy. Simulation results are presented in Section 4. Section 5 draws conclusions and lays out future work directions.

2. WAVELET VIDEO CODER WITH ADAPTIVE MB MODE SELECTION

2.1. Why different MB modes are necessary?

When applying the DWT to motion-compensated video compression, one approach is to treat all MBs as INTER MBs no matter how poor the estimates are, such as the case with Scalable Adaptive Motion Compensated Wavelet (SAMCoW) video coder [9]. In SAMCoW, the motion estimation is block-based with £xed block size, all MBs are INTER MBs, and DWT is then applied on the resulting residue frame, which usually contains a great amount of sharp transitions, object edges, and singularities. On the other hand, in the wavelet video coder proposed by Yang *et al.* in [10], there are three mode blocks, INTRA, INTER and SKIP, and each block is £rst labeled with one of those modes. Different regions are then formed by grouping blocks with the same mode. Upon the initial allocation of bits among the regions, coding is performed region by region. So, intuitively, SAMCoW can be improved by £rst labeling each MB as different mode (INTER/INTER_ENCODE/INTRA in this work), and then separately encoding the different mode MBs. Since the residue frame will be encoded by a wavelet coder, INTRA and INTER_ENCODE mode MBs could be separately encoded to a rough extent beforehand. Then their errors, together with INTER MBs' prediction errors, will form the residue frame, which may be smoother than it would be without using INTRA/INTER_ENCODE mode.

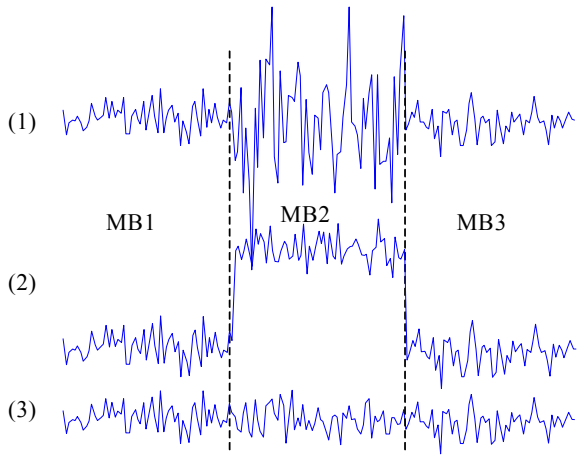


Fig. 1. Illustration of different MB prediction errors in different cases.

To illustrate the above idea, prediction errors of MBs are shown as one-dimensional curves in Fig. 1, where MB1, MB2, and MB3 are neighboring MBs. The case (1) curves denote the prediction errors of the three MBs when they are all treated as INTER MBs. Because MB2's prediction error variance is much larger than that of its neighbors', it will introduce a lot of local discontinuities in the corresponding residue frame. It is easily understandable that the ef£ciency of DWT depends not only on the global variance but also on the local variance [17, 19]. Thus, a better method is to label this MB as INTRA or INTER_ENCODE (depending on whether the prediction error variance of intra MB variance is smaller) and code it separately. This is done in MPEG-2 and H.263. Here, since we will encode the whole residue frame with DWT, it is quite enough to just roughly encode this MB, and group the corresponding differences in the residue frame.

In case (2), the part of MB2 curve stands for either its prediction error or the reference MB itself. Its variance is close to that of its neighbors', while its mean is pretty far from zero. Obviously, MB2 should be coded separately as well in this case. The best way to do that is simply encoding (quantizing) its mean, and putting the remaining error in the residue frame. In case (1) and (2), depending on whether the prediction error variance of intra MB variance is smaller, MB2 will be labeled as INTRA or INTER_ENCODE. Other MBs such as those in case (3) are in INTER mode with no need of separate encoding. After mode labeling, the residue frame will end up with the curve in Fig. 1 (3) for both case (1) and case (2). By separately coding INTRA/INTER_ENCODE mode MBs, the important image characteristics such as object boundaries are maintained. In addition, the resulting smooth residue frames make DWT more ef£cient.

2.2. Quantization stepsize

In our scheme, DCT is employed to encode INTRA and INTER_ENCODE MB due to its simplicity and ef£ciency. An important issue is the determination of the quantization stepsize. The quantization stepsize should be determined based on the variance of the current MB as well as its neighboring MBs so that the prediction error has uniform stochastics across MBs in the resulting residue frame.

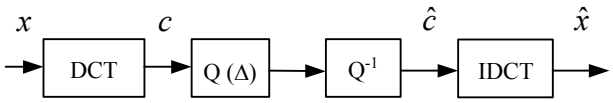


Fig. 2. Encoding model of INTRA and INTER_ENCODE MBs.

The encoding model of INTRA and INTER_ENCODE MB is depicted in Fig. 2. Let x denote the pixel values of an MB to be coded, and c be its DCT coef£cients. After quantization using quantization step Δ and inverse quantization, the reconstructed coef£cients are $\hat{c}$. The corresponding reconstructed pixel values are $\hat{x}$ after inverse DCT is applied. The next step is to determine the quantization stepsize Δ to achieve the desired error variance of x, $Var(x - \hat{x})$, which is dependent

on the neighboring MBs' residue variance. Since the DCT is an orthogonal transform, we have

$$c = AxA^T$$

where A is the DCT transformation vector and $AA^T = 1$. Therefore, the encoding error energy is identical to the quantization noise energy, shown as:

$$|x - \hat{x}|^2 = |c - \hat{c}|^2.$$

Assuming the quantization noise is uniformly distributed within $(-\Delta/2, \Delta/2)$, the expected quantization noise energy is equal to $\Delta^2/12$. If we differentially encode the DC coef£cient of c using the same technique as what is used in MPEG-2 and H.263, we can ignore the DC coef£cient, which is the mean of x, and treat it as zero. In this case, the desired error variance, $Var(x - \hat{x})$, could be roughly described as $\Delta^2/12$. Then the quantization stepsize Δ can be expressed as

$$\Delta = (12 \times Var(x - \hat{x}))^{1/2}. \quad (1)$$

2.3. Overlapped block motion compensation (OBMC)

One problem with block based motion compensation video coding is that it introduces blocky edges in the residue frame, which cannot be ef£ciently coded by using DWT. Although DWT is a global representation of the residue frame and is in itself free from blocking artifacts, block based motion compensation may introduce unpleasant blocking artifacts.

Overlapped block motion compensation (OBMC) is so far one of the simplest and most ef£cient ways to further reduce blocking artifacts. The basic idea of OBMC is that given motion vectors of adjacent MBs, the prediction of the current MB is the weighted sum of its estimates based on motion vectors from adjacent MBs and the current MB. The weighting matrices we use are described in [9, 20]. Based on our simulation results, this method not only improves the visual quality of the picture, but also increases the PSNR. INRA and INTER_ENCODE mode MBs can also use OBMC to reduce the blocky edges that block based motion compensation might introduce.

3. MODE SELECTION STRATEGY

In our scheme, three modes of MB, INTRA, INTER, and INTER_ENCODE are used. The operations of the three modes MB are de£ned below.

INTRA MB: DCT is used to encode this MB. After quantization, inverse quantization, and inverse DCT, the difference from the original frame is put into the residue frame.

INTER MB: The same as what is de£ned in MPEG-2 standard, but only nonzero motion vectors need to be transmitted to the bit stream for INTER MB (zero motion vectors will be skipped). The difference between the predicted MB and the original MB will be put into the residue frame.

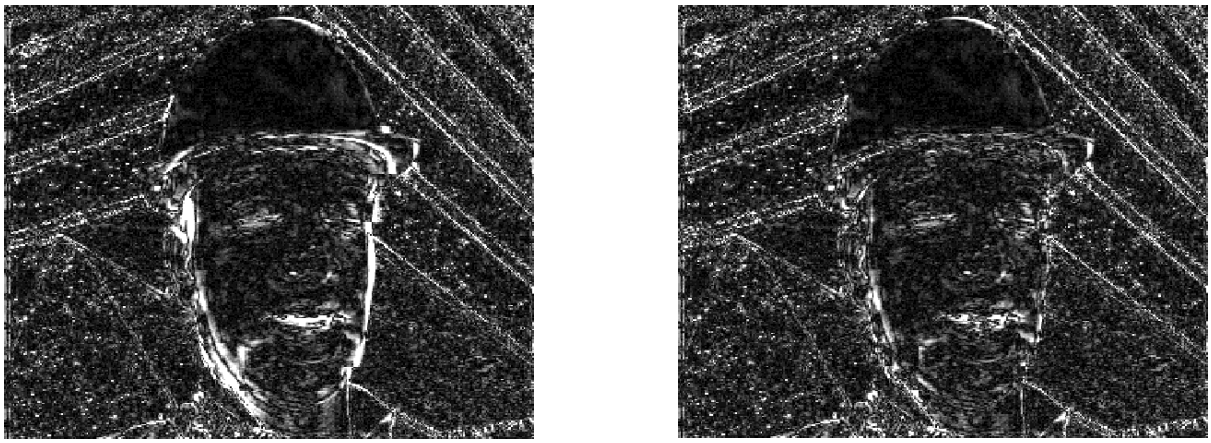
INTER_ENCODE MB: The same as INTER MB, the prediction error is £rst obtained by motion compensation. DCT is then used to further reduce the variance of the prediction error. The re£ned prediction error is then put into the residue frame.

The goal of mode selection is to detect the local discontinuities and smooth down the prediction error frame. The strategy of mode selection for one MB is to adaptively make the decision based on its variance (var_intra), the smallest prediction error variance (var_inter), and its neighboring MBs' residue variance ($var\_neighbor$). Here, we choose $var\_neighbor$ as the average of its left, top and top-right MBs' variances, due to the causality consideration. Let A, B, C, and D be pre-de£ned constants, the proposed mode selection strategy is described in Fig. 4.

The constants A, B, C, and D are chosen to balance the proportions of the three modes. In fact, if A or B is bigger, it favors INTER mode. D is chosen to balance te INTRA and INTER_ENCODE modes. If D is smaller, it favors INTRA mode. In our current simulation, the constants A, B, and D are chosen as 5, 0.5, and 2, respectively. C is closely dependent on the target bit rate, R. The higher the target bit rate, the smaller the C. In our simulations, we have observed that

$$C = (20 - 10^{-2}R) \times 256,$$

where R is in the unit of kbps, can generally achieve good performance when R ranges from 500kbps to 1.5Mbps.



(a) (b)

Fig. 3. Comparison of residue frames (a) without adaptive mode selection (b) with adaptive mode selection.

if $(var\_inter < A \times var\_neighbor \,||\, var\_inter < B \times var\_intra \,||\, var\_inter < C)$
    INTER mode
else if $(var\_inter < D \times var\_intra)$
    INTER_ENCODE mode
else
    INTRA mode

Fig. 4. Adaptive mode selection strategy.

4. SIMULATION RESULTS

Figure 3 depicts the residue frame quality improvement. In Fig. 3(a), all MBs are treated as INTER modes. Large discontinuities could be observed in the residue frame. After adaptive mode selection, and corresponding coding, the residue frame is shown in Fig. 3(b). The proposed mode selection strategy works well to detect those MBs that introduce large discontinuities.

The proposed wavelet video coder is evaluated using the Foreman test sequence with CIF (352×288) format at a frame rate 30 frames per second. Since no frame-level rate control was incorporated yet in the proposed coder, the frame bit budget is the same as the corresponding frame bit budget in TM5 coder to constitute a fair comparison. The coding pattern of 15-frame GOP without a B frame is employed in the TM5 test. In our simulations, we consider EZW to perform DWT on the residue frames. The proposed adaptive mode selection strategy is general, and does not depend on speci£c wavelet coder employed.

Figure 5 shows a realization using TM5 codec and the proposed wavelet codec at the bit rate of 1Mbps. As shown in Fig. 6, simulations on a relatively wide bit rate range between 500kbps and 1.5Mbps show that the proposed wavelet video coder reveals a 0.7-0.8dB average PSNR increase over TM5. We can also see that the average PSNR of the wavelet coder with AMS is merely around 0.02dB lower than that of the wavelet coder without AMS. However, the visual quality of reconstructed frame can be improved by using AMS, especially in the region of relatively high local variance such as the moving object boundaries. Figure 7 shows the visual quality comparison of one frame with relatively slow motion using DCT and different DWT schemes, where Fig. 7(a) shows the original frame, and Fig. 7(b), (c) and (d) illustrate an example of one reconstructed frame using DCT, DWT without AMS, and DWT with AMS respectively. It can be seen that Fig. 7(b) is smoother than (c) and (d), especially when the observer is focus on the object boundaries. However, blocking artifacts are obviously noticeable in Fig. 7(a), which is very annoying to human eyes. By comparing Fig. 7(c) and (d), we can see that AMS is effective in maintaining the important image characteristics such as sharp transitions, edges, singularities. Thus, by using AMS and DCT-coding INTRA MBs, we take advantage of both DCT and DWT to improve the visual quality of the reconstructed video sequence. Figure 8 shows another example for frame 186, which involves a lot of motion.

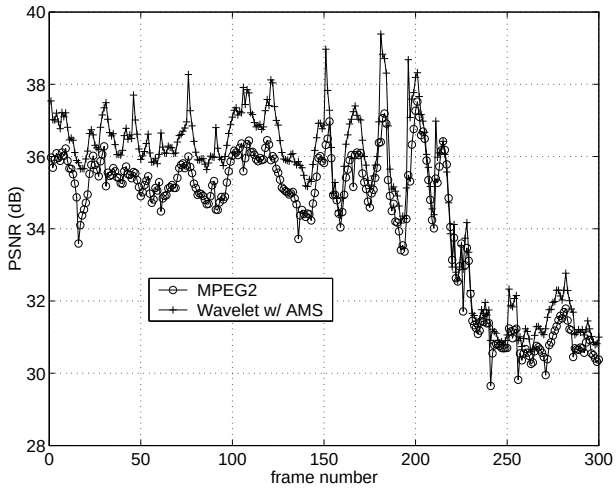


Fig. 5. PSNR comparison of proposed wavelet coder and TM5, (bit rate is 1Mbps).

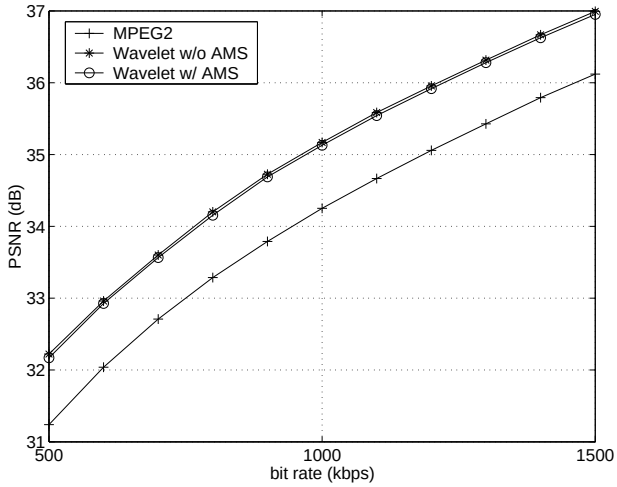


Fig. 6. PSNR vs. bit rate.

5. CONCLUSIONS

This paper introduces a simple adaptive mode selection method in motion-compensated wavelet video coding, which has proved to be effective in smoothing the residue frame, and improving the visual quality of the reconstructed frames. The resulting video coder provides superior coding quality without adding computation complexity. In future research, the constants selected in the proposed mode selection strategy should be further studied and adaptively decided based on the target bit rate. In addition, although OBMC provides certain gain, it does not work consistently for all MBs. Better results might be derived by adaptively enabling or disabling OBMC. A more direct and complete solution should incorporate mode selection within an overall rate-distortion optimization framework to achieve the best performance.

6. REFERENCES

[1] ISO/IEC 12818-2, *Generic Coding of Moving Pictures and Associated Audio*, International Orginization for Standardization, March 1994.

Fig. 7. Comparison of reconstructed frame (frame 16, 700kbps) (a) original frame (b) DCT (c) DWT without AMS (d) DWT with AMS.

[2] ITU, *Video Coding for Low Bitrate Communication,* ITU Telecom. Standardization Sector of ITU, Sept. 1997, Draft ITU-T Recommendation H.263 Version 2.

[3] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coef£cients," *IEEE Trans. Signal Processing*, vol. 41, pp. 3445–3463, Dec. 1993.

[4] A. Said and W. Pearlman, "A new, fast, and ef£cient image codec based on set partitioning in hierarchical trees," *IEEE Trans. on circ. and Syst. for Video Techn.*, vol. 6, pp. 243–250, June 1996.

[5] *JPEG-2000 VM3.1 A Software,* ISO/IECJTC1/SC29/WG1 N1142, Jan. 1999.

[6] Y. Chen and W. A. Pearlman, "Three-dimentional subband coding of video using the zero-tree method," in *Proc. SPIE Visual Communications and Image Processing*, March 1996, pp. 1302–1309.

[7] B.-J. Kim and W. A. Pearlman, "An embedded wavelet video coder using three-dimentional set partitioning in hierarchical trees," in *Proc. of Data Compression Conference*, 1997, pp. 251–260.

[8] J. Xu, Z. Xiong, S. Li, and Y.-Q. Zhang, "Three-dimentional embedded subband coding with optimized truncation (3D ESCOT)," *Applied and Computational Harmonic Analysis 10*, pp. 290–315, 2001.

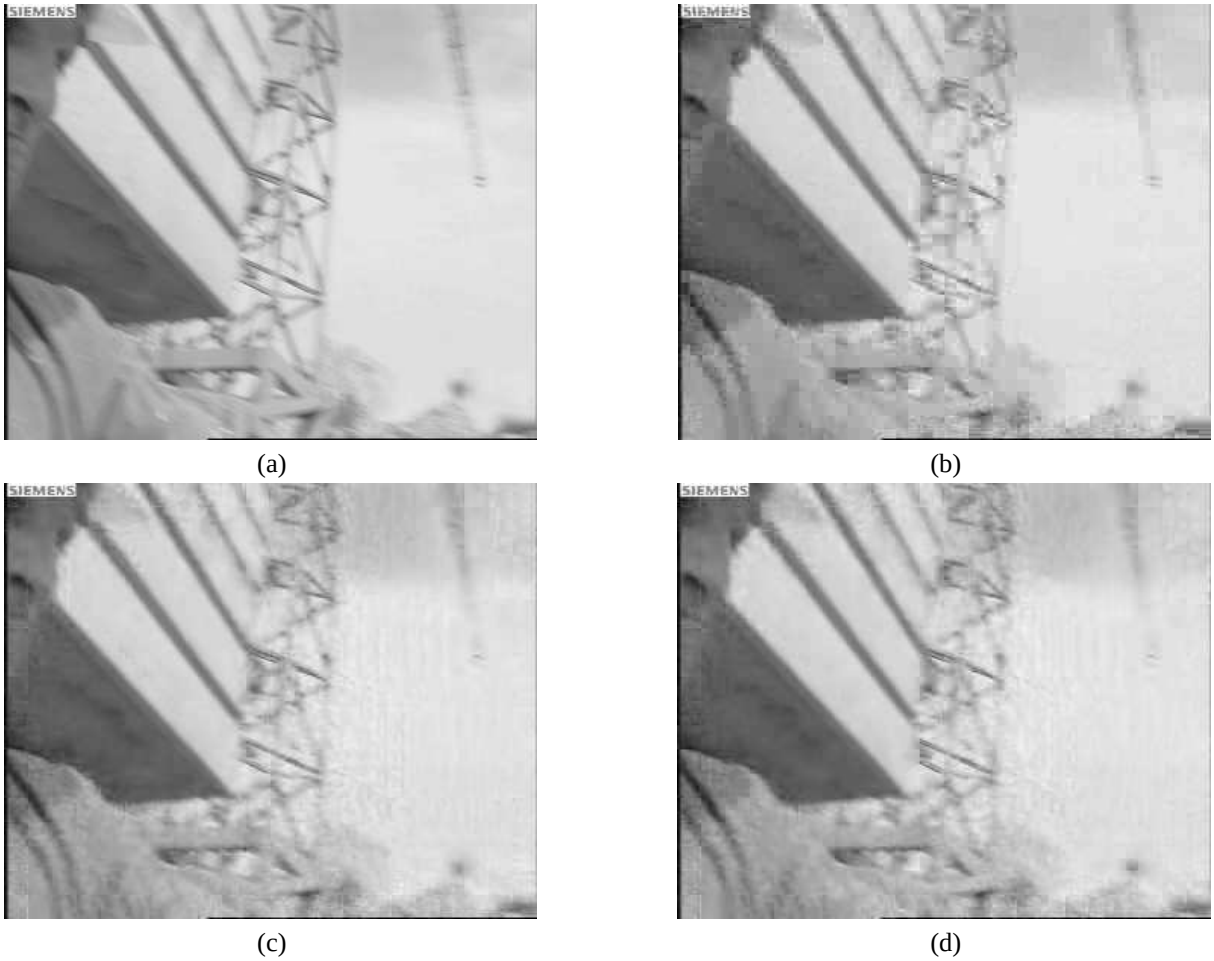


Fig. 8. Comparison of reconstructed frame (frame 188, 700kbps) (a) original frame (b) DCT (c) DWT without AMS (d) DWT with AMS.

[9] K. Shen and E. J. Delp, "Wavelet based rate scalable video compression," *IEEE Trans. circ. and Syst. for Video Techn.*, vol. 9, pp. 109–122, Feb. 1999.

[10] Y. Yang and S. S. Hemami, "Generalized rate-distortion optimization for motion-compensated video coders," *IEEE Trans. on circ. and Syst. for Video Techn.*, vol. 10, pp. 942–955, Sept. 2000.

[11] X. Yang and K. Ramchandran, "Scalable wavelet video coding using aliasing-reducded hierarchical motion compensation," *IEEE Trans. on Image Processing*, vol. 9, pp. 778–791, May 2000.

[12] Y.-Q. Zhang and S. Zafar, "Motion-compensated wavelet transform coding for color video compression," *IEEE Trans. on circ. and Syst. for Video Techn.*, vol. 2, pp. 285–296, Sept. 1992.

[13] J. R. Ohm, "Three-dimentional subband coding with motion compensation," *IEEE Trans. Image Proc.*, vol. 3, pp. 559–571, Sept. 1994.

[14] S. Choi and J. W. Woods, "Motion-compensated 3-D subband coding of video," *IEEE Trans. Image Proc.*, vol. 8, pp. 155–167, Feb. 1999.

[15] S.-C. Han and C. Podilchuk, "Ef£cient encoding of dense motion £elds for motion-compensated video compression," in *Proc. IEEE Int. Conf. Image Processing*, Kobe, Japan, 1999.

[16] H. Heijmans and J. Goutsias, "Nonlinear multiresolution signal decomposition schemes: Part II: morphological wavelets," *IEEE Trans. Image Proc.*, vol. 9, pp. 1897–1913, Nov. 2000.

[17] B. Pesquet-Popescu, H. Heijmans, G. Abhayarantne, and G. Piella, "Quantization of adaptive 2D wavelet decompositions," in *Proc. IEEE Int. Conf. Image Processing*, Barcelona, Spain, Sept. 2003.

[18] Ö. N. Gerek and A. E. Cetin, "Adaptive polyphase subband decomposition structures for image compression," *IEEE Trans. Image Processing*, vol. 9, pp. 1649–1659, Oct. 2000.

[19] B. Tao and M. T. Orchard, "Gradient-based residual variance modeling and its applications to motion compensated video coding," *IEEE Trans. on Image Processing*, vol. 10, pp. 24–35, Jan. 2001.

[20] M. Orchard and G. Sullivan, "Overlapped block motion compensation: An estimation-theoretical approach," *IEEE Trans. Image Proc.*, vol. 3, Sept. 1994.