A NOVEL MOTION ESTIMATION METHOD FOR MESH-BASED VIDEO MOTION TRACKING

Mohammed Sayed and Wael Badawy

Department of Electrical and Computer Engineering University of Calgary
2500 University Drive NW
Calgary, AB, Canada T2N 1N4
{sayed, badawy}@enel.ucalgary.ca

ABSTRACT

This paper presents a novel motion estimation method for mesh-based video motion tracking. The proposed method has been called mesh-based square-matching (MB-SM) motion estimation method. The MB-SM method outperforms the commonly used motion estimation methods in terms of computational cost reduction, efficiency and image quality (i.e. peak signal to noise ratio (PSNR)). It yields comparable PSNR values with the hexagonal matching method [1] while it has lower computation complexity.

1. INTRODUCTION

Mesh-based video motion tracking (VMT) is an advanced VMT method. In mesh-based motion tracking [2], a mesh is established for the video frame and the video motion is represented by the deformation of that mesh. Therefore, the current frame polygonal patches are deformed by the movement of the mesh nodes (control grids) into polygonal patches in the reference frame to represent the video motion. The pixel motion inside each block (patch) is found by interpolating the motion vectors of the mesh nodes (control grids) [3] using one of the spatial transformations (i.e. affine, bilinear or perspective transformations). Then, the texture inside each polygonal patch in the reference frame is warped [4] onto the corresponding current frame polygonal patch.

The main advantage of the mesh-based VMT methods over the block-matching algorithm (BMA) VMT methods is its ability to represent more general types of motions such as rotation, zooming and shear. At the same time, the mesh structure constrains movements of adjacent patches, such that the best matching blocks (patches) do not overlap either in the reference frame or in the current frame. A disadvantage of the mesh-based VMT method is that it does not allow discontinuities in the motion field while the BMA method can handle discontinuities in the motion field. However, such discontinuities may not always coincide with the block borders [1].

There are two types of mesh structures, regular and adaptive meshes. Regular mesh is commonly used due to its simplicity and with regular mesh there is no need to send the mesh topology to the decoder because it is predefined. Rectangular and triangular patches are the two commonly used patches in the mesh structures.

This paper presents a novel motion estimation method for mesh-based video motion tracking. The new method uses the multiplication free algorithm in [5] to compute the affine parameters. The rest of the paper is organized as follows: Section 2 gives a brief overview of the mesh-based motion estimation; Section 3 presents the proposed motion estimation method; Section 4 presents the simulation results; and Section 5 concludes this paper.

2. MESH-BASED MOTION ESTIMATION

Mesh-based motion estimation is performed by estimating one motion vector with two components, horizontal and vertical, for each mesh node. These motion vectors are used to represent the complete 2-D motion field [6]. The simplest method to estimate a mesh node motion vector is to select a block of pixels around this mesh node, then estimate the motion vector for that block using BMA and use the block motion vector as the mesh node motion vector [7] [8]. Since this method depends on the BMA, its implementation is easy using any of the already implemented BMA architectures [7]. In this method, the motion of each mesh node is estimated separately, which may lead to connectivity problems if the mesh nodes motion vectors cross over. In addition this method does not produce high video quality in terms of PSNR values.

A more complicated mesh-based motion estimation method, called hexagonal matching, has been proposed in [1]. In the hexagonal matching method, the motion estimation process is accomplished in two steps: 1) coarse motion estimation using the BMA to find an initial motion vectors; and 2) iterative local minimization of the prediction error to refine the motion vectors [1].

Although this method needs large amounts of computation, it produces higher PSNR values than the mesh-based motion estimation using the BMA method.

3. THE PROPOSED METHOD

The proposed mesh-based motion estimation method is a modified version of the hexagonal matching motion estimation method [1]. It uses the multiplication-free algorithm [5] for computing the affine parameters. The proposed method has been called mesh-based square-matching (MB-SM) motion estimation method. There are three differences between the MB-SM motion estimation method and the hexagonal matching method as follows:

1. Using right angle triangles to benefit from the multiplication-free algorithm introduced in [5] instead of using equilateral triangles, as in the hexagonal matching method.
2. Using address interpolation method instead of the intensity values interpolation method, used with the hexagonal matching method [1].
3. The refinement step is performed inside a square area instead of the hexagonal area, so that it is suitable for memory accessing and single instruction multiple data (SIMD) implementation [9].

In the MB-SM motion estimation method, a new regular triangular mesh is established for every new frame. A backward motion estimation is accomplished for the mesh nodes to find their best matching locations on a previous reference frame, as shown in Figure 1. The mesh nodes on the current frame edges are not allowed to move outside the reference frame, their motion is limited to side motion (i.e. motion on the frame edge) or interior motion (i.e. motion toward the frame center). While the rest of the mesh nodes can move freely in any direction but in a limited region to avoid mesh overdeformation and to prevent mesh nodes from crossing over. The MB-SM motion estimation is performed in two steps:

1. Rough motion estimation using block-matching algorithm (BMA).
2. Fine motion estimation to refine the motion vectors generated from the first step.

The first step can be explained with the help of Figures 2 and 3. For each mesh node in the current frame, a block of pixels surrounding the mesh node is selected such that the mesh node lies in its center, as shown in Figure 2. Then, using the BMA, candidate locations are found for the mesh nodes. The search area for the BMA is limited to avoid excessive motion of the mesh nodes, as shown in Figure 3. The BMA is applied with the full search strategy and with the SAD matching criterion. A fixed mesh size of 16 pixels is applied and a block size of 16x16 pixels is used for the BMA in the first step. The search range in the first step is ±2 and this search range increases to ±3 with the second step (i.e. refinement step).

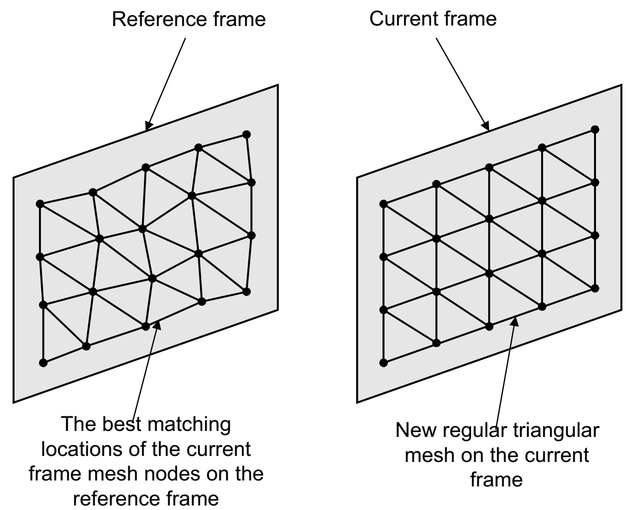

Figure 1. The mesh-based motion estimation process

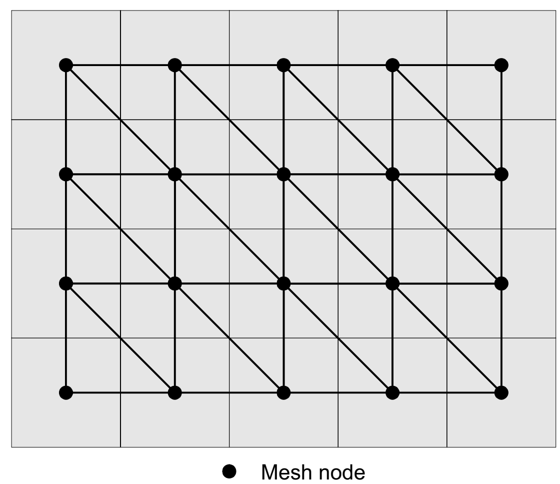

Figure 2. The first step of the MB-SM motion estimation method

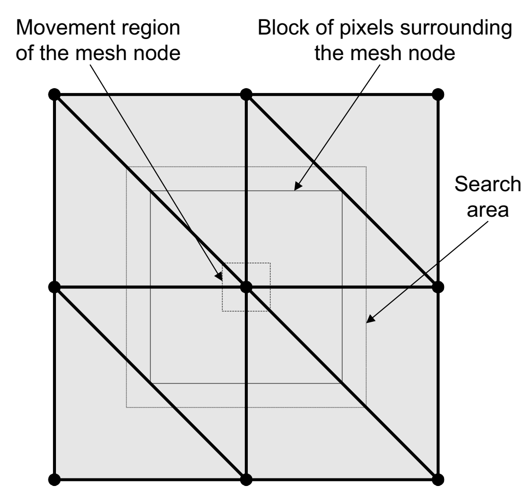

Figure 3. The BMA is applied with limited search area to avoid mesh overdeformation and to prevent mesh nodes from crossing over

The second step is a refinement process using the actual prediction error due to warping the frame area around the mesh node under refinement, as shown in Figure 4. Starting from the mesh nodes candidate locations found by the first motion estimation step, a refinement process is accomplished for each internal mesh node. The refinement process is not accomplished for the mesh nodes on the frame edges because the simulation results show insignificant improvement in the PSNR values using the refinement process for those mesh nodes.

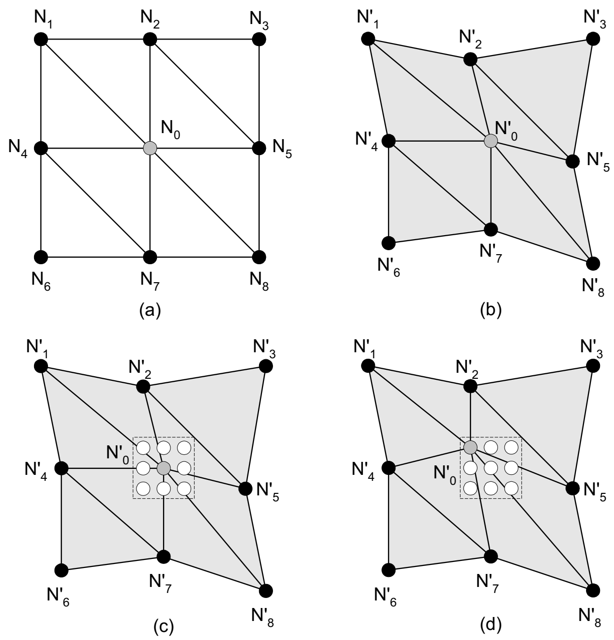

Figure 4. The second step of the MB-SM motion estimation method

The refinement process of one mesh node can be explained with the help of Figure 4. Figure 4a shows the mesh node under refinement (N_0) and the mesh nodes around it (N_1, N_2, N_3, N_4, N_5, N_6, N_7, N_8) in the current frame; Figure 3.4b shows the candidate locations for those mesh nodes (N'_0, N'_1, N'_2, N'_3, N'_4, N'_5, N'_6, N'_7, N'_8) in the reference frame found by the first step. The details of the refinement process are as follows:

1. Fixing the locations of the eight mesh nodes (N'_1, N'_2, N'_3, N'_4, N'_5, N'_6, N'_7, N'_8) around the central mesh node (N'_0) under refinement.
2. The central mesh node (N'_0) will be allowed to move one pixel in each direction, resulting in eight candidate locations, plus the candidate location found by the first step, which gives nine possible locations for that mesh node (i.e. nine possible motion vectors for that mesh node), as shown in Figure 4c. Steps 3 and 4 will be done for each of these possible locations, as shown in Figure 4d.
3. The image inside the shaded region is synthesized by warping the deformed patches of the reference frame onto the undeformed patches of the current frame. The multiplication-free algorithm [5] is used to compute the affine parameters.
4. The sum of absolute difference (SAD) is calculated between the predicted image in step 3 and the corresponding region in the current frame.
5. The optimum location of (N'_0) among the nine possible locations is the location corresponding to the minimum SAD value computed in step 4.
6. The optimum location found in step 5 is registered as the new location of mesh node (N'_0).

After the refinement process, the final motion vectors of the mesh nodes lie between +3 and –3, and are used to represent the complete 2-D motion field.

4. SIMULATION RESULTS

The MB-SM method for mesh-based motion tracking has lower computational cost than the hexagonal matching motion estimation method while it produces almost the same PSNR values. The average PSNR value, in decibel (dB), of the first 30 frames of "Miss America" sequence, for both the MB-SM method and the hexagonal matching method, is shown in Table 1.

	The MB-SM method	The hexagonal matching method
Average PSNR value	38.73	38.76

Table 1. The average PSNR of the first 30 frames of Miss-America" video sequence using the hexagonal-matching and the MB-SM methods

The MB-SM method has a higher computational cost, compared to the mesh-based motion estimation with BMA method, but it yields higher video quality. Figures 5-7 show the PSNR values of the first 30 frames of "Miss-America", "Claire", and "Grandma" video sequences using the BMA, mesh-based motion estimation with BMA (MB-BMA), and the proposed mesh-based square matching method (MB-SM). Table 2 shows the average PSNR values of the first 30 frames of "Miss-America", "Claire", and "Grandma" video sequences using the BMA, MB-BMA, and MB-SM methods.

	BMA	MB-BMA	MB-SM
Miss-America	36.62	37.26	38.73
Claire	30.36	33.98	37.00
Grandma	36.45	42.55	43.07

Table 2. The average PSNR of the first 30 frames of Miss-America", "Claire", and "Grandma" video sequences using the BMA, MB-BMA, and MB-SM methods

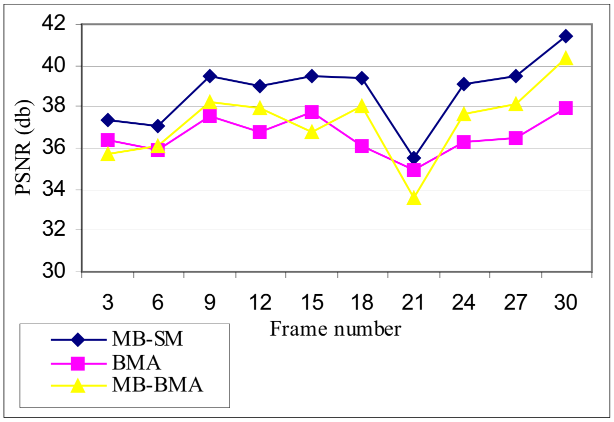

Figure 5. PSNR values of frames 1-30 from “Miss-America” video sequence

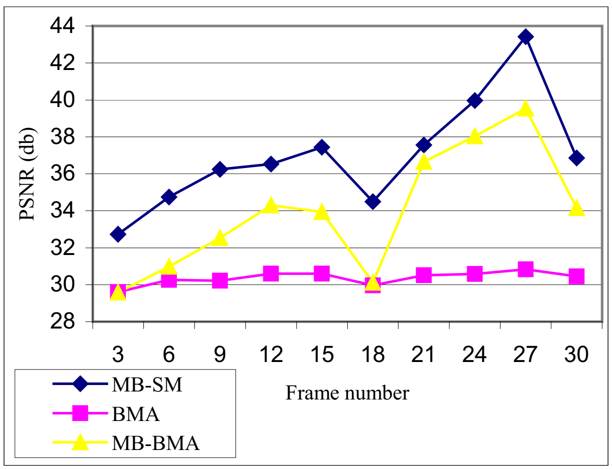

Figure 6. PSNR values of frames 1-30 from “Claire” video sequence

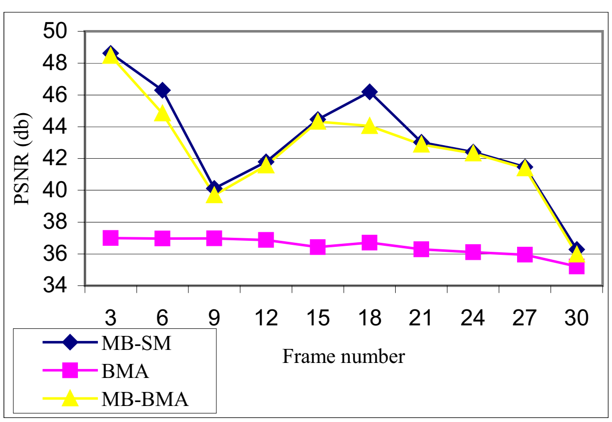

Figure 7. PSNR values of frames 1-30 from “Grandma” video sequence

5. CONCLUSION

This paper presents a novel motion estimation method for mesh-based video motion tracking. The proposed method has been called mesh-based square-matching (MB-SM) motion estimation method. The MB-SM method achieves higher video quality than the BMA and the MB-BMA methods in terms of PSNR values. In addition, the MB-SM method for has lower computational cost than the hexagonal matching motion estimation method while it produces almost the same PSNR values.

6. ACKNOWLEDGEMENT

The authors would like to thank Natural Sciences and Engineering Research Council of Canada (NSERC), Canadian Microelectronics Corporation (CMC), Micronet R&D Canada, Human Resource Canada (HRC), University Research Grant Committee (UGRC) and Department of Electrical and Computer Engineering at University of Calgary for supporting this research.

7. REFERENCES

[1] Y. Nakaya and H. Harashima, “Motion Compensation Based on Spatial Transformations,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 4, no. 3, pp. 339-367, June 1994.

[2] A. M. Tekalp, P. V. Beek, C. Toklu and B. Gunsel, “Two Dimensional Mesh-Based Visual-Object Representation for Interactive Synthetic/Natural Digital Video,” IEEE Proceedings, vol. 86, no. 6, pp. 1029-1051, June 1998.

[3] G. J. Sullivan and R. L. Baker, “Motion compensation for Video Compression Using Control Grid Interpolation,” in Proceedings IEEE International Conference Acoustics Speech Signal Processing, Toronto, Canada, vol. 4, pp. 2713-2716, May 1991.

[4] G. Wolberg, Digital Image Warping, Los Alamitos, CA: Computer Society Press, 1990.

[5] W. Badawy and M. Bayoumi, “A multiplication-free algorithm and a parallel architecture for affine transformation,” Kluwer journal of VLSI signal processing 31, pp. 173-184, 2002.

[6] Y. Wang, O. Lee and A. Vetro, “Use of Two-Dimensional Deformable Mesh Structures for Video Coding, Part II¾The Analysis Problem and a Region_Based Coder Employing an Active Mesh Representation,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 6, no. 6, pp. 647-659, December 1996.

[7] Y. Wang and J. Ostermann, “Evaluation of Mesh-Based Motion Estimation in H.263-Like Coders,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 8, no. 3, pp. 243-252, June 1998.

[8] W. Badawy and M. A. Bayoumi, “A Low Power VLSI Architecture for Mesh-Based Video Motion Tracking,” IEEE Transactions on Circuits and Systems¾II: Analog and Digital Signal Processing, vol. 49, no. 7, pp. 488-504, July 2002.

[9] M. Sayed, “A Computational RAM Architecture for Real-Time Video Motion Tracking,” M.Sc. Thesis, University of Calgary, August 2003.