Importance of quantiser design compared to optimal multigrid motion estimation in video coding

Electronics Letters, Vol. 36, 9, 807–809 (2000)

J. Malo*, F.J. Ferri†, J. Gutierrez† and I. Epifanio†

(*) Dept. d'Òptica, (†) Dept. d'Informàtica, Universitat de València
Dr. Moliner, 50. 46100 Burjassot, València. Spain
jesus.malo@uv.es

Adaptive flow computation and DCT quantisation play complementary roles in motion compensated video coding schemes. Since the introduction of the intuitive entropy-constrained motion estimation of Dufaux et al., several optimal Variable-Size Block Matching Algorithms have been proposed. Many of these approaches put forward their intrinsic optimality, but the corresponding visual effect has not been explored. In this work, the relative importance of optimal multigrid motion estimation with regard to quantisation is addressed in the context of MPEG-like coding. It is shown that while simpler (suboptimal) motion estimates give as good subjective results as the optimal ones, small enhancements in the quantiser have significant visual effects. This suggests that more attention should be paid to the quantiser design.

This work has been supported by CICYT (TIC) projects 1FD97-0279 and 1FD97-1910

Introduction.

In H.263 and MPEG video coders, the original sequence is split into two lower complexity signals that carry information about motion and prediction errors. These are referred to as Displacement Vector Field (DVF) and Displaced Frame Difference (DFD). A more complex and accurate DVF gives rise to a better prediction (lower complexity DFD). However, the relative effectiveness of such a detailed description relies heavily on the DCT quantiser used. In other words, the better the quantiser, the less significant the improvement due to motion estimation.

The so-called *entropy-constrained* multigrid approach [1] explicitly makes use of this trade-off between motion estimation and DFD complexity. Since this approach was introduced, great effort has been devoted to obtaining analytical [2–4] or numerical [5] *optimal* entropy-constrained quadtree DVF decompositions. These approaches criticise the (faster) entropy measure of the DFD in the spatial domain of Dufaux et al. because it does not take into account the effect of the selective DCT quantiser. This necessarily implies a suboptimal bit allocation between DVF and DFD. The literature [2–5] reports the *quantitative* optimality of the proposed methods, but the practical (subjective) effect of this gain on the reconstructed sequence is not analysed. In particular, only (perceptually unweighted) SNR or MSE distortion measures are given and no explicit comparison of the decoded sequences is shown.

In this work, the relative importance of an optimal adaptive motion estimation is addressed. Combinations of Variable-Size Block Matching Algorithms (VSBMA) [1, 5] and different quantisation schemes [6–8] are evaluated. Distortion results with perceptually meaningful measures [9, 10] and explicit comparisons of the reconstructed frames are shown.

Algorithms and distortion measures.

Two different approaches can be followed to design VSBMA. First, the final bit-rate can be optimised taking into account both adaptive motion and error encoding [2–4], given a particular coding scheme. Second, a splitting criterion which is related to the entropy of both motion and error signals can be used for adaptive motion estimation [1, 5]. In the latter case, the corresponding VSBMAs locally increase the resolution if the volume of the resulting signal is reduced, i.e., *a block is split if* $H(DVF_{split}) + H(DFD_{split}) < H(DVF_{nosplit}) + H(DFD_{nosplit})$, where $H(DVF)$ is the entropy of the vector field plus the quadtree, and $H(DFD)$ is a measure of the complexity of the residual error signal. The entropy $H(DFD)$ can be computed either directly from the DFD in the spatial domain (*suboptimal*) [1] or as the entropy of the quantised transform of the DFD (*optimal*) [5]. In the optimal case, motion estimation is directly related to coding effectiveness. The difference between this and other optimal methods [2–4] is in the way this optimisation is stated and performed. However, the best algorithms of each approach are optimal because the particular features of the quantiser are taken into account.

Any VSBMA scheme can be applied using different quantisers. All the DCT quantisers considered here, [6] as well as [7, 8], are based on a perceptually uniform quantisation of the DCT to restrict the maximum perceptual error for every possible input. It has been shown [7] that, under a simple *linear* perception model, the general equations that define the perceptually uniform quantiser reduce to a simple *JPEG-like* quantiser based on the *Contrast Sensitivity Function* (CSF) [6]. If a more accurate *non-linear* perception model is used instead, a different quantiser which gives better subjective results on still images [7, 8] is obtained (hereafter referred to as *enhanced*).

The use of meaningful distortion measures in video coding design is not a simple question [9]. The widely known dependence of human visual sensitivity on spatial frequency makes a fundamental difference between plain Euclidean distortion metrics (e.g MSE or SNR) and any (even the simplest) perceptually weighted metric. In this work, distortion was measured with a non-linear perceptual metric [10] in order to quantify the subjective results obtained by the approaches considered. Simpler [10] or more complex perceptual metrics [9] would lead to the same qualitative results.

Results and Discussion.

The four possible combinations using optimal and suboptimal VSBMA and the two DCT quantisers mentioned above were tried on several standard monochrome QCIF sequences at 200kb/sec. In order to highlight the relative differences among the approaches considered, only the first frame was intracoded and only forward prediction was used in the remaining frames. As a consequence, the results may seem abnormally distorted at this rate. Both Euclidean and perceptual MSE distortions were computed at each frame in all cases. Also, resulting volumes for DVF and DFD signals were obtained. Specifically, Table **1** shows the relative volume of the motion flow with regard to the total volume of the encoded signal using a *JPEG-like* quantiser and different motion estimates including Fixed-Size BMA (FSBMA) as a reference. These and the corresponding results using an *enhanced* quantiser show a dramatic decrease in the DVF volume of both VSBMAs with regard to FSBMA but, at the same time, an almost negligible difference between the two VSBMA approaches.

Fig. **1** shows the distortion in the reconstructed sequence (RUBIK) computed at each decoded frame. Fig. **2** shows the 7th reconstructed frame using the four different approaches considered. The results in Table **1** and Fig. **1** do indeed confirm the quantitative advantages of optimal motion estimation: the flow volume and the distortion (no matter the measure) are consistently reduced. However, this diminution is almost negligible with regard to the benefits of a better quantisation. This can be numerically assessed by looking at the differences in the corresponding curves in Fig. **1** and, more importantly, by looking at the corresponding reconstructed sequences.

In this case, the simple increase in the quantiser band-pass that comes from considering the visual non-linearities [7, 8, 10] may explain the enhancement. Due to the noisy (high frequency) nature of the error signal, wide-band quantisers may be better than narrower-band (CSF-based) quantisers. In fact, the benefits of this enhanced quantiser are more apparent on video frames (Fig. **2**) than in still images [7, 8]. The significant improvement due to a particular enhancement of the quantiser suggests that quantiser design may be more important than optimal motion estimation.

On the other hand, it has been reported [5] that a more robust and meaningful motion estimate is obtained as a side effect of the optimal interaction between the motion estimation and a perceptual quantiser (due to its band-pass nature). As can be seen by comparing the upper and lower frames in Fig. **2**, this advantage in motion estimation has negligible visual effects on the reconstructed sequence. Faster suboptimal multigrid methods are good enough from the visual perspective. However, if this increased robustness effect is exhaustively confirmed, it may be of interest to higher-level uses of motion information such as in model-based video coding.

Table 1: *Relative volume of the DVF (in %)*

	TAXI	RUBIK	YOSEMITE	TREES	Average
FSBMA	12.57	21.71	46.19	39.21	30 ± 8
Subopt. VSBMA	1.89	3.29	8.69	4.97	4.7 ± 1.5
Optimal VSBMA	0.94	1.87	4.32	3.24	2.6 ± 0.7

References

[1] F. Dufaux and F. Moscheni. Motion estimation techniques for digital TV: A review and new contribution. *Proc. IEEE*, 83(6):858–876, 1995.

[2] G.M. Schuster and A.K. Katsaggelos. *Rate-Distortion Based Video Compression*. Kluwer Acad. Publ., Boston, 1997.

[3] G.M. Schuster and A.K. Katsaggelos. A video compression scheme with optimal bit allocation among segmentation, motion and residual error. *IEEE Trans. Im. Proc.*, 6(11):1487–1502, 1997.

[4] J. Lee. Joint optimization of block size and quantization for quadtree-based motion estimation. *IEEE Trans. Im. Proc.*, 7(6):909–912, 1998.

[5] J. Malo, F. Ferri, J. Albert, and J.M. Artigas. Splitting criterion for hierarchical motion estimation based on perceptual coding. *Electr. Lett.*, 34(6):541–543, 1998.

[6] G.K. Wallace. The JPEG still picture compression standard. *Comm. ACM*, 34(4):31–43, 1991.

[7] J. Malo, F. Ferri, J. Albert, and J. Soret. Comparison of perceptually uniform quantization with average error minimization in image transform coding. *Elect. Lett.*, 35(13):1067–1068, 1999.

[8] J.Malo, F.Ferri, J.Albert, J.Soret, and J.Artigas. The role of perceptual contrast non-linearities in image transform coding. *Im. Vis. Comp.*, 18(3):233–246, 2000.

[9] S. Winkler. Issues in vision modeling for perceptual video quality assessment. *Signal Process.*, 78:231–252, 1999.

[10] J. Malo, A.M. Pons, and J.M. Artigas. Subjective image fidelity metric based on bit allocation of the human visual system in the DCT domain. *Im. Vis. Comp.*, 15:535–548, 1997.

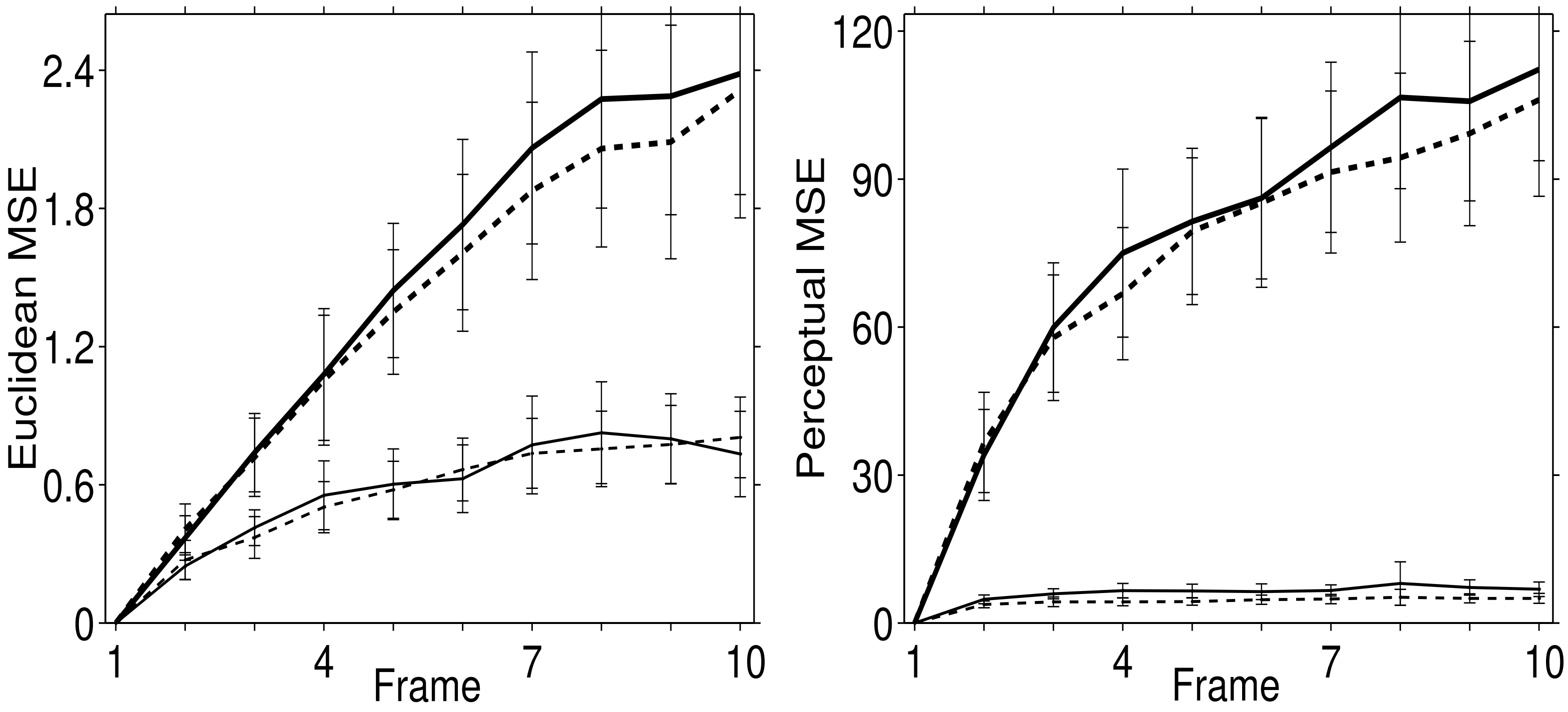

Fig. 1: *MSE measures. Thick lines correspond to* JPEG-like *quantisers, thin lines to* enhanced *quantisers, solid lines to suboptimal VSBMA and dashed lines to optimal VSBMA.*

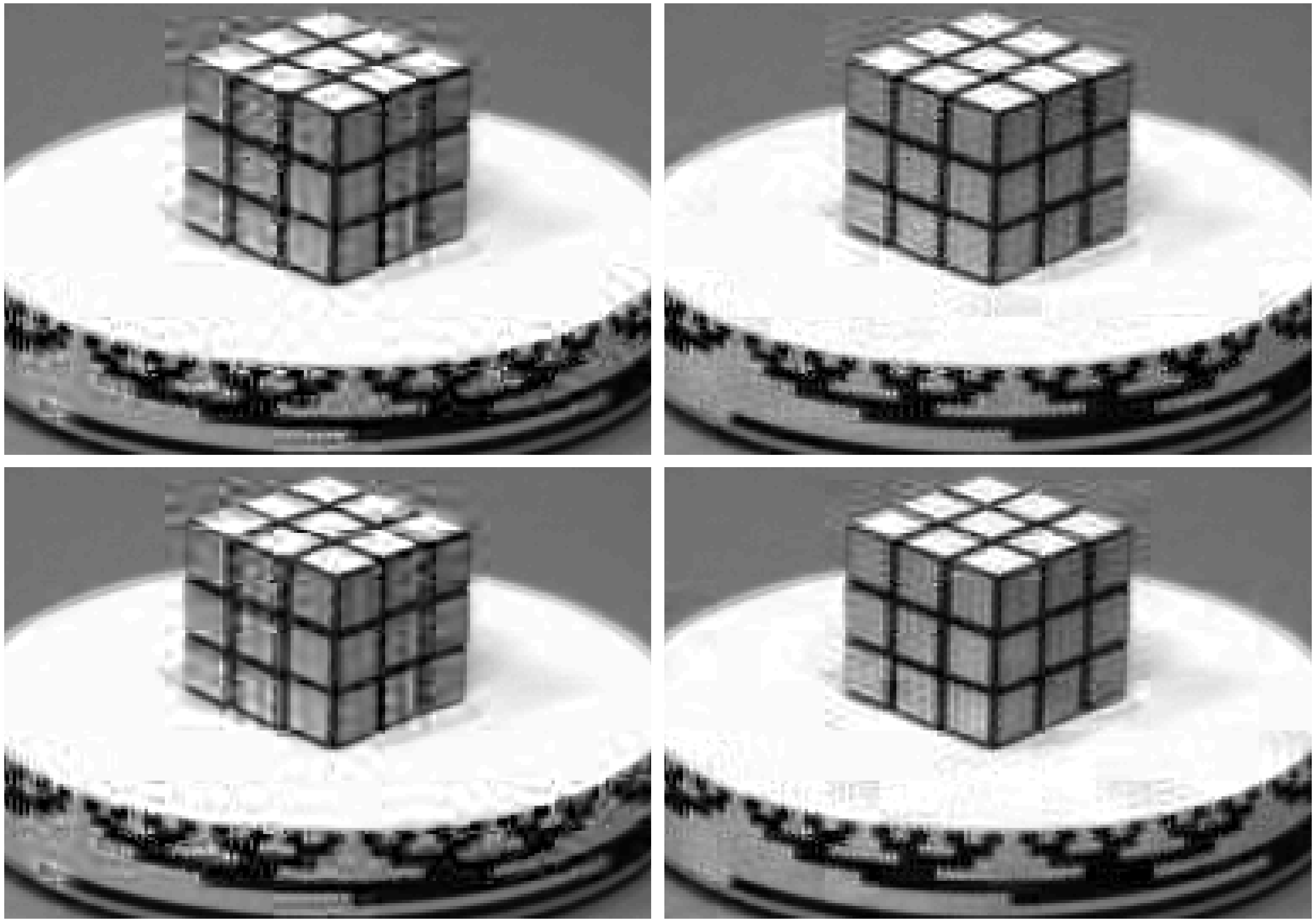

Fig. 2: *Reconstructed frames. The upper and lower frames correspond to suboptimal and optimal VSBMA, respectively. The left and right frames correspond to* JPEG-like *and* enhanced *quantizer, respectively.*