MEDIAN FILTERING-BASED PYRAMIDAL MOTION VECTOR ESTIMATION

Jinwen Zan, M. Omair Ahmad and M.N.S. Swamy

Department of Electrical and Computer Engineering
Concordia University
1455 de Maisonneuve Blvd. West
Montréal, Quebéc, Canada H3G 1M8

ABSTRACT

In this paper, a median filtering-based hierarchical motion vector estimation scheme making use of a pyramidal data structure is proposed. Compared to the conventional hierarchical motion vector estimation schemes, the proposed scheme overcomes the problem of propagation of false motion vectors across resolutions. Simulation studies show that the proposed scheme not only improves the prediction accuracy with respect to the prediction mean square error, but also results in a smoother motion field, which can be encoded with less number of coding bits. It is shown that an improvement in the rate distortion performance is achieved with little increase in the computational complexity. It is also shown that Burt and Adelson's pyramidal data structure provides the best performance among a number of the generating kernels considered in our study.

1. INTRODUCTION

Digital video compression is a driving engine for today's multimedia applications. In order to achieve the goal of high compression gain, video compression techniques rely heavily on efficient motion estimation techniques to remove the temporal redundancy in a video signal. One of the motion estimation techniques is the block matching algorithm (BMA). Full search (FS) BMA, which estimates motion vectors by searching all the candidate positions in a search region, has been widely used for commercial applications because of its good performance and the simplicity of its hardware implementation. It is optimal in terms of the performance measure specified. However, the computational complexity of FS is high.

In order to reduce the computational complexity, hierarchical motion estimation techniques have been proposed [1]-[3]. In a hierarchical motion estimation technique, in which the processing takes place starting from a lower resolution to higher resolutions, the computational complexity is reduced because of the smaller image size at lower resolutions and the utilization of the motion vectors from a lower resolution to predict the motion information at higher resolutions. Pyramidal motion vector estimation technique [2] is one of the hierarchical motion vector estimation techniques. In this method, the original video frames are filtered and downsampled repeatedly to produce a successive images with reduced horizontal and vertical sizes, by a factor of two between adjacent resolutions.

However, all of the hierachical motion vector estimation techniques that have already been proposed [1]-[3], including the pyramidal motion vector estimation technique [2], use properly scaled motion vectors from the corresponding positions at a lower resolution as motion vector predictions. Thus, a serious drawback with these conventional hierarchical motion vector estimation techniques is that, whenever a false motion vector is generated at a certain resolution level, it gets propagated and cannot be recovered along the prediction path, resulting in a degraded prediction performance and a noisy motion field.

In this paper, in order to overcome the problem of propagation of false motion vectors, a median filtering-based hierarchical motion vector estimation scheme making use of the pyramidal data structure is proposed. Although the proposed method is general in nature, as it can be applied to any of the conventional multiresolution motion estimation techniques, we show its effectiveness on the pyramidal motion vector estimation scheme in this study. Besides, a great deal of progress has been made in the field of multirate signal processing since the introduction of Burt and Adelson's pyramidal data stucture [4]. Hence, a performance comparison of various kernels for generating the pyramidal data structure, including wavelets, is also included in this study.

2. MEDIAN FILTERING BASED PYRAMIDAL MOTION VECTOR ESTIMATION SCHEME

A pyramidal motion vector estimation scheme used in this study, showing clearly the motion information relationship,

This work was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada, and Fonds pour la Formation de Chercheurs et l'Aide à la Recherche(FCAR) of Québec.

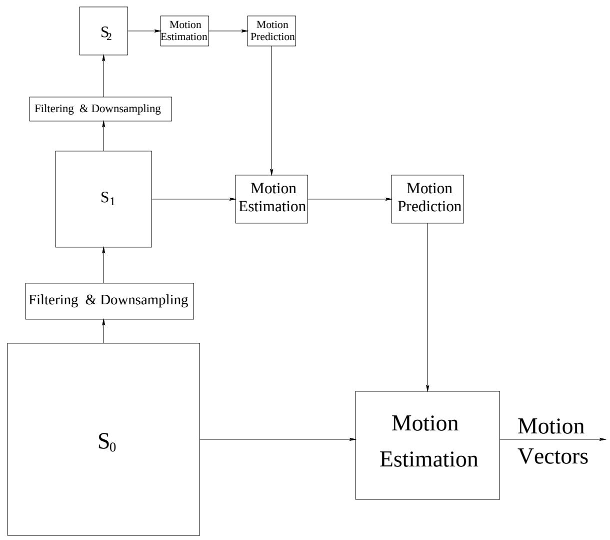

Fig. 1. A 3-level pyramidal motion vector estimation scheme.

is depicted in Fig. 1. The main difference between the proposed median filtering based scheme and the conventional schemes lies in the functional box "motion prediction". In all the conventional schemes [1]-[3], the motion vector prediction across resolution levels can be written as

$$\begin{aligned} MV_L(r,s) &= 2 \times MV_{L+1}(r,s) + \Delta MV_L(r,s) \\ & \quad for \ \ L = 0,1 \end{aligned} \tag{1}$$

where (r,s) denotes the position of the search block and L represents the resolution level. Although this prediction equation is straight forward, it has the problem of propagation of false motion vectors.

Since most motion fields associated with the video signals of natural scenes are smooth and continuous, the false motion vector always presents itself as a discontinuity in the motion field. Thus, a false motion vector can be recovered by comparing to its neighbouring motion vectors that have already been obtained at a certain resolution. The conventional techniques exploiting the prediction equation given by (1) fail to make use of all the motion information that has already been estimated at lower resolutions sufficiently.

In a hierarchical motion vector estimation scheme, we refer to the motion vector which is used to predict another motion vector as the parent motion vector, the motion block corresponding to the parent motion vector as the parent block, and the motion vector to be predicted as the child motion vector. Also, we define the term parent window to be a set of blocks centered around the parent block at a given resolution. This window need not be rectangular. For example, it could be a cross-shaped parent window as shown in Fig. 2.

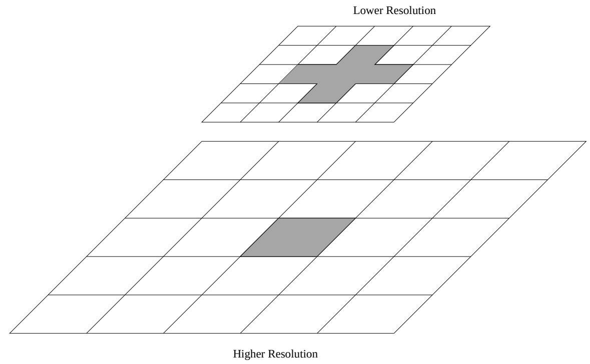

Fig. 2. A cross-shaped parent window.

In the proposed median filtering technique, instead of relying on prediction equation given by (1) to predict motion vectors across resolution levels, the prediction motion vector is generated through median filtering of the candidate motion vector array that consists of motion vectors corresponding to the blocks within the parent window. The x and y components of the motion vectors in the candidate motion vector array are median filtered separately, and the two results combined and appropriately scaled to form the prediction motion vector, which is used as the starting point for further refinement. In this method, the discontinuity in the motion field from a false motion vector, appears either at the beginning or at the end of the sorted x/y array. By choosing the motion vector components at the middle of the sorted x and y arrays, the propagation of a false motion vector is blocked automatically. Note that if there is no false motion vector in the candidate motion vector array, the prediction motion vector generated through the median filtering scheme described above is still a good prediction. In other words, the proposed median filtering method overcomes the problem of the propagation of a false motion vector without sacrificing the motion prediction performance even when no false motion vector is present. Special care must be taken when the candidate motion vector array has an even number of entries. In such a case, the arithmetic average of the two entries in the middle of the sorted array is used as the prediction.

3. CHOICE OF GENERATING KERNELS

There are various methods to generate a pyramidal data structure, the simplest one being the one involving the spatial downsampling of the original video frames. However, it has been pointed out in [2] that downsampling leads to incorrect motion estimation because of the presence of noise. The authors in [2] have suggested the use of the mean pyramid.

In Burt and Adelson's original work of image pyramid [4], a family of generating kernels, which are normalized

Filter or Wavelets	Flower		Football		Tennis		Mobile	
	MSE	Bits	MSE	Bits	MSE	Bits	MSE	Bits
Mean Pyramid	155.80	1559	272.12	2662	84.83	2339	232.72	793
MF-Mean Pyramid	146.20	1294	308.75	2270	87.87	1471	230.58	685
7/9 wavelet	158.88	1576	278.86	2655	82.98	2218	232.80	765
MF-7/9 wavelet	147.10	1412	272.29	2448	79.72	1712	232.13	738
10/6 wavelet	171.16	1681	276.56	2695	84.48	2264	232.97	774
MF-10/6 wavelet	147.36	1429	272.51	2448	80.61	1713	232.06	736
B&A a=0.3	147.67	1412	271.00	2585	76.18	1866	230.25	701
MF-B&A a=0.3	145.05	1349	270.46	2445	77.57	1628	230.23	698
B&A a=0.4	147.16	1410	271.68	2601	76.85	1903	230.70	708
MF-B&A a=0.4	145.09	1352	270.69	2442	78.08	1630	230.28	703
B&A a=0.5	149.47	1446	273.20	2615	79.08	1996	231.16	727
MF-B&A a=0.5	145.54	1367	270.78	2446	78.59	1638	230.72	716
B&A a=0.6	154.71	1532	276.97	2649	82.64	2205	232.31	765
MF-B&A a=0.6	146.32	1392	271.54	2446	79.23	1675	231.64	733

Table 1. Performance comparison of conventional and median filtering based pyramidal motion vector estimation

and symmetric, are used to generate the pyramidal data structure. With a constraint called equal contribution, the one-dimensional generating kernel $\hat{w}$ of length 5, can be written as

$$\begin{array}{rcccl} \hat{w}(0) & = & a & & \\ \hat{w}(-1) & = & \hat{w}(1) & = & 1/4 \\ \hat{w}(-2) & = & \hat{w}(2) & = & 1/4 - a/2 \end{array} \tag{2}$$

Beside the mean pyramid [2] and the Gaussian pyramid [4], wavelet theory has recently attracted a great deal of attention in the image processing literature. A simulation study comparing the motion estimation performance using the mean pyramid, Burt and Adelson's generating kernels and a number of wavelets is carried out in the next section.

4. SIMULATION RESULTS

In this section, simulation studies using both the prediction equation given by (1) as well as median filtering for across resolution motion vector prediction are carried out. Four video sequences of SIF format, namely, 100 frames of the *Football*, *Flower Garden*, and *Mobile & Calendar* sequences, and 60 frames of the *Tennis* sequence are included in the simulation. Mean absolute difference (MAD) is used as the error measure in performing the block matching, while the mean square error (MSE) computed between the original video frame and the motion compensated one is used as the measure of performance evaluation.

The motion vector estimation is performed on a 3-level pyramid, with the block size of 4×4 at the lowest resolution level, and the search range or refinement range of 4 for all the resolutions. A 3×3 parent window is used for the median filtering.

For the encoding of motion vectors, arithmetic coder [5] is used. The motion vectors are DPCM encoded on a row-by-row basis, as required by most of today's video coding standards.

The simulation results using mean pyramid, a couple of well known wavelets [6] pyramids, and the Burt and Adelson's Gaussian pyramids, are shown in Table 1. From these results, we can make the following observations.

- The proposed median filtering technique improves the overall rate distortion performance. Although in a few cases, the median filtering technique results in a higher prediction mean square error compared to the conventional technique. In these cases, however, there is a significant reduction in the number of bits needed in coding of the motion vectors. This is due to the fact that in the conventional technique, emphasis is placed on the matching of the pixel pattern, and thus it is possible to find a best match while introducing a discontinuity in the motion field, which renders more coding bits. In other words, only the prediction performance is emphasized in the conventional technique, instead of the overall rate distortion performance.

- Among the three methods used to generate pyramidal data structures, Burt and Adelson's generating kernels give the best motion estimation performance. It seems that the equal contribution property of these generating kernels, which requires that all the nodes at a given level must contribute the same total weight to the nodes at the next higher level, is a better property than the half-band splitting property of wavelets or the mean pyramid when performing pyramidal motion vector estimation.

Sequence	Complexity	Mean Pyramid	7/9	6/10	Burt(0.3)	Burt(0.4)	Burt(0.5)	Burt(0.6)
Flower	COMPARISON	13	7	7	6	6	6	7
	SWAPPING	10	5	5	5	5	5	5
Football	COMPARISON	19	9	9	10	9	9	9
	SWAPPING	16	7	7	8	7	7	7
Tennis	COMPARISON	16	9	9	8	8	8	9
	SWAPPING	13	7	7	6	6	6	7
Mobile & Calendar	COMPARISON	7	2	2	1	1	1	2
	SWAPPING	5	1	1	1	1	1	2

Table 2. Average computational complexity of median filtering per motion vector component

It should be noted that dozens of wavelets with good reputation in the image processing literature have been investigated in this study, and the two shown in Table 1 are among the few with excellent performance.

5. COMPUTATIONAL COMPLEXITY

It may seem that compared to the conventional pyramidal motion estimation schemes, the proposed median filtering based technique would increase the computational complexity because of the median filtering operation involved. However, we will show in this section that this additional cost is only minimal.

In the case of the 3×3 parent motion vector window that we have used in our simulation studies, there are at most 9 values for either the x or the y component of the motion vector array to be sorted. In view of the small length of the motion vector array, insertion sorting is used for median filtering.

Table 2 shows the computational complexity using various generating kernels. A simple calculation shows that even for the noisy motion field of the *Football* sequence, generated by the mean pyramid, which on the average requires the median filtering process to perform 19 comparisons and 16 swappings for either the x or the y component of a motion vector, the additional computation complexity for one image is less than one-half of one percent of that of computing all the motion vectors at the full resolution level.

6. CONCLUSIONS

In this paper, a median filtering-based pyramidal motion vector estimation scheme has been proposed to overcome the problem of propagation of false motion vectors encountered in the conventional schemes. Based on the assumption that the motion fields associated with the video signals of natural scenes are smooth and continous, the prediction motion vector from a lower resolution is derived by median filtering technique, instead of directly using the corresponding motion vector at the lower resolution.

Simulation studies using the pyramidal motion estimation technique with and without median filtering have shown that the proposed method resolves the problem of propagation of false motion vectors and results in a better rate-distortion performance. It has also been shown that the additional computational complexity introduced by the median filtering is negligible, compared to that required by block matching. Also, the performance of various kernels for generating the pyramidal data structure, including wavelets, is compared. We conclude that the equal contribution property [4] of Burt and Adelson's generating kernels is better than the half-band splitting property of wavelets or mean pyramid in performing pyramidal motion vector estimation.

7. 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, pp. 285–296, September 1992.

[2] K. M. Nam, J.-S. Kim, R.-H. Park, and Y. S. Shim, "A fast hierarchical motion vector estimation algorithm using mean pyramid," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 4, pp. 344–351, August 1995.

[3] S. Kim, S. Rhee, J. G. Jeon, and K. T. Park, "Interframe coding using two-stage variable block-size multiresolution motion estimation and wavelet decomposition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, pp. 399–409, August 1998.

[4] P. J. Burt and E. H. Adelson, "The laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. 31, pp. 532–540, April 1983.

[5] I. H. Witten, R. M. Neal, and J. G. Cleary, "Arithmetic coding for data compression," *Communications of ACM*, vol. 30, pp. 520–540, June 1987.

[6] J. Villasenor, B. Belzer, and J. Liao, "Wavelet filter evaluation for image compression," *IEEE Trans. Image Processing*, vol. 4, pp. 1053–1060, August 1995.