A MOTION ESTIMATION AND IMAGE SEGMENTATION TECHNIQUE BASED ON THE VARIABLE BLOCK SIZE

Hangu Yeo *Yu Hen Hu*

Department of Electrical and Computer Engineering
University of Wisconsin, Madison WI 53706

ABSTRACT

In this paper, we discuss our effort to develop a motion estimation algorithm based on *variable block size*, which reduces computational complexity dramatically while maintaining good picture quality as well as high compression ratio. A key step in this work is to segment each image frame into different regions using a simple binary-level classifier which performs bit-wise comparison. In the second stage, the motion estimation is performed for every block of variable block size within the changed region with a predetermined maximum search range. The proposed technique has been applied to interframe video coding, and it has been shown that this scheme can be a feasible solution for the low bit rate coding application such as video telephony.

1. INTRODUCTION

The image compression techniques rely on two statistical redundandies: *spatial* redundancy and *temporal* redundancy. Coding techniques which exploit the spatial redundancy are referred to as *intraframe* techniques; while *interframe* coding techniques exploit the temporal redundancy. These statistical dependencies can be utilized with a block based hybrid coding structure with a DPCM loop and a discrete cosine transform of the prediction error. Central to most video compression algorithms is the task of motion estimation which removes the temporal redundancy of the video sequence, and the energy in the prediction error is much less than that in the original frame.

Block matching algorithm (BMA) [1] is the most popular motion estimation algorithm which divides an image frame into *macroblocks* of size $N \times N$ and matches each macroblock in an intermediate image frame to the most similar macroblock within a search range in the reference image frame assuming that all the pixels within a macroblock have uniform translational motion. The BMA is the most popular one, because it achieves high picture quality, and has regular formulation which allows straightforward VLSI implementation.

However, the conventional BMA based on the fixed size block still suffers from the following pitfalls, which may decrease the total compression ratio.

1. There is a trade-off between the block size and picture quality as described in figure 1. However, the number of motion vectors that needs to be transmitted is increasing as the block size decreases.

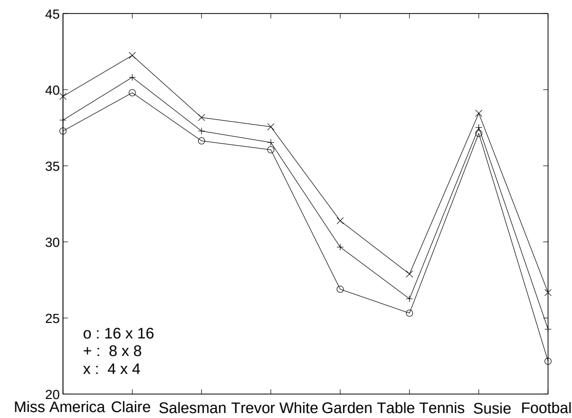

Figure 1. The trade-off between the block size and the picture quality.

2. Motion vectors are coded using a technique called DPCM. However, the motion vector of a block is not well correlated with the motion vector of the neighboring blocks, i.e., the BMA may produce a noisy motion field for the static background as depicted in figure 2.

In this paper, we propose a motion estimation and image segmentation technique based on the variable block size [2]-[3]. The image frames are segmented into *changed* and *unchanged* regions using a simple classifier. Since the human visual system focuses on the moving objects (changed region), we focused on minimizing the prediction error of the changed region.

2. IMAGE SEGMENTATION WITH A SIMPLE BINARY-LEVEL CLASSIFIER

The image segmentation procedure is composed of the following two stages.

1. In the first stage, the image frame is split into two regions, i.e., changed (moving objects) and unchanged (static background) regions using a binary-level matching criterion [4] as a classifier.

   (a) **Binary-level matching criterion:** The binary-level matching criterion (1) classifies a macroblock of size $N \times N$ into changed or unchanged block.

$$BMC = \sum_{i,j=0}^{N-1} \textbf{AND}\{\textbf{XNOR}(x(i,j), y(i,j))\} \quad (1)$$

$x(i,j)$ and $y(i,j)$ are four most significant bits of luminance values of current and previous pixels, respectively. It compares two macroblocks at the same position within the two consecutive image frames, and counts the number of *matching pixels*. A pixel is called a matching pixel if the four most significant bits of $x(i,j)$ and $y(i,j)$ are same, i.e., $\mathbf{AND}\{\mathbf{XNOR}(x(i,j),, y(i,j))\} = 1$.

   (b) **Unchanged region:** A macroblock is classified as *unchanged* if the number of matching pixels within a macroblock (BMC) is greater than a certain threshold (τ_s), i.e., $\tau_s < BMC < N^2$. The block matching procedure for the macroblocks within the changed region are being skipped, and the corresponding motion vectors are set to zero.

   (c) **Changed region:** In the same way, a macroblock is classified as *changed* if the number of matching pixels within a macroblock is less than or equal to τ_s, i.e., $0 \leq BMC \leq \tau_s$. The macroblocks within the changed region need motion estimation based on variable block size or intra-coded in the second stage.

2. In the second stage, the macroblocks within the changed region are subdivided into smaller blocks according to the number of matching pixels as follow.

   (a) $\tau < BMC \leq \tau_s$: Macroblock of size 16×16.

   (b) $\tau_i < BMC \leq \tau$: Subdivide the macroblock into blocks of size 8×8.

   (c) $0 \leq BMC \leq \tau_i$: Intra-coded without motion estimation.

   After the macroblocks within the changed region are subdivided into variable size blocks, the block matching motion estimation is performed for every block with a predetermined search range.

3. SIMULATION RESULTS

In this section, the performance of the proposed motion estimation algorithm based on variable block size has been compared to that of the conventional full search block matching algorithm (FBMA) using the syntax in MPEG-1. Figure 3 depicts the image segmentation results using the same thresholds τ, τ_i, and τ_s between frames 100 and 103 of the sequence "Miss America", frames 100 and 103 of the sequence "Claire", frames 099 and 102 of the sequence "Susie", and frames 033 and 036 of the sequence "Football", respectively.

In table 1 and table 2, we compared the number of macroblocks within each regions using two different thresholds τ_1 and τ_2 where $\tau_2 < \tau_1$ while maintaining same thresholds τ_i and τ_s. Three image sequences, "Miss America (000-119)", "Claire (000-119)", and "Football (000-059)" have been simulated using same thresholds τ, τ_i, and τ_s. It has been shown that the proposed technique skips upto 95% of macroblocks without motion estimation, which may save computational efforts dramatically when compared to the conventional FBMA. Obviously, smaller threshold τ results in smaller number of blocks of size 8×8 within the changed region.

The peak signal to noise ratio (PSNR) in (2) between the predicted frame ($\tilde{x}$) and its original frame (x) is illustrated in table 3.

$$PSNR = 10 \log_{10}\{\frac{255^2}{\frac{1}{MN}\sum_{i=0}^{M-1}\sum_{i=0}^{N-1}(x(i,j) - \tilde{x}(i,j))^2}\} \quad (2)$$

We compared the performance of motion estimation algorithm based on variable block size to that based on fixed block size using same image segmentation technique. For both techniques, the motion field was generated using a maximum search range of ± 16, and block size of 16×16 was chosen in the fixed block size case. Table 4 compares total compression ratio between the two techniques; the conventional FBMA and our proposed technique based on variable block size.

As it can be seen from table 3 and table 4, our proposed technique based on variable block size outperforms the technique based on fixed block size and conventional FBMA in both the PSNR measure and total compression ratio for image sequences such as "Miss America" and "Claire" which do not involve complex motion. However, the total compression ratio decreases for image sequence "Football" which involves complex motion because of the side information for motion vectors and intra-coded blocks. Figure 4 compares the performance of the proposed technique to the conventional FBMA using "Susie" image sequence. The proposed technique outperforms the conventional FBMA within the changed region around the eye area which involves complex motion.

4. CONCLUSION

In this paper, a motion estimation and segmentation technique based on the variable block size has been proposed and simulated. Compared with the existing image segmentation techniques, it could save the computational cost dramatically using a binary-level matching criterion as a classifier. Furthermore, the proposed algorithm allows an improvement of the image quality while achieving higher total compression ratio. The simulation results indicate that this technique can be a feasible solution for the low bit rate coding application such as video telephony.

REFERENCES

[1] H. G. Musmann, P. Pirsch, and H. J. Grallert, "Advances in Picture Coding," *Proc. IEEE*, Vol. 73, pp. 523-548, April 1985.

[2] A. Puri, H. M. Hang, and D. L. Schilling, "Interframe Coding with Variable Block-size Motion Compensation," *GLOBECOM'87* pp. 2.7.1-2.7.3, November 1987.

[3] M. H. Chan, Y. B. Yu, and A. G. Constantinides, "Variable Size Block Matching Motion Compensation with Applications to Video Coding," *Proc. IEE*, Vol. 137, No. 4, pp. 205-212, August 1990.

[4] H. Yeo and Y. H. Hu, "A Novel Matching Criterion and Low Power Architecture for Real-Time Block Based Motion Estimation," will appear in ASAP'96, August 1996.

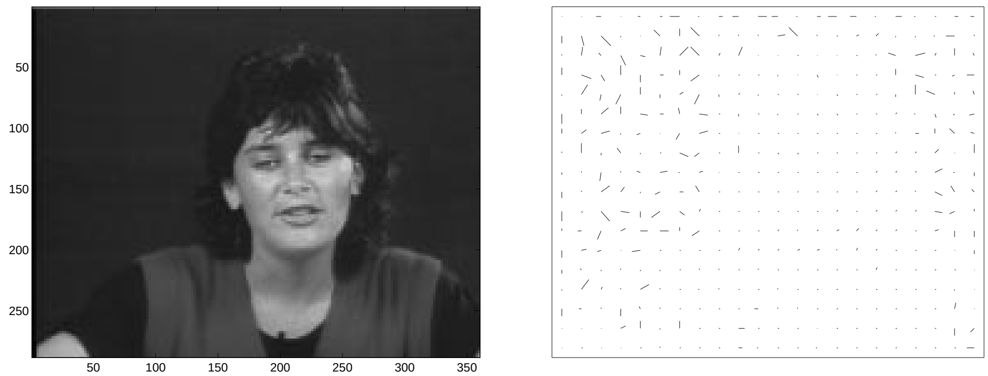

Figure 2. The motion field corresponding to "Miss America" image sequence.

Table 1. Comparison: Number of macroblocks within each region with two different thresholds where $\tau_2 < \tau_1$.

Image Sequence	P Frame											
	Variable Block Size: Case I ($\tau = \tau_1$)						Variable Block Size: Case II ($\tau = \tau_2$)					
	Unchanged		Changed				Unchanged		Changed			
	Skipped	%	16×16	8×8	Intra	Total	Skipped	%	16×16	8×8	Intra	Total
Miss America	12,310	86	916	1,030	0	1,940	12,310	86	1,646	299	1	1,946
Claire	13,007	91	716	533	0	1,249	13,007	91	1,055	193	1	1,249
Football	5,973	25	5,855	11,923	549	18,327	5,973	25	9,327	8,035	965	18,327

Table 2. Comparison: Number of macroblocks within each region with two different thresholds where $\tau_2 < \tau_1$.

Image Sequence	B Frame											
	Variable Block Size: Case I ($\tau = \tau_1$)						Variable Block Size: Case II ($\tau = \tau_2$)					
	Unchanged		Changed				Unchanged		Changed			
	Skipped	%	16×16	8×8	Intra	Total	Skipped	%	16×16	8×8	Intra	Total
Miss America	26,330	92	479	1,703	0	2,182	26,318	92	1,105	1,089	0	2,194
Claire	27,183	95	378	951	0	1,329	27,199	95	616	697	0	1,313
Football	19,738	41	4,441	24,126	295	28,567	19,444	40	9,913	18,921	322	29,156

Table 3. Comparison: PSNR (dB) with two different thresholds where $\tau_2 < \tau_1$.

Image sequence	Variable Block Size							
	P Frame				B Frame			
	Case I ($\tau = \tau_1$)		Case II ($\tau = \tau_2$)		Case I ($\tau = \tau_1$)		Case II ($\tau = \tau_2$)	
	Variable	Fixed	Variable	Fixed	Variable	Fixed	Variable	Fixed
Miss America	38.2	37.4	37.7	37.0	37.3	37.4	37.0	37.0
Claire	39.9	39.1	39.4	39.1	42.4	42.4	42.4	42.4
Football	21.8	18.6	21.5	18.6	27.0	25.2	26.5	25.2

Table 4. Comparison: Total Compression Ratio with two different thresholds where $\tau_2 < \tau_1$.

Image sequence	FBMA	Variable Block Size			
		Case I ($\tau = \tau_1$)		Case II ($\tau = \tau_2$)	
		Variable	Fixed	Variable	Fixed
Miss America	288	342	361	359	361
Claire	383	387	395	394	395
Football	61	54	62	55	62

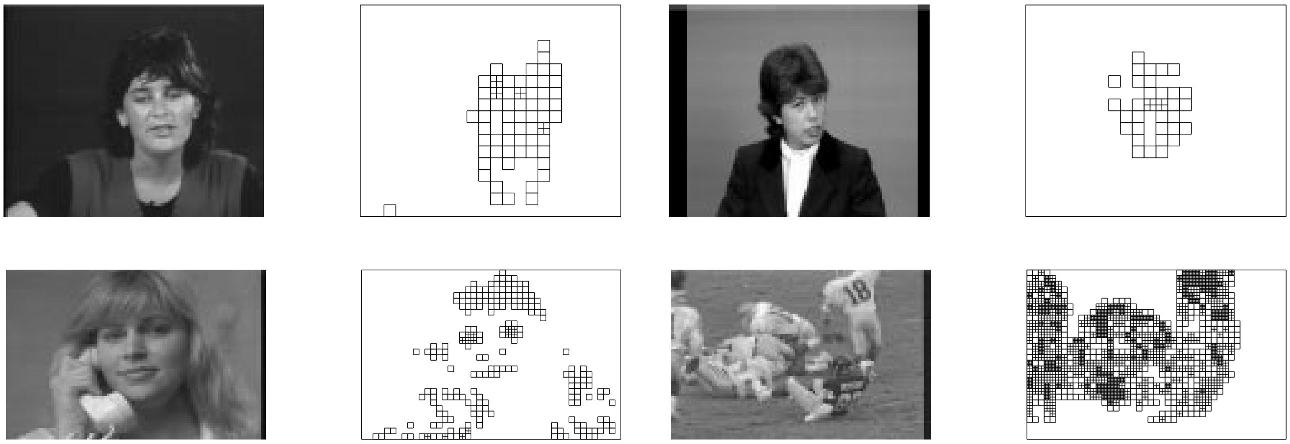

Figure 3. Image segmentation based on the proposed binary-level classifier.

Figure 4. The proposed technique outperforms the conventional BMA within the changed region around the eye area which involves complex motion between two consecutive image frames. (a) Previous frame. (b) Current frame. (c) Fixed block size. (d) Variable block size.