The 2004 International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC2004)
Hotel Taikanso, Sendai/Matsushima, Miyagi-Pref., JAPAN July 6-8, 2004

A New Statistically Adaptive Block-Matching Criterion for Motion Estimation

Seung-Hwan Kim, Hong Kook Kim and Yo-Sung Ho

Gwangju Institute of Science and Technology (GIST)
1 Oryong-dong Buk-gu, Gwangju, 500-712, Korea
{kshkim, hongkook, hoyo}@gist.ac.kr

Abstract: In this paper, we address the issue of motion estimation in real video sequences. While the conventional block-matching criteria for motion estimation is based on the assumption that the average values of the target block and reference blocks are identical. This causes performance degradation for motion estimation because average values are time-varying in real video sequences. In order to alleviate this problem, we propose a new statistically adaptive block-matching criterion (SABMC). In SABMC, the motion vector is estimated by considering time-varying average and shape that represents the fluctuation of pixel values separately, because they are of different importance. Simulation results demonstrate that the proposed SABMC criterion provides higher PSNR values than conventional block-matching criteria at various bit rates.

1. Introduction

For video coding, motion estimation and motion compensation are very effective in removing temporal redundancy in video sequences. Motion estimation is defined as searching the best motion vector, which is the coordinate of the most similar block in the previous frame for the target block in the current frame. In various approaches for motion estimation, the block-matching algorithm (BMA) is very popular. Block-based matching algorithms find the optimal motion vector, which provides the most similarity between the target block and reference blocks. Therefore, the same motion vector is used for all pixels within the block [1-3].

In order to find the optimal reference block that guarantees the most similarity from the target block, we have used the mean absolute difference (MAD) or mean square error (MSE) measure as a matching criterion. However, these conventional block-matching criteria are designed in the assumption that the average values of the target block and reference blocks are the same. Therefore, for video sequences with average variations between the target block and reference blocks, a motion estimation method considering average variations between blocks is needed.

In order to overcome the average variation problem in the conventional BMAs, we propose a new matching criterion called statistically adaptive block-matching criterion (SABMC). In order to improve the coding efficiency, we search the best matching block by considering the statistical distribution of pixel values and average values between the target block and each reference block.

The efficiency of SABMC is verified by comparing the performance of SABMC and that of conventional BMAs in terms of objective and subjective image quality.

2. Matching Criteria for Motion Estimation

2.1 Conventional Matching Criteria

One of the popular motion estimation techniques is the block-matching algorithm [4-7]. In BMA, motion estimation and motion compensation are performed on the block basis, assuming that all the pixels within the block experience the same transnational movement. BMA is widely used in practical applications because of its effectiveness and simplicity for hardware implementation [4-7]. Figure 1 graphically illustrates BMA.

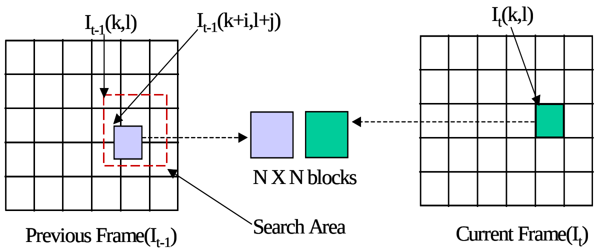

Figure 1. Block-Matching Algorithm

In Figure 1, the dotted line in the previous frame represents the search range. One of the blocks in the search range is selected to estimate the target block in the current frame. The performance of BMA depends on the search algorithm and the matching criterion. The search algorithm is related to the speed while the matching criterion is to the accuracy defined as the distance between the target block and the selected reference block in the search area. One of the popular matching criteria is MSE.

$$MSE(i,j) = \frac{1}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N}\left[I_t(k,l) - I_{t-1}(k+i,l+j)\right]^2 \quad (1)$$

where I_t and I_{t-1} are pixels values in the current and previous frames, respectively. i and j are defined in the range of the predefined search range. Since MSE requires squaring operations, it is a heavy burden for real-time MPEG-4 encoders. In order to reduce computational complexity of MSE, we can use MAD.

$$MAD(i,j) = \frac{1}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N}\left|I_t(k,l) - I_{t-1}(k+i,l+j)\right| \quad (2)$$

2.2 Problems in MSE and MAD

We have defined the cost functions of conventional matching criteria, MSE and MAD. They have been considered as optimal solutions to find the best matching block in terms of numerical similarity. However, it is not always correct for motion estimation to select a numerically similar reference block. The main objective of motion estimation is to compress the video data by reducing temporal redundancy between adjacent frames. Therefore, we should find a matching criterion that provides improved coding efficiency.

Figure 2 explains the problem in the traditional matching criteria. In Figure 2, Current MB is the target block in the current frame and Previous MB1 and Previous MB2 are reference blocks in the previous frame to predict the target block. In the motion estimation process, we find the best matching block using the given matching criterion and calculate the residual block between Current MB and the selected reference block. In the next step, this residual block is discrete cosine transform (DCT), zigzag scanned, variable length coded, and transmitted.

If we use the traditional MAD or MSE as the matching criterion, Previous MB2 will be selected as the reference block because SAD2 and SSD2 are smaller than SAD1 and SSD1. Therefore, we encode Residual MB2. However, if we select Previous MB1, we encode Residual MB1. Intuitively, Residual MB1 is simpler than Residual MB2 in the DCT domain because Residual MB1 has only one component (DC value) in the DCT domain.

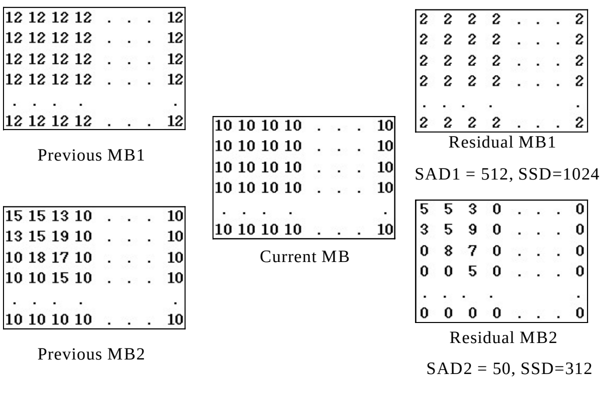

Figure 2. Problems in MAD and MSE

Conventional matching criteria have the average-disparity problem because they only consider numerical similarity to find the reference block. Previous MB1 has a larger different average value than Previous MB2 from Current MB. However, the amount of the required information of the average value (DC) is very small in the DCT domain. When we remove the effect of average values, Previous MB1 is closer than Previous MB2 to Current MB. In other words, Previous MB1 is more similar to Current MB in terms of the statistical distribution of corresponding pixel values. Figure 2 seems to represent the very extreme case. However, the average-disparity problem between blocks usually occurs in real video sequence.

3. A New Matching Criterion

In order to solve the problem mentioned in the previous section, we propose a new matching criterion, called statistically adapted block-matching criterion, which is based on the traditional MSE and MAD. However, SABMC considers the shape that represents the statistical similarity of the intensity distribution between blocks, and average between blocks separately. Figure 3 explains the meaning of the shape and the average.

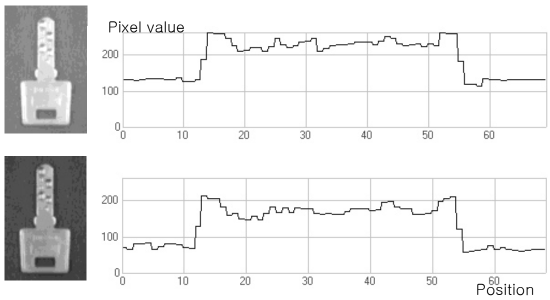

Figure 3. Intensity Distribution at Given Position

In Figure 3, two images are almost the same in the intensity distribution. We regard the distribution form of pixel values in the given position as the shape. In Figure 3, two images have different average values. The lower image in Figure 3 has the lower average value than the above image. With the traditional matching criteria, we may decide that two images are very different. However, in terms of coding in DCT domain, these two images are almost the same because the difference can be recovered by only one DCT coefficient. In other words, the traditional matching criteria do not consider the effect of average differences between blocks.

In order to overcome the average disparity problem, we have to extract the shape and average components from the traditional matching criteria cost functions. We normalize each block to remove the effect of average differences between blocks.

$$I_t(k,l) = \sigma_t \cdot I_t^N(k,l) \qquad (3)$$

where $I_t^N(k, l)$ is the normalized value of $I_t(k, l)$, σ_t is the constant value produced in the normalization process. Using Eq. (3), we can obtain the normalized MSE by

$$NMSE(i,j) = \frac{1}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N}\sigma_t\left(I_t^N(k,l) - \frac{\sigma_{t-1}}{\sigma_t}\cdot I_{t-1}^N(k,l)\right)^2 \qquad (4)$$

NMSE seems to solve the average-disparity problem in MSE. However, NMSE does not provide better performance than MSE because it only considers the shape between two blocks, but not the average. Therefore, we need to consider both the shape and average. In order to consider shape and average separately, we divide each pixel value into the shape and average terms.

$$I_t(k,l) = I_t^{'}(k,l) + \overline{I_t} \qquad (5)$$

where $I_t'(k, l)$ represents the shape term and $\overline{I_t}$ represents the average term in the block. We give different weighting to the shape and the average terms according to their importance. As a result, we can obtain SAMSE by combining Eq. (1) and Eq. (5).

$$SAMSE(i,j) = \frac{1}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N}\alpha(I_t^{'}(k,l) - I_{t-1}^{'}(k+i,l+j))^2 + (1-\alpha)(\overline{I_t} - \overline{I_{t-1}})^2 \qquad (6)$$

By controlling the value of α, we can consider the importance of the shape and the average terms. The traditional MAD is also changed to SAMAD with the same procedures of inducing SAMSE. As a result, SAMAD can be represented by two terms as

$$SAMAD(i,j) = \frac{1}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N}\alpha\,|I_t^{'}(k,l) - I_{t-1}^{'}(k+i,l+j)| + (1-\alpha)\,|\overline{I_t} - \overline{I_{t-1}}| \qquad (7)$$

In order to apply SAMAD to the practical encoding system, we change the weighting factor α considering the importance of the shape and average terms, respectively. We fix the weighting factor of the shape term and change the value of the average term. Therefore, the practical implementation of SAMAD can be expressed by

$$SAMAD(i,j) \approx \frac{1}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N}|I_t - I_{t-1}| + \delta\,|\overline{I_t} - \overline{I_{t-1}}| \qquad (8)$$

where δ is the weighting factor of the average term. In Eq. (8), δ is equal to $(1-\alpha)/\alpha$ in Eq. (7). By Eq. (8), we can effectively compare the measure of matching between blocks, considering importance of the shape and average separately.

In Eq (8), we use the value of δ to give different weighting in the shape and average terms. From the experimental results, we can verify that the value of δ depends on the amount of motion of test image sequences. In general, image sequences, which have complex and large amount of motion, require large value of δ. From experimental results, we observe that the value of δ is optimized in the scope of 0.1 and 0.3. As a result, proposed new matching criterion technique is more efficient than traditional matching criterion and our algorithm can be directly applied to the MPEG-4 video encoder.

4. Experimental Results

In order to evaluate the performance of the proposed matching criterion, we implement the MPEG-4 simple profile encoder. We use various video sequences, such as FOREMAN, NEWS, MOBILE, and COASTGUARD. The resolution of those sequences is 352X288 pixels (CIF). We also use the full search algorithm to prove the efficiency of the proposed matching criterion.

Table 1 shows PSNR values at different bit rates. From Table 1, we can observe that the proposed method provides approximately 0.5dB higher PSNR values than the traditional MAD.

Table 1. Average PSNR Values

Method/ Bitrate \ Frame		100	150	200
SAMAD	306.4	32.72	32.86	32.33
	311.2	32.79	32.87	32.40
	320.0	32.95	32.98	32.53
	333.6	33.02	33.08	32.60
MAD	300.0	31.96	32.10	31.75
	328.0	32.51	32.82	32.35
	334.4	32.97	33.12	32.57

δ=0.16

In Figure 4, MAD uses 328 kbps to encode the video data, and SAMAD uses 320 kbps. However, SAMAD provides higher PSNR values than MAD.

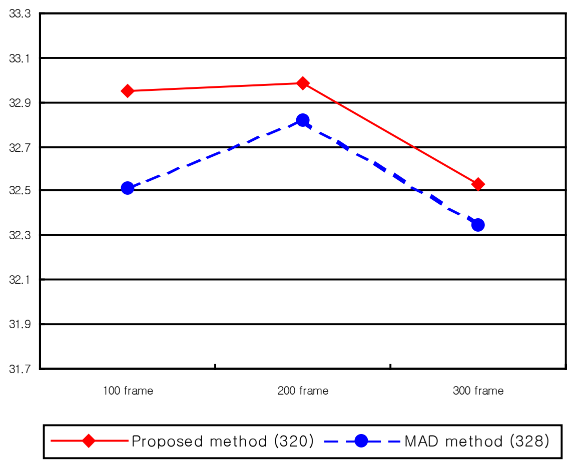

Figure 4. PSNR Comparison

In Figure 5, we can compare subjective image quality of the two matching criteria. The difference is highlighted in near the mouth and ears in the FOREMAN 51st frame, and eyes and the mouth in the FOREMAN 65th frame. In general, eyes, ears, and the mouth are very sensitive regions that require very detailed information and also have a large amount of movement. From Figure 5, we can observe that SABMC provides better performance in complex and large motion areas.

(a): 51st, MAD (b): 65th, MAD (c) 93rd, MAD

(d): 51st, SAMAD (e): 65th, SAMAD (f) 93rd, SAMAD

Figure 5. Subjective Quality Comparison

We apply SAMAD to several test image sequences in the extended search range of 15 pixels. We also use average PSNR values to compare the performance between MAD and SAMAD.

Table 2 shows that four test image sequences have different characteristics. NEWS represents the characteristics of small motion image sequences, and MOBILE and COASTGUARD represent those of large motion image sequences. The human face in FOREMAN is the region of interest (ROI) that is very important in video coding. Table 2 shows that the performance of SAMAD is better than MAD for the test video sequences.

Table 2. Average PSNR Values for Test Images

Images / Method	FOREMAN	MOBILE	COASTGUARD	NEWS
MAD	30.96	23.43	27.61	32.02
SAMAD	31.70	23.55	28.01	32.17

5. Conclusions

In this paper, we have proposed a statistically adaptive block-matching criterion (SABMC). In SABMC, the shape and the average terms of each pixel are considered separately. By applying SABMC, we find SAMSE from MSE, and SAMAD from MAD. Experimental results show that SABMC provides higher PSNR values and improved subjective image quality than the traditional matching criteria at different bit rates.

Acknowledgements

This work was supported in part by Gwangju Institute of Science and Technology (GIST), in part by the Ministry of Information and Communication (MIC) through the Realistic Broadcasting Research Center (RBRC) at GIST, and in part by the Ministry of Education (MOE) through the Brain Korea 21 (BK21) project.

References

[1] J. Zan, M. Omair, and M.N.S. Swamy, “Multiplicationless Burt and Adelson’s pyramids for motion estimation,” IEEE Trans. on Circuits and System for Video Tech., vol. 14, pp. 136-141, Jan. 2004.

[2] ITU-T Rec. H.263, Video coding for low bit rate communication, March 1996.

[3] ISO/IEC DC 14496-2, Information technology coding of audio visual objects: visual, 1997.

[4] T. Koga, K. Linuma, A. Hirano, Y. Iijima, and T. lshiguro, “Motion compensated interframe coding for video conferencing,” Proceedings of the National Conference on Telecommunication Technology, Nov. 29-Dec. 3, 1981.

[5] Y.L. Chan and W.C. Siu, “An efficient search strategy for block motion estimation using image features,” IEEE Trans. on Image Processing, pp. 1-16, Aug. 2001.

[6] Y.L. Chan and W.C. Siu, “Edge oriented block motion estimation for video coding,” IEE Proceedings: Vision, Image and Signal Processing, vol. 144, pp. 136-144, June 1997.

[7] J.Y. Tham, S. Ranganath, M. Ranganath, and A.A. Kassim, “A novel unrestricted center-biased diamond search algorithm for block motion estimation,” IEEE Trans. on Circuits and System for Video Tech., vol. 8, pp. 369-377, Aug. 1998.