PROGRESSIVE MESH-BASED MOTION ESTIMATION USING PARTIAL REFINEMENT

Heechan Park, Andy C Yu and Graham R Martin

Department of Computer Science
University of Warwick, Coventry United Kingdom
email: {heechan, andycyu, grm}@dcs.warwick.ac.uk

ABSTRACT

A technique for performing progressive mesh-based motion estimation in a layered fashion is presented. Motion compensation based on image warping provides a block prediction free of block artifacts. The smooth prediction can be used to identify motion-active regions by comparing with the reference frame and generate a partial denser mesh, thus forming layers of mesh. This approach provides a hierarchical partial refinement according to motion activity without additional cost. Experimental results indicate that the technique shows improvement over a single-layered uniform mesh and advantages over block-based techniques, particularly in scalable and very low bitrate video coding.

1. INTRODUCTION

Block matching motion estimation forms an essential component of inter-frame coding in many video coding standards. The block matching algorithm adopts a translational motion model, but this is intrinsically limited when representing real world motion. In order to cope with complex motion such as rotation and zooming, deformable mesh-based algorithms have been proposed. In general, a mesh consists of polygonal patches, and a spatial transformation function is applied to map the image into a new coordinate system.

Mesh-based motion estimation can be divided into two categories, defined by whether motion is estimated in the forward or backward directions. In backward motion estimation, a mesh is applied to the current frame and deformations are estimated from the current to the reference frame. Forward methods operate in the opposite manner. Backward motion estimation is widely used because of its relative simplicity and the lower computational requirement of the mapping process. Forward methods can provide adaptive deformation of the patch structure to track feature changes in the image, but at the expense of complexity. Mesh-based techniques can be further categorised depending on whether regular or irregular mesh is employed. A regular mesh consists of uniform patches. An irregular mesh is generated according to the image content using Delaunay or Quadtree methods, and where patch size varies with intensity gradient or motion activity. Generally, a regular mesh is associated with backward estimation. The irregular mesh is coupled with forward estimation to avoid transmitting a large overhead for the patch structure. The latter provides better performance than a regular mesh. However it is not popular in real applications due to the high complexity of the forward method or associated overhead for transmitting node positions. In this paper, we present a regular mesh technique with backward estimation that features the advantages of an irregular mesh.

2. ME / MC WITH SPATIAL TRANSFORM

Motion estimation (ME) and compensation (MC) based on a triangular mesh, partitions the image into a number of triangular patches where the vertices are denoted as grid points. Using mesh refinement (Sec. 2.2), the displacement of each grid point is estimated and represented by a motion vector(MV). The displaced grid points define a deformed mesh that describes the underlying motion. The deformed mesh of the reference frame is obtained from estimating the displaced position of the mesh vertices in the current frame. Motion compensation proceeds by retrieving six affine parameters from the displacements of the three vertices of each triangular patch, and synthesizing patch content using a warping operation defined by the six parameters.

2.1. Affine Transform

An affine transform models translation, rotation, and scaling of a patch in the current frame to the corresponding distorted patch in the reference frame. This transformation is represented by six parameters. An intensity value of pixel (x, y) in the i th synthesized patch $\hat{P}$ in the predicted frame K is given by

$$\hat{P_i^k}(x, y) = P_i^{k-1}(f_i(x, y)) \quad (1)$$

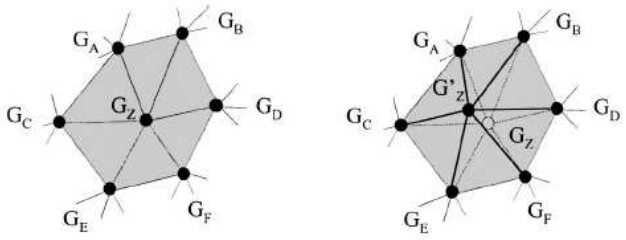


Fig. 1. Hexagon-based mesh refinement : (left) before, (right) after refinement

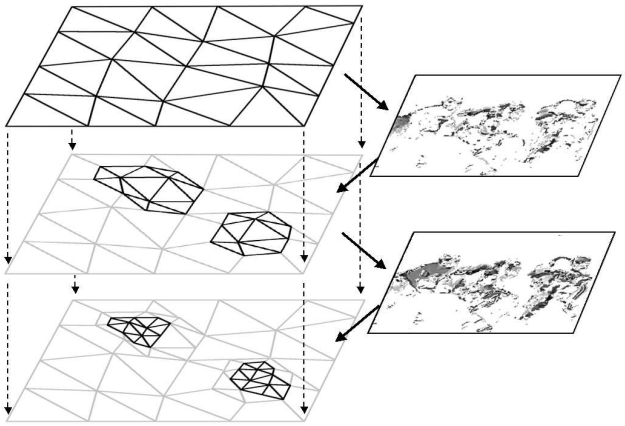


Fig. 2. Progressive mesh refinement layers and approximated frame difference maps

where the affine transform $f(\cdot)$ of the patch is given by

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (2)$$

where (x', y') and (x, y) denote positions in the reference frame and the current frame respectively. There is a one-to-one correspondence between vertices in the current frame and the reference frame, and therefore, the six parameters , $a_1, a_2, a_3, b_1, b_2, b_3$ are obtained by solving equations provided by the motion vectors at the vertices and pixels within corresponding patches can be interpolated accordingly.

2.2. Mesh Refinement

Mesh refinement refers to modification of the grid point locations so that the image intensity distribution within any two corresponding patches in the current and reference frame match under an affine transformation. Finding the optimum combination of each grid point that minimizes the difference between the current and reference frame for all possible combinations of grid point locations is not feasible in practice. Instead, finding a sub-optimum combination by successively visiting each grid point of the mesh and moving the grid point to a new position within range, thus preserving mesh connectivity and minimizing matching error locally has been developed by Nakaya et al. [1].The ME of the grid points proceeds with iterative local minimization of the prediction error to refine the MV as below.

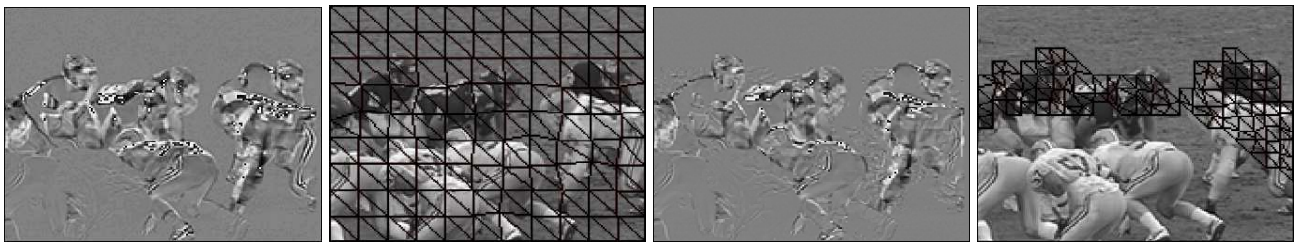
While keeping the location of the six surrounding grid points, $G_A \sim G_F$ fixed(Fig.1), G_Z is moved to an adjacent position G'_Z. This is repeated, and for each move, the six surrounding patches inside the hexagon are warped and compared with the current patch by the mean absolute difference. The optimum position of G'_Z is registered as the new position of G_Z. This procedure is repeated for each grid point until all the grid points converge to either local or global minima.

3. PROGRESSIVE MOTION ESTIMATION

A block-based model leads to severe block distortions while the mesh-based method may cause warping artifacts. In terms of prediction accuracy, the mesh-based model can give a more visually acceptable prediction, particularly in the presence of non-translational motion. The complex motion modelling and block artifact-free characteristic of warping enables identification of the approximate difference region between successive frames using the motion vectors alone. This does not appear to have been explored, and is the motivation behind our proposed algorithm.

As mentioned, regular / backward mesh ME is a popular choice due to its relative simplicity and lower computational cost. However, a uniform patch structure cannot accommodate non-stationary image characteristics nor cope with exact object boundary representation. This is addressed by employing a hierarchical structure using a Quadtree[2]. A mesh of a different density is applied according to motion activity, and this allows more accurate motion modelling where needed. However, the technique requires additional bits to indicate the motion active region and has constraints on the flexible movement of the grid points, which add complexity to the refinement process. We propose a progressive mesh refinement method that is adaptable to motion activity in a layered fashion, which overcomes the limitation above.

Fig.2 illustrates progressive motion estimation in terms



(a) (b) (c) (d)

Fig. 3. Frame difference and partial refinement: (a) difference between current and reference images, (b) deformed regular mesh, (c) difference between frame synthesized by (b) and current frame, (d) partial refinement

of layer and an approximated frame difference between layers to identify motion-active regions and generate the partial mesh of the next layer. A global motion estimation is performed on the top layer while the next layers concentrate on the finer motion activity. A similar approach was suggested in [3]. While our technique provides partial refinement by employing a layered mesh topology without overhead, Toklu et. al. focus on the hierarchical mesh structure to the entire image in one layer, resulting in higher bit rates. Fig.3 shows an example of an approximated difference map. The actual frame difference between the current and reference frames,(a), is quite well approximated in (c). The approximate difference, (c), is obtained from subtracting the (warped) predicted reference frame (b) from the current frame.

The technique proceeds as follows. Firstly we apply mesh refinement for a coarse prediction with a uniform mesh and synthesize an image from the resulting MVs as in the standard mesh-based algorithm. We then identify regions where motion discontinuities exist from the difference map, that is the variance of the patch difference,v_P is greater than the variance of the frame difference, v_F. The variance of the frame difference is given by

$$v_F = \frac{1}{M \cdot N} \sum_{j=1}^{M} \sum_{i=1}^{N} (f(i,j) - \bar{f})^2 \qquad (3)$$

where M and N are the frame dimensions and $\bar{f}$ denotes the mean frame difference. The variance of the patch difference is given by

$$v_P = \frac{1}{K} \sum_{i=1}^{K} (f_P(i) - \bar{f_P})^2 \qquad (4)$$

where K refers to the number of pixels in the patch.

In the next step, a finer regular mesh is applied to the regions containing motion discontinuities, as depicted in Fig.3 (d). Consequently we have layers of mesh, a coarse mesh that covers the entire image and denser partial mesh applied to the moving regions only. This allows a hierarchical refinement without the explicit overhead and constraints on movement of the grid points.

4. EXPERIMENTAL RESULT

The algorithms were evaluated using the QCIF resolution test sequences “Crew”, “Football”, “Ice”, “Suzie” and “Stefan”. According to our experiments, using two layers of mesh with a patch size of 16×16 and 8×8 show the best performance in QCIF resolution in terms of rate-distortion. The hexagonal matching algorithm[1] is applied with a search range of ±7 pixels for the first layer and ±3 pixels for the second layer which preserves mesh connectivity. More grid points in the intial layer do not necessarily lead to either better motion active region identification or a better reconstruction quality when bitrate is considered. This is due to the grid point connectivity constraint that prevents effective estimation when an over-dense mesh covers occluded / discovered areas, and of course the increased number of motion vectors. In this sense, the content-based mesh provides an advantage(see Sec. 5). The motion field is initialized with zero-motion and iteration starts with the grid point closest to the image center. A hexagon that contains at least one patch overlapping with the identified regions is included in the partial refinement process. Note there is no need to transmit the region information as the region can be identified using the MVs transmitted in each layer. Motion vectors are differentially encoded using Exp-Golomb. Wavelet coding is applied to the displaced frame residue. In Table.1, the left column shows the performace of the single-layered mesh refinement and the right column represents the performance with an additional layer. The overall performance is improved in all test sequences at a fixed bitrate (0.2 bpp). The poor improvement in the Crew sequence can be accounted for by frames containing a flashing light which is more efficiently compressed with residual coding.

	One Layer(HMA)		Two Layer	
Sequence	MVs	PSNR	MVs	PSNR
Crew	532	30.93	1190	30.95
Football	645	22.21	1149	22.32
Ice	415	27.11	913	27.68
Susie	349	37.93	719	38.15

Table 1. Experimental Result: bitrate for MV and PSNR (0.2 bpp)

5. FUTURE WORK

Scalable video coding utilizing the wavelet transform applied to both the spatial and temporal domains (3D-DWT) is of current interest in video coding[4]. The mesh-based ME/MC scheme exhibits several merits when deployed in the wavelet coder[5]. The mesh-based estimation can provide scalable coding naturally by controlling the number of grid points or controlling the number of layers in our algorithm. Also, the unique trajectory of each pixel in a mesh-based motion model overcomes the appearance of so-called "multiple/unconnected" pixels occuring in areas not conforming to the rigid translational model. S. Cui et. al. introduced a content-based mesh based on the redundant wavelet[6]. However, it is non-trivial to retrieve the same content-based mesh generated in the encoder when decoding without high overhead, which makes deployment in the wavelet coder prohibitive. In our method, first layer is always initialized with a regular mesh. There is high correlation between deformation of mesh layers. An efficient mesh topology coding strategy can be realised.

Secondly, an effective trade-off between motion coding and residual coding is of prime importance as indicated by the 'Crew' sequence. The layered mesh provides efficient control of the trade-off. Futhermore, intensity control can be introduced using the existing mesh. Each grid point has an additional parameter for intensity scaling by which pixels inside the patch are interpolated.

Lastly, mesh-based coding is also an efficient model for very low bitrate coding with the advantages as mentioned. Adequate subsampling of each layer of mesh leading to a pyramid structure can provide additional improvement in bitrate for the coding of motion information.

6. CONCLUSION

We have described a simple yet effective algorithm that uses the frame difference generated from a mesh-based motion compensated image to identify regions of motion discontinuity. Motion estimation in these regions is refined using a finer mesh structure. It is notable that the proposed approach provides hierarchical refinement without additional overhead, and with no constraint on the movement of grid point positions. The algorithm can be combined with any regular mesh topology. This work shows an improvement over single-layered mesh refinement technique.

7. REFERENCES

[1] Y. Nakaya and H. Harashima, "Motion compensation based on spatial transformations," *IEEE Trans. CSVT.*, vol. 4, no. 3, pp. 339–356, Jun. 1994.

[2] C. Huang and C. Hsu, "A new motion compensation method of image sequence coding using hierarchical grid interpolation," *IEEE Trans. CSVT.*, vol. 4, no. 1, pp. 42–51, Feb. 1994.

[3] C. Toklu, A. Erdem, M. Sezan, and A. Tekalp, "Traking motion and intensity variations using hierarchical 2-d mesh modelling for synthetic object transfiguration," *Graphical Models and Image Process.*, vol. 58, no. 6, pp. 553–573, Nov. 1996.

[4] J. Ohm, M. Schaar, and J. Woods, "Interframe wavelet coding-motion picture representation for universal scalability," *Signal Process.: Image Commun.*, vol. 19, no. 9, pp. 877–908, Oct. 2004.

[5] A. Secker and D. Taubman, "Highly scalable video compression using a lifting-based 3d wavelet transform with deformable mesh motion compensation," in *IEEE ICIP.*, Montreal Canada, Sep. 2002, vol. 3, pp. 749–752.

[6] S. Cui, Y. Wang, and J.E. Fowler, "Mesh-based motion estimation and compensation in the wavelet domain using a redundant transform," in *IEEE ICIP.*, Rochester, New York, Sep. 2002, vol. 1, pp. 693–696.