MOTION VECTOR SMOOTHING FOR TRUE MOTION ESTIMATION

Hai Bing Yin, Xiang Zhong Fang, Hua Yang, Song Yu Yu, Xiao Kang Yang

The institute of image communication and information processing, Shanghai Jiao Tong University, China.

ABSTRACT

This paper proposes a new motion vector (MV) smoothing algorithm to track the real motion in image sequences for MPEG video encoders. First, a pre-checking algorithm is employed to eliminate wrong motion vectors and preserve all possible motion vectors. For each block considered, the motion similarity between the neighboring blocks and the number of candidate motion vectors are jointly exploited to adaptively grow the filtering support, which is supposed to have homogeneous motion and sufficient spatial gradient. Then, all candidate motion vectors are checked within the filtering support using a new motion smoothness-constrained matching criteria. The simulation results show that the proposed algorithm can efficiently track the real motion resulting in smooth motion vector field (MVF).

1. INTRODUCTION

Motion information is one of the most important cues for human to perceive video content [1]. Reliable motion vector (MV) information can considerably aid motion segmentation and object tracking, which is important to implement motion-perception optimized video encoders in the human perception sense. Unfortunately, tracking real motion is an ill-posed problem. Due to the simplicity and the coding efficiency of MVs, block-matching algorithm (BMA) is widely used. However, in the scenes with deformable motion, noise, object occlusion, lighting variation, and existence of multiple local minima in the SAD distribution, the motion field estimated by BMA may be heavily corrupted by noise [2][3]. In this paper, we mainly focus on MV smoothing to aid conventional BMA in tracking the true motion as soon as possible.

Rate-distortion theory had been widely employed to optimize motion estimation (ME) algorithm in [4] and the references therein. These algorithms usually concentrated on reducing the coding bits under a distortion constraint instead of approximating the real motion. Vector median filters were quite effective in reducing impulsive noise in dense MVF generated by optical flow ME algorithm [5]. However, the filtering performance may be unsatisfactory in the case of motion field generated by conventional BMA. In addition, overlapped block motion estimation (OBME) algorithm utilized the motion similarities between the neighboring blocks to smoothen the motion field [6]. Indeed, the OBME algorithm can reduce block artifacts efficiently and smoothen the MVF to some extent. However, a troublesome problem unsolved is the motion edges existing within the blocks located on boundaries of moving objects. In fact, in real image sequences the boundaries of the moving objects seldom coincide with block boundaries, and blocks on these boundaries contain motion edges, thus the MVF obtained may contain serious errors. To circumvent this problem, variable block size BMA with implicit motion segmentation had been proposed in [7]. Blocks containing moving edges are segmented into several variable size subblocks. However, variable block size was not compatible with the MPEG syntax. In [8], a feature-tracking algorithm was proposed using multi-candidate pre-screening to prevent the true MVs from being excluded and eliminate all impossible MVs. In addition, a cost function using motion similarities between the neighboring blocks was used to select the true MVs, which is in effect a method similar to OBME.

In this paper, a new MVF smoothing filter is proposed to aid conventional BMA in tracking the real motion with little additional computation. Pre-checking procedure preserves all candidate MVs and eliminates all wrong MVs. The filtering support is determined adaptively according to the motion similarity between neighboring blocks and the number of its candidate MVs. Then, all candidate MVs are applied to the filtering support and checked using a new MV smoothness constrained matching criteria.

This rest of this paper is organized as follows. Section 2 discusses some common problems of motion smoothing. The proposed algorithm is proposed in section 3. Finally, simulation results and conclusion are given in section 4.

2. SOME PROBLEMS OF MOTION SMOOTHING

The sum of absolute difference (SAD) is the most popular matching criterion in BMA for its simplicity. However, the SAD criterion is challenged by some factors such as noise, deformable motion, motion edges existing within the block, object occlusion, etc. As a result, there will exist multiple local minima in the SAD distribution, and the selected block with minimum SAD may not corresponds to the real motion [8]. Thus, the SAD criterion often results in unreliable MVs.

Object occlusion and deformable motion can be solved using multi-reference frames and adopting generalized/deformable BMA respectively [2]. In this paper, we will mainly focus on the motion edges and the noise factors.

There is an implicit assumption in BMA that all pixels within a block undergo uniform motion. Motion edges existing within a block just violate this assumption. Increasing the motion smoothness of a block is desired for MV smoothing filter to avoid motion edges. In addition, noise usually prevents the BMA from tracking the real motion especially in flat regions with insufficient spatial gradient. Sufficient spatial gradient of a block is also desired to improve the algorithm's robustness to noise.

In fact, how to guarantee a block to have uniform motion and sufficient spatial gradient is a crucial problem in MV smoothing. Theoretically speaking, variable size BMA with implicit motion segmentation and spatial gradient analysis is intrinsically desired for the motion smoothness and spatial gradient constraints [7]. However, variable block size is not compliant with most MPEG standards, motion segmentation and gradient analysis also consumes heavy computation.

To improve the MV accuracy under the compatibility and computation constraints, a practical approach is to segment each conventional block into several subblocks with small size, and block matching is performed on subblock basis. Due to the decreased block size, the chance of moving edges existing in subblocks declines greatly. The subblock-level MVs can be further filtered to track real motion efficiently. With the subblock-level MVs, we can easily estimate the corresponding block's MV, which is used for usual motion compensation and motion vector coding.

In general, the smaller the subblock size is used, the higher motion smoothness the subblocks will have. However, the sufficient spatial gradient constraint may be challenged by the decrease of subblock size. As a result, the BMA's robustness to noise decreases accordingly. We simply segment a macroblock into four subblocks of size 8x8, and the simulation results are basically satisfactory. Adaptive subblock segmentation will be addressed in the future.

The subblock-level MVs estimated by BMA may still be wrong due to existing of multiple local minima caused by insufficient spatial gradient and noise. We have observed in the simulation that the MV with the smallest SAD does not always correspond to the true motion. The object of motion smoothing algorithm is to select the true MV from all possible candidate MVs.

A multi-candidate pre-screening approach was proposed to select several candidate MVs according to the local SAD distribution [8]. In addition, a cost function, which is the sum of the current block's SAD and weighted SADs of the neighboring blocks, is employed to check all candidate MVs to select the desired one by exploiting the motion similarity between neighboring blocks.

There are two major drawbacks in this algorithm. First, the filtering support, i.e. which and how many neighboring blocks are used in the cost function, is fixed. In fact, the chance that moving edges exist between neighboring blocks is very high. This will complicate the determination of the weighting factors. Second, possible motion inhomogeneity between neighboring blocks is not taken into consideration in the cost function. In this paper, we adjust the filtering support adaptively and propose a motion smoothness constrained cost function based on motion similarity to remedy the two drawbacks in [8].

3. THE PROPOSED MV SMOOTHING ALGORITHM

To avoid the motion inhomogeneity between neighboring blocks, we grow the filtering support adaptively at pixel level. The neighboring pixels undergoing the same motion as the pixels in the current subblock are included to grow the filtering support. Correspondingly, a uniform weighting factor is applied to all pixels within the filtering support.

The proposed filtering support structure is illustrated in Fig.1. A foursquare region centering about the current subblock is initialized as the raw filtering support that is marked with white lines. d_{raw} is the distance between the raw filtering support and the current subblock. The final filtering support is enclosed with bold dot line. The final filtering support is composed of the current subblock and four directional overlapped blocks (OB). The left-up (original) pixel and right-down (end) pixel are marked with $\otimes$ and $\otimes$ respectively. In addition, d_l, d_u, d_r, and d_d are the width enlarged at the left, up, right, and down directions with respect to the current subblock.

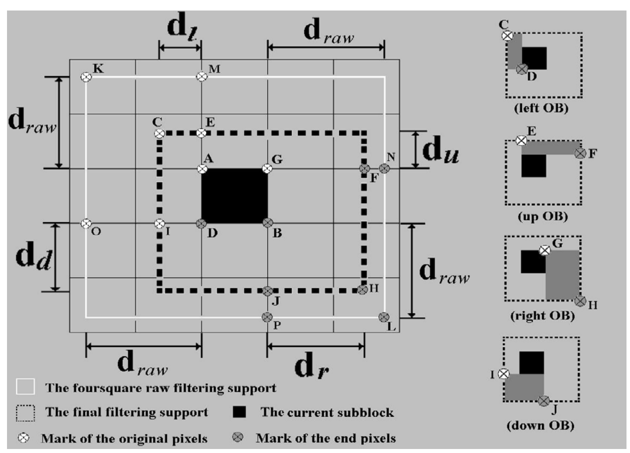

Fig.1. The proposed filtering support structure

An ideal filtering support is supposed to contain sufficient spatial gradient, simultaneously with homogeneous motion. That is, d_l, d_u, d_r, and d_d should be determined adaptively according to the characteristics of the local spatial gradient and motion similarity.

In the following, we will investigate the relationship between the number of the candidate MVs (*N*can) and spatial activity measured with variance. It is crucial to determine *N*can to eliminate wrong MVs and preserve the true MV. In [8], the motion vectors whose corresponding SAD are no more than 1.5 folds of the minimal SAD (SAD_{min}) are admitted into the candidate pool. However, we find that the true MVs are usually eliminated at uniform subblocks adopting the fixed threshold criterion in [8]. This is mainly because that the SAD_{min} are usually very small at uniform subblocks. Thus, we adopt different thresholds at different regions in this paper. 4 folds of SAD_{min} is adopted as the threshold for uniform subblocks, and 1.5 folds of SAD_{min} is adopted the threshold for other subblocks.

The 4^{th} frame of the "Flower Garden" sequence is taken as example frame here. The original example image and the raw MVF estimated by the conventional BMA are sown in Fig.2.

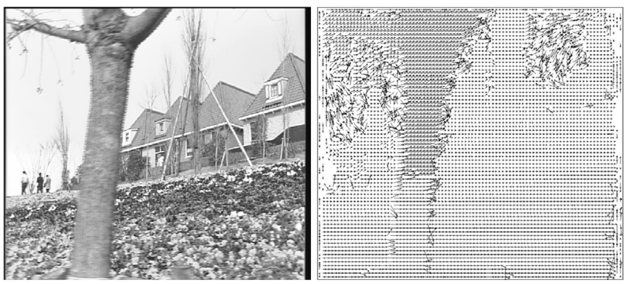

Fig.2. The 4th frame of the "Flower Garden" sequence and the raw MVF.

The variance distribution and the *N*can distribution of all subblocks in the example frame are shown in Fig.3. In this paper, all *N*can are clipped into the range [0, 63]. According to Fig.3, we can find that the subblocks with small variance, i.e. insufficient spatial gradient, generally have large *N*can.

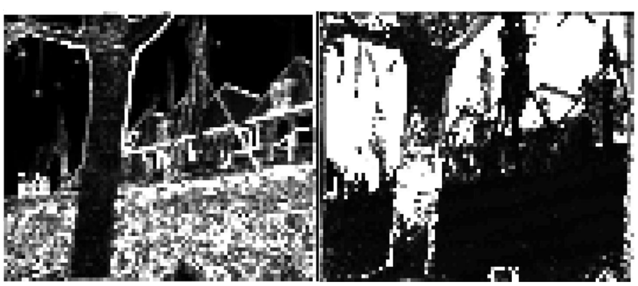

Fig.3. (left): the variance distribution, (right): the *N*can distribution.

Combing the results in Fig.2 and Fig.3, we find that *N*can can directly reflect the confidence of the raw MVs. The smaller *N*can is, the higher the confidence is. Similarly, a subblock with large *N*can generally means that the raw MV is unreliable or wrong. Therefore, *N*can is a very good indicator of the local gradient and spatial activity characteristics to determine the structure of the raw filtering support. As a result, $d_{raw}=N$can is adopted in this paper to determine the raw filtering support. Moreover, motion similarity will be further employed to adjust the raw filtering support to determine the final filtering support, i.e. to determine d_l, d_u, d_r, and d_d adaptively.

The distortion smoothness can efficiently reflect the motion similarity between the current and the reference subblock. However, only using the SAD value can't measure motion smoothness optimally because that SAD is just the total distortion without considering the overall distortion smoothness between the pixels within a subblock. In this paper, we separate each subblock into four granules to facilitate measure of distortion smoothness. Similarly, each directional OB in the filtering support is separated into *L* granules. It is apparent that the granule-level SAD distribution has higher resolution for distortion smoothness differentiation. Foursquare granule of size 4×4 is adopted in both cases. The directional OBs with different shape and size are allowed to have different *L* values. If the height or width of an OB is not an integral multiple of 4, overlapped granules are allowed in the OB. Four granules within a subblock and *L* granules within the directional OBs are shown in Fig.4. The overlapped granules are displayed with dotted lines.

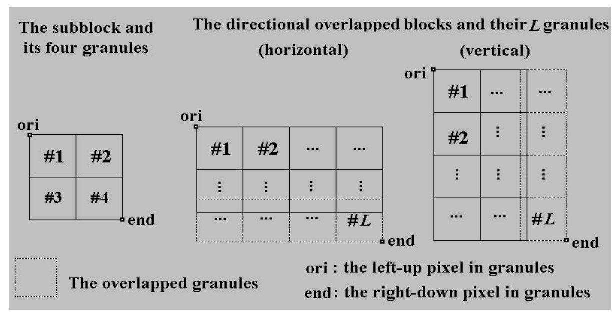

Fig. 4. Left: four granules in a subblock, center: *L* granules in horizontal overlapped blocks, right: *L* granules in verticals overlapped blocks.

The granule-level SAD difference between the *L* granules within a directional OB or a subblock (*L*=4) in the case of k^{th} candidate motion vector ($\boldsymbol{v_k}$) is defined as follows

$$diff_{SAD}(ori,end,\boldsymbol{v_k}) = \sum_{l=1}^{L} \left| L \times SAD_g(l,\boldsymbol{v_k}) - SAD(ori,end,\boldsymbol{v_k}) \right| \quad (1)$$

ori and *end* are the original and end pixels of the directional OB or the subblock respectively. $SAD_g(l,\boldsymbol{v_k})$ is the SAD of the l^{th} granule of $\boldsymbol{v_k}$, and $SAD(ori,end,\boldsymbol{v_k})$ is the SAD of $\boldsymbol{v_k}$. The granule-level SAD distribution and the $diff_{SAD}$ distribution of the subblocks of raw MVs are shown Fig. 5.

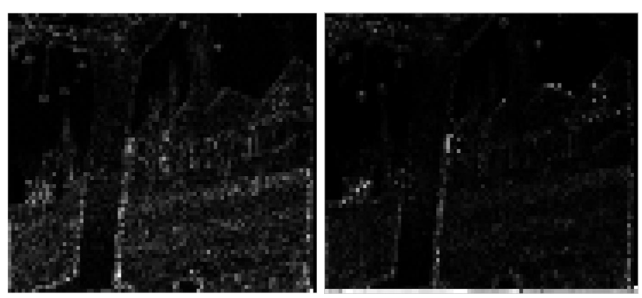

Fig.5. (left): the SAD distribution of all subblocks in the case of raw MVs, (right): the $diff_{SAD}$ distribution in the case of raw MVs.

According to the property of BMA, a true MV is supposed to result in not only small $SAD(ori,end,\mathbf{v_k})$ but also small $diff_{\text{SAD}}(ori,end,\mathbf{v_k})$. Therefore, we define a weighted SAD with motion smoothness constraint as follows

$$WSAD(ori,end,\mathbf{v_k})=\frac{4\times diff_{SAD}(ori,end,\mathbf{v_k})}{L}+SAD(ori,end,\mathbf{v_k}) \quad (2)$$

On the one hand, the weighted SAD (WSAD) will be used as the matching criterion to improve the MV accuracy. On the other hand, WSAD will be incorporated with N_{can} to jointly determine parameters d_l, d_u, d_r, and d_d adaptively.

In the following, we will take d_l as example to explain the proposed method for adaptive filtering support growing. First, d_l is initialized as d_{l1} according to the WSAD of the current subblock and that of the left-up rectangle in the raw filtering support as follows

$$d_{l1}(\mathbf{v_k})=\begin{cases} N_{can}; & if \quad \dfrac{WSAD(A,B,\mathbf{v_k})\times(N_{can}+8)\times N_{can}}{WSAD(K,D,\mathbf{v_k})\times 64}\geq 1 \\ \dfrac{WSAD(A,B,\mathbf{v_k})\times(N_{can}+8)\times N_{can}}{WSAD(K,D,\mathbf{v_k})\times 64}\times N_{can}; & otherwise \end{cases} \quad (3)$$

Then, d_{l2} is determined according to the WSAD of the current subblock and that of the left OB as follows

$$d_{l2}(\mathbf{v_k})=\begin{cases} N_{can}; & if \quad \dfrac{WSAD(A,B,\mathbf{v_k})\times(d_{u1}+8)\times d_{l1}}{WSAD(C,D,\mathbf{v_k})\times 64}\geq 1 \\ \dfrac{WSAD(A,B,\mathbf{v_k})\times(d_{u1}+8)\times d_{l1}}{WSAD(C,D,\mathbf{v_k})\times 64}\times N_{can}; & otherwise \end{cases} \quad (4)$$

Finally, the relatively small value between d_{l1} and d_{l2} is selected for $d_l(\mathbf{v_k})$. Similarly, $d_u(\mathbf{v_k})$, $d_r(\mathbf{v_k})$, and $d_d(\mathbf{v_k})$ can be obtained using the same method.

The method to select the desired MV from N_{can} candidate MVs can be summarized in following procedure form.

1) Set the candidate MV index m=1, record all candidate MVs from $\mathbf{v_1}$ to $\mathbf{v_{Ncan}}$.
2) Calculate $d_{l1}(\mathbf{v_m})$, $d_{u1}(\mathbf{v_m})$, $d_{r1}(\mathbf{v_m})$, $d_{d1}(\mathbf{v_m})$ in the case of $\mathbf{v_m}$ according to (3), and calculate $d_{l2}(\mathbf{v_m})$, $d_{u2}(\mathbf{v_m})$, $d_{r2}(\mathbf{v_m})$, $d_{d2}(\mathbf{v_m})$ according to (4). Then, determine $d_l(\mathbf{v_m})$, $d_u(\mathbf{v_m})$, $d_r(\mathbf{v_m})$, $d_d(\mathbf{v_m})$ to grow the final filtering support.
3) Caculate the WSAD within the final filtering support, and store it to a one-dimension array *wsad* as follows: $wsad(m)$= WSAD(C, H, $\mathbf{v_m}$).
4) Let m=m+1, if m<= N_{can}, go to step 2), otherwise, go to step 5)
5) The candidate MV with the smallest $wsad(m)$ is selected as the final MV.

4. SIMULATION RESULTS

The algorithm in [8] is used as the reference algorithm for performance comparison. The subblock of size 8x8 is also used in the reference algorithm for fair comparison.

The filtered MVF of the reference algorithm and that of the proposed algorithm are respectively shown in the left sub-picture and the right sub-picture in Fig.6. To evaluate the accuracy of MVF in real motion sense, we will segment the image into objects according to the MVF of the reference algorithm and that of the proposed algorithm. The accuracy of motion segmentation can reflect the accuracy of MVF in the real motion sense. Motion segmentation algorithm in [8] is used in our simulation. The motion segmentation results of two algorithms are shown in Fig.7.

Both the MVF results and the motion segmentation results all prove that the proposed algorithm obtains more homogeneous MVF than the reference algorithm do.

Fig.6. The resulting MVF of the reference and the proposed algorithms.

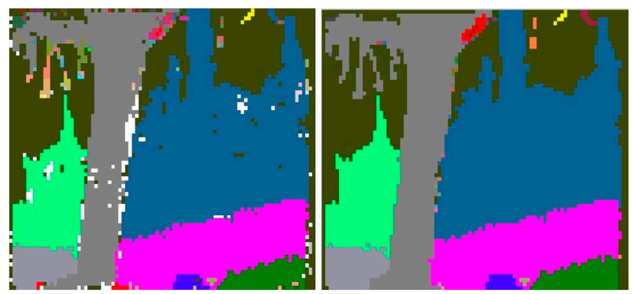

Fig.7. (left): the motion segmentation result of the reference algorithm, (right): the motion segmentation result of the proposed algorithm.

5. REFERENCES

[1] Wilfried Osberger *et al*, "An automatic image quality assessment technique incorporating higher level perceptual factors," *Proceedings of ICIP*, v.3, pp. 414-418, 1998.

[2] F. Dufaux, F. Moscheni, " Motion Estimation Techniques for Digital TV: A Review and a New Contribution," *Proceeding of IEEE*, vol.83, no.6, 1995.

[3] G. de Haan, P. W. A. C. Biezen, H. Huijgen, and O. A. Ojo, "True-Motion Estimation with 3-D Recursive Search Block Matching," IEEE Trans. on Circuits and Systems for Video Technology, vol. 3, no. 5, pp. 368--379,1993

[4] Michael C. Chen, "Rate-distortion optimal motion estimation algorithms for motion-compensated transform video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol.8, No.2, 1998.

[5] A.N. Evans, "Vector area morphology for motion field smoothing and interpretation," *IEE Proc. Vis. Image Signal Process*, Vol. 150, No. 4, August 2003.

[6] H. Watanabe and S. Singhal, "Windowed motion compensation,"in *SPIE Visual Communications and Image Processing,* Vol. 1605, pp. 582-589, Nov. 1991, Boston, MA.

[7] M. H. Orchard, Y. B. Yu, "Variable size block matching motion compensation with applications to video coding, " *IEE Proceedings*, vol.137, no.4, pp.205-212, Aug. 1990.

[8] Y.-K. Chen, Y.-T. Lin, "A feature tracking algorithm using neighborhood relaxation with multi-candidate pre-screening," *Proceedings of ICIP*, vol. 2, pp. 513-516, Sept. 1996.