Analysis of Redundant-Wavelet Multihypothesis for Motion Compensation

James E. Fowler

Department of Electrical and Computer Engineering
GeoResources Institute (GRI)
Mississippi State University, Starkville, Mississippi

Abstract

An analysis is presented that examines multihypothesis motion-compensated video coding using a redundant wavelet transform to produce multiple predictions that are diverse in transform phase. In such redundant-wavelet multihypothesis, the corresponding multiple-phase inverse transform implicitly combines the phase-diverse predictions into a single spatial-domain prediction. The performance advantage of this approach is investigated analytically, invoking the fact that the multiple-phase inverse involves a projection that significantly reduces the power of the noise not captured by the motion model. The analysis predicts that, under the assumption of a simple translational motion model, redundant-wavelet multihypothesis is capable of up to a 7-dB reduction in prediction-error variance over an equivalent single-phase, single-hypothesis approach. Experimental results support the performance advantage for real motion-compensation residuals.

Introduction

In video coding, multihypothesis motion compensation (MHMC) [1] forms a prediction of pixel $s[x, y, t]$ in the current frame as a combination of multiple predictions in an effort to combat the uncertainty inherent in the motion-estimation (ME) process. Assuming that the combination of these hypothesis predictions is linear, we have that the prediction of frame $s[x, y, t]$ is

$$\tilde{s}[x, y, t] = \sum_i w_i[x, y, t]\tilde{s}_i[x, y, t], \tag{1}$$

where the multiple predictions $\tilde{s}_i[x, y, t]$ are combined according to some weights $w_i[x, y, t]$. One approach to MHMC is to implement multihypothesis prediction in the spatial dimensions; i.e., the predictions $\tilde{s}_i[x, y, t]$ are culled from spatially distinct locations in the reference frame (e.g., subpixel accuracy [2], overlapped block motion compensation (OBMC) [3]). Another approach is to deploy MHMC temporally by choosing predictions $\tilde{s}_i[x, y, t]$ from multiple reference frames (e.g., B-frames, long-term-memory motion compensation [4]). Of course, it is possible to combine these two classes by choosing multiple predictions that are diverse both spatially and temporally.

In [5], we introduced a new class of MHMC by extending the multihypothesis-prediction concept into the transform domain. Specifically, we performed ME/MC in the domain of a redundant, or overcomplete, wavelet transform, and used multiple predictions that were diverse in transform phase. We coined the term redundant-wavelet multihypothesis (RWMH) to describe our approach.

This work was funded in part by the National Science Foundation under Grant No. CCR-0310864.

In *Proceedings of the IEEE Data Compression Conference*, J. A. Storer and M. Cohn, Eds., Snowbird, UT, March 2006, pp. 352–361.

In this paper, we investigate the performance advantage that RWMH possesses over single-hypothesis techniques that base ME/MC on merely a single phase, the most prominent of these latter strategies being the system of Park and Kim [6]. The centerpiece of this investigation is an analytical derivation that quantifies the gain of RWMH over single-phase prediction under the assumption of a simple translational motion model. Thanks to the well-known robustness of overcomplete transforms to noise in the transform domain, our RWMH technique substantially reduces the variance of the noise not captured by the translational motion model, leading to a theoretical reduction in the variance of the prediction residual by up to 7 dB with respect to the single-phase system.

In the following, we overview theory behind the redundant discrete wavelet transform (RDWT) [7–9], review our RWMH technique, and present the main contribution of the paper—an analytical derivation of the gain of RWMH over single-phase prediction. We note that a more extensive presentation of this analysis, including a more elaborate motion model, is given in [10].

The Redundant Discrete Wavelet Transform (RDWT)

Let $s[x, y, t]$ be a video sequence sampled spatially on an integer-pixel lattice and temporally at integer times and denote the 2D spatial RDWT of frame $s[x, y, t]$ at time t as the collection of subbands $S^{(k)}[x, y, t]$,

$$\left\{ S^{(k)}[x, y, t] \right\}_k = \mathcal{R}\Big[s[x, y, t] \Big]. \tag{2}$$

In essence, the RDWT [7–9] removes the downsampling operation from the traditional critically sampled discrete wavelet transform (DWT) to produce an overcomplete representation. The well-known shift variance of the DWT arises from its use of downsampling, while the RDWT is shift invariant since the spatial sampling rate is fixed across scale. The RDWT is easily created by employing filtering as in the usual critically sampled DWT; however, all "phases" of downsampled coefficients are retained rather than discarded. Thus, the RDWT consists of a number of distinct, critically sampled DWTs, each one being a separate phase of the RDWT.[1] Furthermore, due to the lack of subsampling, each subband in an RDWT has the same size as the original image. Finally, we observe that, if we resolve values off the integer-pixel sampling lattice using an interpolation operator $I(\cdot)$ which is linear and shift-invariant, the RDWT can be considered to be shift invariant even for non-integer translations (d_x, d_y) in the sense that

$$\mathcal{R}\Big[I\Big(s[x - d_x, y - d_y, t] \Big) \Big] = \left\{ I\left(S^{(k)}[x - d_x, y - d_y, t] \right) \right\}_k. \tag{3}$$

Given subbands $S^{(k)}[x, y, t]$, there are several methods to invert the RDWT. The *single-phase inverse* consists of subsampling the RDWT coefficients to extract one critically sampled DWT and inverting using the corresponding inverse DWT; i.e.,

$$s'[x, y, t] = \mathcal{D}^{-1}\left[\downarrow \left\{ S^{(k)}[x, y, t] \right\}_k \right], \tag{4}$$

[1] In a J-scale 2D RDWT, there exist 4^J distinct critically sampled DWTs; each one corresponds to a unique choice between even and odd phase for the horizontal subsampling as well as the vertical subsampling at each scale of decomposition.

where $\mathcal{D}^{-1}(\cdot)$ is the inverse DWT and $\downarrow$ denotes subsampling to a single phase. Alternatively, one can employ a *multiple-phase inverse* which we denote as

$$s''[x,y,t] = \mathcal{R}^{-1}\left[\left\{S^{(k)}[x,y,t]\right\}_k\right]. \tag{5}$$

In such a multiple-phase inverse, one independently inverts each of the 4^J critically sampled DWTs constituting the J-scale 2D RDWT and averages the resulting reconstructions together (one can equivalently employ a more computationally efficient filtering implementation). We observe that the single-phase inverse (4) is generally not the same as the multiple-phase inverse (5), but the two do yield the same result if the argument is in the range space of $\mathcal{R}(\cdot)$; i.e.,

$$s[x,y,t] = s'[x,y,t] = s''[x,y,t], \tag{6}$$

as long as $\left\{S^{(k)}[x,y,t]\right\}_k$ in (4) and (5) is the RDWT of some $s[x,y,t]$. We note that this will not always be the case since the range space of the RDWT is larger than the original signal domain, the RDWT being overcomplete. Finally, like the forward transform, the multiple-phase inverse is shift invariant under linear fractional-pixel interpolation,

$$\mathcal{R}^{-1}\left[\left\{I\left(S^{(k)}[x-d_x,y-d_y,t]\right)\right\}_k\right] = I\Big(s''[x-d_x,y-d_y,t]\Big), \tag{7}$$

while the single-phase inverse is not.

From the perspective of frame theory, the RDWT is a frame operator, $\mathcal{R}$, while the multiple-phase inverse RDWT, $\mathcal{R}^{-1}$, is the corresponding pseudo-inverse operator. As a pseudo-inverse, $\mathcal{R}^{-1}$ can be considered to consist of an orthogonal projection onto the range space of $\mathcal{R}$ followed by a mapping into the original signal domain. It has been observed [11] that the fact that an inverse frame operator includes a projection makes frame operators such as the RDWT robust to added noise. Specifically, noise in the RDWT domain, when mapped to the original signal domain via the pseudo-inverse, typically undergoes a significant reduction in variance since the component of the noise orthogonal to the range space of the RDWT is eliminated by the projection. Below, we will see that this projection property of the pseudo-inverse and its effect on noise lie at the heart of our proposed RWMH technique.

Redundant-Wavelet Multihypothesis (RWMH)

The RWMH System

The majority of prior work concerning RDWT-based video coding originates in the work of Park and Kim [6], in which the system shown in Fig. 1(a) was proposed. In essence, this system works as follows. An input frame is decomposed with a critically sampled DWT which is matched to an RDWT decomposition of the previous reconstructed frame. The ME procedure of this system amounts to identifying a particular critically sampled DWT in the reference-frame RDWT, and a displacement within that DWT. This ME process is performed on a block basis, with blocks culled across the scales of the DWT.

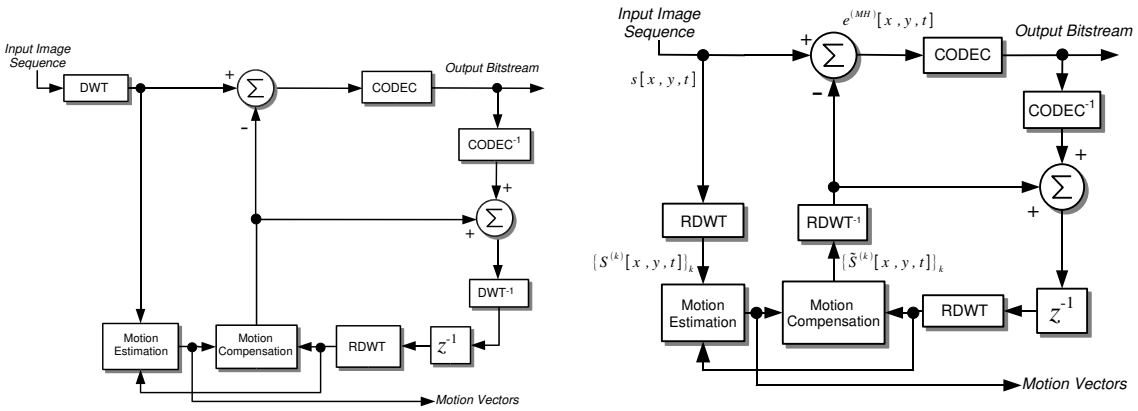

Figure 1: (a) The single-phase, RDWT-based video coder of [6]; (b) The RWMH coder. $z^{-1} =$ frame delay, *CODEC* is any still-image coder operating in the critically-sampled-DWT domain.

In the single-hypothesis system of Fig. 1(a), the redundancy inherent in the RDWT is used exclusively to permit ME/MC in the wavelet domain by overcoming the well-known shift variance of the critically sampled DWT. In [5], we presented an entirely new use for the redundancy in the RDWT; specifically, we employed transform redundancy to yield multiple predictions of motion that were combined into a single multihypothesis prediction. This approach represented a new paradigm in MHMC wherein diversity in transform phase yields multihypothesis predictions that significantly enhance coding performance. The encoder of the resulting RWMH system is depicted in Fig. 1(b).

Intuitively, we observe that each of the critically sampled DWTs within an RDWT will "view" motion from a different perspective. Consequently, if motion is predicted in the RDWT domain, the multiple-phase inverse RDWT, $\mathcal{R}^{-1}(\cdot)$, forms a multihypothesis prediction in the form of (1). Specifically, for a J-scale RDWT, the reconstruction from DWT i of the RDWT is $\tilde{s}_i[x, y, t]$, $0 \leq i < 4^J$, while $w_i[x, y, t] = 4^{-J}$, $\forall i$.

Performance Gain of RWMH

In this section, we show analytically that the multihypothesis nature of our RWMH coder of Fig. 1(b) offers substantial performance gain over the single-hypothesis system of Fig. 1(a). In Fig. 1(b), the current and reference frames are transformed into RDWT coefficients, and MC takes place in the RDWT domain, yielding $\left\{\widetilde{S}^{(k)}[x, y, t]\right\}_k$, the RDWT-domain prediction of the current frame. The spatial-domain prediction error is then

$$e^{(\mathrm{MH})}[x, y, t] = s[x, y, t] - \mathcal{R}^{-1}\left[\left\{\widetilde{S}^{(k)}[x, y, t]\right\}_k\right] = \mathcal{R}^{-1}\left[\left\{S^{(k)}[x, y, t] - \widetilde{S}^{(k)}[x, y, t]\right\}_k\right], \tag{8}$$

where the superscript (MH) denotes that this is the prediction error for the multihypothesis system of Fig. 1(b); we will determine an equivalent quantity for the single-hypothesis system of Fig. 1(a) shortly.

Within RDWT subband k, we adopt the simple translational motion model from [12]. Specifically, we assume that the current frame at time t is a simple displacement of the previous frame plus residual noise not captured by the translational motion; i.e.,

$$S^{(k)}[x,y,t] = I\Big(S^{(k)}[x-d_x, y-d_y, t-1]\Big) + N^{(k)}[x,y,t], \tag{9}$$

where (d_x, d_y) is the unknown translation and $I(\cdot)$ is a linear interpolation operator used to resolve fractional-pixel values. The motion-estimation process estimates (d_x, d_y) as $(\hat{d}_x, \hat{d}_y)$, yielding

$$\widetilde{S}^{(k)}[x,y,t] = I\Big(S^{(k)}[x-\hat{d}_x, y-\hat{d}_y, t-1]\Big). \tag{10}$$

Thus, we have

$$e^{(\mathrm{MH})}[x,y,t] = \mathcal{R}^{-1}\left[\left\{I\Big(S^{(k)}[x-d_x, y-d_y, t-1] - S^{(k)}[x-\hat{d}_x, y-\hat{d}_y, t-1]\Big) + N^{(k)}[x,y,t]\right\}_k\right], \tag{11}$$

which, due to (6) and (7), becomes

$$e^{(\mathrm{MH})}[x,y,t] = I\Big(s[x-d_x, y-d_y, t-1] - s[x-\hat{d}_x, y-\hat{d}_y, t-1]\Big) + n^{(\mathrm{MH})}[x,y,t], \tag{12}$$

where

$$n^{(\mathrm{MH})}[x,y,t] = \mathcal{R}^{-1}\Big[\big\{N^{(k)}[x,y,t]\big\}_k\Big]. \tag{13}$$

For a prediction error such as given by (12), Girod [2, 12] derives the 2D power spectral density to be

$$\Phi_{ee}^{(\mathrm{MH})}(\omega_x,\omega_y) = \Phi_{nn}^{(\mathrm{MH})}(\omega_x,\omega_y)\Big|I(\omega_x,\omega_y)\Big|^2 + \Phi_{ss}(\omega_x,\omega_y)\Big(1 + \Big|I(\omega_x,\omega_y)\Big|^2 - 2\Re\Big[I(\omega_x,\omega_y)P(\omega_x,\omega_y)\Big]\Big), \tag{14}$$

where $\Phi_{ss}(\omega_x,\omega_y)$ and $\Phi_{nn}^{(\mathrm{MH})}(\omega_x,\omega_y)$ are the 2D power spectral densities of $s[x,y,t]$ and $n^{(\mathrm{MH})}[x,y,t]$, respectively; $P(\omega_x,\omega_y)$ is the 2D Fourier transform of the probability density function of the displacement error, $(\Delta_x, \Delta_y) = (d_x, d_y) - (\hat{d}_x, \hat{d}_y)$; $I(\omega_x,\omega_y)$ is the frequency response of the interpolation filter used to resolve fractional-pixel values; and $\Re(\cdot)$ denotes the real part of a complex number.

As in [2], we focus on the prediction-error variance to gauge coding performance. From (14), the prediction-error variance is

$$\nu_e^{(\mathrm{MH})} = \nu_n^{(\mathrm{MH})} + \Gamma_{ss}, \tag{15}$$

where $\nu_n^{(\mathrm{MH})}$ is the variance of $n^{(\mathrm{MH})}[x,y,t]$,

$$\Gamma_{ss} = \frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\Phi_{ss}(\omega_x,\omega_y)\Big(2 - 2\Re\big[P(\omega_x,\omega_y)\big]\Big)\,d\omega_x\,d\omega_y, \tag{16}$$

and we have assumed $I(\omega_x, \omega_y) = 1$ (i.e., sinc interpolation [2]) in order to simplify the analysis.

The key to our RWMH technique is embodied by (13)—as we have discussed previously, the multiple-phase inverse RDWT is a pseudo-inverse frame operator which is tantamount to a projection onto the range space of the RDWT following by a mapping back into the original spatial domain. Since the noise $N^{(k)}[x, y, t]$ not captured by the motion model of (9) is almost certainly not in the range space of the RDWT, the mapping of the RDWT-domain noise back to the spatial domain via (13) will result in a reduction in noise variance.

Following the lead of [2, 12], let us assume that the noise $N^{(k)}[x, y, t]$ not captured by the translational motion model of (9) as applied in the RDWT domain is zero-mean, white, and of variance ν_N. This white-noise model is an obvious oversimplification; however, this assumption permits us to invoke the following theorem to quantify the noise-variance reduction. A more thorough investigation into the character of the residual noise is conducted in [10].

Theorem 1 *If noise present in the RDWT domain is zero-mean, white, and of variance ν_N, and the wavelet filters underlying the J-scale 2D RDWT are orthonormal, then the variance of the spatial-domain noise given by (13) is*

$$\nu_n^{(MH)} = \frac{\nu_N}{5}\left[1 + 4\left(\frac{1}{16}\right)^J\right]. \tag{17}$$

The proof of Theorem 1 is given in [13]. From this point on, we assume orthonormal filters, although Theorem 1 will approximately hold if the filters are "near-orthonormal" biorthogonal filters, as is common in practice.

Now, suppose that, in Fig. 1(b), rather than mapping the RDWT-domain prediction to the spatial domain using the multiple-phase inverse RDWT, we instead use the single-phase inverse of (4). It is straightforward to see that this single-phase MC process is equivalent to the single-hypothesis system shown in Fig. 1(a). In this case, (8) becomes

$$\begin{aligned} e^{(\mathrm{SH})}[x, y, t] &= s[x, y, t] - \mathcal{D}^{-1}\left[\downarrow \left\{\widetilde{S}^{(k)}[x, y, t]\right\}_k\right] \\ &= \mathcal{D}^{-1}\left[\downarrow \left\{S^{(k)}[x, y, t] - \widetilde{S}^{(k)}[x, y, t]\right\}_k\right], \end{aligned} \tag{18}$$

where last equality is due to (4) and (6). Furthermore, (11) becomes

$$\begin{aligned} &e^{(\mathrm{SH})}[x, y, t] = \\ &\mathcal{D}^{-1}\left[\downarrow \left\{I\Big(S^{(k)}[x - d_x, y - d_y, t - 1] - S^{(k)}[x - \hat{d}_x, y - \hat{d}_y, t - 1]\Big) + N^{(k)}[x, y, t]\right\}_k\right], \end{aligned} \tag{19}$$

which, through use of (3), can be expressed as

$$\begin{aligned} &e^{(\mathrm{SH})}[x, y, t] = \\ &\mathcal{D}^{-1}\left[\downarrow \mathcal{R}\Big[I\Big(s[x - d_x, y - d_y, t - 1] - s[x - \hat{d}_x, y - \hat{d}_y, t - 1]\Big) + N^{(k)}[x, y, t]\Big]\right]. \end{aligned} \tag{20}$$

Applying (4), (6), and the linearity of $I(\cdot)$ then gives

$$e^{(\mathrm{SH})}[x,y,t] = I\Big(s[x-d_x, y-d_y, t-1] - s[x-\hat{d}_x, y-\hat{d}_y, t-1]\Big) + n^{(\mathrm{SH})}[x,y,t], \quad (21)$$

with

$$n^{(\mathrm{SH})}[x,y,t] = \mathcal{D}^{-1}\left[\downarrow \left\{N^{(k)}[x,y,t]\right\}_k\right]. \quad (22)$$

Since the DWT is a unitary transform, and the RDWT-domain noise is equally present in all phases, the spatial-domain variance of the noise in this case is

$$\nu_n^{(\mathrm{SH})} = \nu_N, \quad (23)$$

assuming, as before, that the noise not captured by the motion model in the RDWT domain is zero-mean, white, and of variance ν_N. The prediction-error variance is then

$$\nu_e^{(\mathrm{SH})} = \nu_n^{(\mathrm{SH})} + \Gamma_{ss}, \quad (24)$$

where Γ_{ss} is once again given by (16).

Experimental Results

To quantity the gain of the multihypothesis approach over the single-hypothesis approach, let us assume, as was done in [2, 12], an isotropic signal power spectrum,

$$\Phi_{ss}(\omega_x, \omega_y) = \frac{2\pi}{\omega_0^2}\left(1 + \frac{\omega_x^2 + \omega_y^2}{\omega_0^2}\right)^{-\frac{2}{3}}, \quad (25)$$

where $\omega_0 = -\ln(0.93)$, and an isotropic Gaussian displacement-error density of variance ν_Δ such that

$$P(\omega_x, \omega_y) = \exp\left[-\frac{\nu_\Delta}{2}\left(\omega_x^2 + \omega_y^2\right)\right]. \quad (26)$$

Under these models, we numerically evaluate $\nu_e^{(\mathrm{MH})}$ from (15) and $\nu_e^{(\mathrm{SH})}$ from (24) versus displacement-error accuracy β for several noise variances, where $\beta = \frac{1}{2}\log_2(12\nu_\Delta)$ such that $\beta = -1$ for half-pixel accuracy, $\beta = -2$ for quarter-pixel accuracy, etc. [14]. We define the difference in prediction-error variance in dB between the two approaches as

$$\gamma = 10\log_{10}\left(\frac{\nu_e^{(\mathrm{MH})}}{\nu_e^{(\mathrm{SH})}}\right) \quad (27)$$

and numerically evaluate γ as β varies in Fig. 2.

The numerical results of Fig. 2 leads to the following observations: 1) RWMH systematically reduces prediction-error variance over that of single-phase ME/MC, although the performance gain may be small if ME is inaccurate, or the power of the residual noise is low; 2) as ME becomes highly accurate, RWMH produces a 7-dB reduction in prediction-error variance as compared to single-phase ME/MC, regardless of the residual-noise power; 3) for a given ME accuracy, the larger the power of the residual noise, the more effective

In *Proceedings of the IEEE Data Compression Conference*, J. A. Storer and M. Cohn, Eds., Snowbird, UT, March 2006, pp. 352–361.

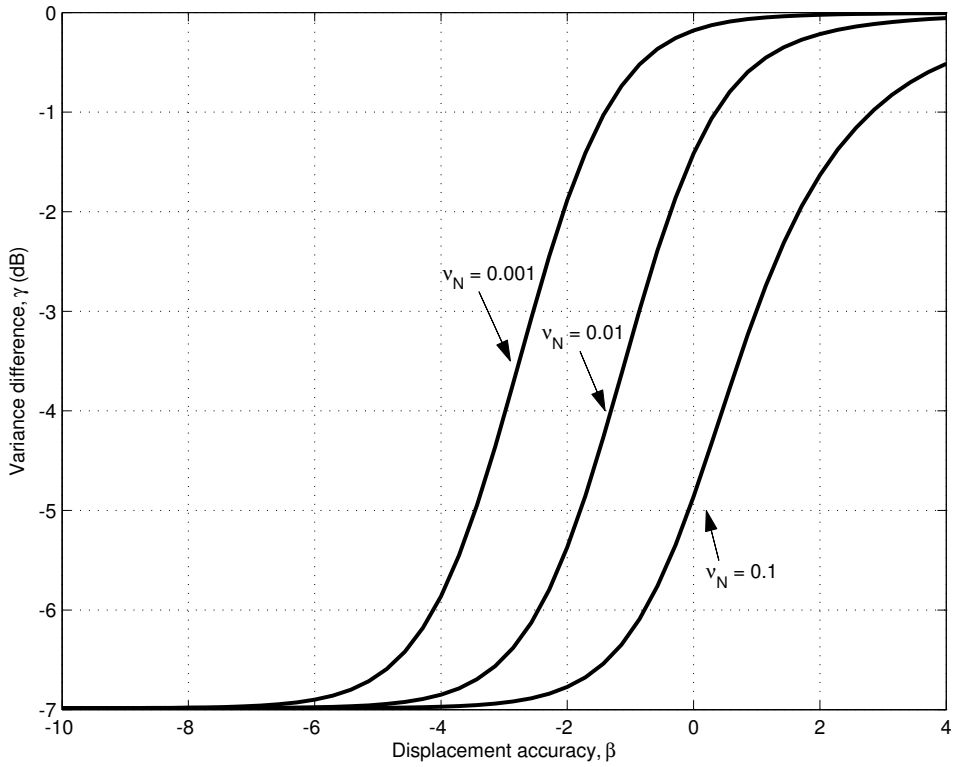

Figure 2: Difference, γ, in prediction-error variance between the multihypothesis and single-hypothesis systems as the displacement accuracy β varies at various noise variances ν_N. The number of transform scales is $J = 3$.

RWMH is at reducing the prediction-error variance; and, 4) there is a "threshold" ME accuracy required for RWMH to produce substantial gain over the single-phase system—less accurate ME is required for substantial gain as the residual-noise power increases.

Fig. 2 and the above observations are derived under the assumption that the residual noise $N^{(k)}[x, y, t]$ is white. This is an obvious oversimplification as it is well-known that real MC residuals may exhibit substantial correlation spatially. Consequently, neither the white-noise assumption nor Theorem 1 hold in a strict sense in practice. However, we have observed empirically that the residual noise in our RWMH technique does in fact undergo a substantial reduction in variance as predicted by Theorem 1. For example, in Table 1, we consider a single frame from a video sequence and use the previous frame to create a prediction of the first frame using block-based MC in the RDWT domain. We then invert the resulting RDWT-domain residual using both the single-phase inverse of (4) and the multiple-phase inverse of (5), calculating γ via (27) using the resulting spatial-domain variances $\nu_e^{(\mathrm{MH})}$ and $\nu_e^{(\mathrm{SH})}$. In order to divorce the effects of ME from those of single- and multiple-phase MC, we fix the motion-vector field for both inverse transforms to be the same field, estimated separately in the spatial domain between the two original frames. The results are shown in Table 1 for several block sizes B, several scales J of transform decomposition, several subpixel accuracies β, and a pair of frames from two sequences.

We see empirically in Table 1 that RWMH does, in fact, reduce the prediction-error

Table 1: γ calculated on the MC residual between frames 52 and 53 of the "NYC" sequence and frames 60 and 61 of the "Football" sequence for MC blocks of size $B \times B$, an RDWT of J decomposition scales, and ME accuracy of β.

		"NYC" γ (dB)			"Football" γ (dB)		
J	B	$\beta = -1$	$\beta = -2$	$\beta = -3$	$\beta = -1$	$\beta = -2$	$\beta = -3$
1	2	-4.29	-4.45	-4.65	-4.34	-4.45	-4.62
2	2	-5.20	-5.27	-5.39	-4.81	-4.88	-5.00
3	2	-5.68	-5.75	-5.84	-5.08	-5.14	-5.24
4	2	-5.96	-6.01	-6.09	-5.21	-5.27	-5.35
3	2	-5.68	-5.75	-5.84	-5.08	-5.14	-5.24
3	4	-4.86	-5.29	-5.50	-3.87	-4.36	-4.56
3	8	-2.02	-2.25	-2.35	-1.96	-2.26	-2.34
3	16	-0.92	-0.99	-1.05	-0.93	-1.06	-1.12

variance over that of single-phase MC as predicted by the analysis above. Specifically, as the number of scales of decomposition is varied with a fixed block size, we see an exponential relationship between γ and J as indicated by (17)—as J increases, γ decreases by a decreasing amount. Additionally, as J is held fixed, we see that the gain γ varies significantly with the block size used in MC. This is as expected, since we anticipate that smaller MC block sizes result in a prediction residual that is less correlated, at least spatially, and is thus more "noise-like." We see that, for extremely small MC block sizes, the reduction in prediction-error variance approaches 5–6 dB—close to the maximum gain of 7 dB as anticipated by Fig. 2. Additionally, we observe that the reduction in prediction-error variance increases with increasing ME accuracy, again as predicted by Fig. 2. Thus, we conclude that RWMH does indeed lead to variance reduction for real MC residuals in much the same way as indicated by the analysis above despite the fact that the white-noise assumption underlying Theorem 1 does not strictly hold in practice.

Conclusions

In this paper, we have examined a new class of MHMC—phase-diversity multihypothesis—by analyzing MHMC in the domain of a redundant wavelet transform. Recognizing that RDWT coefficients with different phases view motion from different perspectives, this RWMH system treats each critically sampled DWT within the RDWT as a separate hypothesis prediction. A multiple-phase inverse RDWT operation implicitly combines the multiple predictions into a single spatial-domain prediction. The primary contribution of the work we report here is an analytical derivation that quantifies the performance gain of RWMH over single-phase prediction. Key to this analysis is that noise in the RDWT domain undergoes a substantial reduction in variance when the multiple-phase inverse RDWT is applied due to the well-known fact that this pseudo-inverse contains a projection onto the range space of the forward transform. Consequently, noise not captured by the motion model is greatly reduced in an RWMH system, leading to substantial reduction in the prediction-residual variance and higher coding efficiency. In fact, our analysis predicts that

RWMH can reduce the prediction-residual variance by up to 7 dB regardless of the power of the noise as ME becomes highly accurate. Our experimental observations support the analysis in that performance gains seen for real motion residuals approximately follow the predicted behavior.

References

[1] 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*, vol. 5, Minneapolis, MN, April 1993, pp. 437–440.

[2] B. Girod, "Motion-compensating prediction with fractional-pel accuracy," *IEEE Transactions on Communications*, vol. 41, no. 4, pp. 604–612, April 1993.

[3] M. T. Orchard and G. J. Sullivan, "Overlapped block motion compensation: An estimation-theoretic approach," *IEEE Transactions on Image Processing*, vol. 3, no. 5, pp. 693–699, September 1994.

[4] 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.

[5] 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*, vol. 2, Barcelona, Spain, September 2003, pp. 53–56.

[6] 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.

[7] M. Holschneider, R. Kronland-Martinet, J. Morlet, and P. Tchamitchian, "A real-time algorithm for signal analysis with the help of the wavelet transform," in *Wavelets: Time-Frequency Methods and Phase Space*, J.-M. Combes, A. Grossman, and P. Tchamitchian, Eds. Berlin, Germany: Springer-Verlag, 1989, pp. 286–297.

[8] P. Dutilleux, "An implementation of the "algorithme à trous" to compute the wavelet transform," in *Wavelets: Time-Frequency Methods and Phase Space*, J.-M. Combes, A. Grossman, and P. Tchamitchian, Eds. Berlin, Germany: Springer-Verlag, 1989, pp. 298–304.

[9] M. J. Shensa, "The discrete wavelet transform: Wedding the à trous and Mallat algorithms," *IEEE Transactions on Signal Processing*, vol. 40, no. 10, pp. 2464–2482, October 1992.

[10] J. E. Fowler, S. Cui, and Y. Wang, "Motion compensation via redundant-wavelet multihypothesis," *IEEE Transactions on Image Processing*, submitted.

[11] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1992.

[12] B. Girod, "The efficiency of motion-compensating prediction for hybrid coding of video sequences," *IEEE Journal on Selected Areas in Communications*, vol. 5, no. 7, pp. 1140–1154, August 1987.

[13] J. E. Fowler, "The redundant discrete wavelet transform and additive noise," *IEEE Signal Processing Letters*, vol. 12, no. 9, pp. 629–632, September 2005.

[14] B. Girod, "Efficiency analysis of multihypothesis motion-compensated prediction for video coding," *IEEE Transactions on Image Processing*, vol. 9, no. 2, pp. 173–183, February 2000.

In *Proceedings of the IEEE Data Compression Conference*, J. A. Storer and M. Cohn, Eds., Snowbird, UT, March 2006, pp. 352–361.