



Czerepinski, P.J., & Bull, DR. (1996). Coder-oriented matching criteria for motion estimation. In *Proc. 1st Intl workshop on Wireless Image and Video Communications* (pp. 38 - 42). Institute of Electrical and Electronics Engineers (IEEE).
<https://doi.org/10.1109/WIVC.1996.624640>

Peer reviewed version

Link to published version (if available):
[10.1109/WIVC.1996.624640](https://doi.org/10.1109/WIVC.1996.624640)

[Link to publication record on the Bristol Research Portal](#)
PDF-document

University of Bristol – Bristol Research Portal

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/brp-terms/>

CODER-ORIENTED MATCHING CRITERIA FOR MOTION ESTIMATION

P.J. Czerepinski and D.R. Bull

Image Communications Group, Centre for Communications Research,
University of Bristol, Merchant Venturers Bldg., Woodland Rd., Bristol BS8 1UB, UK
email: dave.bull@bristol.ac.uk p.j.czerepinski@bristol.ac.uk
Tel. +44 117 9545195; Fax: +44 117 9255265

Abstract *Classical matching criteria for motion estimation determine match quality by operating in the spatial domain. The resulting DFD signal is, however, compressed by means of frequency transformations, and any optimal matching criteria should estimate the cost of coding the composite bit stream. Such criteria are presented in this paper and shown to offer some improvement in the PSNR of the reconstructed sequence.*

1 INTRODUCTION

In the context of video coding, motion estimation and compensation [1][2][3][4] represent a well established means of removing the temporal redundancy from a signal. Motion information is generally extracted by comparing the contents of adjacent video frames. For simplicity, a block-based translational motion model is often assumed, and motion vectors are found by determining a match between corresponding regions in the current and previous frames. 'Classical' matching criteria such as MSE or MAD minimise a measure of the prediction error, with the aim of producing an estimate which is visually pleasing. What pleases the eye however, need not be pleasing for the coder and suboptimal (in the MSE sense) predictions may exist which produce a differential signal more amenable to coding. A simple example, involving a transform coder is shown in figure 1. The current frame (fig. 1b) is predicted using the previous frame (fig. 1a). In the MSE sense, the previous frame's central block is the best prediction for the current frame's all-black block. Due to sharp edges in the resulting displaced frame difference signal (DFD), (fig. 1c), ringing artefacts are bound to appear in the reconstruction if transform coefficients are quantised. On the other hand, the choice of any of the all-white blocks as a prediction would produce an all-black differential signal (fig. 1d), which could be coded with the *DC* coefficient alone.

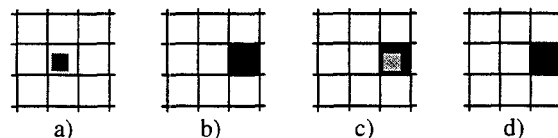


Figure 1 *The current frame (b) is predicted using the previous frame (a). Displaced frame differences produced using different matching criteria: (c) (d)*

For video coding, especially at lower bit rates, it is important that the DFD signal can be coded efficiently. In this paper we discuss features that make a good matching criterion and we reflect on the relevance of the popularly used MSE or MAD measures. As an alternative to these, we investigate a coder-oriented approach to the problem.

2 WHAT IS AN IDEAL MATCHING CRITERION ?

A block diagram of a typical block discrete cosine transform (DCT) video coder is shown in figure 2.

For a pre-determined quantiser and entropy coding algorithm, motion estimation with its associated matching criterion remains the only degree of freedom in the compression process. The quality of the reconstructed image will entirely depend on the content of the displaced frame difference signal, since it will affect the quantisation (some signals will produce smaller quantisation errors than others), the residual entropy and the motion field entropy. An ideal matching criterion should therefore satisfy the following, sometimes mutually exclusive, conditions:

- *good match*: the discrepancy between the original frame and its prediction should be minimised, i.e. blocks that minimise the MAD or MSE should be chosen;
- *good compression of the residual signal*: the entropy of the DCT-coded displaced frame difference should be as low as possible;
- *good reconstruction quality* (for a fixed compression ratio): the distortion of the reconstructed signal should be minimised, i.e. blocks that produce the smallest quantisation error should be chosen;
- *good compression of the motion field*: the motion field should be as uniform as possible to facilitate efficient coding.

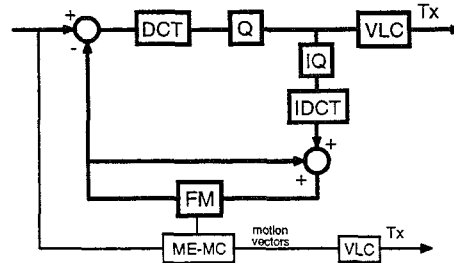


Figure 2 Coder block diagram. DCT — discrete cosine transform, Q — quantiser, IQ — inverse quantiser, IDCT — inverse discrete cosine transform, FM — frame memory, VLC — variable length coding, ME-MC — motion estimation and compensation.

Of the above list, the first condition (minimum MAD or MSE) was found by the authors to be of lowest relevance to the compression efficiency: it estimates match quality in the spatial domain, whereas the quantisation and entropy coding processes operate on the transform domain data. The widespread application of these criteria can however be justified by their simplicity. We conclude, that the efficiency of the motion estimation process is a function of the quantisation error Q_e and the bit rates B_r and B_{mv} of the DCT-coded residual signal and the motion field respectively. For optimum performance, Q_e , B_r and B_{mv} must be taken into account in the design of the matching criterion.

3 MATCHING CRITERION DESIGN

The following matching criteria have been proposed for testing:

$$(1) \quad D_1 = \sum_i \sum_j \hat{F}_{ij} \quad ,$$

i.e. minimising the sum of quantised DCT coefficients \hat{F}_{ij} of the displaced frame difference signal.

$$(2) \quad D_2 = B_r + B_{mv} \quad ,$$

i.e. minimising the bit-rate required to code the residual signal and the motion vectors. Conditional statistical models are used when coding both. In the case of the DCT-coded DFD, coefficient values are conditioned on coefficients' positions in an 8×8 block, whereas the motion vectors are coded in respect to the mean of their three causal half-plane neighbours.

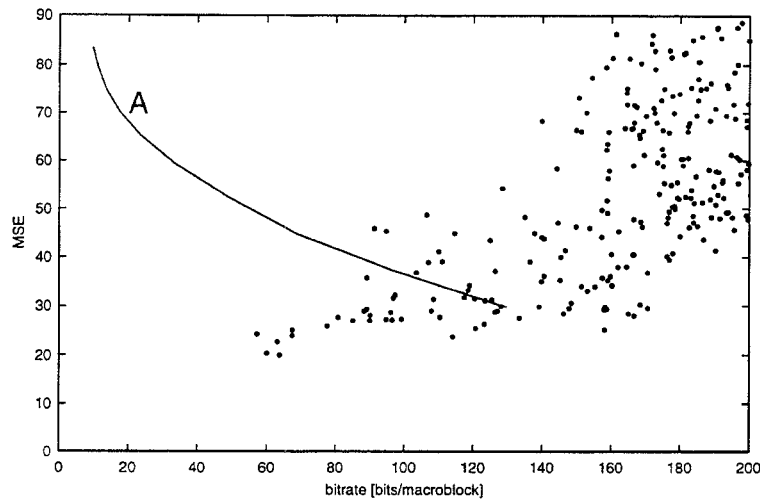


Figure 3 Mean square error of the reconstruction in respect to the bit rate; 'Table Tennis' sequence.

Next, consider the curve A in figure 3. This illustrates the relationship between the average number of bits used to code a displaced frame difference macroblock (16×16) and the resulting mean squared error for the 'Table Tennis' sequence. 'A' is just a coder performance curve, where, for the sake of this article, the axes have been labelled bits/macroblock-MSE rather than the usual bits/second-PSNR.

The clustered (B, MSE) vectors correspond to the number of bits required to encode a single frame difference macroblock and its resulting MSE . Every (B, MSE) vector is associated with a candidate motion vector, evaluated during an example search. The goal of this approach is to choose the translation that offers the best B - MSE trade-off. In order to accomplish this, bit rate and distortion must be weighted, dependent on their contribution to reconstruction quality. These weights (k and l respectively) can be derived from the averaged coder performance curve, A. Their ratio at a point $(B_0, MSE_0) \in A$ is equal to the differential $\frac{k}{l} = \frac{\Delta MSE_0}{\Delta B_0}$. Since this value varies along the curve, the

investigation is restricted to the bit rate interval 30-130 bits/macroblock, over which $\frac{k}{l}$ remains approximately constant (≈ 4). In this context, two additional matching criteria are proposed:

$$(3) \quad D_3 = k(B_r + B_{mv}) + lQ_c \quad ,$$

i.e. minimising a weighted sum of the bit rates required to code the residual signal and motion vectors and the quantising distortion. The weighting of the components corresponds to their contribution to objective reconstruction quality. Distortion is measured in terms of the mean square error.

$$(4) \quad D_4 = \sqrt{(k(B_r + B_{mv}))^2 + (lQ_c)^2} \quad ,$$

i.e. selecting the motion vector MV_0 that minimises the norm of the corresponding $(k \times B, l \times MSE)$ vector.

4 CODING

The criteria described in section 3 have been applied to the 'Table Tennis' and 'Akiyo' sequences (240×352×8bpp, 30fps). The codec employed half-pixel block matching motion estimation, with 16×16 macroblocks and 8×8 DCT subblocks. Full search has been performed, with the search area restricted to ± 15 pixels ('Table Tennis') and ± 6 pixels ('Akiyo').

DCT Coefficient Coding

Match evaluation requires that the DCT is performed on DFD blocks during the search. Additionally, criteria (2)(3)(4) require an estimate of the bit rate associated with a given coefficient bit stream. After quantisation, a zig-zag scan is performed on DCT subblocks. Our entropy coder uses a fixed conditional model, where the states are identified with coefficients' positions within the subblock. In order to speed up the bit rate calculation for motion estimation purposes, the model is converted into a look-up table. This does not affect the precision, since eventually arithmetic coding is employed to compress the quantised coefficients, and this method is known to introduce very little overhead and operate in close accordance to the model.

Motion Vector Coding

Motion vectors are coded dependent on the prediction formed as the mean of their three causal half-plane neighbours. As is the case with the DCT coefficients, look-up tables are used during motion estimation and arithmetic coding is finally applied to the motion field.

5 RESULTS

Results achieved by applying the proposed matching criteria to the two test sequences are shown in figure 4. In the case of 'Table Tennis', criteria (1), (2) and (4) yield a slightly better performance, compared to MSE alone, at higher bit-rates but deteriorate at lower bit-rates. Criterion (3) is superior to MSE at all bit rates considered and therefore has been selected for further evaluation using 'Akiyo'. The coding gain, however, does not exceed 0.2 dB in the case of 'Table Tennis' and is smaller in the case of 'Akiyo'. This is due to low motion content in the latter, which gives the criterion fewer opportunities to demonstrate its capabilities.

6 CONCLUSIONS

In this article, a number of alternative matching criteria for motion compensated DCT coding have been presented. Their performance has been compared to that of the conventional MSE criterion for both high and low activity sequences. These experiments demonstrate the performance bounds of block matching motion estimation and show that only modest quality improvements are possible with more complex criteria. In addition, computational cost required to achieve these bounds is unlikely to be justified by quality improvements.

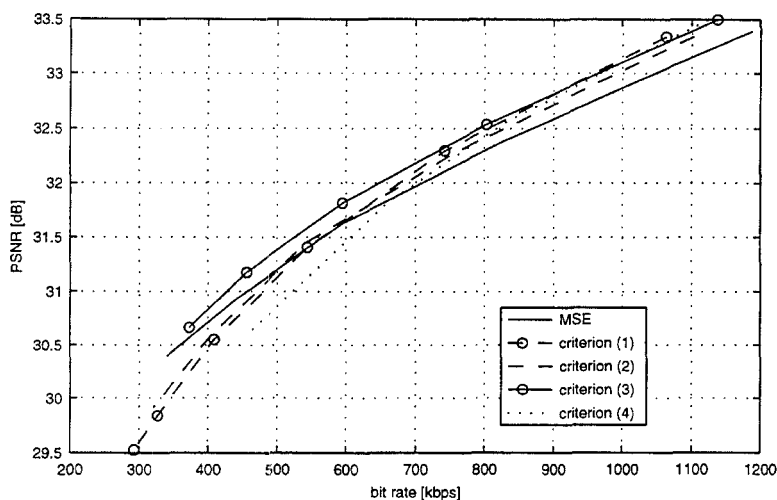


Figure 4a. Performance of different matching criteria, 'Table Tennis'.

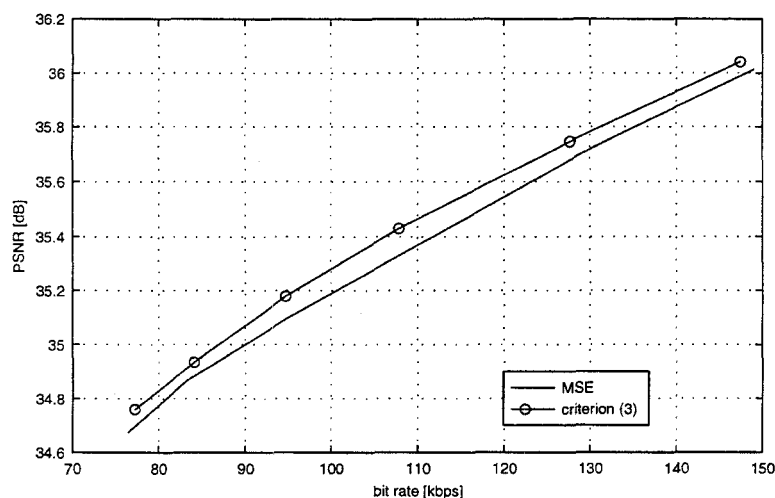


Figure 4b. Performance of different matching criteria, 'Akiyo'.

Acknowledgements

The authors wish to acknowledge the support of Sony Broadcast and Professional Europe, the CVCP and the University of Bristol for their support of this work.

References

- [1] Le Gall D. *MPEG: A Video Compression Standard for Multimedia Applications*, Communications of the ACM, vol.34 no.4 Apr.'91, pp.47-58
- [2] Musmann H.G., Pirsch P., Grallert H.J. *Advances in Picture Coding*, Proc. IEEE vol.73 no.4 Apr.'85 pp.523-548
- [3] Netravali A.N., Robbins J.D. *Motion Compensated Television Coding - Part I*, Bell Syst. Tech. J. vol.58 March'79 pp.631-70
- [4] Wang Q., Clarke R.J. *Motion Estimation and Compensation for Image Sequence Coding*, Signal Processing: Image Communications 4 (1992) 161-174
- [5] Witten I.A., Neal R.M., Cleary J.G. *Arithmetic Coding for Data Compression*, Communications of the ACM, vol.30 no.6 June '87, pp.520-540