A Low-complexity Wavelet Based Algorithm for Inter-frame Image Prediction

S. Usama, B. Šimák

In this paper, a novel multi-resolution variable block size algorithm (MRVBS) is introduced. It is based on: (1) Using the wavelet components of the seven sub-bands from two layers of wavelet pyramid in the lowest resolution; (2) Performing a block matching estimation within a nine-block only in each sub-band of the lower layer; (3) Scaling the estimated motion vectors and using them as a new search center for the finest resolution. The motivation for using the multi-resolution approach is the inherent structure of the wavelet representation. A multi-resolution scheme significantly reduces the searching time, and provides a smooth motion vector field. The approach presented in this paper providing an accurate motion estimate even in the presence of single and mixed noise. As a part of this framework, a comparison of the Full search (FS) algorithm, the three-step search (TSS) algorithm and the new algorithm (MRVBS) is presented. For a small addition in computational complexity over a simple TSS algorithm, the new algorithm achieves good results in the presence of noise.

Keywords: video compression, motion estimation, wavelet transform, multi-resolution.

1 Introduction

Video image compression plays an important role in transmission and storage of digital video data. The applications include multimedia transmission, teleconferencing, videophones, high-definition television (HDTV), CD-ROM storage, etc. A large body of work in image/video processing has involved motion estimation [1] and [6]. Applications of motion estimation exist in image sequence filtering and restorations, video coding, target tracking, robot navigation, monitoring and surveillance, biomedical problems, and the human-computer interface.

The most effective technique for motion estimation makes use of block matching algorithms (BMA). The full search algorithm (FS) is the most obvious candidate for a search technique for finding the best possible weight in the search area. Kago et al [7] use a three-step motion vector search (TSS) to compute displacements up to 6 pel/frame. This method, for $W = 6$ pel/frame, searches 25 positions to locate the best match. The three-step search (TSS) algorithm is one of the best fast search algorithms, and provides a good estimation.

To reduce the computational complexity, hierarchical and multi-resolution fast block matching is used. One family of fast block motion estimation algorithms relies on the idea of predicting the approximate large-scale motion vectors in the coarse-resolution video and refining the prediction motion vectors to find the final values. These are called hierarchical [5], [2] or multi-resolution methods [3] and [4]. Hierarchical methods use the same image size but different block sizes at each level. Multi-resolution methods use different image resolutions with a smaller image size at a coarser level. The wavelet transform has recently emerged as a promising technique for image processing applications, due to its flexibility in representing non-stationary image signals, and its ability in adapting to human visual characteristics. Zhang and Zafar [8] applied wavelet theory to real-time video compression, and proposed multi-resolution motion estimation (MRME). This scheme exploits the cross correlation among all layers of the wavelet pyramid structure in order to reduce the computational complexity of the motion estimation process.

We present a novel multi-resolution variable block size algorithm (MRVBS) based on wavelet decomposition. The approach presented in this paper provides an accurate motion estimate even in the presence of noise. We utilize a wavelet component of the seven sub-bands from two layers of a wavelet pyramid in the lowest resolution. In each sub-band we perform the block matching estimation within a nine-block only. The simulation results are analyzed to assess the proposed algorithm with and without influence of noise. Noise in a sequence not only degrades the visual quality, but also hinders the subsequent analysis and processing (e.g., compression, estimation and coding). The problem of removing noise from image sequences has attracted a number of researchers. However, the noise cannot be completely removed from the image sequences.

This paper is organized as follows. In section 2, the proposed algorithm based on wavelet decomposition is briefly described. Section 3 presents simulation results of the new algorithm without influence of noise. Section 4 presents simulation results under the influence of a single noise and mixed noise. A few concluding remarks are given in section 5.

2 The proposed algorithm

Fig. 1 sets out the structure of the algorithm. The MRVBS algorithm is summarized as follows:

Step 1: To begin the motion vectors estimation process, the original image frame is decomposed into two layers using the two-dimensional discrete wavelet transform (DWT2). The motion vectors for the lowest low-pass band are estimated by central search. The search is performed at the center and its eight neighboring blocks with a block size of 4×4.

Step 2: These motion vectors are then used as a new center for other three-bands in the same layer (layer number 2). For these three-bands, the search is

http://ctn.cvut.cz/ap/

also performed by the same method (central search) using the same block size. We used a block size of $p \cdot 2^{-j} \times q \cdot 2^{-j}$ for the *jth* layer in the wavelet pyramid, where p and q are the sizes of the block required at the highest resolution ($p = 16$ and $q = 16$).

Step 3: The current motion vectors, estimated from layer 2, are scaled and used as a new center for the three highest frequency bands in layer 1. In layer 1, the search is performed using a block size of 8×8.

Step 4: The estimated motion vectors are then scaled and used as a center for the final central search process. Also, the final central search is performed at the center and its eight neighboring blocks. This process is performed for the original frame using block size of 16×16.

The computational cost for MRVBS, without the wavelet complexity, is $(36*p2+27*p1+9*p0)$, where $p2$, $p1$, and $p0$ are the block sizes used in layer 2, layer 1, and layer 0, respectively.

3 Simulation results without influence of noise

Experimental results using the proposed algorithm are reported in this section. The algorithm is applied to three famous video sequences in the QCIF format: Carphone, Foreman, and Miss America. These video sequences have a three-kinds of motion. The experimental results are evaluated using the luminance component of each sequence. The results are based on the peak-signal-to-noise ratio (PSNR) function, and use the mean absolute difference (MAD) in performing block matching. The error terms are not used in the frame reconstruction. Only forward prediction is implemented in the experiments. No threshold value is used in the search process.

A performance comparison of MRVBS, TSS, and FS in terms of PSNR between the estimated frames and the original frames is carried out for these video sequences. The comparison is made among the first 30-frame of each sequence.

The PSNR comparisons show that the MRVBS usually provides a performance similar to the TSS and FS algorithms, especially in the case of slow motion with a stationary background.

As an example, Fig. 2 shows the performance comparison for the Carphone sequence (this was the worst result).
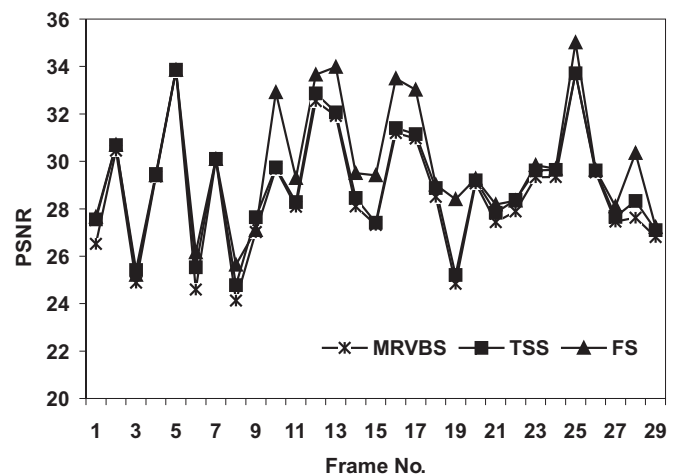

Fig. 2: PSNR comparisons of MRVBS, TSS, and FS algorithms for the "Carphone" sequence without influence of noise
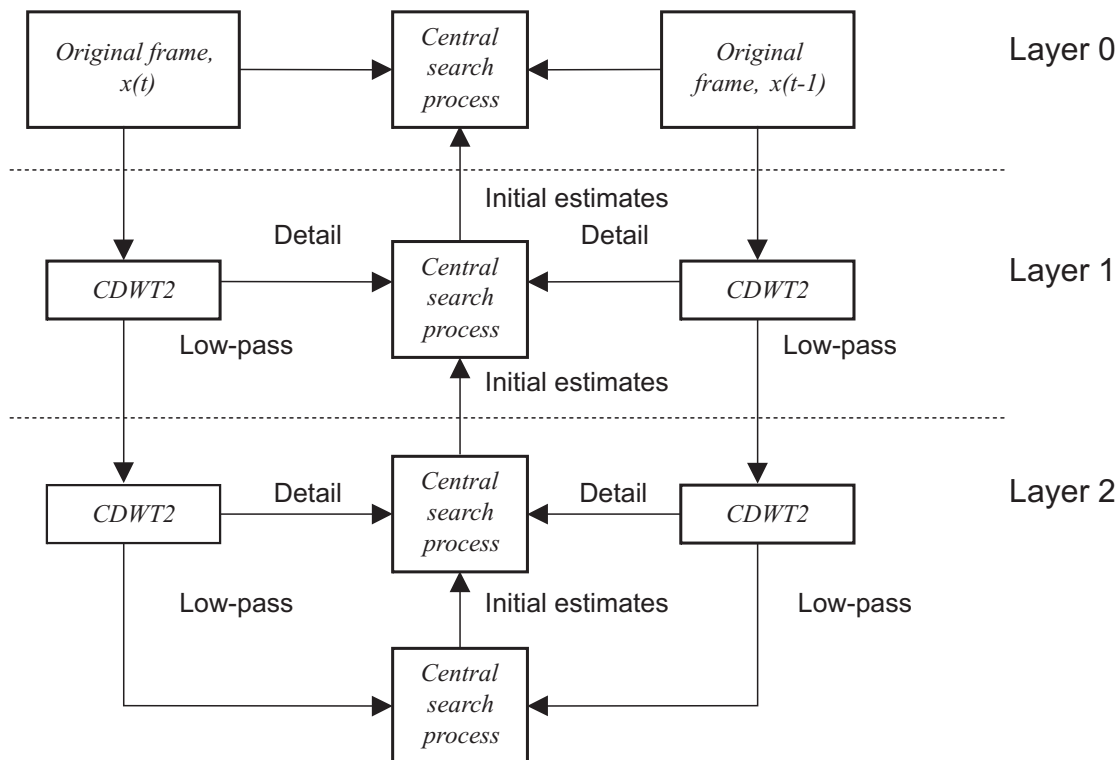

Fig. 1: Motion Estimation Using Wavelet Decomposition

4 Simulation results under influence of noise

To demonstrate the performance of our algorithm, the average PSNR (across all input frames) is plotted against input noise density and signal-to-noise ratio (SNR). The average PSNR, PSNRavg, is given as

$$PSNR_{avg} = \frac{1}{F} \sum_{i=1}^{F} PSNRi \qquad (1)$$

where $PSNRi$ is the measured PSNR for frame i, and F is the total number of frames. We shall compare the MRVBS algorithm against the TSS and FS algorithms. In addition, the PSNR comparison among the three algorithms will be introduced.

4.1 Simulation results under the influence of Gaussian noise

An additive Gaussian noise with a different signal-to-noise ratio (SNR) degraded the three video sequences. We applied the new motion estimation algorithm (MRVBS) to these sequences. Fig. 3 shows the performance comparison for the Miss America sequence with a Signal-to-noise ratio (SNR) of 10 dBs. FS and TSS are performed at layer 0 using a block size of 16×16. The PSNR comparison shows that the MRVBS usually performs better than the TSS and FS algorithms. Under normal operating, e.g., input SNR between 30 to 50 dBs, the performance of MRVBS is similar to the performance of the TSS and FS algorithms. For extremely noisy sequences, e.g., for SNR of 10 dBs, the performance of MRVBS is as much as 2 dBs better than the other two algorithms. In Fig. 4, the average PSNR, PSNR_{avg} (across all input frames) is plotted against input noise level for the "Carphone" sequence. These results indicate that the motion estimation techniques used approximately at low level of noise have the same performance. The performance of MRVBS is as much as 2 dBs better than the performance of FS and TSS under a high level of Gaussian noise. In addition, for the "Foreman" sequence the performance comparison is similar to the results of the "Carphone" sequence. For the "Miss America" sequence, the performance of MRVBS is as
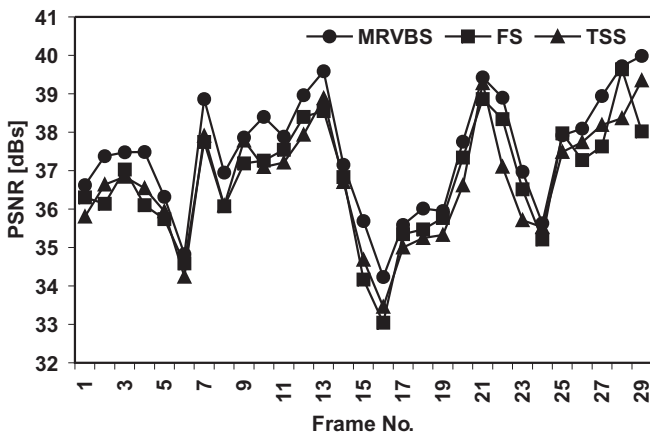

Fig. 3: PSNR comparisons of the MRVBS, TSS, and FS algorithms for the "Miss America" sequence with Gaussian noise (SNR = 10 dBs)
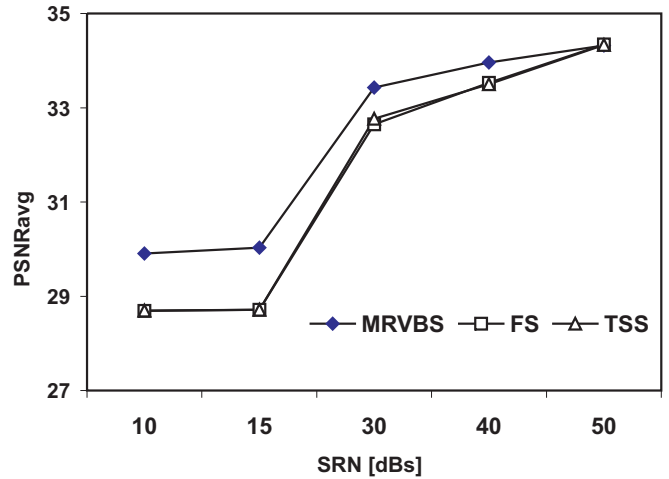

Fig. 4: PSNR (average) vs. source SNR for the "Carphone" sequence

much as 3 dBs better than the other performance of FS and TSS under a high level of Gaussian noise.

4.2 Simulation results under the influence of salt & pepper (impulse) noise

Additive salt & pepper noise with different noise density degraded the three video sequences. We applied the new motion estimation algorithm (MRVBS) to these sequences.

Fig. 5 shows the performance comparison for the "Foreman" sequence with noise density of 40 %. FS and TSS are performed at layer 0 using a block size of 16×16.
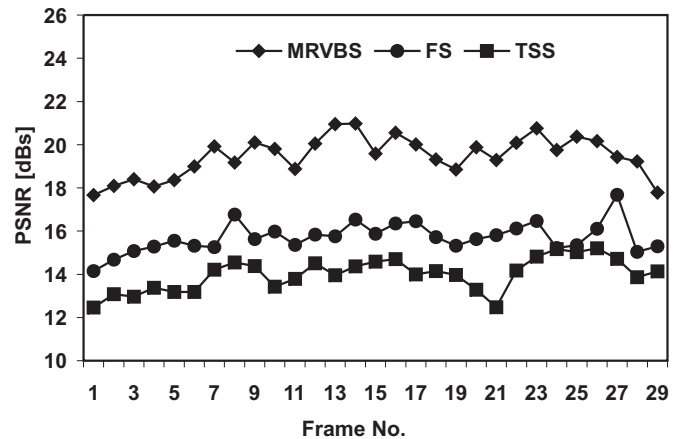

Fig. 5: PSNR comparisons of the MRVBS, TSS, and FS algorithms For the "Foreman" sequence with salt & pepper noise (40%)

As an example, the average PSNR, PSNR_{avg} (across all input frames) is plotted against the input noise level for the "Miss America" sequence in Fig. 6.

The performance of MRVBS is as much as 8 dBs better than the other performance of FS and TSS under a high level of salt & pepper noise. MRVBS performs well in the case of salt & pepper noise, better than the presence of Gaussian noise.
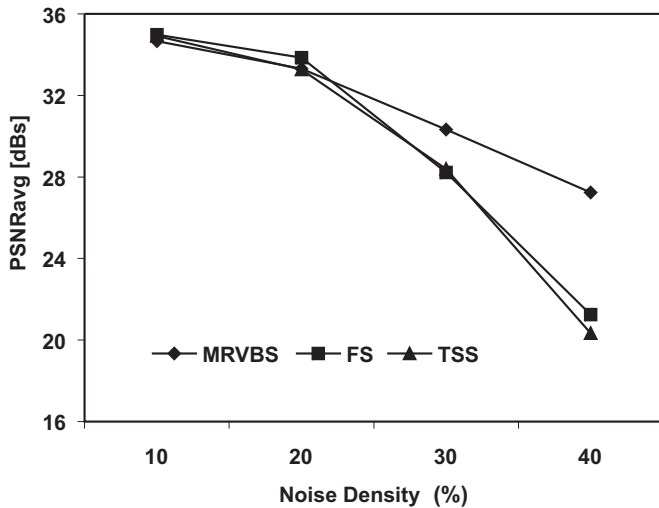
Fig. 6: PSNR (average) vs. source for the "Miss America" sequence

4.3 Simulation results under the influence of mixed noise

We will now assess the performance of the proposed algorithm with respect to mixed Gaussian noise and impulse noise. The restoration result for the "Miss America" videosequence is shown in Fig. 7 for mixed Gaussian (Variance = 200) and impulse noise (20 %). From these tests, we conclude that our algorithm works extremely well for video sequences corrupted with single or mixed noise. In addition, for mixed Gaussian (Variance = 200) and impulse noise (20 %), Fig. 8 shows reconstructed frame number 30 of the "Carphone" sequence from frame number 29 and motion vectors estimated using the current algorithms.
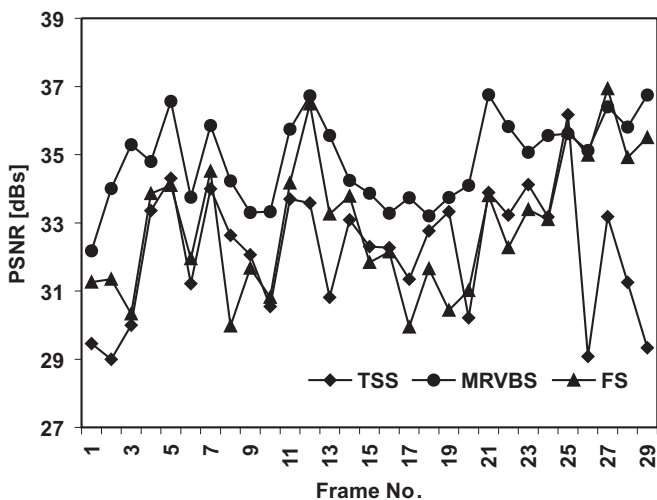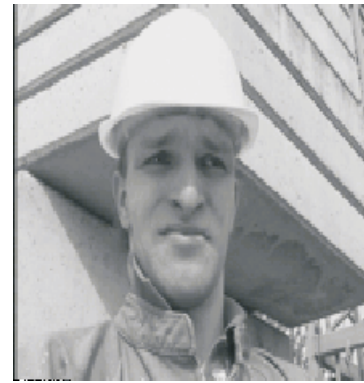

Fig. 7: PSNR comparisons of the MRVBS, TSS, and FS algorithms for the "Miss America" sequence under the influence of mixed noise
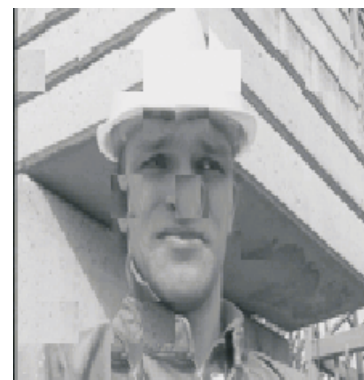
5 Conclusion

We introduced the new multi-resolution variable block size (MRVBS) algorithm. This algorithm is based on a central search in each layer of the wavelet pyramid. Three QCIF format video sequences corrupted by single and mixed

a) original frame no. 30
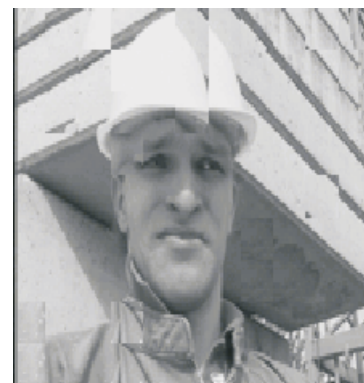
b) reconstructed using MRVBS

c) reconstructed using FS

d) reconstructed using TSS

Fig. 8: "Foreman" sequence: Reconstructed frame no. 30 using the MRVBS, TSS, and FS under the mixed noise

noise used for performance evaluation. The results show that MRVBS is usually better than the FS and TSS algorithms, especially with slow motion video sequences. The PSNR comparisons show that, the best performance is in the case of the Miss America sequence (slow motion with stationary background). Experimentally, the proposed algorithm has been shown to significantly outperform the motion estimation for these three types of video sequences for several distinct noise types, including impulsive, Gaussian, and mixed impulsive Gaussian noise.

From the experimental results, under the influence of mixed noise, the maximum improvement within the first 30-frame was about 6 dBs with the Miss America sequence. We observe that the maximum improvement in case of panning and object translation in the Foreman sequence is 2.5 dBs in comparison with the FS algorithm. This supports our claim that MRVBS can be effectively used with noisy sequences to get a better estimation.

The simulation confirms that the proposed algorithm performs better than the FS and TSS algorithms with the three types of motion used. These gains can be observed in terms both of the perceptual quality and of the PSNR of the restored images. It should also be noted that since the MRVBS algorithm can contain a regular data flow through the entire search procedure, it is suitable for hardware implementation.

References

[1] Sezan, M. I., Langendijk, R. L. (ed.): *Motion Analysis and Image Sequence Processing*. Kluwer Academic Publishers, 1993.

[2] Dufaux, F., Kunt, M.: *Multi-grid Block Matching Motion Estimation with an Adaptive Local Mesh Refinement*. In Proc. SPIE Visual Commun. and Image Processing '92, Vol. 1818, p. 97–109.

[3] Li, J., Lin, X., Wu, Y.: *Multiresolution Tree Architecture with its Application in Video Sequence Coding: A New Result*. In Proc. SPIE Visual Communications and Image Processing '93, Vol. 2094, p. 730–741.

[4] Uz, K. M., Vetterli, M., Legall, D.: *Interpolative Multiresolution Coding of Advanced Television with Compatible Sub-channels*. IEEE Trans. Circuits Syst. Video Technol., March 1991, Vol. 1, p. 86–99.

[5] Bierling, M.: *Displacement Estimation by Hierarchical Block Matching*. In Proc. SPIE Visual Communications and Image Processing '88, Vol. 1001, p. 942–951.

[6] Netravali, A. N., Robbins, J. D.: *Motion Compensated Television Coding: Part I*. The Bell System Technical Journal, March 1979, Vol. 58, p. 631–670.

[7] Kago, T., Iinuma, K., Hirano, A., Iijima, Y., Ishiguro, T.: *Motion compensated interframe coding for video conferencing*. In Proc. Nat.: Telecommun. Conf., New Orleans, LA, Nov. 29–Dec. 3, 1981, p. G5.3.1–5.3.5.

[8] Zhang, Y. Q., Zafar, S.: *Motion-compensated Wavelet Transform Coding for Color Video Compression*. IEEE Trans. Circuits Sys. Video Technol., September 1992, Vol. 2, p. 285–296.

Doc. Ing. Boris Šimák, CSc.
phone: +420 2 2435 2203
e-mail: simak@feld.cvut.cz

Dept. of Telecommunications Engineering
Czech Technical University in Prague
Faculty of Electrical Engineering
Technická 2
166 27 Prague 6, Czech Republic

Ing. Sayed Usama, Ph.D.
phone: +420 2 2435 2088
e-mail: usama@feld.cvut.cz

Deparment of Electrical Engineering
Faculty of Engineering
Assiut University
Assiut, Egypt