Comparative Study of Partial Closed-loop Versus Open-loop Motion Estimation for Coding of HDTV

Jeffrey S. McVeigh[1] and Siu-Wai Wu[2]

[1]Carnegie Mellon University
Department of Electrical and Computer Engineering
5000 Forbes Ave.
Pittsburgh, PA 15213

[2]AT&T Bell Laboratories
600 Mountain Ave.
Murray Hill, NJ 07974

ABSTRACT

Motion estimation is a vital function of video coders that utilize motion compensated prediction to exploit the temporal redundancy of video signals. The reference images used in the motion estimation process can be either original images (open-loop architecture) or reconstructed images (closed-loop architecture). While the closed-loop architecture is intuitively superior to the open-loop architecture, closed-loop motion estimation is more difficult to implement in a real-time video encoder. A compromise is to perform open-loop integer-pel accurate motion estimation followed by closed-loop half-pel accurate estimation. In this paper we present a comparative study of the performance of half-pel closed-loop versus open-loop motion estimation for coding of High Definition Television with the MPEG algorithm. Simulation results indicate that, on average, closed-loop half-pel accurate motion estimation provides slightly better performance. This improvement depends on the bit rate of the coded video stream. For relatively high bit rates, the closed-loop implementation often provides no improvement and sometimes an actual degradation of performance. Analysis of why a closed-loop implementation does not guarantee a better performance is presented.

1. INTRODUCTION

Motion estimation is a necessary but computationally intensive process of motion compensated predictive coders. An important application of motion estimation is found in the Motion Picture Expert Group (MPEG) standard which employs block-based motion compensation to predict an input frame from one or more previously coded frames [1, 3].

A typical MPEG compatible encoder consists of the following basic processes: 1). estimate motion vectors for the current input frame from one or more reference images, 2). perform motion compensated prediction with the motion vectors and previously reconstructed images, 3). generate and encode the prediction error signal, and 4). reconstruct the frame from the encoded error signal, motion vectors and decoded reference frames. To avoid error accumulation at the decoder, motion compensation must operate on the reconstructed images. However, the choice of reference images for the motion estimation process is not as straightforward.

There are two obvious choices for the reference images to be used in the motion estimation process: the original images or the reconstructed images. For a given input image block, the goal of motion estimation is to find the motion vector which minimizes the cost of encoding the block. This cost is usually measured as the distortion between the input block and the corresponding block displaced by the motion vector in the reference image. Since reconstructed images are used as references in the prediction process, the closed-loop architecture is intuitively superior to the open-loop motion estimation architecture. However, a real-time encoder with closed-loop motion estimation is more difficult to implement than one with an open-loop architecture. Therefore one must consider the trade-off between performance and the complexity in the selection of a motion estimation scheme. A compromise between entirely open-loop and closed-loop implementations would be to perform part of the estimation based on the original images and the remainder on reconstructed images. A possible example of this compromise, termed half-pel closed-loop motion estimation, entails open-loop coarse (full-pel accurate) motion estimation followed by closed-loop fine (half-pel accurate) motion estimation.

In this paper we present a comparative study of the performance of closed-loop versus open-loop motion estimation for coding High Definition Television (HDTV) with the MPEG-2 algorithm. For practical considerations, our study is focused on the process of estimating the fractional parts of the motion vectors, assuming that the integer part is always performed in the open-loop fashion.

2. MOTION ESTIMATION

Block-based motion estimation proceeds by initially sectioning the input image into equally sized blocks, called macroblocks in MPEG jargon. A search technique is then employed to find a displacement vector, or motion vector, which minimizes a pre-defined distortion function between the current macroblock and the displaced macroblock in the reference image. Common distortion functions are positive, non-decreasing functions of the difference between the luminance components of the original macroblock and the predicted macroblock, for example, the mean absolute error (MAE) and the mean squared error (MSE).

The minimum distortion block is searched for over a pre-specified search area. If an exhaustive search is performed, i.e. each location within the search area is examined, the per-pel complexity of estimating motion vectors will be $O([2h+1][2v+1])$, where the search ranges covers $\pm h$ horizontal times $\pm v$ vertical locations. Constrained search techniques exist, such as logarithmic, cojugate direction, and hierarchical, which attempt to reduce the amount of computations compared with an exhaustive search by limiting the number of locations examined. These techniques compromise the coder's performance for a reduction in complexity of motion estimation. Since all locations within the search area are not examined, the minimum distortion block can not be guaranteed.

Although pels are located at integer locations over the image grid, non-integer motion vector values can be obtained through interpolation. We shall consider half-pel accurate motion estimates where bilinear interpolation is used to estimate the non-integer pel location values. A form of hierarchical searching is usually performed, where full-pel motion estimation is followed by half-pel accurate motion estimation. This technique is utilized to provide half-pel accurate motion vectors without requiring each half-pel location within the search area to be examined [2].

A basic block diagram of an MPEG encoder capable of performing both open and closed loop motion estimation is shown in Fig. 1. While the MPEG standard provides for both intraframe and interframe coding, only predictively coded frames will be considered in this paper.

The encoding process begins with motion estimation. If the current input frame is tagged as a forward predicted frame (P frame) a previous reference frame is used. On the other hand, if the current frame is to be bidirectionally predicted (B frame), both previous and future reference frames are used in the motion estimation. The reference frames stored in Frame Buffers 1 or 2 are used for open or closed-loop motion estimation, respectively. The motion vectors are losslessly encoded and are sent to the decoder. Motion compensation uses these motion vectors to predict the current input image from the reconstructed reference frames stored in Frame Buffer 2. The predicted image is then subtracted from the original image to form the error signal. The Discrete Cosine Transform (DCT) is applied on this error signal to reduce spatial redundancy. The DCT coefficients are quantized and variable-length encoded for transmission to the decoder. To complete the prediction loop, the dequantized and inverse transformed error signal is added to the motion compensated prediction to reconstruct a coded image which is stored in Frame Buffer 2.

In a real-time encoder we wish to estimate motion vectors for a current input frame in parallel with the rest of the encoding operations. An implementation can be envisioned which performs closed-loop motion estimation in parallel with the other operations by encoding and reconstructing an image block as soon as the motion vector for the block is available. The resulting new reference frame can be stored in Frame Buffer 2 with a double-buffering scheme. This set-up alleviates the need to wait until all motion vectors are calculated for an image before beginning the rest of the encoding operations. However, the processing of such a closed-loop implementation must be fast enough to ensure that the reconstructed image is available within one frame period, for use in the next frame's motion estimation. Also, an additional frame buffer capable of storing one complete frame is required.

A simpler implementation, which also performs the motion estimation in parallel with the rest of the encoding operations, is the open-loop motion estimation structure. The motion estimation for the current frame proceeds while the previous frame is being encoded and reconstructed. The drawbacks of using an open-loop implementation are the added frame buffer required to store the original reference images and the possible degradation in performance as compared to a closed-loop implementation. The half-pel closed-loop architecture attempts to maintain low hardware complexity while hopefully providing performance better than the completely open-loop structure.

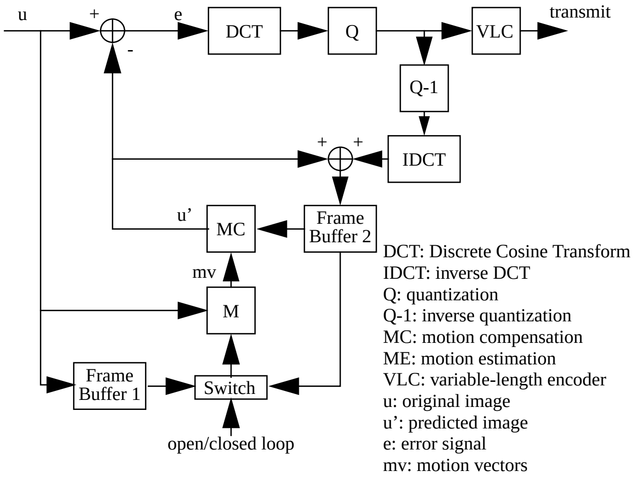


Figure 1. MPEG Encoder Block Diagram with Closed-loop and/or Open-loop Motion Estimation

3. RESULTS AND ANALYSIS

Three separate 1280×720 HDTV test sequences were encoded with the MPEG-2 algorithm using both open-loop and half-pel closed-loop motion estimation. The motion estimates were calculated for standard luminance macroblocks of size 16×16 pels. Full-pel accurate motion estimation was followed by half-pel accurate estimation.

All test sequences used in the experiments were 60Hz progressively scanned and therefore only frame motion estimation was performed. The minimum distortion block was defined as the reference block which yielded the minimum MAE value,

$$MAE = \sum_{i=1}^{16} \sum_{j=1}^{16} \left| u(i, j) - \hat{u}(i+p, j+q) \right| \qquad \text{(EQ 1)}$$

where $u(i, j)$ denotes the current input luminance macroblock under consideration and $\hat{u}(i+p, j+q)$ denotes the displaced reference image block. The components of the motion vector are the values of p and q which minimize (EQ 1). For bidirectional frames, motion vectors were estimated independently in both forward and backward directions. The bidirectional estimate was then obtained by averaging the forward and backward predicted macroblocks. The prediction type decision, i.e. forward, backward, bidirectional, or intra, was based on the minimum MSE between the original image and the reconstructed reference images.

Performance comparisons were obtained by calculating the peak signal-to-noise ratio of the reconstructed luminance images,

(EQ 2)

where σ_e^2 equals the mean squared prediction error. YSNR values were calculated and compared before and after quantization to observe the effect of quantization on the performance relationship between the different motion estimation schemes. Per-macroblock prediction errors were also compared to evaluate performance gain on an individual macroblock basis.

In the first experiment, 90 frames of the raft sequences were coded. The spacing between coded frames was specified as 30 frames between intra-coded images ($N = 30$) and 5 frames between forward predicted images ($M = 5$). The remaining frames were bidirectionally predicted. The sequence was coded at a target bit rate of 17 Mbits/sec. The YSNR's after quantization for both open and half-pel closed-loop motion estimation are shown in Fig. 2.

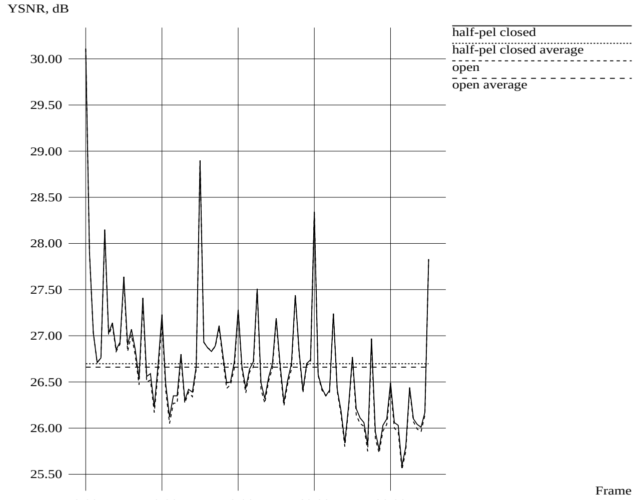


Figure 1. YSNR of raft sequence after quantization for open and half-pel closed-loop motion estimation

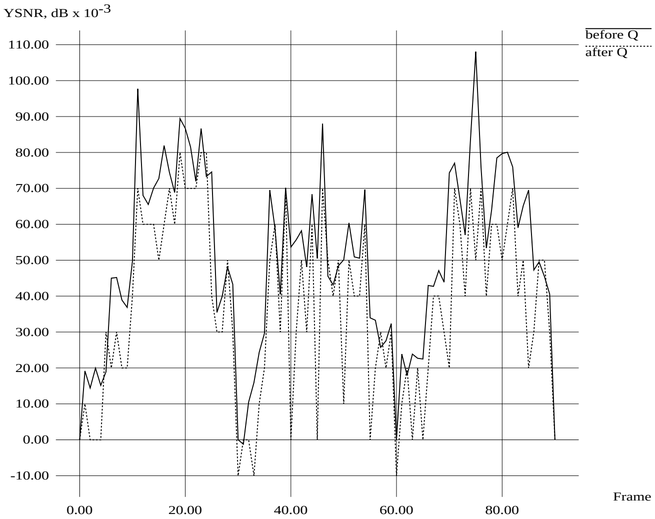


Figure 2. Raft sequence: half-pel closed-loop motion estimation gain, before and after quantization (Q)

The gain of closed-loop over open-loop half-pel motion estimation, given by (EQ 3), is depicted in Fig. 3, before and after quantization.

$$gain = YSNR_{closed} - YSNR_{open} \qquad \text{(EQ 3)}$$

On average, half-pel closed-loop motion estimation provides better performance than the open-loop implementation, with greater gain observed prior to quantization. The performance loss from quantization is most likely due to the fact that quantization has an equalizing effect on the SNR's through the introduction of errors. Since the motion estimation procedure minimizes the specified distortion function, the prediction error energy may be less with the closed-loop implementation, but the entropy may increase; hence, coarser quantization would need to be performed to maintain the desired bit rate.

While the closed-loop implementation usually provides better performance, the gain is marginal with a maximum increase in YSNR of 0.11 dB before quantization and 0.08 dB after quantization. Upon viewing the reconstructed images, it was extremely difficult to perceive any difference in quality between video coded with the two motion estimation architectures. Although the half-pel closed-loop implementation provides a better YSNR for most frames, an actual degradation in performance (negative gain) is observed for some frames. The causes of this counter-intuitive result will be analyzed in detail subsequently.

To observe the relationship between target bit rate and the performance gain of using half-pel closed-loop motion estimation over open-loop estimation, thirty frames of the sequence bulls were encoded at 3, 10, and 17 Mbits/sec with $N = 30$ and $M = 3$. The results are shown in Fig. 4.

Once again the half-pel closed-loop implementation provides slightly better performance than the open-loop implementation, with the pre-quantization gain being greater than the gain after quantization. The amount of gain for the encoded sequence with a bit rate of 17 Mbits/sec is similar to that obtained for the raft sequence. The gain is a function of the bit rate of the sequence, with lower bit rates achieving greater gain. Needless to say, as the bit rate decreases so does the overall YSNR of the encoded sequence.

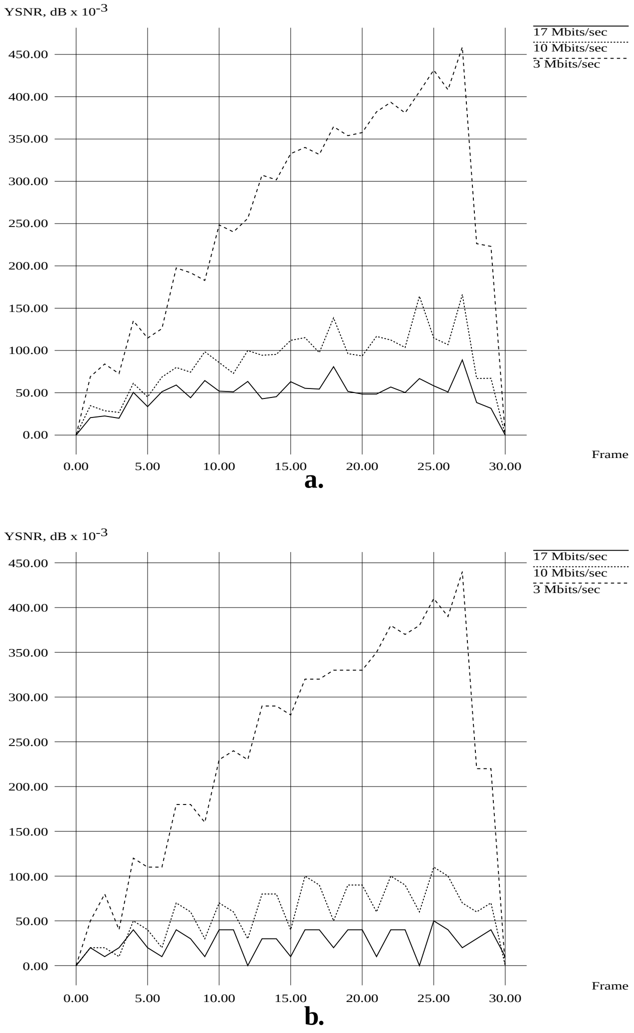


Figure 3. Gain for bulls sequence at varying bit rates: (a). before quantization, (b). after quantization

The dependence of the gain on the bit rate of the sequence is due to the fact that for high bit rates the original and reconstructed images bear little difference. Therefore, little or no gain is obtained by using the reconstructed images for the motion estimation procedure. However, as the bit rate decreases the quality of the reconstructed images suffers and the reconstructed images often differ greatly from the original images. A closed-loop structure can more effectively handle the image degradations of low bit rate sequences.

In an attempt to understand why half-pel closed-loop motion estimation provides only a slight gain, and sometimes even a loss in performance, the macroblock prediction errors were examined for 15 frames of the tulips sequence coded at 17 Mbits/sec with $N = 30$ and $M = 3$. The gain experienced per frame before and after quantization is shown in Fig. 5a. Here, the gain prior to quantization is comparable to that obtained for the two previous sequences. However, after quantization the gain is actually negative for all frames except for the first, intra coded, frame (Frame 0).

Figure 5b plots the per-macroblock gain, in dB, of the prediction error before quantization for the half-pel closed-loop implementation over the open-loop architecture. For intra-coded macroblocks, the variance of the original pixels within the macroblock are used in the computation of the gain. The resolution of the images was 1280×720, which results in 3600 macroblocks per image. The first 3600 macroblocks are from Frame 9, followed by the 3600 macroblocks from Frame 10, and so on. The prediction types were P, B, B, and P for frames 9, 10, 11, and 12, respectively. Approximately half of the prediction errors for the open-loop estimation are less than the half-pel closed-loop values. This validates the per-frame results which showed only minimal gain for the closed-loop structure. Since the prediction error is defined as the difference between original and reconstructed images, it would appear that a smaller macroblock prediction error would be produced using the reconstructed reference frames than using original reference images. Unfortunately, the results of Fig. 5b indicate that this is not always the case. Four distinct causes of why the half-pel closed-loop estimation does not guarantee a better performance were examined.

The first and most obvious reason deals with the discrepancy in the distortion functions used in the motion estimation process and the performance measure. While the motion vectors are selected on the basis of the minimum MAE, the macroblock type decision and the calculation of gain are based on the MSE criteria. A probability exists such that the half-pel motion estimation selects a block which has a lower MAE value but a greater MSE value. The performance, measured by the MSE, would then be degraded for the closed-loop procedure. This can be avoided if all operations and performance evaluations use the same distortion measure.

Secondly, the open-loop prediction error may be less than the closed-loop error when macroblocks are selected to be intra-coded within a predictively encoded frame. In a typical MPEG encoder, such as the one used in our experiments, if the motion compensated prediction error of a macroblock is greater than both the variance of the original pixels in the macroblock and a predefined threshold, the macroblock is intra-coded. A scenario may exist, given by (EQ4), where both closed-loop error and open-loop error are greater than the variance but the threshold lies in between the two prediction error values.

$$error_{open} > threshold > error_{closed} > variance \qquad \text{(EQ 4)}$$

Assuming that the cost of encoding an intra-coded macroblock is measured by the variance of the original pixels in the macroblock while that of encoding a nonintra-coded macroblock is measured by the mean squared prediction error, the condition in (EQ 4) leads to the following relationship.

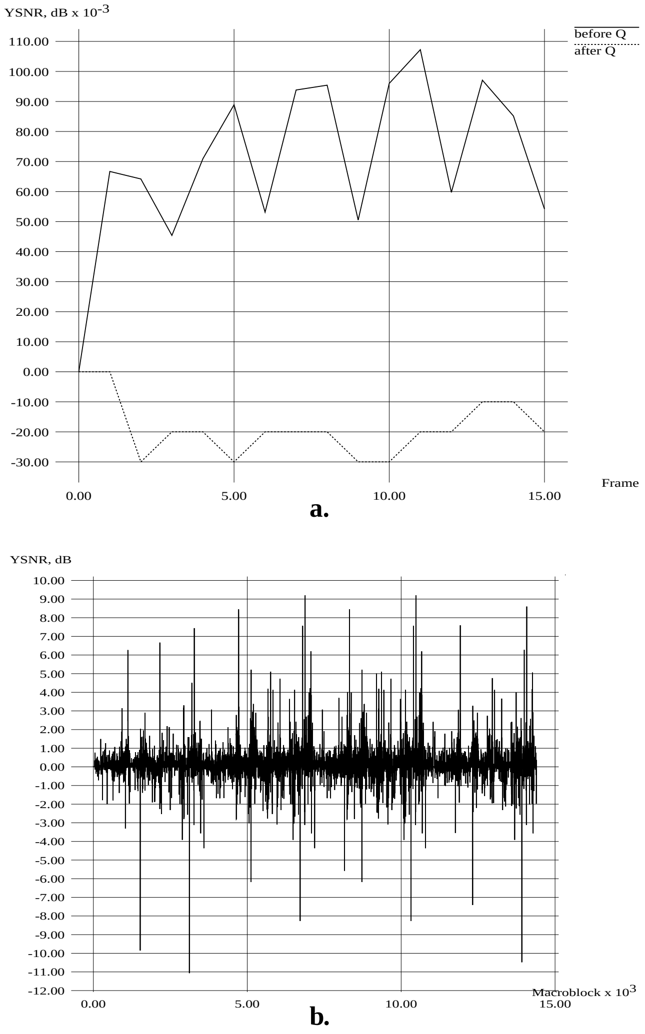
(EQ 5)

where $Cost_{closed}$ and $Cost_{open}$ are the costs of coding the macroblock when closed-loop and open-loop motion estimations are employed, respectively. In this case, the prediction error is smaller for the closed-loop implementation but the macroblock type decision results in a higher cost for coding the macroblock.

Using a non-exhaustive search for the bidirectional estimate for B frames is the third reason why closed-loop motion estimation does not guarantee a performance gain. A bidirectionally estimated block obtained from open-loop motion estimation may produce a lower prediction error, even though the prediction error is greater for the individual forward and backward prediction blocks. Gain can be ensured only if the bidirectionally predicted macroblock selected by the open-loop estimation is also examined by the closed-loop procedure. This macroblock will be examined if both open and closed-loop motion estimation are performed or if the search for the bidirectional estimate is exhaustive. An exhaustive half-pel accurate search entails finding the minimum distortion block over 81 locations (9 locations for each forward and backward reference images).

The final cause of possible degraded performance for closed-loop motion estimation deals with the reference images used in the motion compensated prediction process. Due to the fact that the predictive encoder possesses memory, the reconstructed reference images in the two implementations are not necessarily identical. This mismatch in the reconstructed reference images is the result of the half-pel motion estimation coupled with variations between quantizer step sizes imposed by the rate control mechanism. Therefore, the motion compensated prediction for the closed-loop and open-loop cases do not operate on identical reference images, hence it is possible for the open-loop architecture to yield better performance for some macroblocks.

The influence of each of the above factors on the performance of closed-loop estimation was examined by removing the



a.

b.

Figure 3. Tulips sequence: (a). gain, per frame, before and after quantization, (b). macroblock gain for half-pel closed-loop over open-loop motion estimation for frames 9-12.

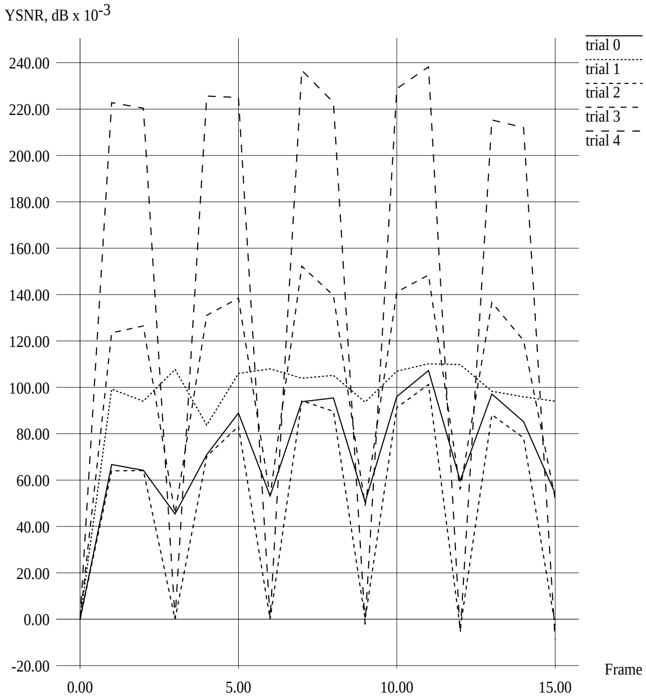


Figure 3. Performance gain for tulips sequence before quantization. Trial 0: standard half-pel closed-loop motion estimation, Trial 1: MSE distortion function used and intra threshold set to zero, Trial 2: half-pel closed loop ME on B-frames only, open-loop ME performed on P frames, Trial 3: exhaustive half-pel closed-loop bidirectional search, Trial 4: combination of Trials 1-3.

described discrepancies and re-encoding the tulips sequence. A constant quantization scale was not imposed so the experiments which specified equal reconstructed reference images in fact varied slightly. The performance gains, before quantization, of the various trials are shown in Fig. 6. The maximum gain was observed for the B frames in Trial 4, where all dependencies were removed. Nevertheless, the gain was approximately zero for the P frames of this trial. Also, the maximum gain experienced of roughly 0.24 dB produced little noticeable subjective improvement.

4. CONCLUSION

We have presented a comparative study of the performance of half-pel closed-loop versus open-loop motion estimation for the coding of HDTV. On average, half-pel closed-loop estimation provided slightly better performance than completely open-loop estimation. The gain in performance was shown to depend on the bit rate of the video sequence, with lower bit rates yielding larger gain. Since the reconstructed images closely resemble the original images for high bit rates, little or no advantage is attained by using a closed-loop motion estimation architecture. The encoding of the prediction errors diminished the gain obtained with a closed-loop architecture. Cases were observed where the encoding process actually caused frames with gain before quantization to result in a loss of performance. This was attributed to the fact that the motion estimation procedure attempts to minimize the prediction error energy, which may or may not decrease the entropy of the DCT coefficients.

While the half-pel closed-loop estimation usually produced a performance gain, any gain observed was marginal and for a few cases the prediction error actually increased. A performance gain for the closed-loop implementation could not be guaranteed for the following reasons: 1). different distortion functions used in motion estimation and performance evaluation, 2). the use of a threshold value for intra/nonintra macroblock type decision, 3). non-exhaustive bidirectional search, and 4). differing reference images used in the prediction process for closed-loop and open-loop cases.

For a bit rate of 17 Mbits/sec, the gain experienced after quantization for the three sequences examined was always less than one-tenth of a dB. Since the gain of using a half-pel closed-loop architecture was shown to be quantitatively and subjectively insignificant, we suggest that system designers may chose the architecture which yields a smaller implementation complexity subject to the design specification of the encoder.

5. REFERENCES

[1]. ISO/IEC JTC1/SC29/WG11, “Information Technology - Generic Coding of Moving Pictures and Associated Audio, Recommendation H.262, ISO/IEC 13818-2, Draft International Standard,” March 1994.

[2]. Iu, S. L., “Comparison of Motion Compensation using Different Degrees of Sub-pixel Accuracy for Interfield/interframe Hybrid Coding of HDTV Image Sequences,” ICASSP, vol. 3, pp. 465-68, 1992.

[3]. ISO/IEC JTC1/SC29/WG11 Test Model Editing Committee, “MPEG-2 Video Test Model 5", ISO/IEC JTC1/SC29/WG11 Doc. N0400, April 1993.

Acknowledgment

The authors would like to thank Rajesh Hingorani and Wendi Rabiner for contributive discussions and suggestions which helped to verify and enhance the work presented in this paper.