Region-Based Video Coding Using A Multiscale Image Segmentation

Seung Chul Yoon, Krishna Ratakonda and Narendra Ahuja
Department of Electrical and Computer Engineering
Beckman Institute, University of Illinois, Urbana, IL 61801
scyoon@stereo.ai.uiuc.edu

Abstract

This paper proposes a novel region-based video coding technique using a multiscale image segmentation method thus obtaining better quality at the same bit rate. In most of the previous region-based video coding techniques, occlusion caused degradation in terms of both PSNR and perceptual video quality. We propose a new motion estimation and compensation algorithm which solves occlusion related problems effectively. The proposed motion estimation and compensation is a two stage procedure: the first stage uses a coarse motion model while the second stage uses a dense motion model. The coarse motion model generates region level motion vectors which are then fine tuned by the dense motion model which produces pixel level motion vectors. A fusion of these concepts leads to a gain of $2 \sim 3$ *dB in PSNR over the block-based algorithm for a variety of test sequences using a fully functional video coder.*

1 Introduction

This paper is concerned with region-based video coding for very low bit rates ($\sim$ 10 kbits/s). 2-D region-based coding techniques such as [1] need region shape information to be sent, which is a major bottleneck on a very narrow bandwidth channel. In order to avoid this shortcoming, a region shape prediction algorithm [2] was proposed, which predicts arbitrary region shapes in the current frame by identifying their counterpart in the previous frame and sending the motion parameters. The general problem with such an approach is the instability of region segmentation across a video sequence.

To stabilize region shape prediction, we propose a two-stage motion estimation and compensation algorithm. Regions are obtained from a recently proposed multiscale segmentation transform [3]. The first stage of the proposed algorithm includes selection of the segmentation scale (which determines fineness/coarseness of detected regions) and assignment of motion parameters to each predicted region. The second stage involves the refinement of motion field using Wiener based pel-recursive motion estimation/compensation [4, 5] and linear causal models [6].

Most of the previous segmentation methods [1, 2] for video coding purposes have ignored the multiscale structure inherent in images and hence it was difficult to make efficient region shape prediction. In this paper, we propose a new region shape prediction scheme based on a multiscale image segmentation technique [3]. In Section 2, the rationale behind choosing the optimal segmentation scale is described. In order to select the appropriate scale of segmentation (according to appropriate optimization criteria), theoretically we need to loop over the entire encoding process for each scale recursively. In practice we restrict ourselves to $2 \sim 3$ scales. This process is usually termed pseudo-coding (Section 3.1). Based on the segmentation of the previous frame, the region segmentation of the current frame is predicted by simple contour motion estimation and compensation. In Section 3.2, a new method is introduced that addresses the problem of occlusion (overlapped regions and uncovered regions). This method modifies region shape by declaring these occluded regions as new regions. Motion parameters for these new regions are then transmitted as usual. This completes stage 1 of the proposed scheme. In Section 3.3, the pel-recursive motion estimation and compensation techniques [4, 5, 6] are incorporated into the proposed coding algorithm in order to increase the accuracy of the prediction of intensity values. In Section 4, the scheme used for efficient coding of the prediction error is described. Some simulation results are given in Section 5 using a *fully functional video coder*(Figure 1).

2 Region Shape Prediction and Multiscale Image Segmentation

For coding purposes, multiscale region segmentation based on gray level similarity is useful for extracting visually important information from images. Since most of the previous segmentation methods for

0-8186-8183-7/97 $10.00 © 1997 IEEE

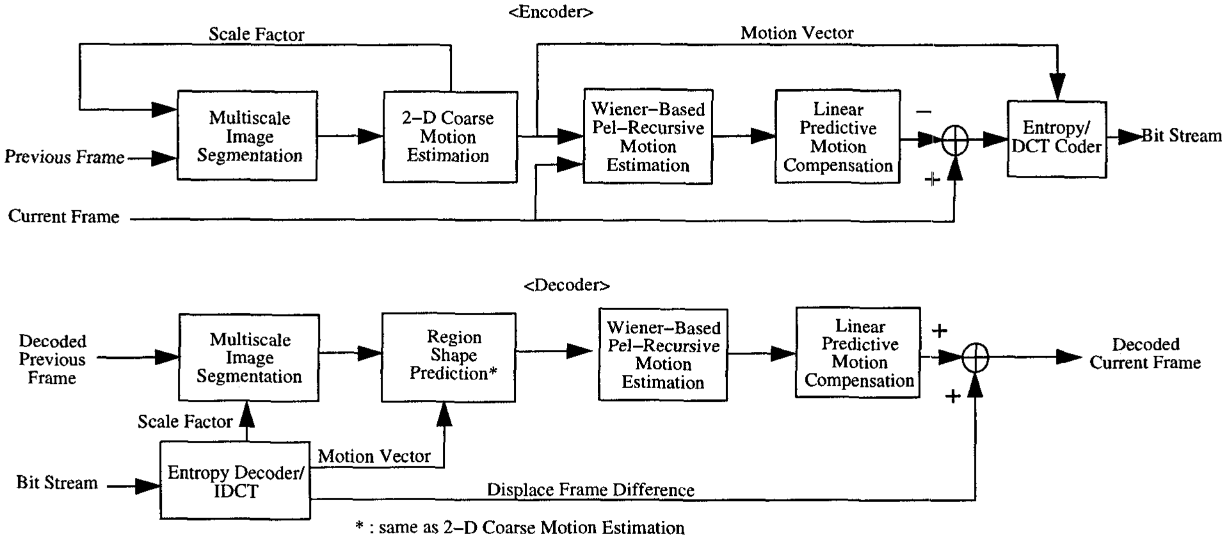

Figure 1: Overall Block Diagram of Proposed Video Coding Algorithm

video coding purposes have ignored the multiscale nature of images, it was difficult to trade off coding cost and coding efficiency. If a multiscale (e.g. tree) representation of an image is available, we can more easily control the coded bit rate by adjusting the number of regions and the accuracy of motion estimation. This is achieved by going up and down through the image segmentation tree. In this paper, we choose a multiscale image segmentation method [3] which allows us to perform a tree-search for the optimal scale from a coding view point.

Since we have no prior information about image structure, we set the initial scale parameter to the coarsest scale. The initial segmentation is obtained from the *previously decoded frame.* Region matching between successive frames is then performed to reduce temporal redundancies using motion estimation followed by compensation. If the coding cost for the current scale is less than that for the previous scale, the next finer scale of segmentation is performed. Otherwise the algorithm terminates and encoding is done at the present scale.

3 Motion Estimation and Compensation

3.1 Pseudo-Coding

The main purpose of pseudo-coding is to decide which scale of region segmentation is to be sent. Pseudo coding refers to the process of finding out whether the current set of parameters (in our case segmentation scale) leads to optimal performance; we encode once we are certain of optimality. This procedure is not a part of the decoder but a part of the encoder. The decision is based on (1) the estimation of bit rate and (2) reconstructed image quality. If the current region segmentation satisfies the pre-defined conditions such as bit rate and image quality, we transmit the information at the current segmentation scale. Otherwise, region segmentation is updated to the next finer scale.

3.2 Coarse Motion Field Model

Motion estimation solves the region matching problem, and reduces temporal redundancies between frames. The number of regions is usually much less than the number of blocks in the conventional block matching methods like H.261 and MPEG-1. Thus, motion estimation in this procedure leads to a coarse motion field. As mentioned earlier, the reference frame is the current frame. Motion vectors are transmitted by a lossless coding technique.

We assume that all the regions in the previous frame move to locations within the current frame. Even though more complex motion models can be considered, we use the translational motion model which can be implemented easily. We will refine this motion field by using a dense motion field model later on. The best matched region is found within a $+16 \sim -16$ neighborhood. The criterion to find the best match

	Mean Squared Error	PSNR
Proposed Method	19.64	35.42 dB
Block-Based Method	39.47	32.45 dB

(a) "Miss America"

	Mean Squared Error	PSNR
Proposed Method	26.06	34.16 dB
Block-Based Method	41.40	32.15 dB

(b) "Claire"

Table 1: Results from the proposed method and the conventional block-based method

is to minimize the absolute summed error. Since we use forward motion estimation, both unassigned pixels and overlapped pixels are encountered. We assume that most of the prediction error is due to such pixels. To reduce such errors, an overlapped region is first viewed as the result of occlusion between differently moving objects. If this does not suffice, then the region is declared as a newly emergent object in the current frame. Intensity values of such pixels affected by occlusions are predicted using backward motion estimation.

At this stage, small regions formed by error pixels are removed by a median filter. The remaining uncertain regions are given motion vectors using the same method as before. Motion information for the new regions is sent separately as it is available after the first contour motion compensation at decoder.

3.3 Dense Motion Field Model

Dense motion field estimation produces a more accurate motion field which reduces the prediction error. Pel-recursive motion estimation algorithms [4, 5, 6] are attractive in that no explicit motion information needs to be sent to decoder and they provide motion vectors with sub-pixel accuracy. The motion vectors transmitted by the forward motion estimation of Section 3.2 are used as initial estimates in Wiener pel-recursive motion estimation [4, 5] which is followed by linear prediction [6]. Uncertain regions due to region overlapping and uncovered region are more accurately predicted well here.

4 Residual Image Coding

A residual image between the reconstructed frame and the original frame is obtained and divided into 8 × 8 blocks for DCT coding. A uniform quantizer is used on the DCT coefficients. The quantized DCT coefficient blocks are sorted in descending order according to their energy values [2]. We transmit blocks till the desired bit rate is achieved [2]. In this method, the position of the block and their number are also sent. However this overhead has negligible effect on the bit rate.

5 Simulation Results

We evaluate the proposed method by using 150 frames of "Miss America", and 124 frames of "Claire" sequences. Only the luminance signal of QCIF(176 × 144) is used. Every fourth frame is coded (so we code 35 frames of the "Miss America" sequence and 31 frames of the "Claire" sequence). The target bit rate is fixed at 9.6 kbits/s. The proposed method uses the same Huffman coder as H.261 in order to code motion information and DCT coefficients. Side information like the block number is coded losslessly. We compare the proposed method with the conventional block-based method. Table 1. (a) and Table 1. (b) show the average mean squared error and PSNR for the decoded frames. The block-based method uses full search conventional motion estimation and compensation based on 16 × 16 blocks. Figure 2 shows that the PSNR gain of the proposed method over the block-based method is about 2 dB for "Claire" video sequence. Figure 3 and Figure 4 show the decoded frames of the proposed method and the conventional method, respectively.

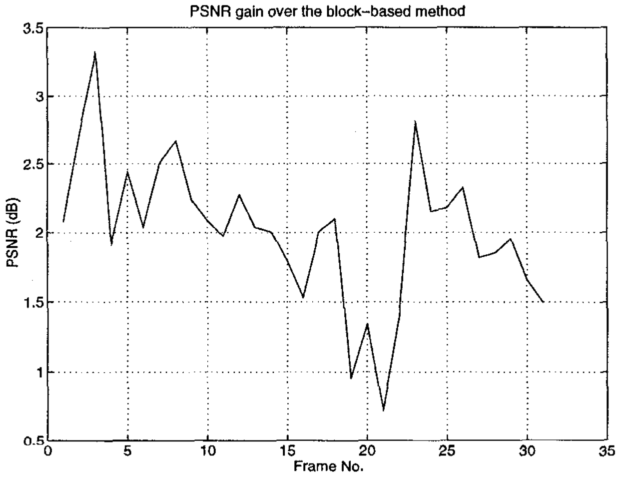

Figure 2: PSNR gain of the proposed method over the block-based method. Test Sequence is "Claire".

6 Conclusion

The key ideas which aid the proposed scheme are: (a) we use multiscale segmentation information and

Figure 3: Proposed Region-Based Method : "Miss America" (149th decoded frame).

Figure 4: Conventional Block-Based Method: "Miss America" (149th decoded frame).

select an optimal scale of segmentation from a compression view point, (b) a novel method is introduced to deal with occluded regions which normally degrade the performance of region based techniques, (c) we employ pel recursion and linear prediction methods to fine tune our region estimation, and (d) we perform region segmentation on the *previously decoded frame* (so we do not need to encode any segmentation information). A fusion of these important ideas leads to a gain of about $2 \sim 3$ dB in PSNR over the block matching algorithm for a variety of test sequences using a fully functional video coder (when the bit rate is constrained to be the same for both schemes).

References

[1] P. Salembier, L. Torres, F. Meyer, and C. Gu. Region-based video coding using mathematical morphology. *Proceedings of the IEEE*, Vol. 83(No. 6):843-857, June 1995.

[2] Y. Yokoyama, Y. Miyamoto, and M. Ohta. Very low bit rate video coding using arbitrarily shaped region-based motion compensation. *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 5(No. 6):500:507, December 1995.

[3] N. Ahuja. A transform for mutiscale image segmentation by integrated edge and region detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 18(No. 12):1211-1235, December 1996.

[4] J. Biemond, L. Looijenga, D.E. Boekee. A pel-recursive Wiener-based displacement estimation algorithm for interframe image coding applications. *SPIE Visual Communications and Image Processing II*, Vol. 845:424-431, 1987.

[5] A.N. Nosratinia and M.T. Orchard. New pel-recursive motion estimation algorithms based on novel interpolation kernels. *Proc. of SPIE Conf. Visual Communications and Image Processing '92*, Vol. 1818:85-96, November 1992.

[6] R.D. Brandt and M.T. Orchard. Motion field modeling for motion estimation and compensation. *Proc. of the 36th Midwest Symposium on Circuits and Systems*, Vol. 1:249-252, August 1993.