Motion Vector Quantization in a Rate-Distortion Framework

Shankar L. Regunathan and Kenneth Rose
Department of Electrical and Computer Engineering
University of California, Santa Barbara
CA, 93106*

Abstract

Motion Compensation is formulated as quantization of the true motion field and optimized within an overall rate-distortion framework. The desired accuracy in motion estimation is achieved by designing the codebook of the motion quantizer for the target rate. We propose an iterative algorithm for the codebook design which optimizes the compression performance of the coder by directly minimizing the overall rate-distortion cost. A multistage structure is imposed on the codebook to obtain a practical motion quantization scheme, which is optimized through a joint design of the stages. Simulation results on video sequences are presented which demonstrate that RD optimized motion vector quantization outperforms conventional motion compensation and can achieve significant gains of 1.3 dB in the PSNR of the reconstructed picture. Further, the multiresolution nature of multistage motion quantization makes it applicable for scalable video coding applications.

1 Introduction

Video compression algorithms rely on prediction to remove temporal redundancies. The current frame is approximated by a motion compensated version of the previous frame and the prediction (or approximation) error is compressed.

The most common method of motion estimation is the block matching algorithm [7]. In block matching, the current frame is divided into a set of disjoint blocks. Each block is compared to blocks in the previous reconstructed frame, within a given search region, to find the best match. The matching criterion is the prediction error, commonly defined as the mean squared error or mean absolute difference. The motion vector is defined as the displacement of the current block from its best match. A motion vector for each block is transmitted to the decoder, along with the quantized prediction error (commonly called the residual). The decoder uses this information to reconstruct the current frame.

It is well known that the performance of the coding scheme depends, to a large extent, on achieving the "right" level of accuracy in motion representation. A more accurate rendering of the motion field leads to a better prediction, thus requiring a lower rate for adequate compression of the prediction error. On the other hand, a greater side information rate is needed for specifying to the decoder this accurate motion field. Conventionally, the accuracy of the encoded motion field has been selected as a heuristic compromise between these two conflicting objectives.

In this work, we optimize the accuracy in motion representation so as to maximize the compression efficiency of the overall coding scheme. To this end, we attack the problem of motion compensation in an overall rate-distortion (RD) framework by viewing it as quantization of the true motion field.

The organization of the paper is as follows: In section 2, we review some existing methods to optimize the motion estimation accuracy. In section 3, motion estimation is reformulated as quantization of the motion field via a codebook of motion vectors. We show that the desired accuracy in motion field representation can be achieved via a well designed codebook for the motion quantizer. Section 4 outlines an iterative algorithm to design the motion vector codebook so as to optimize its rate-distortion performance. To obtain a motion codebook with feasible storage requirements, we introduce multistage quantization of motion. Section 6 presents simulation results which illustrate the gains obtained by RD optimized motion quantization over conventional motion estimation techniques.

2 Conventional Approach to Motion Estimation

In conventional motion estimation algorithms, the accuracy of the motion field representation is limited by the block sizes and the precision used to encode the motion vectors. Smaller block sizes allow encoding the motion field with greater spatial resolution and hence lower prediction error. However they require transmission of a larger number of motion vectors per frame. Typically, a block size of 16X16 is chosen as a reasonable compromise for most video coding applications. The effect of using smaller block sizes has been investigated in [1], [12] where a criterion based on prediction error was used to split the blocks and the resulting motion vectors were quadtree encoded. The objective in these studies was to minimize the prediction error

21

without significantly increasing the number of motion vectors.

An alternate approach to increasing the spatial resolution, is to use greater precision to encode the motion vectors so as to obtain better motion field representation. Interpolation on the previous reconstructed frame allows motion estimation to be performed at subpixel precision and leads to better prediction. However, more bits are spent on each motion vector in this case. The efficiency of motion estimation at subpixel precision has been studied in [4].

Previous studies on motion estimation [1] [4] [12] have focussed on optimizing the tradeoff between prediction error cost and the side information rate spent on motion vectors. To the best of our knowledge, there has been no explicit analysis of the relation between the the accuracy of motion field representation and the compression performance of the coding scheme. In this work, we propose to optimize the motion field encoding in an *overall RD* framework with the objective of directly maximizing the *compression efficiency*.

3 Motion Vector Quantization

We first formulate motion estimation from a signal compression perspective. We view the selection of motion vectors as a quantization operation applied to the true, dense motion field. We therefore propose to quantize this dense motion field directly as described next.

The current frame is divided into a set of small blocks (4X4, 8X8). These blocks represent a high resolution partition of the frame. A lower spatial resolution is simultaneously obtained by grouping several adjacent small blocks into a superblock (whose size is the same as the conventional block size 16X16). We represent the motion field of each superblock by a single "super motion vector"(SMV) index which accounts for the motion of all the small blocks in the superblock. Note that, in principle, the SMV can allow each block in a superblock to have a different motion vector. A motion codebook contains the candidate SMVs from which the best match is selected to represent the motion field of the super block. Thus we quantize the motion of all the blocks in a superblock together (or vector quantize the motion field of the superblock).

The accuracy of the motion field quantization depends on the contents of the motion codebook. At the extreme, if the codebook entries are constrained to have uniform motion for all the blocks in the superblock, we will be estimating motion with spatial resolution corresponding to superblock size, i.e., the motion estimation is equivalent to conventional block matching. The motion field can be quantized at higher spatial resolutions by allowing the codebook to consist of entries which allow different motion for the individual blocks in a superblock. Similarly, the precision of the codebook entries determines the precision of motion estimation. Thus, the problem of optimizing the accuracy of motion estimation is now one of designing a codebook of SMVs, tailored to maximize the overall compression performance of the coding scheme.

4 Rate-Distortion Optimization of Codebook

A design algorithm for a fixed rate motion codebook was described in [8]. But the method suffered from statistical mismatch and the authors found it unsuitable for coding applications. An adaptive extension of this idea was proposed in [11], where a variable rate motion codebook was designed online in a backward adaptive manner. However, these design methods focussed only on minimization of the prediction error instead of optimizing the overall rate-distortion performance.

Our objective is to design a codebook of SMVs which optimizes the compression efficiency of the coding scheme at the target rate. This is achieved by minimizing the true RD cost associated with the motion codebook. Here, the rate R represents the total number of bits needed for compression, which includes both the side information rate for encoding the motion vectors and the rate spent on quantizing the prediction error. The distortion D represents the *overall* reconstruction error and not merely the prediction error after motion compensation.

We design the motion codebook so as to minimize the overall distortion D, subject to a constraint on the rate $R \le R_{max}$. (For a generic VQ codebook design procedure for such constrained optimization see [2]. Another work that is particularly relevant here is [10].) The optimization problem can be rewritten as an unconstrained minimization of the Lagrangian, $D + \lambda R$, where λ is the Lagrange multiplier. We next derive a design algorithm which minimizes this Lagrangian cost.

4.1 Design Algorithm

Start with an initial SMV codebook which encodes the motion field with a very high degree of accuracy. A correspondingly high side information rate is associated with this codebook. We iteratively refine this codebook to achieve the desired level of accuracy in motion field representation.

Given a training set of video sequence, iterate the following steps:

Step 1. *To each super block, assign the codebook entry which minimizes the Lagrangian cost of encoding this block; $D + \lambda R$, where D and R are the error and total rate associated with encoding this block. This assignment forms a partition of the training set among the codebook entries.*

Step 2. *Redesign the entropy code matched to the population of the training subsets assigned to the indices, so as to reduce the average side information rate.*

Step 4. *If the convergence criterion is met, stop. Otherwise go to Step 1.*

The algorithm descends in the Lagrangian cost in practice and we obtain a locally optimal motion codebook.

During the iterative steps of the algorithm some motion vector entries in the codebook may not have

any blocks assigned to them. These motion vector indices are pruned out and the final codebook is a smaller subset of the initial codebook. The algorithm tailors the motion vector codebook and its entropy code to optimize the RD performance, while achieving the desired accuracy in motion estimation.

The Lagrange multiplier can be adjusted to meet the target rate constraint. A higher value for Lagrange multiplier results in the design of a compression scheme with a reduced overall rate. Further, we observed that the design algorithm resulted in a smaller final codebook, with reduced accuracy in motion estimation and lower side information rate. These results validate the well known fact that lower motion estimation accuracies are desirable at lower target rates because: (i) the higher quantization noise in the previous reconstructed frame results in poorer prediction performance and (ii) the side information rate expended on encoding the motion vectors is relatively more expensive. Thus the design performs an implicit rate allocation for the motion quantization at a given overall compression ratio.

Note that step 1 of the algorithm is computationally intensive as it requires compressing the prediction error of the superblock for every candidate motion vector in order to estimate the Lagrangian cost. In practice, however, we used a fast multiple-survivor technique. As a first step, a small number of candidate motion vector indices were chosen from the codebook for each superblock by using a suboptimal criterion which is computationally simple. A criterion based on prediction error, such as mean absolute difference, can be used for this purpose. We perform the residual quantization and compute the Lagrangian cost for each of the survivors and choose the best motion vector index to represent the block. With a modest number of survivors ($N = 5 - 10$), a drastic reduction in complexity was obtained with a negligible drop in performance.

5 Multistage Motion Vector Quantization

The design algorithm produces an unstructured SMV codebook which, is locally optimal, may have an impractically large number of entries, particularly at high target bit rates. For example, a motion estimation scheme with M blocks in a superblock and a motion vector search range of $-B$ to B with integer pel accuracy, might have as many as $(2B + 1)^M$ entries. The storage requirements for such a large codebook would often make the unstructured motion quantizers infeasible. A standard method for reducing the VQ storage requirements is to use codebooks which satisfy a structural constraint.

We impose a multistage structure on the SMV codebook in order to obtain a practical motion quantization scheme. The first stage of the motion quantizer performs motion estimation at the lowest spatial resolution (superblock size) and integer pel precision. Subsequent stages encode the motion field representation at higher accuracies through increasing levels of spatial resolution (smaller blocks) and

subpixel precision. Each stage of the quantizer has a codebook of motion vector indices associated with it. The overall SMV codebook is the product code formed by the individual codebooks. The multistage SMV codebook has a memory requirement comparable to that of conventional motion estimation schemes. Our basic design algorithm is extended to enforce the structural constraint of multistage motion codebook design. All stages are jointly designed to optimize the rate-distortion performance. The simulation results presented in the next section compare the compression performance of RD optimized multistage motion quantization to that conventional motion estimation.

We note that multistage motion quantization has a natural hierarchy of motion estimation accuracy. A coarse representation of the motion field is provided by the first stage of the motion quantizer, while the subsequent stages refine this representation. Multistage motion quantization is therefore particularly appropriate for scalable video coding, where different motion estimation accuracies are needed for the different layers of compressed video [5]. The motion representation produced by the multistage motion quantizer can be specifically optimized for such applications as in [9]. For example, the first stage of the motion codebook, can be optimized for the rate of the base layer. and the subsequent codebook stages can be optimized for the total rate. Thus applications such as layered coding for ATM networks [3] can utilize RD optimized multistage motion quantization profitably.

6 Results

To test the performance of optimal motion quantization, we carried out simulations on video sequences corresponding to *sif* resolution images. The training sequence was obtained from the sequence *Table Tennis*. We use the training set to design a two stage motion quantizer. The first stage of the quantizer performs motion estimation with a 16 X 16 block size and integer pel precision. The second stage of the quantizer corresponds to an increased spatial resolution by using a 8 x 8 block size and integer pel precision. We restricted the codebook entries to be confined to a search range of (-15,15) for the first stage and (-2,2) for the second stage.

The design algorithm was used to jointly design the codebook of both the stages of the quantizer so as to minimize the rate-distortion cost. The Lagrangian multiplier was varied to meet the target rate constraint. The prediction error was encoded by using a residual quantization similar to H.263 [6]. The RD optimized motion codebook was then used to compress the test sequence *Football*. For comparison we used standard motion compensation corresponding to block size of 16 X 16 and integer pel precision to compress the test sequence to the same total rate. Table 1 shows the average PSNR in dB of reconstructed frames of the sequence *Football*. A sample reconstructed frame is shown in figure 1. The total rate is 0.5 bpp or equivalently 335 kbps (at a frame rate of 8 fps). The results demonstrate gains of 1.3 dB in average PSNR of the reconstructed picture over conventional block matching algorithms over 1 second of encoded video.

PSNR (in dB) of reconstructed "Football"	
Standard Motion Compensation	30.29
RD Optimized Motion Quantization	31.59

Table 1: Performance of Standard Motion Compensation and RD Optimized Motion Quantization on the Sequence "Football" at a rate of 0.5 bpp. The sequence "Table Tennis" was used to optimize the motion quantizer

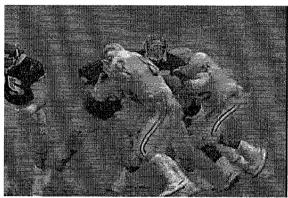Further gains can be achieved by incorporating backward adaptive design of the codebook as described in [11].

Figure 1: Sample reconstructed frame of "Football" for conventional motion estimation (top), and RD optimized motion quantization (bottom) for rate $R = 0.5$ bpp.

7 Conclusions

We use an overall rate-distortion framework to optimize the performance of motion compensation in video coding schemes. A codebook of motion vectors was used to directly quantize the dense motion field and the accuracy of motion estimation was optimized by designing the codebook for the target rate. An iterative algorithm was proposed for the design of the motion codebook which minimizes the Lagrangian rate-distortion cost. A practical motion quantization scheme is developed by imposing a multistage structure on the codebook, which is then optimized by joint

design of the stages. Simulations on benchmark video sequences demonstrate that RD motion quantization significantly outperforms conventional motion estimation and gains 1.3 dB in PSNR of the reconstructed picture. Scalable video coding applications can also profit from the hierarchical motion representation provided by multistage motion quantization.

References

[1] M. H. Chan, Y. B. Yu and A. G. Constantinides, "Variable size block matching motion compensation with applications to video coding," *IEE Proceedings I*, vol.137, pp. 205-12, Aug. 1990.

[2] P. A. Chou, T. Lookabaugh and R. M. Gray, "Entropy-constrained vector quantization," *IEEE Trans. on Accous. Speech and Sig. Proc.*, vol. 37, pp. 31-42, Jan. 1989.

[3] M. Ghanbari, and V. Seferidis, "Efficient H.261-based two-layer video codecs for ATM networks," *IEEE Trans. on Circuits and Systems for Video Tech.*, vol. 5, pp. 171-5, April 1995.

[4] B. Girod, "Motion-compensating prediction with fractional-pel accuracy," *IEEE Trans. on Com.*, vol. 41, pp. 604-12, April 1993.

[5] U. Horn and B. Girod, "Performance Analysis of Multiscale Motion Compensation Techniques in Pyramid Coders," *Proc. ICIP 1996*.

[6] ITU-T Recommendation H.263 "Video coding for low bit rate communication," Dec. 1995.

[7] J. R. Jain and A. K. Jain, "Displacement measure and its application in interframe image coding," *IEEE Trans. on Comm.*, vol. COM-29, pp. 1799-808, Dec. 1981.

[8] Y. Y. Lee and J. W. Woods, "Macro motion vector quantization," *Proc. of the SPIE*, vol. 2501, pt.2, pp. 778-86, 1995.

[9] K. Ramchandran, A. Ortega, and M. Vetterli, "Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders," *IEEE Trans. on Image Proc.*, vol. 3, pp. 533-45, Sept. 1994.

[10] R. P. Rao and W. A. Pearlman, "Alphabet-constrained vector quantization," *IEEE Trans. on Inform. Theory.*, vol. 39, pp. 1167-79, July 1993.

[11] S. L. Regunathan and K. Rose, "Motion Compensation by Adaptive Vector Quantization," *Proceeding of HDTV'96*.

[12] G. J. Sullivan and R. L. Baker, "Rate-distortion optimized motion compensation for video compression using fixed or variable size blocks," *GLOBECOM '91*.