

# Combining Phase-diversity with Spatial-diversity Multihypothesis Motion Compensation

Suxia Cui, Yonghui Wang

Department of Engineering Technology  
Prairie View A&M University, Prairie View, TX

James E. Fowler

Department of Electrical and Computer Engineering  
Engineering Research Center  
Mississippi State University, Mississippi State, MS

**Abstract**—Discrete Wavelet Transform (DWT) has been proved to be a very efficient tool for data compression. Redundant Wavelet Transform (RDWT) conquers the “shift-variant” problem in DWT and enables the motion estimation and motion compensation (ME/MC) in wavelet domain. Also, RDWT produces growing number of phases at each level of the transform, each phase contributes to the ME/MC procedure by providing a unique hypothesis. Based on the previous successful usage of phase-diversity multihypothesis in redundant wavelet transform domain, this paper presents a technique to combine the spatial-diversity multihypothesis with the transform domain multihypothesis. We take the overlapped block motion compensation (OBMC) as one example from spatial domain multihypothesis. Experiment results show that the combination of the two classes of multihypothesis outperforms either method implemented alone. Further more, choose different subbands for OBMC in order to simplify the system is also discussed.

## I. INTRODUCTION

The modern video coding scheme is based on decorrelation among video sequence. Since a sequence of images are highly correlated both temporally as well as spatially, employ motion estimation and motion compensation (ME/MC) and then code the residual image in transform domain instead of in spatial domain are two efficient temporal and spatial decorrelation methods that are widely used in current video-coding systems. In the video coding standards, such as H.264[1] and MPEG-4 [2], discrete-cosine transform (DCT) is used to decorrelate the residual image. With the recent wavelet-based still-image compression algorithm, such as set partitioning in hierarchical trees (SPIHT) [3], several wavelet-based video coders are proposed. Especially after the “shift-variant” problem existed in DWT is conquered by the redundant discrete wavelet transform (RDWT)[4, 5], how to implement the RDWT in video coding to improve the current video coding technique has attracted more and more attention.

It is obvious that besides the “shift-invariant” property which RDWT has, RDWT also provides different phases during the transform because it gets rid of the subsampling procedure of DWT. So all the DWT coefficients of different phases are preserved and they make their own contribution in motion estimation process. Thus provide more hypotheses in motion compensation compared to the coders in which only one critically sampled DWT remained for compensation. Our previous work successfully developed a coder that deployed the ME/MC by using the multiple predictions in RDWT domain. We name our system redundant wavelet multihypothesis (RWMH) [6, 7]. That is an contribution to the multihypothesis motion compensation (MHMC) [8] by extending the different hypothesis into transform domain.

With the achievement in exploring the phase-diversity multihypothesis, in this paper, we further enhance the performance of our RWMH system by increasing the number of hypothesis. Our research shows that to combine our phase-diversity multihypothesis with spatial-diversity is possible and easy to implement since the RDWT preserves the spatial relation of the original image. Each subband of an RDWT

transform will have the same size as the original image. Any spatial-diversity multihypothesis method can be easily deployed in RDWT domain. OBMC can better represent spatial-diversity and coincide with RWMH. Our experimental results illustrated that by adding OBMC to RWMH system; we can get substantial gain for both low-motion sequences as well as high-motion sequences. Further more, to get rid of some of the calculation complexity, there are choices of implementing OBMC on selected subbands instead of all the RDWT subbands.

This paper is organized as follows. In Sec. II, we provide an overview of the theory behind MHMC. In Sec. III, we introduce our coder that combines the phase -diversity with spatial-diversity multihypothesis motion compensation followed by experiment results in Sec. IV which shows the combination yields performance superior to that of either class alone. Finally, the conclusion is given in Sec. V.

## II. MULTIHYPOTHESIS MOTION COMPENSATION

MHMC combines different prediction possibilities occurred in the ME process. As shown in the equation below:

$$\tilde{s}(x, y) = \sum_i w_i(x, y) \tilde{s}_i(x, y), \quad (1)$$

the prediction of a pixel  $s(x, y)$  in the current frame is obtained by summing up all the multiple predictions  $\tilde{s}_i(x, y)$  according to some weights  $w_i(x, y)$ . Several methods can be used to obtain multiple predictions in motion compensation. We classify them into three categories: spatial-diversity, temporal-diversity, and phase-diversity multihypothesis. Spatial-diversity multihypothesis means the diverse predictions are generated from spatial domain of the same reference frame. Examples are fractional-pixel MC [9] and OBMC. In the class of temporal-diversity multihypothesis, such as long-term-memory motion compensation (LTMMC) [10], the different hypothesis reside in multiple reference frames in the temporal direction. In [6], we drove our motion compensation in RDWT domain and took into account the diverse predictions provided by different phases existed in RDWT transform, thus our redundant-wavelet multihypothesis (RWMH) coder explored the phase-diversity motion compensation in the transform domain.

### A. Phase-diversity Multihypothesis in the RDWT Domain

To illustrate our previous RWMH system, let's first look at the structure of RDWT and its multiple phases explained in terms of filter-banks. Based on the DWT transform, after each level of filtering, a downsampling will be performed to get rid of half of the coefficients, either even-phase or odd-phase. This procedure assigns two phase choices to the wavelet coefficients before subsampling. As DWT retains only one single phase of coefficients and discards all the others, no matter how many levels of transform is performed, the coefficients size of a J level DWT will always be the same as the

original signal. But each subband will shrink by half after each level of filtering. On the contrary, RDWT keeps all the wavelet coefficients of different phases by removing the downsampling procedure from DWT. As a result, each subband in a RDWT domain transform retains the same size of the original signal, the total number of coefficients will grow as more and more level of filtering is implemented. For example, after a  $J$ -scale RDWT decomposition, there will be  $3J+1$  subbands, each of the same size as the original signal. For a 1D signal,  $J$ -scale RDWT will produce  $2^J$  phases. For a 2D image, after a  $J$ -scale RDWT, there will be  $4^J$  phases all together. We illustrate the different phases and the different subbands information of a 3 scale 2D RDWT transform in Fig. 1. As we can see there are 10 subbands and 64 different phases existed. Coefficients retain their correct spatial location within each subband, and each subband is the same size as the original image.  $B_j$ ,  $H_j$ ,  $V_j$ , and  $D_j$  denote the baseband, horizontal, vertical, and diagonal subbands, respectively, at scale  $j$ . The shaded coefficients are from phase 0, and they indicate a critically sampled DWT.

### B. Redundant Wavelet Multihypothesis (RWMH) System

RWMH extends the MHMC into transform domain, which is based on the phase-diversity introduced in Sec. II-A. Since the number of phases will increase following the number of decomposition, we can build our RWMH system based on different level of RDWT. The RWMH introduced in [6] chooses one level decomposition and small vectors for each phase, while here we use 3 level of decomposition and abandon the small vectors to save bits on overhead information. So the ME matching equation is revised to fit more than one level of decomposition.

We choose the block matching technique for the ME. In a  $J$ -scale RDWT decomposition, each  $B \times B$  block in the original spatial domain corresponds to  $3J + 1$  same size blocks in RDWT domain, one in each subband. ME in RDWT domain considers all the coefficients reside in the spatial block in all the subbands for block matching. A window  $[-W, W]$  is used to define the searching range. The motion vector,  $(\Delta_x, \Delta_y)$ , is determined by minimizing the mean absolute error (MAE) for all the coefficients of all phases. Specifically,

$$(\Delta_x, \Delta_y) = \arg \min_{-W \leq \Delta_x, \Delta_y \leq W} \text{MAE}(x, y, \Delta_x, \Delta_y) \quad (2)$$

where

$$\text{MAE}(x, y, \Delta_x, \Delta_y) = \frac{1}{B^2} \sum_{k=1}^B \sum_{l=1}^B \text{AE}(x+k, y+l, \Delta_x, \Delta_y), \quad (3)$$

and the absolute error (AE) is

$$\begin{aligned} \text{AE}(x, y, \Delta_x, \Delta_y) = & \sum_{j=1}^J 2^{-j} \left\{ \left| V_j^{\text{cur}}(x, y) - V_j^{\text{ref}}(x + \Delta_x, y + \Delta_y) \right| \right. \\ & + \left| H_j^{\text{cur}}(x, y) - H_j^{\text{ref}}(x + \Delta_x, y + \Delta_y) \right| \\ & + \left. \left| D_j^{\text{cur}}(x, y) - D_j^{\text{ref}}(x + \Delta_x, y + \Delta_y) \right| \right\} \\ & + 2^{-J} \left| B_j^{\text{cur}}(x, y) - B_j^{\text{ref}}(x + \Delta_x, y + \Delta_y) \right|, \quad (4) \end{aligned}$$

where *cur* and *ref* denote subbands from the current and reference frames, respectively, and  $B_j$ ,  $H_j$ ,  $V_j$ , and  $D_j$  are the baseband, horizontal, vertical, and diagonal subbands, respectively, at scale  $j$ .

### III. COMBINING PHASE-DIVERSITY WITH SPATIAL-DIVERSITY MULTIHYPOTHESIS

Different class of multihypothesis can be combined together for better performance. Current video coding standards, such as H.264[1] provides several options as sub-pixel accuracy, OBMC, and B-frame to increase coding efficiency. After successfully depolyed the multihypothesis in the transform domain, we are thinking of go beyond the limit by combining the phase-diversity with the spatial-diversity MHMC. While concerning the combination of different types of multihypothesis, we must consider the tradeoffs between possible gain and the increasing overhead information. In our proposed scheme which adds OBMC to our RWMH system, one important feature leads to substantial gain is that neither of these two multihypothesis methods requires any side information, such as extra vector or weighting matrix, send to the decoder. Because in RWMH, the weighting of the individual predictions is carried out implicitly in the form of an inverse transform, and OBMC uses fixed weights matrices.

OBMC has other merits other than no overhead burden that suitable to be adopted to our RWMH system. Since our RWMH system uses block-based ME, blocking artifact can not be avoid. In a wavelet-based coder, blocking artifacts will greatly hinder the coding efficiency. OBMC is to use the vectors from the nearby blocks and the vector from itself to predict the motion. From a MHMC point of view, OBMC uses more than one hypothesis in the spatial domain to predict the motion of the current pixel. It is a good example of spatial-diversity MHMC. Different pre-defined weighting matrices will be applied on the pixels within the block and its nearby blocks on the four sides during MC. And OBMC has been proved to be very successful in reducing blocking artifacts.

#### A. RWMH\_OBMC System

We combine RWMH with OBMC, and call the new system RWMH\_OBMC. The diagram of our RWMH\_OBMC system is illustrated in Fig. 2.

Both current and reference frames are transformed into RDWT domain, so both ME and MC are performed in RDWT domain. After OBMC, we subtract the predicted image from the current image, and get the residual image in RDWT domain. Then an inverse RDWT mixes all the multihypothesis in different transform phases into one spatial residual image. It is then coded by a still image coder. In our approach, we use SPIHT [3], but any still-image coder, wavelet-based or otherwise, would suffice.

At the decoder side, after we first decode the bitstream to get the vectors and the residual image, we transform both the residual image and the reference image to the RDWT domain and do the same OBMC as at encoder side. Finally the reconstructed image is transformed to spatial domain by an inverse RDWT.

#### B. OBMC on Selected Subbands

In our proposed RWMH\_OBMC system, we apply OBMC to all the subbands of RDWT coefficients. For a  $J$ -scale decomposition, the calculation complexity spends on OBMC will be  $3J + 1$  times of the spatial domain case. In order to reduce the calculation burden, we may select some of the subbands for OBMC, and leave the rest subbands stay with the normal MC procedure. Since the high frequency information resides mainly at high band of a wavelet transform, and the blocking artifact in high band will cause more trouble in coding efficiency. So we choose different levels of high bands  $V_j$ ,  $D_j$ , and  $H_j$  for OBMC and denote the system as RWMH\_OBMC\_HIGH. In this 3 scale RDWT coefficients, when we do OBMC on all the 3

TABLE I  
DISTORTION AVERAGED OVER ALL FRAMES OF THE SEQUENCE.

	PSNR (dB)				
	RWMH	Spatial OBMC	RWMH OBMC	RWMH.OBMC all highbands	RWMH.OBMC 3 highbands
Susie	40.4	41.2	41.6	41.6	41.4
Football	27.8	28.8	29.2	29.2	29.0
Mother & Daughter	44.5	44.6	45.2	45.2	45.1
Carphone	36.8	37.2	37.9	38.0	37.7

Rate is 0.5 bpp.

levels of high bands, leave only  $B_3$  for normal MC, the performance will be nearly equivalent to the all bands RWMH.OBMC. Then we can go even further to implement OBMC on the highest level of subbands  $V_1$ ,  $H_1$ , and  $D_1$ . So three out of ten are chosen. The time for MC will be reduced proportionally especially at the decoder side.

#### IV. EXPERIMENTAL RESULTS

We choose four sequences to show the performance of comparative methods. They are 75-frame “Carphone” sequence, with a spatial resolution of  $176 \times 144$  pixels, 71-frame “Susie” sequence, 100-frame “Football” sequence, both with a spatial resolution of  $352 \times 240$  pixels, 300-frame of “Mother-daughter” sequence with a spatial resolution of  $352 \times 288$  pixels. All the sequences are of grayscale. We code the first frame intra-coded (I-frame) while all subsequent frames use ME/MC (P-frames). All wavelet transform use the popular 9-7 biorthogonal filter with symmetric extension, and all ME/MC methods use integer-pixel accuracy. The block-based ME/MC procedure uses  $B = 16$ , and  $W = 15$ . All motion-vector overhead are included in the bitrate.

Table I lists the PSNR from different methods we choose for comparison. They are RWMH, Spatial OBMC, RWMH.OBMC, RWMH.OBMC all highbands, and RWMH.OBMC 3 highbands.

These results illustrate that, first, compare the first three methods, our proposed RWMH.OBMC system achieves at least a 0.4-dB gain over RWMH and Spatial OBMC. A frame-by-frame PSNR of “Carphone” sequence is given by Fig. 3. The rate-distortion performance of “Football” sequence from 0.1bpp to 1.0bpp is shown in Fig. 4. Second, compare the last three methods, OBMC on selected highbands only degrade the performance slightly, so that even only OBMC the highest subbands, we still have at least 0.5-dB gain over RWMH system for all the sequences. The rate-distortion performance of “Mother-daughter” sequence from 0.1bpp to 1.0bpp is shown in Fig. 5. We can expect the curve of 6 highbands OBMC lies between the 9 bands and 3 bands OBMC.

#### V. CONCLUSIONS

In this paper, we present RWMH.OBMC system which combines phase-diversity with spatial-diversity multihypothesis. Experimental results show that our RWMH.OBMC outperforms either RWMH or OBMC method implemented alone. OBMC on selected subbands may save the calculation cost and still have the promising performance increment.

#### VI. REFERENCES

- [1] ITU-T, *Advanced Video Coding for Generic Audiovisual Services*, May 2003, ITU-T Recommendation H.264.
- [2] ISO/IEC 14496-2, *Information Technology—Coding of Audio-Visual Objects—Part 2: Visual*, 1999, MPEG-4 Coding Standard.

- [3] A. Said and W. A. Pearlman, “A new, fast, and efficient image codec based on set partitioning in hierarchical trees,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, no. 3, pp. 243–250, June 1996.
- [4] H.-W. Park and H.-S. Kim, “Motion estimation using low-band-shift method for wavelet-based moving-picture coding,” *IEEE Transactions on Image Processing*, vol. 9, no. 4, pp. 577–587, April 2000.
- [5] S. Cui, Y. Wang, and J. E. Fowler, “Mesh-based motion estimation and compensation in the wavelet domain using a redundant transform,” in *Proceedings of the International Conference on Image Processing*, Rochester, NY, September 2002, vol. 1, pp. 693–696.
- [6] S. Cui, Y. Wang, and J. E. Fowler, “Multihypothesis motion compensation in the redundant wavelet domain,” in *Proceedings of the International Conference on Image Processing*, Barcelona, Spain, September 2003, vol. 2, pp. 53–56.
- [7] Y. Wang, S. Cui, and J. E. Fowler, “3D video coding using redundant-wavelet multihypothesis and motion-compensated temporal filtering,” in *Proceedings of the International Conference on Image Processing*, Barcelona, Spain, September 2003, vol. 2, pp. 755–758.
- [8] G. J. Sullivan, “Multi-hypothesis motion compensation for low bit-rate video coding,” in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Minneapolis, MN, April 1993, vol. 5, pp. 437–440.
- [9] B. Girod, “Motion-compensating prediction with fractional-pel accuracy,” *IEEE Transactions on Communications*, vol. 41, no. 4, pp. 604–612, April 1993.
- [10] T. Wiegand, X. Zhang, and B. Girod, “Long-term memory motion-compensated prediction,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 1, pp. 70–84, February 1999.

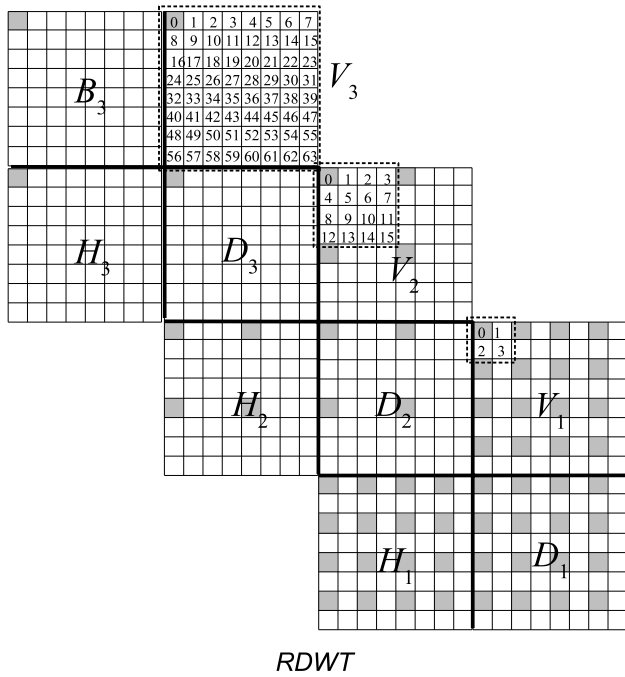


Fig. 1. Coefficients of different phases from a 3 scale 2D RDWT.

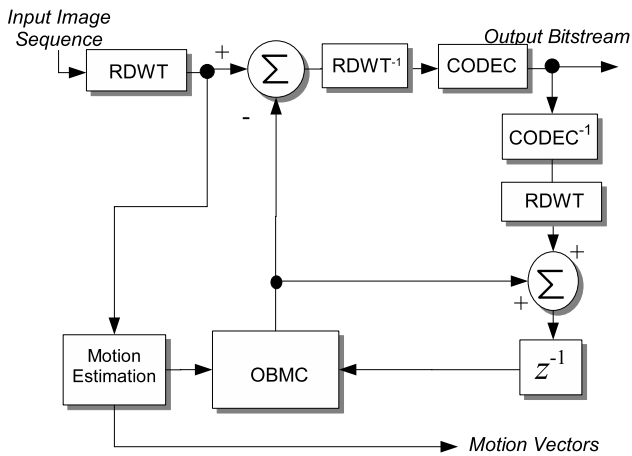


Fig. 2. The RWMH\_OBMC coder.  $z^{-1}$  = frame delay, CODEC is any still-image coder.

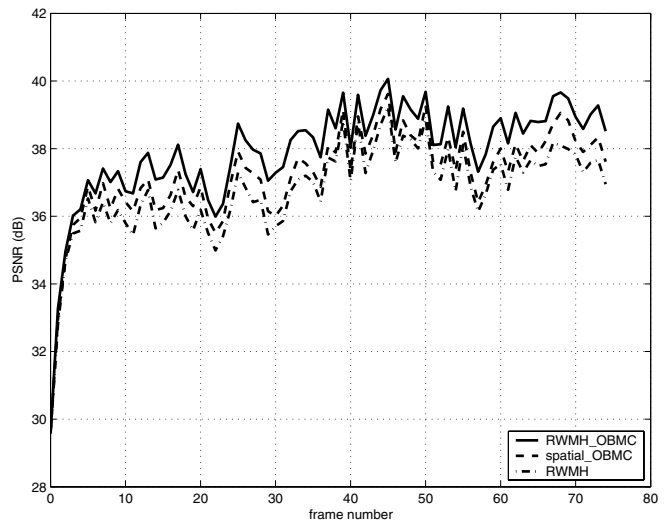


Fig. 3. Frame-by-frame PSNR for "Carphone" at 0.5 bpp (380 kbps). Comparing RWMH\_OBMC, spatial\_OBMC, and RWMH.

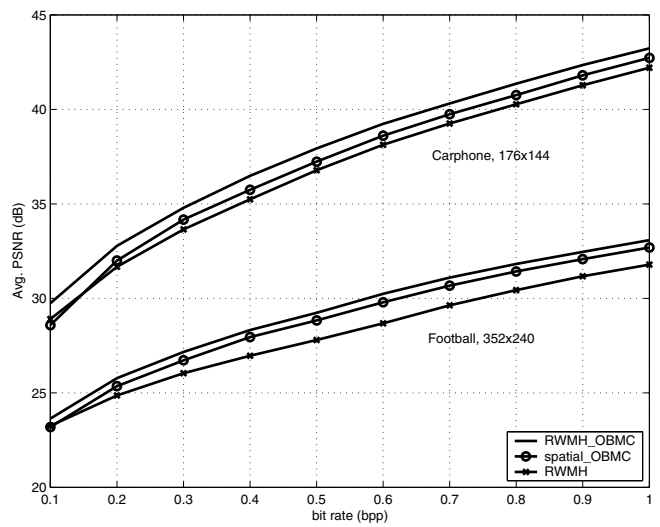


Fig. 4. Rate-distortion curves for "Carphone" and "Football".

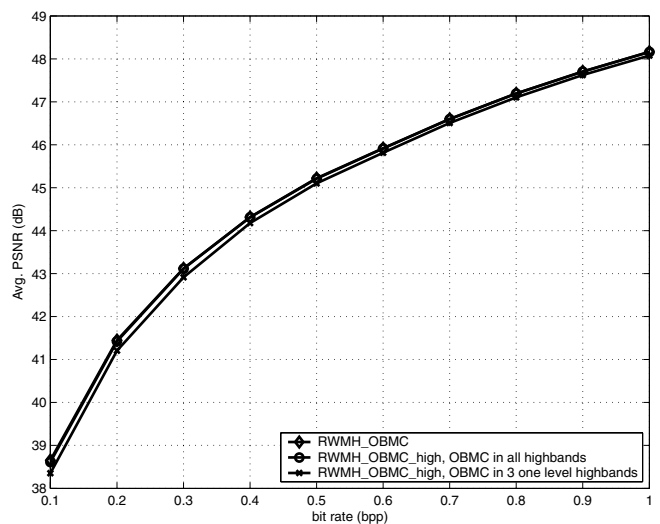


Fig. 5. Rate-distortion curves for "Mother & Daughter".