A Mesh-based Method for Wavelet Video Coding using Edge-Detection in Low Frequency Subband

Miok Kim, Nam Ling,
Department of Computer Engineering
Santa Clara University
Santa Clara, California 95053, USA

John D. Ralston
Yosot, Inc.
400 Capitol Mall Ste 3000
Sacramento, CA 95814, USA

Li Song
Institute of Image Communication and Information Processing
Shanghai Jiao Tong University
Shanghai, China

***Abstract*—In this paper, we propose a mesh-based method for motion estimation and compensation over low frequency subband for wavelet video coding. We concentrate on edge-detection information and apply different parameters using edge magnitude and edge angle values to indicate curvature values for objects. This helps to detect motion patches between adjacent frames for meshes. We analyze and compare coding efficiency with a variety of video test sequences. Experimental results show that our mesh method improves coding performance for videos with less globally distributed motions compared to local motions.**

***Keywords*-Motion estimation, video coding, visual communications, wavelet video coding, mesh-based motion estimation.**

I. INTRODUCTION

Wavelet video coding with multi-resolution theory for signal processing allows some characteristics to be easily detected at one resolution than at others [1]. In [2] and [3], motion estimation and compensation methods for wavelet video coding which focus on high frequency subbands such as a mesh-based motion estimation method and overlapped block motion compensation (OBMC) are presented, especially for real-time low resolution video compression. One of the key challenges in wavelet coding is to reduce high-band signals from the motion estimation and compensation step.

Various wavelet video coding schemes based on triangular meshes have been proposed [4, 5, 6]. In [4], triangular meshes are applied to redundant-wavelet domain in order to address a shift variant property problem. In [5], uniform triangular meshes are used with different node-point topology and meshes in [6] are based on 3D wavelets.

To improve the process of generating a mesh for video coding, several mesh refinement methods have been applied; examples include progressive mesh refinement structure in layers [7], occlusion adaptive content-based meshes [8], and object-based approach of triangular meshes [9]. Although these techniques were proposed to the segmentation of object or polygon shapes in video frame sequences, their models do not concern about previous object shapes in the reference frame.

In this paper, we propose edge-detection-based mesh motion estimation and compensation method for wavelet video coding by using motion features on low frequency subband in order to enhance coding efficiency and visual quality. Some edge detectors cannot remove salt and pepper noise so that wavelet transform is utilized in order to decrease noise [14]. However, the key idea of the proposed method is to utilize spatial edge gradients that leverage low-band signals generated via wavelet coefficient in the low frequency subbands.

The rest of paper is organized as follows. Section II describes existing edge detection methods in image processing. In Section III, we propose a content-based mesh motion estimation and compensation method. Experimental results are given in Section IV. Section V concludes the discussion.

II. BACKGROUND

In this section, we describe wavelet video coding, mesh-based motion estimation, and edge and corner detection in order to present motion characteristics between video frames.

A. Wavelet Video Coding

Motion estimation and compensation on spatial domain is used in wavelet video coding in order to exploit the spatial correlation present in the video sequences. A discrete wavelet transform (DWT) is applied to generate a set of wavelet coefficients for each subband which is generally coded separately.

We denote the relationship between DWT and subbands and focus on low frequency subbands:

$$\begin{aligned} &\mathrm{DWT}(F_{\mathrm{source}}(i, j)) \\ &\quad = \mathrm{LL}(F_{\mathrm{source}}(i, j)) + \mathrm{LH}(F_{\mathrm{source}}(i, j)) \\ &\qquad + \mathrm{HL}(F_{\mathrm{source}}(i, j)) + \mathrm{HH}(F_{\mathrm{source}}(i, j)) \end{aligned} \tag{1}$$

where L denotes a low pass filter function, H denotes a high pass filter function, F_{source} represents an original input frame, and (i, j) is the block location on a frame.

B. Mesh-based Motion Methods

Generally, a mesh-based motion estimation and compensation technique assumes that each frame is partitioned into several non-overlapping triangular patches. This model takes advantage of the spatial and temporal block continuity by using triangular meshes and affine transformations, respectively.

In mesh-based motion estimation, mesh generation methods require accurate representation of motion by using triangular meshes. Particularly, a Delaunay triangulation method is popular [13]. In this paper, we utilize an incremental insertion algorithm which is an $O(n^2)$ time algorithm, where n is the number of vertices.

Another issue with triangular meshes is the affine transformation, which models the motion at each pixel (x, y) within a triangular mesh. It is modeled as follows:

$$\begin{aligned} x' &= a_1 x + a_2 y + a_3 \\ y' &= a_4 x + a_5 y + a_6 \end{aligned} \quad (2)$$

where $\{a_1, a_2, a_3, a_4, a_5, a_6\}$ denote affine transform parameters and (x', y') is the pixel location in the reference frame offset by the motion vector. Furthermore, we can use similarity transform to approximate this affine transform, that is

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = s\begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix}\begin{bmatrix} x-x_0 \\ y-y_0 \end{bmatrix} = s\begin{bmatrix} \cos\alpha(x-x_0)+\sin\alpha(y-y_0) \\ -\sin\alpha(x-x_0)+\cos\alpha(y-y_0) \end{bmatrix}$$

$$\begin{aligned} x' &= (s\cos\alpha)x + (s\sin\alpha)y - s(x_0\cos\alpha + y_0\sin\alpha) \\ y' &= (-s\sin\alpha)x + (s\cos\alpha)y + s(x_0\sin\alpha - y_0\cos\alpha) \end{aligned} \quad (3)$$

C. Edge and corner detection

Edge detection and corner detection are essential parts in various computer vision and image-understanding systems [5, 7]. The approach extracts edges of an image and then identifies corner points with gradient strength and direction.

We compute the directional gradient value for each pixel by using the sobel mask which is selected for this purpose due to smoothing characteristics. $Mag(x, y)$ is defined as the magnitude of vector ∇f:

$$Mag(x, y) = Mag(\nabla f) = \sqrt{g_x^2 + g_y^2} \quad (4)$$

which is the value of the rate of change in the direction of the gradient vector. In addition, $Angle(x, y)$ is defined as the direction of vector ∇f:

$$Angle(x, y) = tan^{-1}\begin{bmatrix} g_y \\ g_y \end{bmatrix} \quad (5)$$

which is measured with respect to the y-axis.

There exist a number of approaches which detect the corners on extracted edges [10]. Usually, boundary-based methods which detect edges first and then identify the curvature along edges are used. However, edge detection and corner detection of the current frame is likely to result in region recognition and is not related to reference frame information. Moreover, edge detection methods for corner detection are sensitive to noise and texture so denoising is an important issue when dealing with derivatives for edge detection. In addition, the criterion of edge strength is not well suited for video compression, since each boundary in a frame could lead to false motion estimations, and increase motion estimation error.

The proposed algorithm tackles the problems with the local maximum of gradient angle and low frequency subband on wavelet transform; we focus on motion recognition with edge detection and corner detection and discuss irregular triangular meshes.

III. CONTENT-BASED MESH MOTION METHOD

In this paper, we present a content-based triangular mesh-based algorithm, including directional gradient value-based mesh node generation over low frequency subband between two adjacent frames. The underlying assumptions is that low frequency subbands possess large structure of whole frame and edge angle differences reflect huge motion changes or new regions between two adjacent frames. Moreover, when we utilize only low-frequency subbands, the noise of the original frame is removed using wavelet filters.

A. Mathematicl model

Edge detection is the method used frequently for segmenting images based on local abrupt changes. In order to distinguish between background and foreground regions, edge angle for spatial and temporal directions are considered. Note that regions in a frame have angular appearance. In practice, edges have edge strengths that are blurred and noisy so that edge angle is more reliable than edge strength while taking into account factors such as image noise and the nature of edges themselves.

For motion estimation and compensation, if two regions have widely differing spatial and/or temporal characteristics, meshes should be generated. Let us define a linear combination of the motion parameters of the regions as follows:

$$Motion(x, y) = \lambda \cdot SpatialAngle(x, y) + (1-\lambda) \cdot TemporalAngle(x, y) \quad (6)$$

where *SpatialAngle* is the angle of gradient and *TemporalAngle* is the angle difference between current frame and reference frame of the pixel (x, y). That is, this measure has to be modified to take into account the curvature. As a consequence, the weighting factor λ allows more importance to be given to the spatial edge or to the edge difference.

B. Procedure

This triangulation method assumes that low-band subbands contain overall information for motion estimation and compensation. To make use of information in the low-band signals, we utilize corner-detection method over low frequency subbands.

Figure 2 shows how we determine corner points using the edge angle constraints defined by using Eq. (7). Using Eqs. (8) and (9) below, the steps used to obtain meshes are as follows:

Step 1: Perform DWT and calculate low frequency subbands for the current input frame and previous reconstructed frame, as in Figure 1. Calculate edge gradient detected over low-bands.

Step 2: Compare it with that of reference frame and give an angle a weight according to the status of edge strength and edge angle as follows:

$$\begin{aligned} &AngleValue(i,j,t) = a(Angle(i,j,t)); \\ &\text{if } Angle(i,j,t) - Angle(i,j,t-1) \le Th, \text{ set } a < 1 \text{ (case 1)}; \\ &\text{if } Angle(i,j,t) - Angle(i,j,t-1) > Th, \text{ set } a > 1 \text{ (case 2)}. \end{aligned} \quad (7)$$

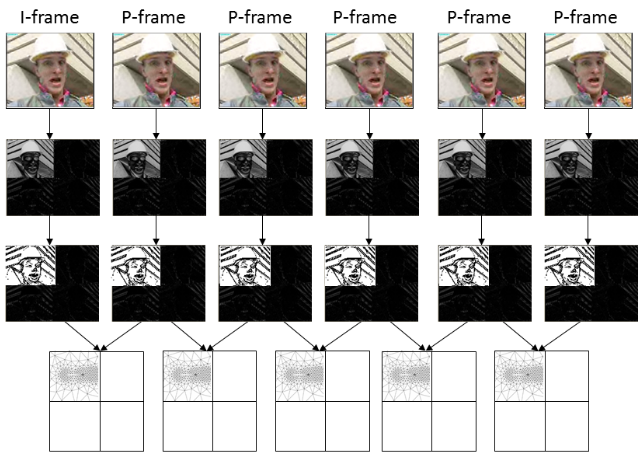

Figure 1. Proposed hierarchical mesh-based algorithm. Pixels above the threshold value are in black.

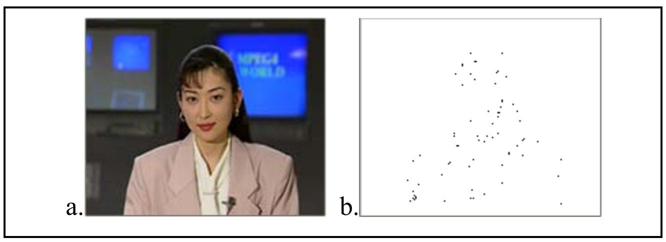

Figure 2. a: original frame and b: candidate corner points

where *Mag*(*i, j*) as defined in Eq. (4) denotes pixel magnitude value, *AngleValue* (*i, j*) denotes edge angle value at (*i, j*) in a frame, *Th* is a threshold value and *a* is a constant.

In case 1, *Mag*(*i, j, t*) is identical to *Mag*(*i, j, t*-1) and *Angle*(*i, j, t*) is similar to *Angle*(*i, j, t*-1) if the angle difference is less than or equal to a threshold. *t* and *t*-1 refer to current and previous frames, respectively. The *Mag*(*i, j, t*) in case 2 has the same edge strength in *Mag*(*i, j, t*-1) and *Angle*(*i, j, t*) is different from *Angle*(*i, j, t*-1). In this paper, we use *a* = 0.1 for case 1 and 5 for case 2, by empirical method. *Th* is set close to 0.

Step 3: Let *m* = predefined number of vertices, *Max_X* = maximum *x* position of a vertex, *Max_Y* = maximum *y* position of vertex, of the input frame, *dist* = predefined minimum distance between two vertices in a mesh triangle, and *depth* = predefined DWT depth, we obtain the angle values (*AngleValue*(*i, j, t*)'s) and the intensity values (*MagValue*(*i, j, t*)'s) using Eqs. (7) and (8) for each pixel (*i, j*), where $0<i<(Max\_X)/2^{depth}$ and $0<j<(Max\_Y)/2^{depth}$. In the low-subband, we find a *CurvatureValue*(*i, j, t*) in each scan, for a total of *N* scans for *N* mesh nodes. For each scan, we choose the pixel with the largest *CurvatureValue*(*i, j, t*) (defined in Eq. (9)) excluding the selected *CurvatureValue's* in the previous scans. These *CurvatureValue*(*i, j, t*) form our mesh nodes, where $(i'-i) \le dist$ and $(j'-j) \le dist$, and (i', j') already belongs to the set of mesh nodes, then the pixel (*i, j*) is selected as a new mesh node.

$$MagValue\ (i,j,t) = c \times Mag\ (i,j,t) + d \times \sum_{k=-1}^{1}\sum_{l=-1}^{1} Mag\ (i+k, j+l, t) \quad (8)$$

$$CurvatureValue(i,j,t) = AngleValue\ (i,j,t) \times MagValue\ (i,j,t) \quad (9)$$

MagValue(*i, j*) denotes a filtered edge magnitude value, and *CurvatureValue* (*i, j*) denotes an edge curvature value at (*i, j*) in a frame and *c* and *d* are constant values. In this paper, we use *c* = 7/8 and *d* = 1/72, from empirical method.

Step 4: We use the Bowyer/Watson Algorithm [13]. The algorithm outlined above allows for denoising information in such a way that low frequency bands are produced. After the triangular meshes are generated for the low frequency subband, motion estimation and compensation are performed over the meshes.

We apply our method and generate Delaunay triangles.

IV. Experimental Results

We implemented two mesh-based motion estimation and compensation methods and tested them with OBMC based on the Dirac wavelet codec software [11].

A. Experiment setup and results

The experimental codec was implemented using MS Visual Studio C++ on an Intel® Xeon® CPU 2.4 GHz with 24 GB of RAM, and MS Windows 7 OS. A LEGALL5_3 DWT [15] was selected, and the wavelet depth was 2. A group of pictures (GOP) length of 20 frames and a frame rate of 30 fps were used throughout our experiments.

Pixel-level motion vector precision was applied. One video sequence of CIF resolution (352 × 288) and three QCIF (176 × 144) sequences were tested: Grandma, Hall, Bridge-close, and Salesman [12].

In this paper, our triangular mesh-based method uses irregular triangles generated by a Delaunay triangulation and affine transformation. While utilizing edge detection methods, it is possible to design content-based triangular meshes that allow tracking motions using overall frame information. Recall that the performance of mesh-based techniques is based on how triangular meshes are designed in terms of occlusion.

If the video sequence contains no global motions the proposed method works well. That is, the background has no motion and object has locally motions (such as head and shoulders of Gandma). In such case there is not much occlusion.

Two proposed mesh-based methods were tested: edge mesh-based method (*Edge-Mesh*) and low-band signals of edge differences mesh-based method (*Hier_ED*). In *Edge-Mesh* method, we implemented our edge-based technique on the spatial domain to produce conventional 2D triangular meshes. For *Hier_ED*, DWT was applied first to obtain frequency domain subband frames. Our edge-based technique was applied to the low subband in each P frame. Every P frame utilized a different irregular mesh structure and the mesh structure was formed based on the edge differences between two adjacent frames and the DWT depth as explained in the previous section. Experiments show that *Hier_ED* significantly reduces the time complexity compared to those of *Edge-Mesh*, because they are based on low-band signals of frame, and applied after the DWT is performed, as shown in Figure 1. Figure 3 shows the rate-distortion results. The method was based on luminance magnitude values of all frames.

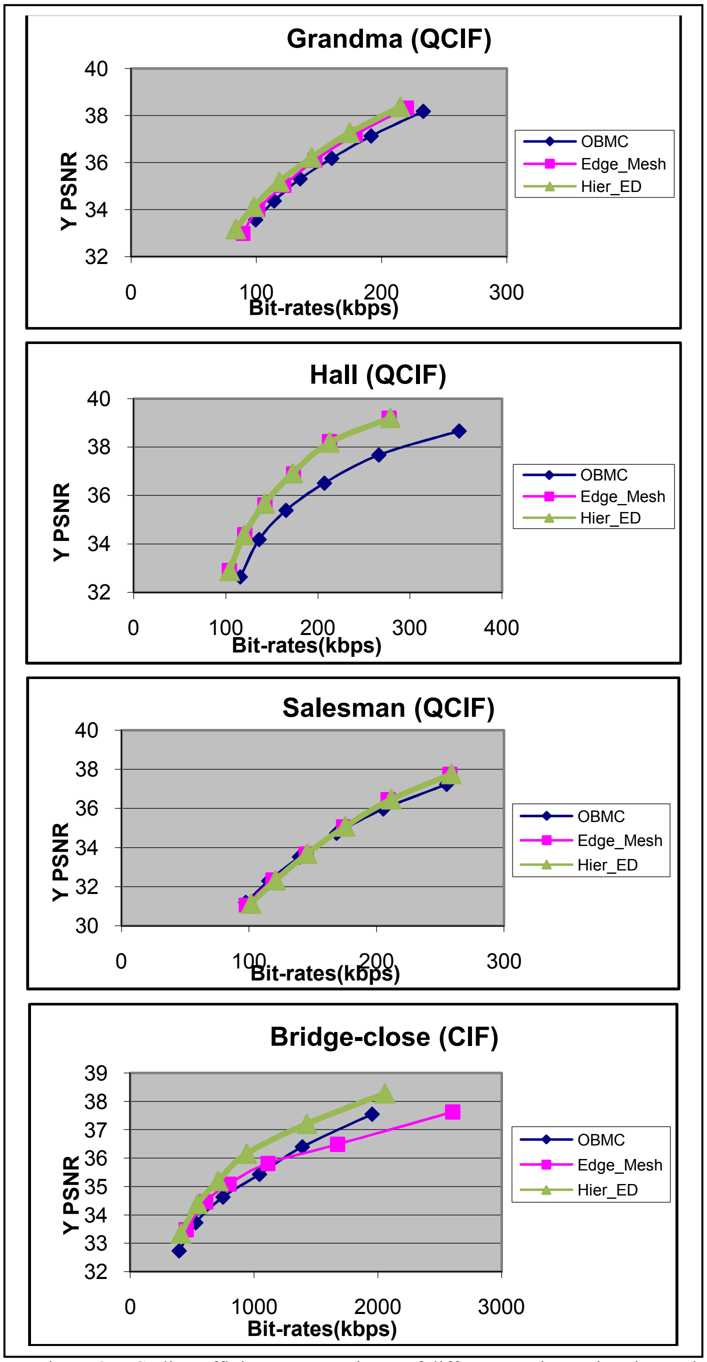

Figure 3. Coding efficiency comparisons of different motion estimation and compensation methods.

B. Analysis

In our experiment, the average number of triangles was 24. Search range was ±8 pixels.

We have also tested other sequences with globally distributed motions, such as Foreman and Highway. Mesh-based approaches produce poor results when there are globally distributed motions, due to sudden occlusion regions.

In video sequences with a mixture of different motions (e.g. Hall), *Hier_ED* and *ED_Mesh* method requires lower bit-rates for higher peak-signal-to-noise-ratio (PSNR) than OBMC because triangular meshes are flexible to generate high-band signals for prediction errors. It is more difficult to track motions using triangular meshes for video sequences with fine details in the frame (e.g., Suzie). In addition, even though some video sequences contain head and shoulders (e.g. Silent and Suzie) they show different performances due to the levels of fine detail and motion.

From our results, a Delaunay hierarchical mesh-based method utilizing low-band signals of frame achieves the best rate-distortion in video sequences containing less detailed regions with only local motions. In other cases (e.g. Suzie, Highway and Foreman) not shown in Figure 3, OBMC was found to deliver better rate-distortion results for video sequences containing lots of fine details and large globally distributed motions.

V. Conclusion

This paper proposes a new low frequency approach based content mesh for wavelet video coding. Our algorithm is especially useful for low bit-rate applications when videos contain little global motions compared to local ones, and without much fine details in motion.

References

[1] Rafael C. Gonzales, and Richard E. Woods, Digital Image Processing, third edition, Prentice Hall, 2008.

[2] Nikola Bozinovic and Janusz Konrad, "Mesh-based motion models for wavelet video coding," Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 3, pp. 141-144, May 17-21, 2004.

[3] Jonathan K. Su and Russell M. Mersereau, "Motion estimation methods for overlapped block motion compensation," IEEE Transactions on Image Processing, Vol 9, No 9, pp. 1509-1521, Sep. 2000.

[4] Yonghui Wang, Suxia Cui, and James E. Fowler, "3D video coding with redundant-wavelet multihypothesis," IEEE Transactions on Circuits and Systems for Video Technology, vol. 16, pp. 166-177, February 2006.

[5] N. Božinović and J. Konrad, "Mesh-based motion models for wavelet video coding," Proceedings of the IEEE International Conference Acoustics Speech Signal Processing, vol. III, pp. 141-144, May 2004.

[6] N. Cammas and S. Pateux, "Fine grain scalable video coding using 3D wavelets and active meshes," SPIE Visual Communications and Image Processing, VCIP 2003, Santa Clara, USA, Jan. 2003.

[7] H. Park, A. Yu, and G. Martin, "Progressive mesh-based motion estimation using partial refinement," Proceedings of the International Workshop on Very Low Bit-rate Video 2005, pp. 92-98, Sep. 2005.

[8] Hak Soo Kim, Taekhyun Yun, and Kyu Tae Park, "Fast motion estimation for 2D triangular mesh based image coding," Electronics Letters, vol. 34, No. 6, pp. 540-541, Mar. 1998.

[9] M. Dudona, O. Avaroa, and C. Roux, "Triangular active mesh for motion estimation," Signal Processing: Image Communication, vol. 10, Issues 1-3, pp. 21-41, July 1997.

[10] Song Yang, Siew Kei Lam, and Thambipillai Srikanthan, "An efficient edge and corner detector," International Conference on Control, Automation, Robotics and Vision, pp. 1628-1631, Dec. 2010.

[11] Tim Borer, "BBC research white paper WHP 159," Nov. 2007.

[12] YUV video sequences, http://trace.eas.asu.edu/yuv/

[13] Henrik Zimmer, "Voronoi and Delaunay techniques," Lecture Notes, Computer Sciences VIII, RWTH Aachen, July 2005.

[14] Wenshuo Gao, Xiaoguang Zhang, Lei Yang and Huizhong Liu, "An improved Sobel edge detection," IEEE ICCSIT, 2010.

[15] M. Unser and T. Blu, "Mathematical properties of the JPEG2000 wavelet filters," IEEE Transactions on Image Processing, vol. 12, no. 9, pp. 1080-1090, Sep 2003.