A HYBRID MOTION ESTIMATION APPROACH BASED ON NORMALIZED CROSS CORRELATION FOR VIDEO COMPRESSION

Wei-Hau Pan, Shou-Der Wei and Shang-Hong Lai

Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan
{whpan, greco, lai}@cs.nthu.edu.tw

ABSTRACT

In this paper we propose a new hybrid approach for block based motion estimation (ME) by adaptively using the normalized cross correlation (NCC) and sum of absolute differences (SAD) measures. We use the SAD value and gradient sum as the criterion to determine which similarity measure to be used for motion estimation. In general, using the NCC as the similarity measure in the motion estimation leads to more uniform residuals than those of using the SAD. This leads to larger DC terms and smaller AC terms, which yields less information loss after DCT quantization. However, NCC is not suitable for homogeneous regions since the best match may have a high NCC value but with large average gray level difference. Thus, we propose to alternatively use the SAD and NCC as the ME criterion for homogeneous and inhomogeneous blocks. Experimental results show the proposed hybrid motion estimation algorithms can provide superior PSNR and SSIM values than the traditional SAD-based ME method.

Index Terms—Motion estimation, normalized cross correlation, SSIM

1. INTRODUCTION

Motion estimation (ME) is widely used in many applications related to computer vision and image processing, such as object tracking, object detection, pattern recognition and video compression, etc. Especially, block-based motion estimation is very essential for motion-compensated video compression, since it reduces the data redundancy between frames to achieve high compression ratio. Many block-based ME algorithms have been proposed in the past. All of the block-based motion estimation algorithms were developed for finding the block with the smallest matching error. In terms of block distortion measure, the sum of absolute difference (SAD) is commonly used and it is defined by

$$SAD_{x,y}(u,v) = \sum_{j=0}^{N-1}\sum_{i=0}^{N-1}\left|I_t(x+i,y+j) - I_{t-1}(x+u+i,y+v+j)\right| \quad (1)$$

where the block size is *NxN*, *(u,v)* is the motion vector, and I_t and I_{t-1} denote the current and reference images, respectively.

In addition to SAD and SSD, the NCC is also a popular similarity measure. The NCC measure is more robust than SAD and SSD under uniform illumination changes, so it has been widely used in object recognition and industrial inspection. The definition of NCC is given by

$$NCC(x,y) = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} I(x+i,y+j)\cdot T(i,j)}{\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N} I(x+i,y+j)^2}\cdot\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N} T(i,j)^2}} \quad (2)$$

In this paper, we propose two hybrid block-based motion estimation methods by using NCC as the similarity measure. Our experimental results show that the proposed algorithms provides superior PSNR and SSIM values compared to the traditional SAD-based ME. The rest of this paper is organized as follow: we first describe the reason of combining the similarity measure of NCC and SAD for ME in section 2. The SSIM is adopted as the video quality measure and is briefly reviewed in section 3. The experimental results are given in section 4. Finally, we conclude this paper in the last section.

2. HYBRID MOTION ESTIMATION WITH NCC AND SAD SIMILARITY MEASURES

The NCC is more robust than the SAD, especially under uniform illumination change. If we apply the NCC as the similarity measure in motion estimation, we can obtain more uniform residuals between the current MB and the best MB. Here is an example of applying the NCC and SAD measures for ME on the MB (the block 8-by-8 square) to the frame shown in Figure 1 and the corresponding residuals are depicted in Figure 2. Because the SAD is to find the best match with the lowest matching error and the NCC is to find the best MB whose overall intensity variations is most similar to the current MB, the error of SAD (539) is less than that of NCC (623). Although the error of NCC is larger than SAD, the residuals obtained by using NCC for ME are more uniformly distributed than those obtained with SAD as shown in Figure 2. Thus, we obtain larger DC terms and smaller AC terms after DCT, which leads to less information loss after the DCT quantization and better quality of the reconstructed frame.

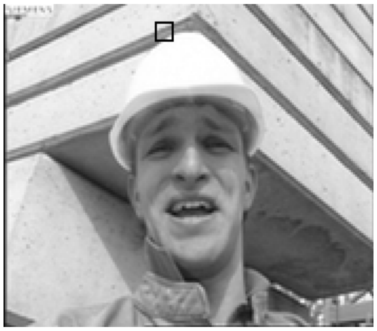


Figure 1: An example MB in frame 58 of Forman sequence.

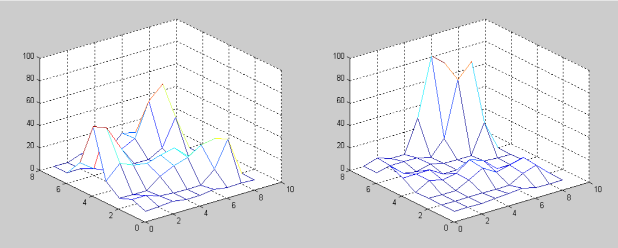
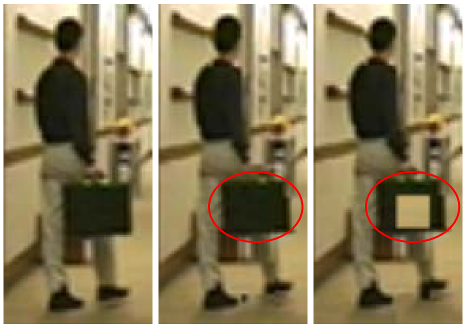


Figure 2: The residuals between the current MB and best MB in Figure 1. The best MBs are determined by using (a) NCC and (b) SAD as the matching criterion. The SAD values for (a) and (b) are 623 and 539, respectively.

Although the NCC is a more robust similarity measure than SAD, but, for a flat MB, using NCC as the matching criterion may find a wrong flat candidate with large different average gray level as the best match. Figure 3 shows the motion compensation result by applying SAD and NCC as the matching measures, respectively. In Figure 3(c), the best match for the black MB of the suitcase by using the NCC matching measure has large differences in the intensities.

In this paper, we proposed two hybrid motion estimation algorithms to achieve high image quality of the reconstructed frame. The first method, called gradient-thresholding, is to use the sum of gradient magnitudes in the macroblcok to determine which matching criterion for motion estimation. If it is greater than a predefined threshold, we use the NCC as the matching criterion, otherwise we use the SAD instead. The second method, i.e. the SAD-thresholding method, first applies the SAD as the matching measure for each macroblock to find the best match. If the SAD value of the best match exceeds a predefined threshold, then we set the NCC as the similarity measure and apply the full search motion estimation again. Figure 4 shows the pseudo-code of these two proposed methods.



(a) (b) (c)

Figure 3: (a) A frame of the test sequence. The motion compensation results by using (b) SAD and (c) NCC as the matching criterion for ME.

```
Algorithm1: Gradient-thresholding
    For each macroblock
    1. Calculate gradient_sum for the macroblock
    2. If (gradient_sum >T) set NCC as similarity measure
       else set SAD as similarity measure
    3. Apply full search block-based ME.
    4. End
```

```
Algorithm2: SAD-thresholding
    For each macroblock
    1. Set SAD as similarity measure
    2. Apply Full search block-based ME.
    3. If ($SAD_{best}$< T) goto end
       else set NCC as similarity measure
    4. Apply Full search block-based ME.
    5. end
```

Figure 4: the pseudo-code of two proposed algorithms.

3. VIDEO QUALITY ASSESSMENT

The peak signal-to-noise ratio (PSNR) is a traditional image quality measure, which computes the difference between two input signals by mean-square error. It is defined as follows:

$$PSNR = 10 \times \log(\frac{MAX_I^{\,2}}{MSE}), \tag{3}$$

where MAX_I is the maximum intensity value of the signal, MSE is the mean-square error of the two given signals. However, previous research [10] argues that PSNR may not well represent the perceptual quality evaluated by human perception since PSNR only considers the mean-square error of two given signals.

Recently, Wang et al. [11] proposed a measure based on image structural distortion, called SSIM, which is more consistent with human perception. In [11], the luminance, contrast and structure measures are defined as:

$$l(\mathbf{x},\mathbf{y}) = \frac{2\mu_x\mu_y + C_1}{\mu_x^{\,2} + \mu_y^{\,2} + C_1},\ c(\mathbf{x},\mathbf{y}) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^{\,2} + \sigma_y^{\,2} + C_2},\ s(\mathbf{x},\mathbf{y}) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}, \tag{4}$$

where **x** and **y** are two vectors obtained from the image in the corresponding local windows, μ_x, μ_y are the sample means of **x** and **y**, respectively, σ_x^2 and σ_y^2 are the variances of **x** and **y**, respectively, σ_{xy} is the covariance between **x** and **y**, C_1, C_2 and C_3 are constants. The SSIM is defined by [11]:

$$SSIM(x,y) = l(x,y) \cdot c(x,y) \cdot s(x,y) \quad (5)$$

In this paper, we evaluate the image quality of the reconstructed frames with both the PSNR and SSIM measures to show the superior video compression quality by using the proposed methods.

4. EXPERIMENTAL RESULTS

In our experiments, we used the integral image scheme [7][8][9] to reduce the computation in the denominator of the NCC. We implemented the proposed NCC-based MEs with the gradient-thresholding and SAD-thresholding schemes in JM 11.0 and the platform is equipped with the 2.01GHz AMD64 X2 CPU. The threshold for gradient-thresholding is set to 1500, 2000, and 3500 for QP value 28, 32, and 36, respectively. The threshold for SAD-thresholding is set to 200, 250, and 300 for QP value 28, 32, and 36, respectively. These threshold values are determined empirically. The experiments were carried out on 5 sequences in QCIF format and the total numbers of encoded frames are 300, 300, 300, 150 and 125 for Foreman, M&D, Silent, Suzie and Football, respectively. The search range was set to +/-16 pixels, and only the 8x8 transform was enabled. Since we focused on the image quality obtained with different ME methods, only the intermode 8x8 block type was set to active to perform ME, followed by the 8x8 DCT transform. We compared the two proposed hybrid ME algorithms and the conventional SAD-based ME algorithm with PSNR and SSIM [11] image quality measures.

Table 1 and 2 depict the video quality comparison between the traditional SAD-based ME method and the proposed hybrid methods with different QP values. It is clear that not only the PSNR but also the SSIM of the proposed methods are much better than the traditional SAD-based ME method. They also demonstrated the encoding quality increment is better in more high QP value, since the quantization/dequantization error of NCC-based ME reduces more when QP value is larger.

Table 3 shows the average bitrates between the traditional SAD-based ME method and the proposed methods. The reason of the bitrate increment is due to the larger DC value of the NCC-based ME after DCT. However, the increment is negligible in different QP value considering the outperforming encoding quality.

Table 4 describes the ME encoding time per frame. Our first method has similar computational speed compared to the SAD-based ME, but our second method requires more encoding time than the SAD-based ME since it performs SAD-based ME first and applies NCC-based ME subsequently if needed.

Table 1: COMPARISION OF PSNR

QP = 28			
Sequence	SAD (dB)	Gradient-TH (dB) (Difference)	SAD-TH (dB)(Difference)
Foreman	36.100	0.136	0.113
Silent	36.055	0.119	0.191
Suzie	37.839	0.189	0.095
Football	34.865	0.085	0.069
M&D	37.717	0.098	0.086
Average	36.515	0.125	0.111
QP = 32			
Foreman	33.608	0.186	0.207
Silent	34.197	0.17	0.202
Suzie	35.755	0.225	0.278
Football	32.196	0.065	0.096
M&D	35.739	0.125	0.138
Average	34.299	0.156	0.184
QP = 36			
Foreman	31.720	0.189	0.213
Silent	32.702	0.221	0.236
Suzie	34.273	0.188	0.242
Football	29.954	0.149	0.172
M&D	34.384	0.155	0.094
Average	32.606	0.180	0.192

Table 2: COMPARISION OF SSIM

QP = 28			
Sequence	SAD	Gradient-TH (Difference)	SAD-TH (Difference)
Foreman	0.93815	0.291%	0.208%
Silent	0.93527	0.213%	0.171%
Suzie	0.93471	0.432%	0.334%
Football	0.89938	0.310%	0.205%
M&D	0.94272	0.173%	0.099%
Average	0.930046	0.284%	0.203%
QP = 32			
Foreman	0.90498	0.584%	0.522%
Silent	0.90532	0.463%	0.543%
Suzie	0.90534	0.844%	0.789%
Football	0.83653	0.679%	0.594%
M&D	0.91405	0.385%	0.330%
Average	0.893244	0.591%	0.556%
QP = 36			
Foreman	0.8701	0.793%	0.848%
Silent	0.87565	0.767%	0.839%
Suzie	0.88135	0.804%	0.977%
Football	0.77764	0.778%	1.115%
M&D	0.8866	0.613%	0.693%

Average	0.858268	0.751%	0.894%

Table 3: COMPARISION OF AVERAGE BITRATES

QP	SAD(Kbits)	Gradient-TH (%)	SAD-TH (%)
28	276.336	1.48%	0.86%
32	185.986	4.24%	2.86%
36	145.444	4.20%	4.36%

Table 4: COMPARISION OF AVERAGE ME TIME

QP	SAD(ms)	Gradient-TH (ms)	SAD-TH (ms)
28	388.73	365.53	604.36
32	390.23	367.18	571.77
36	393.15	374.52	546.51

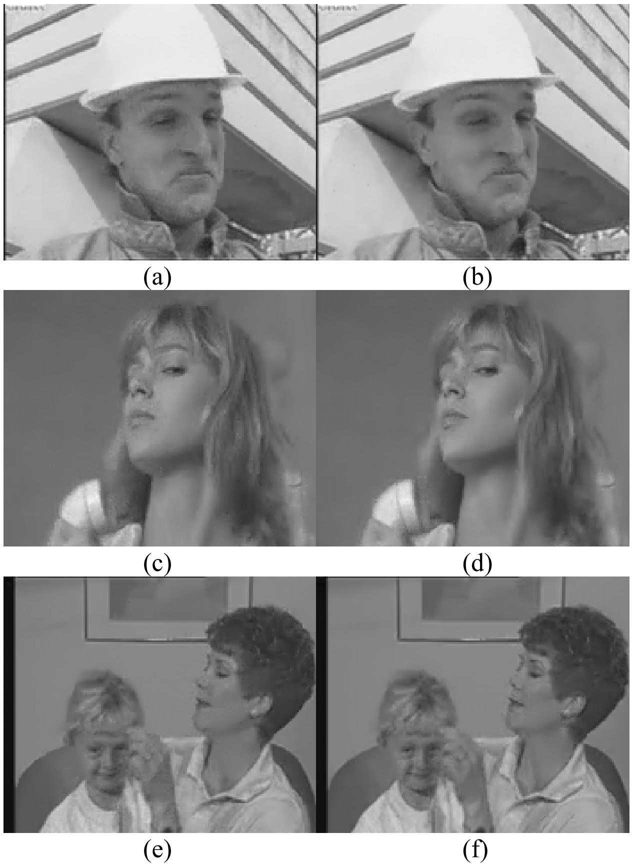

(a) (b) (c) (d) (e) (f)

Figure 5: The reconstructed frames from Foreman, Suzie and MD sequences with QP=28 by using (a)(c)(e) SAD-based ME and (b)(d)(f) gradient-thresholding ME methods.

To demonstrate the superior video quality provided by the proposed hybrid ME methods, the reconstructed frames sampled from three different sequences are depicted in Fig. 5. The PSNR differences between (a)(b), (c)(d) and (e)(f) are 0.39, 0.69, and 0.23, respectively. The SSIM differences between (a)&(b), (c)&(d) and (e)&(f) are 0.71%, 1.05%, and 0.46%, respectively. It is obviously that the proposed methods improve the perceptual video quality.

5. CONCLUSION

In this paper, we proposed two hybrid motion estimation algorithms by combining the NCC and SAD similarity measures. Applying NCC as similarity measure yields more uniform residuals, which leads to larger DC values and smaller AC values and thus better reconstructed frame. However, the SAD is better than NCC as the similarity measure in the homogeneous regions. We determine which similarity measure to use for block matching in each block from its associated SAD value and the local gradient sum. The experimental results show the proposed hybrid ME methods can provide higher PSNR and better SSIM in the reconstructed frame than the traditional SAD-based ME method. In the future, we will develop more efficient NCC search algorithms to reduce the computational cost in the NCC-based ME.

11. REFERENCES

[1] S. Zhu and K.K. Ma, "A new diamond search algorithm for fast block-matching motion estimation," IEEE Trans. Image Processing, Vol. 9, No. 2, pp. 287-290, Feb. 2000.

[2] R. Li, B. Zeng and M.L. Liou, "A new three-step search algorithm for block motion estimation," IEEE Trans. Circuits Systems Video Tech., Vol. 4, No. 4, pp. 438-442, 1994.

[3] L.M. Po and W.C. Ma, "A novel four-step search algorithm for fast block motion estimation," IEEE Trans. Circuits Systems Video Technology, Vol. 6, No. 3, pp. 313-317, 1996.

[4] W. Li and E. Salari, "Successive elimination algorithm for motion estimation," IEEE Trans. Image Processing, Vol. 4, No. 1, pp. 105-107, 1995.

[5] X. Q. Gao, C. J. Duanmu, and C. R. Zou, "A multilevel successive elimination algorithm for block matching motion estimation," IEEE Trans. Image Processing, Vol. 9, no. 3, pp. 501-504, 2000.

[6] C.-H. Lee and L.-H. Chen, "A fast motion estimation algorithm based on the block sum pyramid," IEEE Trans. Image Processing, Vol. 6, No. 11, pp. 1587-1591, 1997.

[7] J.P. Lewis, "Fast template matching," Vision Interface, pp. 120–123, 1995.

[8] M. Mc Donnel, "Box-filtering techniques," Computer Graphics and Image Processing, vol. 17, pp. 65–70, 1981.

[9] P. Viola and M. Jones. Robust real-time object detection. In Proceeding of Workshop on Statistical and Computation Theories of Vision, 2001.

[10] Wang, A.C. Bovik, L. Lu, "Why is image quality assessment so difficult," IEEE Intern. Conf. Acoustics, Speech, and Signal Processing, Orlando, May 2002.

[11] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Trans. Image Processing, vol. 13, no. 4, 2004.