BLOCK MOTION ESTIMATION WITH THE WREATH PRODUCT TRANSFORM AND SUBBLOCK ROTATION TECHNIQUE

Xuling Luo and Gagan Mirchandani

Department of Electrical and Computer Engineering
The University of Vermont
Burlington, Vermont 05405

ABSTRACT

A novel multiresolution block motion estimation algorithm for video coding is proposed in this paper. The algorithm estimates motion vectors in a coarse-to-fine scheme, using the wreath product transform (WPT). All of its subband coefficients are used to predict the motion vectors, generating an estimate close to that obtained from the standard exhaustive search method. The multiplication free characteristic of the WPT, provides a greater computational advantage over other multiresolution algorithms. In addition, a subblock rotation technique is proposed to compensate for rotational movements, which is a novel attempt in improving the estimate. When a rotational motion occurs frequently within the video, this technique substantially improves the estimate, yet only needs a little extra computation load within the multiresolution framework.

1. INTRODUCTION

Video coding is important for effective transmission and storage in this information society. Motion compensated interframe coding is an effective video coding technique which is widely adopted in video compression standards such as H.261 and MPEGs [1]-[5]. Motion compensation is usually carried out using estimated motion vectors on a block-by-block basis. The estimated motion vector is the one which minimizes a given distortion function. Motion vectors and corresponding quantized prediction error values are then transmitted.

The accuracy and robustness of the motion estimator is paramount in producing a successful estimate. There are some different matching criteria available for estimating, such as mean-square-error (MSE) and mean-absolute-error (MAE). MAE is usually used as the estimation error metric, since it requires no multiplication and produces results not dissimilar to that with the MSE criterion. As is well known, motion estimation is a computationally intensive task. Therefore, computational simplicity is another important requirement for the motion estimation method.

Many motion estimation methods have been proposed in the literature [3]-[7]. They can be generally divided into two categories, namely, hierarchical methods and gradient based methods. Hierarchical methods use a coarse-to-fine strategy, search a subset of eligible candidate blocks and typically have a lower computation load. When the search strategy is carefully selected, these methods are usually robust. Gradient based methods search for a minimum of an error surface, using some optimization techniques. The optimization algorithm determines the compromise between accuracy and the computation load.

The theory of wavelets and filter banks theory has developed considerably in the past decade and found many applications. Motion estimation algorithms with these multiresolution analysis techniques have been explored in [8]-[12]. Both estimation errors and computational loads comparable to that obtained with the full search algorithm (FSA) are obtainable with some of these algorithms. In this paper, we develop a novel multiresolution block motion estimation algorithm, based on the quadtree-based WPT [13, 14, 15], which is a multiplication free wavelet transform. With this algorithm, we can achieve estimation errors similar to that with the FSA, but with only 1/7th the computational load. We further propose an additional technique called subblock rotation (SBR), to estimate rotational movements between video frames, thus further reducing the estimation error. Within the WPT's multiresolution framework, it needs only a little more computation time for fast implementation.

The paper is organized as follows. In Section 2, we review the block motion estimation problem and the three step search algorithm, which we use for comparison. The quadtree-based WPT multiresolution algorithm is proposed in Section 3, and its performance and computational complexity are examined. In Section 4 we describe the subblock rotation technique. Discussion and conclusion are given in Sections 5 and 6.

2. BLOCK MOTION ESTIMATION

In block (matching) motion estimation, the current frame is divided into blocks of size $N \times N$. The estimation algorithm seeks to determine the minimum error motion vector between the $N \times N$ block in the current frame and its counterpart, within a prescribed neighborhood, in the previous frame. If the neighborhood is described by a maximum displacement in horizontal and vertical directions by w pixels, then the exhaustive search method (FSA) checks all $(2w+1)^2$ possible motion vectors. The three step search (TSS) algorithm [4] looks for the motion vector in multiple steps: first determine the best match from nine uniformly distributed points in the search space; then check another nine points with a smaller search range centered at the position found in the first step; finally check the nine point neighborhood of the position found in the second step. The FSA and TSS are well known and repeatable, have no adaptive parameters, and hence are used for comparison in this study.

The MAE used as the prediction error metric is defined as follows:

$$MAE(m,n) = \sum_{k=1}^{N}\sum_{l=1}^{N} |f_c(k,l) - f_p(k+m, l+n)|, \quad -w \le m,n \le w, \quad (1)$$

where (m,n) is the candidate motion vector, and f_c and f_p refer to the current and previous frames. While the MAE is used as the measure for determining the motion vector, we use MSE to evaluate the performance of the algorithms.

3. MOTION ESTIMATION IN WREATH PRODUCT TRANSFORM DOMAIN

3.1. WPT spectra

The WPT, a cyclic (wreath product) group-based transform [13, 14, 15], is a 1-dimensional block transform that provides a multiresolution representation of an image. Equivalently, the quadtree spectrum of the image is obtained by a nested wreath product Fourier series decomposition, which is multiplication free. Using the notation of the above references, a two-level decomposition of an image can be expressed as follows:

$$\begin{aligned} L(1,2) &= Q_0^1 \oplus Q_1^1 \oplus Q_2^1 \oplus Q_3^1, \\ Q_0^1 &= Q_0^2 \oplus Q_1^2 \oplus Q_2^2 \oplus Q_3^2, \end{aligned} \quad (2)$$

where $L(1,2)$ is the two-level quadtree-based vector space representing the image and Q_k^l is the vector space placed in the kth quadrant of scale l in the decomposition. Q_1^l and Q_3^l are complex conjugate. For our purposes, we replace Q_3^l with Q_0^l, which is used in our analysis.

3.2. WPT-based motion estimation algorithm

The WPT-based multiresolution motion estimation algorithm (WBA) employs a coarse-to-fine searching scheme. First, a block in the current frame is transformed. (We assume that the wreath product spectrum of the previous frame is available). Then, the Q_1^2 coefficients of the current block and previous frame are used to choose a candidate motion vector, using the FSA with a maximum displacement of $(w+1)/4$. Next, the corresponding higher resolution coefficients in the Q_1^1 subspace of the current block and the previous frame, are used to choose a new candidate motion vector $\mathbf{v_1}$, searching through all nine points surrounding the previous (lower resolution) candidate motion vector. Steps 2 and 3 are repeated for vector spaces Q_2^l and Q_3^l for $l = 2, 1$, thus generating two more candidate motion vectors $\mathbf{v_2}$ and $\mathbf{v_3}$. All three candidate motion vectors are then checked in the original spatial domain, and the best one is kept. The new candidate's nine point neighborhood is then checked, and the vector which achieves the minimum MAE is then saved as the final estimate. Lastly, the transform coefficients of the current block are stored for future use.

3.3. Simulation results and computational complexity

Two examples are given to show the performance of the above algorithm. The FOOTBALL and STEFAN sequences are a set of frames for two popular sports, publically available [1]. In our experiments, we use $N = 16$ and $w = 7$ to specify the block size and search range. After estimation over 50 frames using our method and that of FSA and TSS, the MSEs obtained for the two video sequences are shown in Figure 1. From the figure, it is clear that the performance of the new algorithm is better than TSS and close to FSA.

The WPT-based algorithm is computationally efficient. Assume that checking one possible motion vector requires one operation unit time. Then obtaining the motion vector for a normal block (a block not at the edges) with the FSA requires 225 operation units while the TSS needs 27. Our method requires less than 31 operation units, including the overhead needed for the transform. Comparing with other multiresolution algorithms, which usually are comparable to the FSA, this algorithm has a considerable advantage in computational complexity.

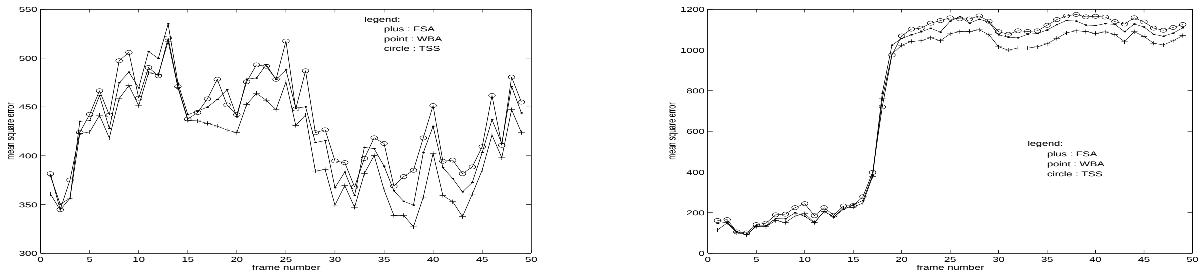
4. SUBBLOCK ROTATION TECHNIQUE

4.1. Rotation compensation

Block matching motion estimation in motion compensated video coding attempts to predict movements between contiguous frames of the video. There are some fundamental assumptions in this structure. First, that all pixels and objects within such an artificially formed block move in the same direction. Secondly, that such a uniform movement is translational. These assumptions may not necessarily be true. Consider a block containing a moving object. Here we may have more than two regions (object and background) moving in different directions. Furthermore, a movement other than translational, such as rotational, occurs frequently at various scales. When these fundamental assumptions are not satisfied, the result is inaccurate compensation. Thus increased bit rate is needed to encode the significantly increased residual error. In [1], the authors presented a method of segmenting the motion field of blocks and applying different motion vectors to each segment, thus relaxing the constraint of constant translational motion at blocks located on boundaries of moving objects. This partially solves the inaccurate compensation due to the first category of assumptions.

We try to improve the performance of block motion compensation by using a subblock rotation technique, which estimates subblock rotation vectors and compensates for some specific rotational movements, thus relaxing the second assumption. We note that rotational movement can occur at different scales and in various combinations. For example, a block can contain a small object rotating 90 degree clockwise, and lying within a larger object rotating 90 degree counterclockwise, in the context of a previous frame.

[1] *www.image.cityu.edu.hk/imagedb*



(a) FOOTBALL sequence

(b) STEFAN sequence

Figure 1: MSEs of FOOTBALL and STEFAN

Searching and compensating for this kind of movement can be computationally prohibitive. In our work, we consider a subset of rotations illustrated in Figure 2(a). A $N \times N$ block, called scale 0 block is divided into four subblocks, that are of size $N/2 \times N/2$ and called scale 1 blocks. The scale 1 blocks are again divided into four smaller subblocks, of size $N/4 \times N/4$ and called scale 2 blocks. All four subblocks within lower scale subblocks can rotate 0, 90, 180 and 270 degrees clockwise. Thus a total of $4\times4\times4\times4\times4 = 1024$ rotations are possible. That is the rotation set we consider. In Figure 2(b) and 2(c), we show the MSEs obtained with the FSA when rotational movements are compensated for, in the FOOTBALL and STEFAN sequences. The prediction error is greatly reduced in these two examples. For rotation compensation with the WPT multiresolution representation, a fast implementation is derived and its effectiveness is shown in the following subsection.

4.2. Rotation estimation with the WBA

A fast algorithm to search through the subset of rotational movements is implemented as follows. The first 5 steps are the same as in the motion estimation algorithm for translation compensation in Section 3.2; the best candidate among $\mathbf{v_1}$, $\mathbf{v_2}$ and $\mathbf{v_3}$ is kept. Rotation estimation is considered in scale 1 coefficients as follows: if $\mathbf{v_k}$ is the selected vector, then the Q_k^1 spectral coefficients are used. Scale 1 subblock rotations are checked first, after which the four scale 2 subblock rotations are checked. This way, we assume lower scale (high resolution) rotation priority over higher scale rotation. After the rotation vector $\mathbf{r_k}$ is obtained, we proceed to the next step. The selected vector $\mathbf{v_k}$ is checked at the nine point neighborhood, and the vector which achieves minimum MAE is saved. This new vector and its rotated version with vector $\mathbf{r_k}$ compete for the best (translation and rotation vector) solution.

Searching in scale 1 coefficients makes the algorithm faster, but with a degraded performance compared to the original subblock rotation method. The fast algorithm needs about 2.5 operation units, thus is quite computationally efficient. The performance of this motion estimation algorithm with fast subblock rotation technique is presented in Figure 2(b) and 2(c). Although the fast SBR algorithm is not as good as the original SBR, it still achieves a better performance than FSA, which can be observed from the figure.

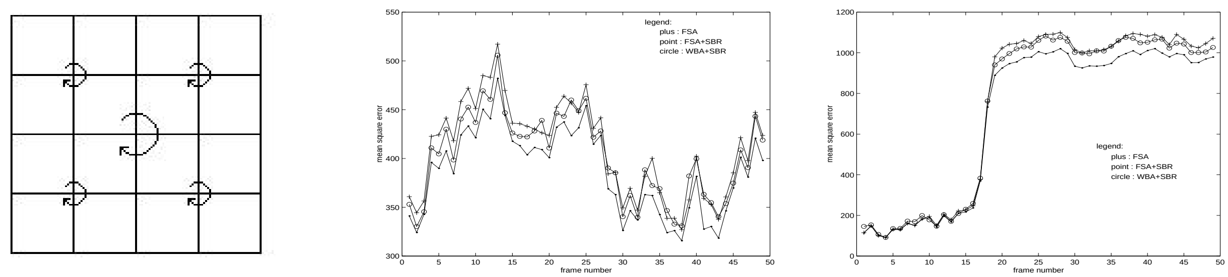
5. DISCUSSION

Comparing the WBA to other wavelet based multiresolution motion estimation algorithms, we note differences other than computational simplicity. Most multiresolution algorithms use coefficients in only some of the subbands. In our experiments, we observe that such an implementation can retrieve at most 85% motion vectors found by FSA. With all the lowpass and bandpass branches used, we achieve above 95% motion vectors found by FSA. This appears to hold for most wavelet decompositions. Depending on the algorithm, the lowpass subband may not necessarily be available in the multiresolution decomposition. For the WPT, this can be overcome by replacing Q_3^l with Q_0^l, since Q_3^l and Q_1^l are complex conjugate, and there is no need to keep both of them.

Rotational vectors are generated and need to be transmitted in the algorithm proposed in Section 4.2. We estimate the overhead needed. For the scale 1 subblock rotation, there are 4 possibilities; thus 2 bits are needed for representation of the rotation vector. For the four scale 2 subblock rotations, 4x2 bits are needed for representation of the rotation vectors. In total, we need 10 bits to represent the rotation for a 16x16 pixels' block. Therefore, the bit rate is increased 0.039 bpp (no compression case). If a video signal contains no rotational movements, then the rotation vector will contain long zeros, which can be compressed efficiently with run-length coding. An interesting question is when the overhead needed for coding rotational vectors, exceeds the bit rate reduction achieved by the more precise prediction of rotation. This deserves further investigation.

6. CONCLUSION

A fast WPT-based multiresolution block motion estimation algorithm has been presented. Coarse motion vectors are predicted from high scale coefficients with exhaustive search



(a) Subblock rotation pattern

(b) FOOTBALL sequence

(c) STEFAN sequence

Figure 2: Motion estimation with subblock rotation technique

in the WPT representation. They are refined in the lower scales gradually, by searching through specific neighborhoods. Three candidates are then obtained from three subbranches, which then compete for the final estimate. This algorithm has a computational complexity comparable to the three step search algorithm, but with a better performance.

The algorithm can be improved with the subblock rotation technique, which imposes very little computation load when implemented within the multiresolution framework. This combination is quite effective, especially when the video signals incorporate rotational movements.

7. REFERENCES

[1] M.Orchard, "Predictive motion-field segmentation for image sequence coding", IEEE Trans. Circuits and Systems for Video Technol., vol.3, no.1, pp.54-70, 1993.

[2] H. G. Musmann, P.Pirsch and H.J.Grallert, "Advances in picture coding", Proc. IEEE, vol.73, pp.523-548, Apr. 1985.

[3] S. Kappagantula and K.R.Rao, "Motion compensated interframe image prediction", IEEE Trans. on Commun., vol.33, no.9, pp.1011-1015, 1985.

[4] T.Koga, K.Iinuma, A.Hirano, Y.Iijima and T.Ishiguro, "Motion-compensated interframe coding for video conferencing", Proc. Nat. Telecommun. Conf., pp.G.5.3.1-G.5.3.5, 1981.

[5] Y.L.Chan and W.C.Siu, "Reliable block motion estimation through the confidence measure of error surface", Signal Processing, vol.76, no.2, pp.135-146, 1999.

[6] R.Li, B.Zeng and M.L.Liou, "A new three-step search algorithm for block motion estimation", IEEE Trans. Circuits and Systems for Video Technol., vol.4, no.4, pp.438-442, 1994.

[7] B.Liu and A.Zaccarin, "New fast algorithms for the estimation of block motion vectors", IEEE Trans. Circuits and Systems for Video Technol., vol.3, no.2, pp.148-157, 1993.

[8] M.H.Ahmad Fadzil and T.J.Dennis, "Video subband VQ coding at 64 kbit/s using short-kernal filter banks with an improved motion estimation technique", Signal Processing: Image Communication, vol.3, no.1, pp.3-21, 1991.

[9] M.K.Mandal, E.Chan, X.Wang and S. Panchanathan, "Multiresolution motion estimation techniques for video compression", Optical Engineering, vol.35, no.1, pp.128-136, 1996.

[10] Y.Zhang and S.Zafar, "Motion-compensated wavelet transform coding for color video compression", IEEE Trans. Circuits and Systems for Video Technol., vol.2, no.3, pp.285-296, 1992.

[11] J.Magarey and N.Kingsbury, "Motion estimation using a complex-valued wavelet transform", IEEE Trans. Signal Processing, vol.46, no.4, pp.1069-1084, 1998.

[12] R.W.Young and N.G.Kingsbury, "Frequency-domain motion estimation using a complex lapped transform", IEEE Trans. Image Processing, vol.2, no.1, pp.2-17, 1993.

[13] D.Healy, G.Mirchandani, T.Olson and D.Rockmore, "Wreath products for image processing", Proceedings of ICASSP, vol.6, pp.3582-3586, 1996.

[14] R.Foote, G.Mirchandani, D.Rockmore, D.Healy and T.Olson, "A wreath product group approach to signal and image processing: part I — multiresolution analysis", to appear in IEEE Trans. on Signal Processing, January, 2000.

[15] G.Mirchandani, R.Foote, D.Rockmore, D.Healy and T.Olson, "A wreath product group approach to signal and image processing: part II-convolution, correlation, and applications", to appear in IEEE Trans. on Signal Processing, 2000.