FAST MOTION ESTIMATION BASED ON THE REDUNDANT WAVELET FOR 2D+T WAVELET CODER

Heechan Park, Graham R Martin and Andy C Yu

Department of Computer Science
University of Warwick, Coventry United Kingdom
email: {heechan,grm, andycyu}@dcs.warwick.ac.uk

ABSTRACT

Scalable video coding utilizing the wavelet transform applied to both the spatial and temporal domains (3D-DWT) is of current interest in video coding. In the case of the 2D+t scheme, motion estimation is required to be conducted in the wavelet domain. We explore approaches that overcome the well-known shift-variant property of the wavelet transform so that motion is estimated accurately. Specifically, an algorithm using the redundant wavelet and phase-multi-hypothesis shows PSNR superiority but computational complexity is high. We propose a fast motion search method that significantly reduces computational complexity to as low as 2% of the existing algorithm yet still shows better performance than full-spatial motion estimation.

1. INTRODUCTION

Conventional video compression is based on the motion compensated prediction from reconstructed frames in the decoder in a recursive fashion and residue coding by a frequency based technique such as the discrete cosine transform. Due to the recursion and the inherent drift, scalability is difficult to establish in this so-called hybrid coder. Inter-frame wavelet coding overcomes this limitation by replacing the prediction along the time axis by a wavelet filter[1]. Depending on the order of wavelet filters, wavelet coding systems are defined with different characteristics. We focus on the 2D+t system where a spatial filter is followed by a temporal filter. This system provides a more flexible and adaptive system; for instance, it allows a different level of temporal filtering depending on the spatial subband. However, motion estimation must be performed in the wavelet domain that exhibits shift-variance. There are a couple of solutions that overcome the shift-variance of the wavelet. Kim and Park[2] suggested that a low-band-shift method that uses the wavelet macroblock from the redundant wavelet representation. S. Cui *et. al.*[3] proposed a similar but direct and simple method performed in the redundant wavelet domain. This enables a multi-hypothesis paradigm using phase diversity in the wavelet domain, which in turn offers substantial improvement over Kim and Park's approach. Both methods are computationally demanding compared to the spatial motion compensation scheme. We focus on S. Cui's algorithm and attempt to reduce computational complexity.

This paper is organized as follows. Sec. 2 introduces the inter-frame wavelet coder to highlight the 2D+T structure. Sec. 3 discusses the use of the redundant wavelet transform, and Sec. 4 proposes a fast and accurate motion estimation technique. Sec. 5 provides experimental results and finally conclusions are drawn in Sec. 6.

2. INTER-FRAME WAVELET CODER

The objective of scalable video coding is to introduce an effective decoding mechanism to suit different end-users. Ideally, the new video compression architecture should provide scalability in all aspects, that is spatially, temporally, and in quality and complexity. The multi-resolution paradigm behind the wavelet transform that can be applied not only spatially but also temporally provides an important approach. However, there is an issue regarding the order of wavelet filters for different dimensions. Although the order of spatial and temporal filters appear insignificant, they provide an important insight into the wavelet video coder and different properties are observed. We briefly mention a couple of options enabled by the 2D+t scheme. Firstly, the accuracy of the motion estimation and GOF per subband can be adaptively determined while this is not viable in t+2D. Secondly, different temporal filters can be utilized at each subband. For instance, bi-directional temporal filtering can be used for the low bands, while only forward temporal filtering can be used for the higher bands. The flexible choice of temporal filtering options makes the 2D+t framework deviate from the strict decomposition scheme as performed in the t+2D to provide a flexible 3D decomposition scheme as shown in Fig. 1. Finally, motion estimation/compensation combined with the redundant wavelet, which we shall explore in the next section, provides better PSNR performance

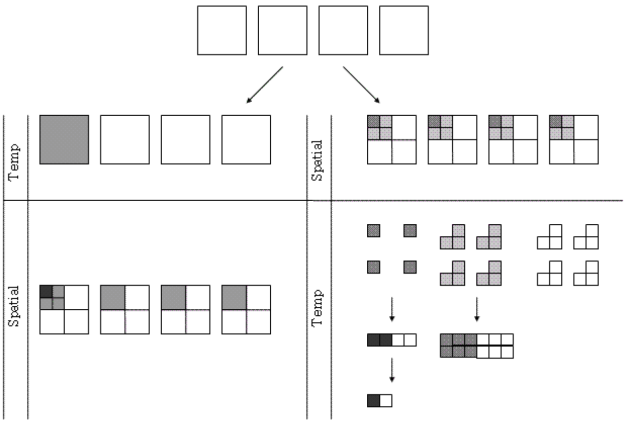

Fig. 1. t+2D(left) and 2D+t(right) architectures

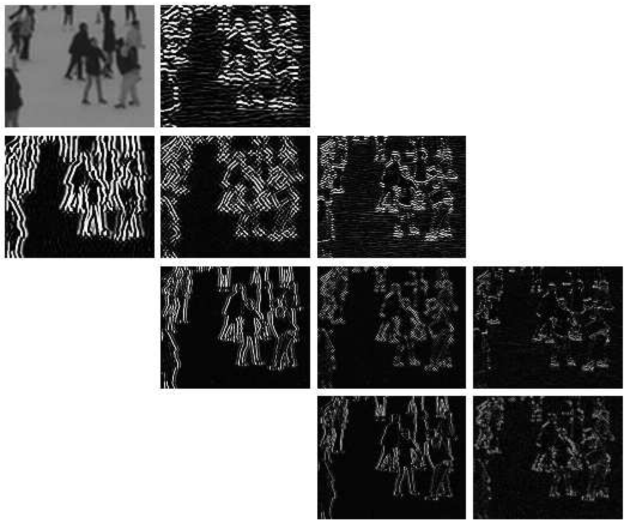

Fig. 2. Redundant wavelet decomposition

than if conducted in the spatial domain.

3. REDUNDANT WAVELET AND MULTI-HYPOTHESIS

In essence, the redundant discrete wavelet transform removes the downsampling procedure from the 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. Fig. 2 illustrates the redundant wavelet decomposition (3 levels). The most direct implementation of the algorithms is a trous wavelet; resulting in subbands that are exactly the same size as the original signal. Due to its fixed sampling rate, each RDWT coefficient across scale offers a spatially coherent representation, which in turn guarantees shift invariance. With appropriate subsampling of each subband of RDWT, one can produce exactly the same coefficients as a critically sampled DWT applied to the same input signal. There are several methods to invert the RDWT. The single-phase inverse consists of subsampling the RDWT coefficients to extract one critically sampled DWT from the RDWT and application of the corresponding inverse DWT,

$$Recon[x, y] = InvDWT(sub(RDWT[x, y])) \quad (1)$$

where *InvDWT()* is the inverse DWT and *sub()* denotes subsampling to a single-phase. Alternatively, one can employ a multiple-phase inverse, denoted as

$$Recon[x, y] = InvRDWT(RDWT[x, y]) \quad (2)$$

In the multiple-phase inverse, one independently inverts each of the 4J critically sampled DWTs constituting the J-scale 2D RDWT and average the resulting reconstructions, or one can employ an equivalent but more computationally efficient filtering implementation. In either case, the single-phase inverse is generally not the same as the multiple-phase inverse. The multiple-phase inverse is shift invariant under linear fractional-pixel interpolation, while the single-phase inverse is not. S. Cui. *et. al*[3] extended the above-mentioned multiple-phase inverse operation to a multi-hypothesis paradigm that takes advantage of the diversity of phase in the redundant wavelet representation (RWMH). This system achieves a substantial gain over the single-phase prediction system(RWSH)[2].

4. FAST MOTION SEARCH

A redundant wavelet representation shows merits for deployment in a scalable video coder. However, motion estimation in the redundant wavelet domain increases computational complexity because it requires calculating the absolute difference of blocks across all subbands for each position while the number of subbands increases as the level of decomposition. For instance, 3 levels of decomposition produces 10 subbands. As information in the subbands is redundant, it might not be necessary to consider all subbands when calculating the absolute difference. Also, the wavelet representation possesses useful properties such as a hierarchical structure and decomposition into horizontal, vertical and diagonal subbands. This is the motivation behind our investigation for a fast motion estimation technique.

We propose a fast motion search method combining the

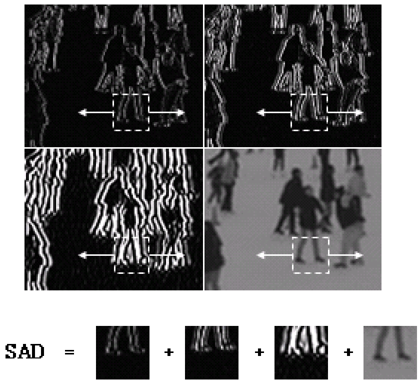

Fig. 3. LLL and vertical high subbands perpendicular to search direction

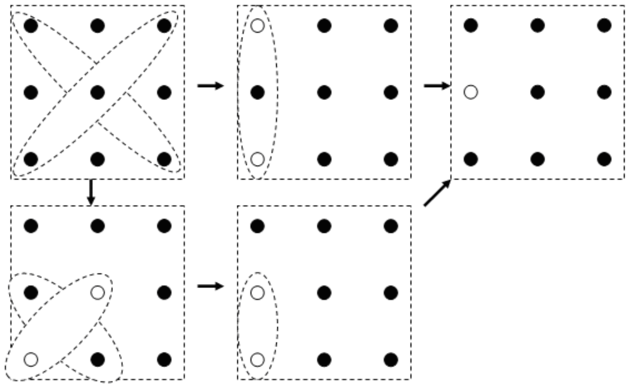

Fig. 4. Proposed search: two possible scenarios after diagonal search

three-step search and a subband-based search. The three-step search is an effective and fast motion search algorithm[4]. Computation can be further enhanced by selectively considering subbands that are perpendicular to the search direction. For example, for horizontal motion search (x component), the vertical high bands are likely to contribute more than other subbands, thus taking more vertical subbands into account while reducing other subbands as shown in Fig. 3.

Our algorithm shares the same overall procedure as the three-step search that determines the optimal search point among nodes of a lattice and repositions a half-sized lattice about the optimal point found. This process is repeated until it reaches 3×3 lattice size. In our algorithm, the full-search method in each level is replaced by the subband-based search. Fig. 4 illustrates the subband-based search. The algorithm starts by making two diagonal search point groups and determines an optimum point for each diagonal group. There are two possible scenarios onwards. One includes the center point as optimum in either diagonal group and the other does not. As illustrated in Fig. 4, the top flow shows the latter case and the bottom flow shows the former case. For the latter case, the search finishes by comparing the two optimal points and a point in-between. For the former case, a similar diagonal search is performed again with two optimal points found and two additional points, and finally a comparison between the local optimum point from each diagonal search group determines the final optimum point. Note that the horizontal, vertical and diagonal subbands are related to the vertical, horizontal and diagonal searches respectively, in the sense that a subband perpendicular to the search direction contributes more than other subbands.

5. EXPERIMENTS

We evaluated several algorithms along with our algorithm using common settings as follows: CIF resolution, a fixed bitrate of 0.5 bpp and blocksize 16×16. The aim of the experiment was to determine, for each algorithm, the effectiveness in terms of PSNR against the efficiency in terms of computation. Computation was measured by counting the number of block difference operation.
RWSH[2], RWMH, RWMH with one-level decomposition (ONELEVEL), full-spatial search (SPATIAL), Three-step search (TSS), Diamond search (DS) and Hierarchical search (HIER) are compared and results are tabulated in Table. 1.
The RWSH system takes more time than RWMH due to the subsampling procedure to construct the tree structure from wavelet macroblocks (for more details, see [3]). The hierarchical method starts searching from the lowest level using only subbands at the current level and proceeds to the higher level with a reduced search window and more subbands included. The hierarchical search shows similar performance to RWMH but with 50% reduced computation. However, the hierarchical search requires five times as much computation as the full-spatial search. We also attempted a hierarchical search considering subbands at only the current level, motion vectors are easily mis-selected in the highpass band due to the scattered representation. The RWMH with one-level decomposition shows comparable performance to the hierarchical search. Surprisingly, the TSS shows better performance than DS with comparable computation which contrasts the TSS and DS comparison found in the literature. As a motion-compensated redundant wavelet representation undergoes multi-phase to single-phase conversion[Eq(2)], the estimation of motion is not necessarily obvious. From another angle, one might consider block matching on high frequency filtered images as correlation. Empir-

	RWSH	RWMH	HEIR	ONELEVEL	SPATIAL	TSS	DS	PROPOSED
ICE	39.55	**40.29**	40.28	40.05	**38.94**	39.87	40.15	**39.77**
CREW	37.71	**37.81**	37.80	37.79	**36.62**	37.82	37.79	**37.73**
FOOTBALL	27.82	**27.89**	27.86	27.85	**27.11**	27.72	27.75	**27.72**
SUSIE	39.15	**39.79**	39.79	39.79	**39.28**	39.79	39.79	**39.79**
STEFAN	26.83	**27.33**	27.32	27.28	**26.81**	27.18	27.30	**27.06**
COMPUTATION	>>	**100%**	51%	44%	**11%**	5%	5%	**2%**
					100%	42%	40%	**19%**
						100%	95%	**44%**

Table 1. PSNR results at a fixed bitrate of 0.5 bpp

ically, we observe that optimal motion vectors found in the redundant wavelet domain are much closer to those found by correlation than those obtained by spatial block matching; a wavelet representation decomposes an input into sparse highpass and dense lowpass components, and weight factors between subbands can affect performance.

According to the results of Table. 1, our algorithm achieves consistently better performance than full-spatial search and with only 19% of the computation. PSNR deterioration is only 0.1dB compared with the TSS while computation is reduced by 56%. Our experiments show that utilizing perpendicular subbands to search direction is an effective technique to reduce computation, however the hierarchical structure remains difficult to use and futher work needs to be done.

6. CONCLUSION

We have discussed the role of the redundant wavelet in scalable video coding, the advantages of a phase-multi-hypothesis and its high computational requirement. We propose a fast motion search technique that utilizes directional subbands in combination with the three-step search. The technique reduces computation by 98%, 81% and 56% compared to RWMH, full-spatial search and the three-step search respectively, and yet still results in comparable performance to the three-step search method in the wavelet domain. Further improvement may be possible if the hierarchical structure can be utilized efficiently.

7. REFERENCES

[1] J. Ohm, M. Schaar, and J. Woods, "Interframe wavelet coding-motion picture representation for universal scalability," *Signal Process.: Image Commun.*, vol. 19, no. 9, pp. 877–908, Oct. 2004.

[2] H. Park and H. Kim, "Motion estimation using low-band-shift method for wavelet-based moving-picture coding," *IEEE Trans. Image Process.*, vol. 9, pp. 577–587, Apr. 2000.

[3] S. Cui, Y. Wang, and J.E. Fowler, "Multihypothesis motion compensation in the redundant wavelet domain," in *IEEE ICIP.*, Barcelona Spain, Sep. 2003, vol. 2, pp. 53–56.

[4] T. Koga, K. Iinuma, A. Hirano, Y. Iijima, and T. Ishiguro, "Motion-compensated interframe coding for video conferencing," *IEEE NTC*, pp. 531–534, 1981.