An Enhancement to MRMC Scheme in Video Compression

Jie Wei and Ze-Nian Li

***Abstract*— Zhang and Zafar proposed a video compression scheme based on the wavelet representation and multiresolution motion compensation (MRMC). In this letter, an additional masking module will be created to further enhance its efficiency. Specifically, between the modules of wavelet decomposition and MRMC, the masking module will be inserted which will construct binary images based on the difference of the wavelet coefficients. The binary images will serve as masks to facilitate a more efficient motion compensation. Experiments show that the processing time could be significantly reduced from what is required for a full search in the original MRMC algorithm.**

***Index Terms*— Image analysis, image coding, motion compensation, wavelet transforms.**

I. Introduction

Zhang and Zafar [1] proposed a video compression scheme based on the wavelet representation and multiresolution motion compensation (MRMC) where the size of the block in which the motion vector is computed is adapted to its level in the wavelet pyramid [2]. Their motion compensation is based on the block-matching algorithm (BMA) where each $k \times k$ block of the current video frame is compared to a block of the same size in the previous (reference) frame in the vicinity of its corresponding position. Due to the nature of the multiresolution wavelet decomposition of images, the size of the block in different subbands in the wavelet pyramid is varied according to the resolution; the higher the resolution, the larger the block size. This scheme takes advantage of the discrete wavelet transform (DWT) with a multiresolution approach to the motion estimation. Test results in [1] have shown good performance of the algorithm.

In the BMA, the full-search is computationally intensive and hence quite time consuming. In order to reduce it, many methods aimed at cutting its computation have been proposed, such as the logarithmic search procedure [3], three-step search [4], and conjugate direction search [5]. The multiresolution motion estimation (or hierarchical motion estimation) [6], [7] is based on the Laplacian pyramids or subband coefficients. It first estimates a coarse displacement by employing the lower resolution images and then refines them utilizing the detailed images. The MRMC proposed in [1] was on the same track of this idea applied to the wavelet representations. Due to the advantages of wavelets, namely, localizations in both space (or time) and frequency [8], and spatial orientation selectivity [2], this method of motion estimation has a potential of being more efficient. To enhance the efficiency of the MRMC, in this letter we propose a new method for reducing the computation overhead by exploring the temporal redundancy between video frames. Since the DWT is a linear localized transform, after wavelet decomposition, the above redundancy and locality are preserved between the corresponding coefficients of the consecutive frames. Therefore, if we can locate (predict) the areas where motion has likely occurred in the wavelet subimages before the actual motion estimation, the subsequent search for the motion vectors will be conducted only in these areas—a sure savings in computational cost. This constitutes the basis of the *enhanced MRMC* proposed in this letter.

II. MRMC and the Enhanced MRMC

This section describes the original MRMC scheme [1] and the proposed enhanced MRMC. Without loss of generality, a wavelet decomposition pyramid of three levels will be used in this letter.

A. MRMC

As shown in Fig. 1, the total number of wavelet subimages is ten. At the highest level (Level 3) of the pyramid are the wavelet subimages W_8^1, W_8^2, W_8^3, and S_8—the subimage generated by applying the scaling function three times. If a block size of 2 × 2 for the BMA is chosen at this level, then as shown in Fig. 1, for subimages at Level 2: W_4^1, W_4^2, and W_4^3, the block size is 4 × 4; whereas for subimages at Level 1: W_2^1, W_2^2, and W_2^3, it is 8 × 8.

Let $\mathbf{V}(x, y)$ represent the motion vectors of the block whose lower left corner is located at position (x, y) and whose size varies according to its level. What follows is the scheme of the MRMC [1]. The motion vectors $\mathbf{V}_8(x, y)$ of those 2 × 2 blocks in S_8 are first estimated by the BMA. Let $\mathbf{V}_j^i(x, y)$ represent the motion vectors of those blocks with the size $16/j \times 16/j$ for the wavelet subimage W_j^i (where $i = 1, 2, 3$ and $j = 2, 4, 8$), and $\mathbf{V}_8^i(x, y)$ is simply copied from $\mathbf{V}_8(x, y)$. The motion vectors in subimages at lower levels are refined using the corresponding vectors at one level above as the bias with a scaling factor of two. The following procedure MRMC illustrates the hierarchical implementation of this scheme with a full search where each block in W_j^i is examined:

```
PROCEDURE MRMC
    estimate $\mathbf{V}_8(x, y)$ in $S_8$;
    for ($i = 1$; $i \leq 3$; $i$++) $\mathbf{V}_8^i(x, y) = \mathbf{V}_8(x, y)$;
    for ($l = 3$; $l > 1$; $l - -$)
        $j = 2^l$; $j' = 2^{l-1}$;
        for ($i = 1$; $i \leq 3$; $i$++)
            for all $(x, y)$
                if ($x$ mod $2^{4-l}) = 0$ and ($y$ mod $2^{4-l}) = 0$
                    $\mathbf{V}_{j'}^i(x', y') = 2\mathbf{V}_j^i(x, y) +$
                        $\mathbf{\Delta}_{j'}^i(x', y')$,
                            where $x' = 2x$ and $y' = 2y$
END MRMC
```

The refinement $\mathbf{\Delta}_{j'}^i(x', y')$ is defined [1] as

$$\mathbf{\Delta}_{j'}^i(x', y') = \Bigg\{ (\delta x, \delta y) \Bigg| \text{Min} \Bigg[\frac{1}{XY} \sum_{p=-(X/2)}^{X/2} \sum_{q=-(Y/2)}^{Y/2} \Big| C_{j'}^i(x'+p, y'+q) - R_{j'}^i(x'+p+d_x+\delta x, y'+q+d_y+\delta y) \Big| \Bigg] \Bigg\}$$

where δx and δy are the refinements, $X \times Y$ is the size of the search window Ω, $(d_x, d_y) = 2\mathbf{V}_j^i(x, y)$ is the bias inherited from one level above, and (x', y') is the current position in $R_{j'}^i$ (the subimage of the reference frame) and $C_{j'}^i$ (the current frame), $i = 1, 2, 3$, $j' = 2, 4$.

Manuscript received September 23, 1996; revised January 21, 1997. This paper was recommended by Associate Editor Y.-Q. Zhang.

The authors are with School of Computing Science, Simon Fraser University, Burnaby, BC, V5A 1S6, Canada.

Publisher Item Identifier S 1051-8215(97)04250-X.

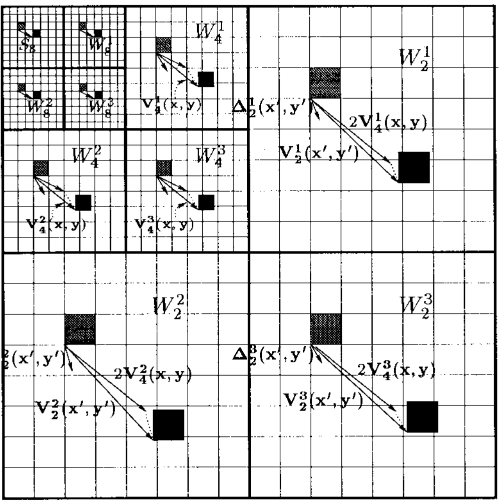


Fig. 1. Multiresolution motion compensation (MRMC) using wavelet subimages.

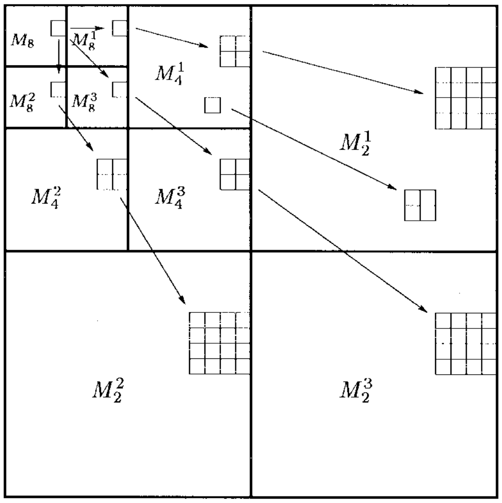


Fig. 2. Mask propagation for the enhanced MRMC.

B. Enhanced MRMC

In order to reduce the high cost of motion estimation, we propose to identify the *potential motion areas* (PMA's) in the wavelet subimages and conduct the MRMC only in the PMA's. Given the fact that the two consecutive video frames are often very similar, i.e., large portions of the two frames are identical, the PMA's are often sparse and small in size. Because of the spatial orientation selectivity of the wavelet decomposition, a motion along a particular direction will be mainly reflected in one (or two) of the three subimages, hence the PMA's in the other subimages are even smaller. Moreover, motion is confined to a small area because of the small temporal interval, therefore the search for motion vectors can be fulfilled locally and efficiently.

Binary images M_j^i $(i = 1, 2, 3, j' = 2, 4, 8)$ are introduced for marking the PMA's. Each M_j^i is of the same size as the respective W_j^i. If $(x, y) \in$ PMA then $M_j^i(x, y) = 1$; otherwise 0. An additional binary image M_8 is also introduced for S_8.

Every element of M_j^i is initialized to zero. Marking starts at Level 3 by comparing S_{C8} and S_{R8}, i.e., the S_8 images for the current frame and the reference frame. If the absolute difference $|S_{C8}(x, y) - S_{R8}(x, y)| > \tau_0$, where τ_0 is a predefined threshold, then $M_8(x, y) = 1$. If some position in M_8 is marked as one, then corresponding positions in M_8^i $(i = 1, 2, 3)$ will automatically be marked as one. The PMA's in M_j^i at two lower levels of the pyramid will be marked through a top-down process called *propagation*.

```
PROCEDURE **propagation** (start_level, end_level)
    for ($l = start\_level$; $l > end\_level$; $l - -$)
        $j = 2^l$; $j' = 2^{l-1}$;
        for ($i = 1$; $i \leq 3$; $i$++)
            for all $(x, y)$ at level $l$
                $M_{j'}^i(x', y') = M_j^i(x, y)$,
                        where $2x \leq x' \leq 2x + 1$
                          and $2y \leq y' \leq 2y + 1$
END **propagation**
```

The result of propagation is shown in Fig. 2.

Although the top-down propagation is efficient, it relies heavily on the adequacy of the marking at M_8. It is quite possible that some PMA's are not marked in M_8 because one of the following is too small: a) the magnitude of the motion, or b) the size of the moving object. In either case, since the resolution of S_8 is low, the process will fail in marking the PMA's at Level 3. Therefore, for Level 2 and Level 1, after the propagation process, a *post-marking* process is applied. For (x, y) where $M_j^i(x, y) = 0$, if the absolute difference of the corresponding $W_j^i(x, y)$ in the two consecutive video frames exceeds a threshold τ_1 (generally it can be chosen such that $\tau_1 \geq \tau_0$), then $M_j^i(x, y) = 1$. If (x, y) is not at the lowest level, propagate to the lower level following the procedure **propagation**.

Now, the binary images M_j^i which reflect the potential motion areas for the two consequent frames are obtained. Based on M_j^i, we will conduct the MRMC. The main difference between the method proposed here and that of [1] is that the motion vectors are now only estimated in the PMA's which are often a small portion of the entire image.

The following is the algorithm for the enhanced MRMC:

```
**Algorithm**
    { initialization }
    $\forall(i, j, x, y)$
        $\mathbf{V}_j^i(x, y) = 0$;   $M_j^i(x, y) = 0$;

    { marking the PMA's in wavelet subimages }
    if   $|S_{C8}(x, y) - S_{R8}(x, y)| > \tau_0$
        then $M_8(x, y) = 1$; else $M_8(x, y) = 0$;
    for ($i = 1$; $i \leq 3$; $i$++) $M_8^i(x, y) = M_8(x, y)$;
    **propagation** (3, 1);
    At Level 2 and Level 1, do post-marking;

    { enhanced MRMC }
    call MRMC only for those blocks marked as PMA
END;
```

Some possible improvements of the above algorithm already implemented are as below.

- In order to make the motion estimation less sensitive to noise, in each subimage after wavelet transform, if the absolute values of the coefficients are less than a predefined small number (e.g., five), they are set to zero.
- "Isolated" ones in M_j^i are deleted and not propagated to the lower levels. "Isolation" is defined as fewer than n_l ones in an eight-connected neighborhood, where n_l varies depending on the level l, e.g., $n_3 = 1$, and $n_1 = 9$.

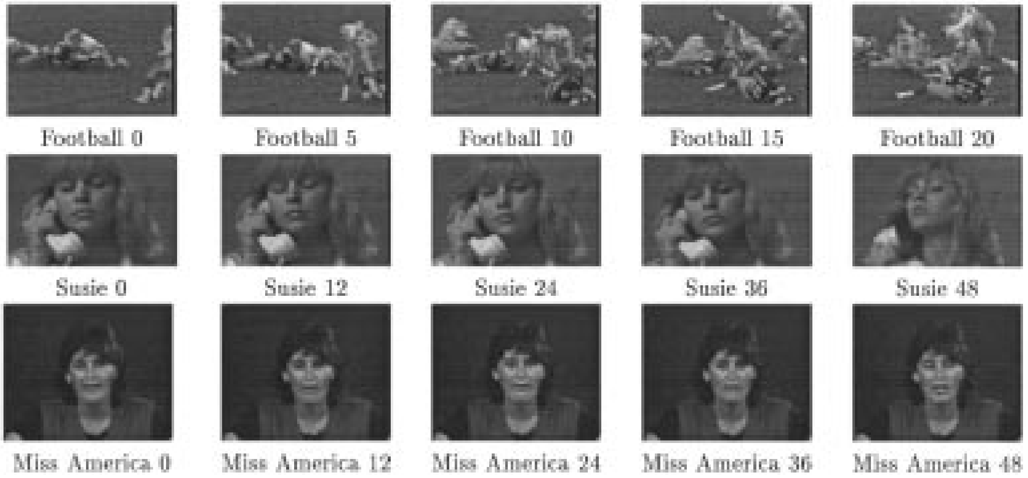


Fig. 3. The image sequences to which the proposed method is applied.

- If the number of ones in all M_j^i's is less than a predefined small number, which means the absolute difference between the consecutive video frames is extremely small, then the time-consuming motion estimation is skipped.
- If the number of ones in all M_j^i's is greater than a predefined large number, i.e., almost every pixel moves (for example due to a quick camera pan) then the binary masks are useless. A boolean flag $mask\_skip$ is set in order to reduce the unnecessary overhead of the subsequent checking about whether or not the block contains points in PMA.

III. Implementation and Test Results

The enhanced MRMC scheme with the masking module is tested using several test video clips. The motion estimation uses only the Y-frames (the luminance images) since they contain almost all necessary information for motion estimation in a color video. Wavelet decomposition is first performed on the Y-frames. Because this letter only compares the results of motion compensation using the original and the enhanced MRMC schemes, the issue of multiscale quantization or entropy encoder is not addressed here. The comparison is on the displaced residual subimages (DRS's) between two consequent video frames, since the DRS's will show the difference (and the "correctness") of the two motion estimation schemes and they are further encoded for video compression. The test result explained below in detail is from the "Miss America" video which was used very often in video compression literatures. The enhanced MRMC performed successfully on various pairs of frames in the video. The following is a test result from Frames 130 and 131.

If one uses the full-search as specified in the procedure MRMC in estimating motion vectors, and the window for the motion vector search is 5 × 5, the running result reads:

Motion estimation done in 11.85 seconds[1]

This was evidently a very slow process.[2]

While using the proposed method with the masking module, if the same search window size of 5 × 5 is used, and $\tau_0 = \tau_1 = 5$, the running script is as follows:

TABLE I
Timing Results and PSNR's of the Enhanced MRMC Method

Threshold ($\tau_0 = \tau_1$)	t_1 (sec.)	t_2 (sec.)	PSNR (dB)
2	0.64	2.80	54.90
3	0.64	1.75	53.78
5	0.64	0.95	52.64
8	0.65	0.44	51.60
10	0.65	0.19	51.35

Binary motion mask images done in 0.64 seconds.

Motion estimation done in 0.95 seconds.

Obviously, as far as processing time is concerned, the enhanced method is quite efficient. Specifically, it consumes approximately 13% of the time compared to the original MRMC method, when in both cases the step for refining the motion vectors by $(\delta x, \delta y)$ is not bypassed. Subjectively and visually, hardly any difference can be perceived from the resulting DRS's of the two different methods. To measure the quality of the result of the proposed method quantitatively, the parameter peak signal-to-noise ratio (PSNR) is used, which is defined as

$$\text{PSNR} = 10 \log_{10}\left(\frac{255^2}{\text{MSE}}\right)$$

where MSE is the mean square error between two signals/images.

The PSNR of the two DRS's from the enhanced MRMC and the original MRMC is 52.64, which is quite satisfactory.

Table I depicts the results when different thresholds are used. For simplicity, τ_0 is set to be equal to τ_1. t_1 is the processing time for obtaining the binary mask images, and t_2 is the time for motion estimation in the enhanced MRMC.

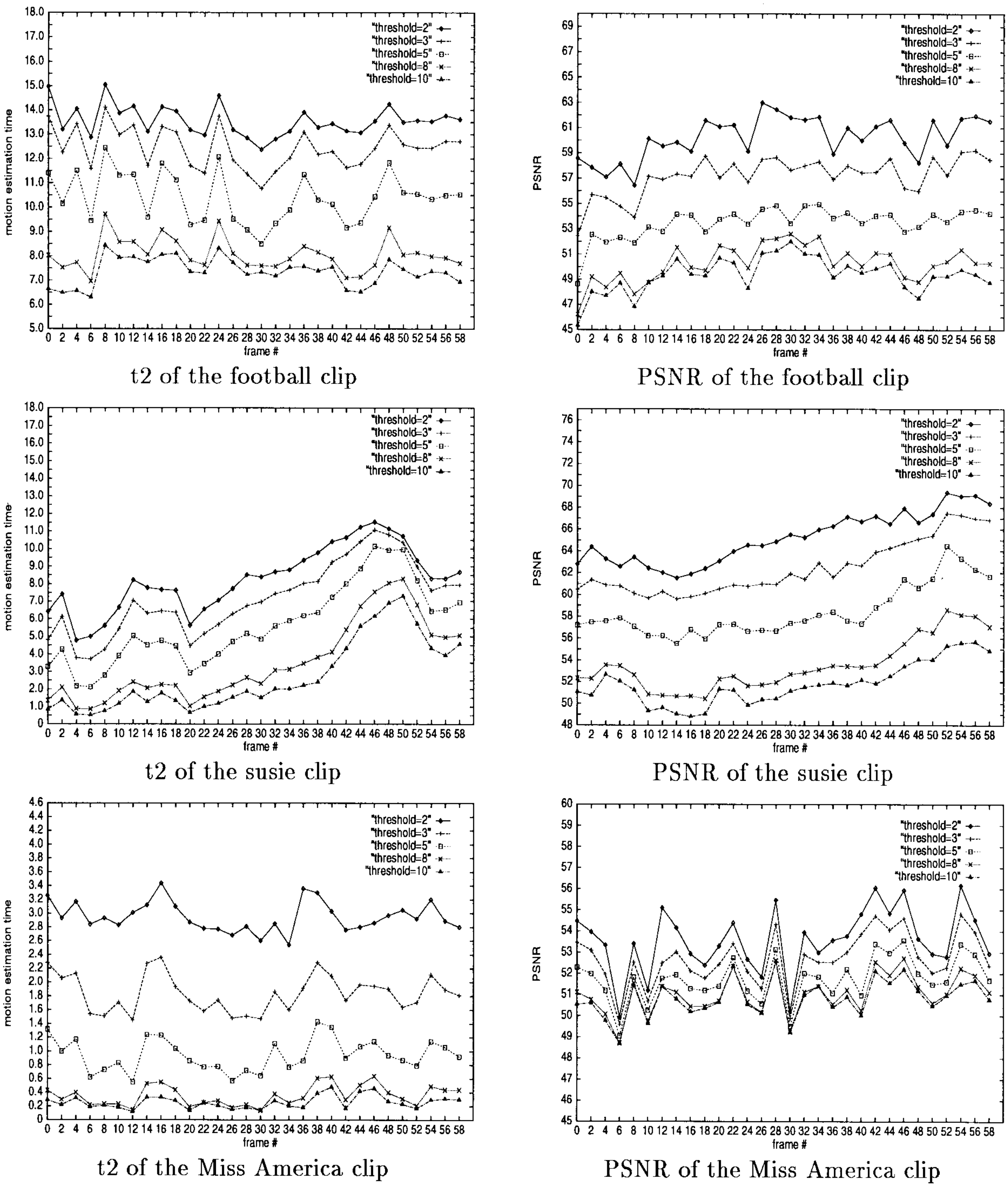
In the same manner, extensive experiments have been conducted on different clips. The running results of three of them, namely, the football clip, the Susie clip, and the Miss America clip, are shown in Fig. 3. In each of those sequences, the first 60 (30 pairs of) images are employed as input to our program. The average values of t_1, t_2, and PSNR are listed in Table II, while the detailed results for each clip are illustrated in Fig. 4. The size we opted here for the search window is 5 × 5. By trial and test, it is observed that slight variations around our choice will produce similar results.

[1] All timing results are generated from a SPARC 4/SS4 work station.

[2] As a matter of fact, considering the high computational cost of the full search method, one of the remedies Zhang and Zafar [1] recommended was the bypass of the refinement step in the MRMC by always setting $\delta x = \delta y = 0$. The search time could thus be reduced. However, the quality of the motion estimation and consequently the compression ratio would be affected.

TABLE II
TIMING RESULTS AND PSNR'S OF THE THREE CLIPS

	Football			Susie			Miss America		
$\tau_0 = \tau_1$	t_1 (sec.)	t_2 (sec.)	PSNR	t_1 (sec.)	t_2 (sec.)	PSNR	t_1 (sec.)	t_2 (sec.)	PSNR
2	2.29	13.57	60.25	2.30	8.27	65.20	0.64	2.95	53.59
3	2.33	12.46	57.19	2.32	7.17	62.42	0.65	1.84	52.69
5	2.33	10.41	53.49	2.35	5.62	58.41	0.65	0.95	51.74
8	2.35	8.06	50.28	2.35	3.48	53.51	0.65	0.36	51.03
10	2.35	7.35	49.36	2.34	2.66	51.80	0.65	0.27	50.83
original MRMC		16.00			16.67			11.89	



t2 of the football clip

PSNR of the football clip

t2 of the susie clip

PSNR of the susie clip

t2 of the Miss America clip

PSNR of the Miss America clip

Fig. 4. The statistics of the results of the first 30 pairs of frames (0–1, 2–3, 3–4, · · ·, 58–59) for the three clips.

From Table II and Fig. 4, some apparent data-dependent phenomena can be observed.

- In the football clip, which contains scenes of busy motions, when τ_0 and τ_1 are small (e.g., two or three), a relatively large portion of the image will be marked as PMA. This has two consequences: first, the time saving is very limited and second, because the PMA is large, the difference between the DRS of the original method and that proposed here is very small, thereby the PSNR's in these two cases are fairly high.
- The images in the Susie clip are of high qualities. Generally, the moving area is much smaller than in the football clip and a bit larger than in the Miss America clip. That is the reason why the corresponding PSNR's are fairly high and its process time falls between the other two clips. From Fig. 4 one can observe that, from frame 4 to 13 and from frame 20 to 47, the time consumed in motion estimation and the values of PSNR both increase monotonically, which stems from the fact that the PMA's in the two cases increased monotonically. The PMA's of frames in the vicinity of frame 46 is so large that they are almost as "bad" as in the case of the football clip, very limited time can be saved. When we take a look at the original image sequence, it can be found that the contents are in the process of a dramatic change—she is putting down the phone—which aligns with our results quite well.
- In the Miss America clip, since the motion is restricted in some small areas, the PMA is only a very small portion. Hence, even with smaller thresholds, as far as processing time is concerned, high efficiency is still accomplished. On the other hand, because the images in this clip are considerably noisy, the PSNR is not so high.

As can be seen from the tables, higher thresholds can be used to save time at the expense of lowered PSNR. It is because when the threshold is too high, very few PMA's would be marked and many motion vectors would be left out.

From our experiments, it is observed that the percentage of savings in total processing time is data-dependent. If most areas of the frame are moving at very high speed and magnitude, as is the case of the football clip, then a very large portion of the image would be marked as PMA, little savings can be realized.

Through our experiments, when setting τ_0 and τ_1 to a value ranging from five to eight, in most cases a good compromise between efficiency and image quality can be accomplished. Nevertheless, the choice of them is also data-dependent.

IV. Conclusion and Discussions

In this letter, an enhancement to the MRMC based on the wavelet decomposition is proposed. The proposed method takes advantage of the temporal redundancies between consequent video frames. A masking module is created prior to the MRMC module to obtain multiresolution binary images reflecting the PMA's in the wavelet subimages. In the subsequent MRMC module the binary image is used as a mask to determine whether or not the motion vector search will proceed. Since the threshold for generating the PMA's can be adjusted with ease, a tradeoff can be made between the quality of the MRMC and the computational cost. Our experiments have shown good results from several sequences of test video clips. When relatively small motion was encountered, the processing time was reduced significantly from what was needed for the original full search MRMC method. Our experiments show that more often than not this method produces good results when the thresholds are appropriately set up. Due to its efficiency, the enhancement has the promise of rendering the MRMC scheme more applicable in daily video compression.

The masking method can be further exploited in other ways. For example, if all ones are marked in the lowest level and almost in the same position at consecutive frames, that would indicate that motion is minor at this moment. Frame skipping as suggested in [3] can then be employed, which will further reduce the computation as well as the storage/transmission cost. Moreover, because of the spatial localization and orientation selectivity in wavelet images, the binary images also convey such useful information such as the location, direction and velocity of motion, and the estimated contour of the motion area. They could be further explored to aid important tasks of motion analysis in video such as object tracking [9].

References

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

[2] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. 11, no. 7, pp. 674–693, 1989.

[3] J. Jain and A. K. Jain, "Displacement measurement and its applications in interframe image coding," *IEEE Trans. Commun.,* vol. 29, no. 12, pp. 1799–1808, 1981.

[4] T. Koga, K. Iinuma, A. Hirano, Y. Iijima, and T. Ishiguro, "Motion compensated interframe coding for video conferencing," in *Proc. NTC'81,* New Orleans, LA, 1981, pp. G5.3.1–G5.3.5.

[5] R. Srinivasan and K. R. Rao, "Predictive coding based on efficient motion estimation," *IEEE Trans. Commun.,* vol. 33, no. 8, pp. 888–896, 1985.

[6] K. M. Mutch and W. B. Thompson, "Hierarchical estimation of spatial properties from motion," in *Multiresolution Image Processing Analysis,* A. Rosenfeld, Ed. Berlin: Springer-Verlag, 1984.

[7] T. Naveen and J. W. Woods, "Motion compensated multiresolution transmission of high definition video," *IEEE Trans. Circuits Syst. Video Technol.,* vol. 4, no. 1, pp. 29–41, 1994.

[8] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Commun. Pure Appl. Math.,* vol. XLI, pp. 909–996, 1988.

[9] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision.* Reading, MA: Addison-Wesley, 1993, vol. II.