VIDEO CODING BASED ON TRUE MOTION ESTIMATION

EePing Ong, Hua Wang, Ping Xue**

Laboratories for Information Technology, 21 Heng Mui Keng Terrace, Singapore 119613
*School of Electrical & Electronic Engineering, Nanyang Technological University, Singapore 637722
Email: epong@lit.a-star.edu.sg, epxue@ntu.edu.sg

ABSTRACT

In this paper, a new approach for MC-DCT (motion compensated – Discrete Cosine Transform) hybrid video coding based on true motion estimation is proposed. The true motion estimation technique employs the least-median-squares (LMedS) matching criterion in block matching process for motion estimation. The rationale for using such true motion estimator is that at low bit-rates, very few bits are available for coding the motion-compensated DCT errors, and thus the detection of the true underlying motion of the picture block becomes much more important than just trying to find a block-match that satisfies the traditional least-mean-squares (LMS) matching criterion. Without detecting the true motion of the picture block, the errors due to insufficient bits for coding the motion-compensated DCT errors will be propagated to succeeding inter-frames, resulting in a progressive degeneration on perceptual quality of the pictures. An important advantage of using true motion estimation is that it results in better perceptual video quality compared to the conventional block-matching algorithm that utilizes least-mean-squares matching criterion. Experimental results obtained had supported our proposition on using the LMedS-based matching criterion in block-matching motion estimation for low bit-rate video coding.

1. INTRODUCTION

Designing video coders that provide high compression ratio and particularly best achievable video quality represents one of the biggest challenges in the field of image and video processing [1].

Until recently, improvements on the designs of video codec have been focused on optimizing the speed of the block-matching motion estimator which has been used in a hybrid MC-DCT (motion compensated - Discrete Cosine Transform) video codec. Various fast motion estimation techniques (in addition to the tradition full-search block-matching) have been proposed, such as 3-step search, logarithmic search, cross-search, etc [1,8].

Current block-matching motion estimation techniques in hybrid video encoder employs mean squared error (or sometimes the mean absolute error) as the matching criterion to measure the similarity between a block of the current frame and a candidate block of the reference frame. This matching criterion adds up the (absolute or squared) intensity differences of all the pixels in the searching block [1,8] and chooses the motion vector which gives rise to the block's location in the searched image with the smallest summed error as the best match (hence this is in effect using LMS criterion). By using the above approach, video compression can be achieved whereby the smallest least-mean-squares error and hence the best peak signal-to-noise ratio (PSNR) can be attained.

However, one factor that has been overlooked is that at low bit-rates, even with current state-of-the-art, it was observed that current MC-DCT (Motion Compensated Discrete Cosine Transform) hybrid video compression technique does not provide very good perceptual quality. It should also be noted that video quality subjective tests have also shown that PSNR may not correlate well with the perceptual video quality [7], and this is particularly true when coding at low bit-rates.

One of the major problems identified is that the use of mean-squares error (and hence PSNR) is not a good measure of the perceptual quality of compressed video images, particularly at low bit-rates. As such, subjective tests need to be carried out for comparing proposed algorithms and techniques during the MPEG-4 competitive core experiments to determine techniques that will be suitable for inclusion into the MPEG-4 video coding standard [16]. The above problem has also resulted in numerous works on developing an objective perceptual distortion metric [12,13,14,15] to replace the tedious and time-consuming task of performing human subjective tests.

In [10], the adaptive quantization in MPEG2-TM5 has been modified to improve the perceptual picture quality in MPEG2 encoded video sequences. Here, the variance measure in adaptive quantization has been replaced by an 8x8 block-based perceptual activity measure computed from a non-perceived noise sequence which is derived from the output of a spatio-temporal computational model of human visual system by using the original sequence and a uniform noise sequence as its input. In [11], the non-perceived noise sequence has been replaced by a perceptual visibility predictor, which is essentially a similar spatio-temporal computational model of human visual system but using the original sequence and the distorted sequence as the input.

As LMS matching criterion has been used in the block-matching motion estimation in a conventional hybrid video coder, we postulated that such motion estimator may have also contributed significantly to the poor perceptual quality of the low bit-rate compressed video images.

In our previous work on computation of optical flow (dense motion field) [4], it has been shown that with the use of LMedS regression with affine motion model, highly robust and accurate motion field that depicts the true motion of underlying object can be estimated.

Based on these findings, we studied the effect of motion estimation on perceptual quality of video sequences encoded using hybrid video coder and subsequently proposed a new true motion estimation method using LMedS as matching criterion in the block-based motion estimator. Experimental results show that our true motion estimator can enhance perceptual video quality.

2. TRUE MOTION ESTIMATION

Current state-of-the-art MC-DCT hybrid video compression technique does not provide very good perceptual quality at low bit-rates. One of the major problems identified is that the mean-squares error (and hence PSNR) is not a good measure of the perceptual quality of low bit-rate compressed video images. As such, we postulated that such block-matching motion estimator using a LMS matching criterion may have contributed significantly to the poor perceptual quality of the compressed video images at low bit-rates.

To see why this is so, firstly we have to take the note that the block-matching motion estimation technique that uses LMS criterion currently being employed in hybrid video codec (be it full-search, 3-step search, logarithmic search etc) does not really detect the true motion of the block, since it only chooses the motion vector of the searched location that minimizes some function of the displaced frame difference (DFD) (usually the sum of the squared or absolute DFDs) over a block of pixels, where DFD is defined as:

$$DFD(\mathbf{p},\mathbf{d}) = I(\mathbf{p},\mathbf{t}) - I(\mathbf{p}-\mathbf{d},t-1) \qquad (1)$$

where $\mathbf{d}$ is the displacement of a pixel at location $\mathbf{p}=(x,y)^T$ and $I(\mathbf{p},\mathbf{t})$ is the intensity of a pixel at location $\mathbf{p}$ at frame t. Thus for a typical block-matching motion estimation algorithm, either $|DFD|$ or $(DFD)^2$ is summed over all the pixels in a block to obtain a measure of the similarity or closeness of the block-matching.

Effects of noise (especially the presence of spike noise – when CMOS camera is being used to capture the video sequences) and also when the block straddles on motion boundaries (containing 2 different moving regions) will cause this motion estimator to produce such motion vectors which do not really portray the true motion of the single object or the real motion of the majority pixels (if the block straddles on 2 differently moving objects) contained within the block.

When high bit-rates are available, using motion vectors that cannot portray or try to capture the true motion of the block does not pose much problem because the errors are compensated at the decoding stage by information contained in the coding of the motion-compensated residual errors using DCT. However, at low bit-rates, since available bits are not sufficient for more detailed coding of the motion-compensated residual errors using DCT, these errors will propagate through the subsequent P-frames or B-frames, causing the picture quality to degenerate.

Thus, at low bit-rates, it is perceived that the detection of the true underlying motion of the picture block becomes much more important than just trying to find a block-match that satisfies the LMS matching criterion.

Based on this finding, we proposed a new true motion estimation method using LMedS as matching criterion (which will be described in the pursuing section). Experimental results obtained had supported our proposition on using the LMedS-based matching criterion for low bit-rate video coding. An important advantage of using true motion estimation is that it results in better perceptual video quality compared to the conventional block-matching algorithm that employs LMS matching criterion.

The LMedS block-matching motion estimator first estimates the variance of the DFD values within some block b and for some displacement $\mathbf{d}$ using the median variance estimator [5]:

$$\hat{\sigma}_{DFD}(\mathbf{d}) = med_{\mathbf{p}\in S(b)}\left|DFD(\mathbf{p},\mathbf{d}) - med_{\mathbf{p}\in S(b)}(DFD(\mathbf{p},\mathbf{d}))\right| / \Phi^{-1}(0.75) \qquad (2)$$

where $S(b)$ represents the spatial space of all the pixels in a block b, $1/\Phi^{-1}(0.75)$ = 1.4826 is the reciprocal value of the inverse of the standard normal function at 0.75 and makes $\hat{\sigma}_{DFD}(\mathbf{d})$ a consistent estimator of the variance σ^2 when the DFDs in the block are distributed as a standard normal distribution $N(0,\sigma^2)$. The median variance estimator provides robustness to the estimate of the variance in the presence of up to 50% outlying DFD values [2,3] which may be caused by the presence of noise or the block straddling a motion discontinuity (i.e. a region containing 2 or more differently moving sub-regions).

The displacement vector assigned to block b is then determined from minimizing the mean-absolute DFD for pixels inlying to the block:

$$\hat{\mathbf{d}}_b = \arg\min_{\mathbf{d}} \frac{\sum_{\mathbf{p}\in S_I(b,\mathbf{d})} \left|DFD(\mathbf{p},\mathbf{d})\right|}{\sum_{\mathbf{p}\in S_I(b,\mathbf{d})} 1} \quad (3)$$

where:

$$\mathbf{p} \in S_I(b,\mathbf{d}) \quad \text{if } |DFD(\mathbf{p},\mathbf{d}) - med_{\mathbf{q}\in S(b)}(DFD(\mathbf{q},\mathbf{d}))| < 2.5\hat{\sigma}_{DFD}(\mathbf{d}) \quad (4)$$

$S_I(b,\mathbf{d})$ represents the spatial space of all the pixels that are considered as inlying pixels (with respect to the displacement **d**) in the block *b*.

Note that the LMedS block-matching motion estimator here does not actually uses the procedure described in [2,3] and calling it "LMedS" is a misnomer. However, our approach has been conceptualised based on the concept described in [2,3] and the median absolute deviation (MedAD) calculation (eqn (2)) in [5] has been used instead. In addition, the MedAD calculation is further followed by a weighted-averaging step (equation (3) and (4)). This is akin to the reweighted least-squares step as suggested in [2,3] in order to achieve better efficiency in the presence of Gaussian noise.

Literally, what the LMedS motion estimator does is to assign the motion vector of the estimated motion of the majority of the pixels (in the block) to each block. By doing so, we try to reduce the DFD errors for the majority of the moving pixels in the block, thus enabling these region to achieve better perceptual picture quality since there are limited number of bits available for inter-frame DCT coding at low bit-rates, and thus reducing the errors that may be propagated through subsequent P-frames. Even in the presence of large noise (especially the presence of spike noise), the LMedS motion estimator is still able to provide a correct estimate of the true motion in the block. This has the effect as above in providing better picture quality as well.

3. SIMULATIONS

In order to demonstrate that the true motion estimation method has better performance in the sense of subjective video quality, a set of simulations has been carried out. Here we produce coded video sequences (of size 176x144 pixels, i.e. QCIF-sized) using H.263 encoder [6] with the traditional LMS motion estimation and the proposed LMedS motion estimation and then compare the subjective qualities.

To measure the subjective qualities of the video sequences, the Double-Stimulus Impairment Scale (DSIS) method as defined in ITU-R BT. 500 [7] was used. Here, two video sequences are presented to the assessors (i.e. the subjects), where the first sequence is the source (original unimpaired) sequence and the second sequence is the processed (impaired via video coding) sequence, and then the sequences are repeated again in sequel. The assessors are required to vote using a five-grade impairment scale: Imperceptible (5), Perceptible but not annoying (4), Slightly annoying (3), Annoying (2), and Very annoying (1). A total of 20 different subjects are involved in the subjective video quality assessment.

4. RESULTS

The subjective tests have been performed on numerous QCIF-sized video sequences at various bit-rates and in the following, the results for three different source video sequences, namely Foreman, Claire, and Trevor, coded using a H.263 encoder with the same parameters' set-up (except at different bit-rates), will be shown. Figure 1 shows one frame of each source video sequence.

Figure 1: One frame of each source test video sequence: (a) Foreman sequence, (b) Claire sequence, (c) Trevor sequence.

Subjective scores (i.e. Mean Opinion Scores, MOS) of coded image sequences using least-mean-squares motion estimation (LMS-ME) and least-median-squares motion estimation (LMedS-ME) are shown in Table 1.

Sequences	LMS-ME MOS	LMedS-ME MOS
Foreman (64 kbps)	2.11	2.65
Foreman (48 kbps)	1.66	2.32
Claire (64 kbps)	2.97	3.39
Claire (48 kbps)	2.49	3.02
Trevor (64 kbps)	2.82	3.28
Trevor (48 kbps)	2.26	2.84

Table 1: Mean Opinion Scores of coded sequences using LMS and LMedS motion estimation.

From Table 1, we can see that LMedS motion estimation coded sequence has better performance than original LMS motion estimation coded sequence. It was also found that the more motion the video contains (at the same bit-rate), the better the perceived video quality for the case of using LMedS as the matching criterion in the block-matching motion estimator compared to the case of using LMS.

5. CONCLUSION

In this paper, a new approach for MC-DCT hybrid video coding based on true motion estimation has been proposed. The true motion estimation technique employs the LMedS matching criterion in block matching process for motion estimation. The rationale for using such true motion estimator is that at low bit-rate, very little available bits are available for coding the motion-compensated DCT errors, and thus the detection of the true underlying motion of the picture block becomes much more important than just trying to find a block-match that satisfies the least-mean-squares matching criterion. Without detecting the true motion of the picture block, the errors due to insufficient bits for coding the motion-compensated DCT errors will propagate to succeeding inter-frames, resulting in a progressive degeneration on perceptual quality of the pictures. An important advantage of using true motion estimation is that it results in better perceptual video quality compared to the conventional block-matching algorithm that utilizes least-mean-squares matching criterion.

Experimental results obtained support our proposition on using the LMedS-based matching criterion for low bit-rate video coding. Further work will look into incorporating some adaptive quantization method together with the true motion estimator to seek possible further enhancement in the perceptual quality of video sequences coded using hybrid video coders.

6. ACKNOWLEDGEMENT

Authors would like to thank the PMP program of EDB, Singapore and the LIT grant 2002-4 for the support of this work.

7. REFERENCES

[1] Tekalp, M., *Digital Video Processing*, Prentice Hall PTR, 1995.

[2] Rousseeuw, P.J., “Least Median of Squares Regression”, *Journal of America. Statistical Assoiation*, 79 (1984), pp. 871-880.

[3] Rousseeuw, P.J., and Leroy, A.M., *Robust Regression and Outlier Detection*, John Wiley, New York, 1987.

[4] Ong, E.P., and Spann, M., "Robust Optical Flow Computation Based on Least-Median-Of-Squares Regression", *International Journal of Computer Vision*, Vol. 31, No. 1, pp. 51-82, Jan. 1999.

[5] Holland, P.W., and Welsch, R.E., “Robust regression using iteratively reweighted least squares”, *Communication Statistics – Theoretical Methods*, Vol. A6, pp. 813-828, 1977.

[6] ITU Telecom. Standardization Sector of ITU, “Video coding for low bitrate communication”, ITU-T Recommendation H.263, Mar. 1996.

[7] ITU-T Rec. BT. 500, “Methodology for the subjective assessment of the quality of television pictures”, Geneva, Switzerland, 1998.

[8] Jain, J.R. and Jain, A.K., “Displacement measurement and its application in interframe image coding", *IEEE Transactions On Commununications*, Vol. 29, pp. 1799-1806, Dec. 1981.

[9] VQEG (Video Quality Expert Group), “Final Report from the Video Quality Expert Group on the validation of Objective Models of Video Quality Assessment”, www.vqeg.org, March 2000.

[10] Verscheure, O., et. al., “Perceptual Bit Allocation for MPEG-2 CBR Video Coding”, *Int. Conference on Image Processing (ICIP)*, Lausanne, Switzerland, (Sep. 1996).

[11] Verscheure, O., and Lambrecht, C., "Adaptive Quantization Using a Perceptual Visibility Predictor," *International Conference on Image Processing (ICIP)*, Santa Barbara, CA, October 26-29 1997.

[12] Yu, Z, Wu, H.R., Winkler, S. and Chen, T., “Vision-model-based impairment metric to evaluate blocking artifacts in digital video”, *Proc. IEEE*, Vol. 90(1), pp. 154-169, 2002.

[13] Tan, K.T. and Ghanbari, M, “A multi-metric objective picture-quality measurement model for MPEG video”, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 10, No. 7, Oct. 2000, pp. 1208-1213.

[14] Watson, A.B., et. al., “DVQ: A digital video quality metric based on human vision”, *J. of Electronic Imaging*, Vol. 10, No.1, 2001, pp. 20-29.

[15] Winkler, S., “A perceptual distortion metric for digital color video”, *SPIE Proc. Human Vision and Elect. Imaging IV*, Vol. 3644, pp. 175 – 184, 1999.

[16] ISO/IEC JTC1/SC29/WG11/N0999, "MPEG-4 testing and evaluation procedures document", MPEG-4 document, Final version, Tokyo, July 1995.