Flexible frame-reordering and multi-temporal motion estimation for scalable MPEG encoding in mobile consumer equipment

DOI:
10.1109/ICCE.2002.1014057

Document status and date:
Published: 01/01/2002

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Flexible frame-reordering and multi-temporal motion estimation for scalable MPEG encoding in mobile consumer equipment

Stephan Mietens[1], Peter H.N. de With[1,2] and Christian Hentschel[3]

[1]Eindhoven University of Technology, EESI / [2]CMG Eindhoven, 5600 MB Eindhoven, The Netherlands
[3]Philips Research Labs., 5656 AA Eindhoven, The Netherlands

ABSTRACT

We present a new scalable three-stage motion estimation technique that allows scaling of the computational effort and memory bandwidth usage by a factor of 14. At full processing, our technique slightly outperforms a 128×128 full search motion estimation.

1. INTRODUCTION

In this paper, it is our objective to design a scalable MPEG encoding system, featuring scalable video quality and a corresponding scalable resource usage [1]. Such a system enables advanced video encoding applications on a plurality of low-cost or mobile consumer terminals, having limited resources (available memory, computing power, stand-by time, etc.) as compared to high-end computer systems. A computationally expensive corner stone of an MPEG encoder is motion estimation (ME), which is used to achieve high compression rates by reducing the temporal redundancy. The search of motion vectors in ME requires significant computation power, because high temporal distances between reference frames lead to large search areas.

Section 3 presents a new technique to perform ME so that considerable computing power and memory bandwidth can be saved for resource-constrained (e.g. mobile) applications. These savings are obtained by performing an initial ME with the video frames at the entrance of the encoder. This estimation is exploited to derive efficiently the desired motion vector fields needed for the final MPEG encoding process. Furthermore, the quality of full-search ME can be obtained with an optional refinement stage. Section 4 shows experimental results and Section 5 concludes the paper.

2. PROBLEM STATEMENT

When compared to the full search, popular algorithms like the New Three Step Search and the Center-Biased Diamond Search [2] provide a good quality of motion vector fields at low cost. Further reduction of the computational effort is achieved by using recursive motion estimation (RME) [3][4] that derives candidate motion vectors from previously computed motion vectors in both the current motion vector field (so-called "spatial" candidates) or the previous motion vector field (so-called "temporal" candidates). Currently, RME algorithms are used in such a way that a vector field is used only once for the computation of the next field and without modification. A more sophisticated approach for ME is presented in [5], featuring a two-step estimation process and enhanced vector-field prediction. The first step of this approach is a coarse RME to estimate forward vector-fields. The second step uses the vector fields computed in the first step as prediction and performs an additional RME. Vector fields used for prediction are scaled to the required temporal distance.

The problem of the aforementioned algorithms is that a higher frame distance between reference frames hamper accurate ME. Furthermore, the algorithms do not provide sufficient scalability. For these reasons, we introduce a new three-stage ME that works not only for the typical MPEG GOP structures, but also for more general cases. Furthermore we introduce a new technique called *multi-temporal* ME. These new techniques give more flexibility for a scalable MPEG encoding process.

3. NEW THREE-STAGE MOTION ESTIMATION

Our new ME is carried out in three stages as follows.

- *Stage 1.* Prior to defining a GOP structure, we perform a simple RME for every received frame X_n and compute the forward motion-vector field $(X_{n-1} \rightarrow X_n)$ and than the backward field $(X_{n-1} \leftarrow X_n)$. Since succeeding frames are used (temporal distance is 1), the RME provides high quality/accurate vector-fields.

Stage 1 is scalable in computational effort and memory by e.g. omitting the computation of the backward vector fields or computing only significant parts of a vector field.

- *Stage 2.* After defining a GOP structure, all vector fields required for MPEG encoding are approximated by appropriately accessing multiple available vector fields and combining them. For example, $(I_0 \rightarrow B_2) = (I_0 \rightarrow B_1) + (B_1 \rightarrow B_2)$.

Stage 2 introduces a new concept called *multi-temporal* ME. It means that the computation of a vector field is based on two other vector fields (as shown above) and/or

0-7803-7300-6/02 $17.00 © 2002 IEEE

the total prediction is based on various vector fields. The latter option can be used for estimating special motions like real velocity and acceleration.

- *Stage 3.* For final MPEG motion estimation in the encoder, the computed approximated vector fields from the previous stage are used as an input. Beforehand, an optional refinement of the approximations can be performed with a second iteration of RME.

Stage 3 is intended for high-quality applications to reach the quality of a conventional MPEG motion estimation algorithm. On the other hand, if this refinement stage is omitted, the new technique can take advantage of reduced computational effort, because the processing of vector fields in stage 1 and 2 is much simpler than ME itself when applied to frames with a high temporal distance.

4. EXPERIMENTS AND RESULTS

Figure 1 shows a proof-of-concept of the new scalable motion-estimation technique, based on a GOP size of $N = 16$ and $M = 4$ (thus "IBBBP" structure). We use the RME taken from [3] (limited to pixel-search) in stage 1 and 3 because of its simple design. Note that the RME for this proof-of-concept does not yet take advantage of the reduced temporal distance during the initial stage at the entrance of the encoder. The area with the white background shows the scalability of the quality range that results from downscaling the amount of computed motion-vector fields. Each vector field requires 14% of the effort compared to a 100% simple RME based on 4 forward vector fields and 3 backward vector fields when going from one to the next reference frame. If all vector fields are computed and the refinement stage 3 is performed, the computational effort is 200% (not optimised). The numerical results of the scalability in quality and effort in our experiment are as follows. The average SNR of our ME technique of the reconstructed P- and B-frames prior quantisation are 25.16 dB, 24.55 dB, 23.52 dB, 22.48 dB and 18.58 dB for 200%, 157%, 100%, 57% and 14%, respectively. For a full quality comparison (200%), we consider full-search block-matching. A full-search ME with a search window of 128×128 pixels results in an average SNR of 25.05 dB. Thus our new technique outperforms full search by 0.11 dB for this test sequence, although the full search was applied on an exhaustive 128×128 search window.

5. CONCLUSIONS

We have presented a new scalable technique for ME in MPEG encoding. The scalability can be exploited to reduce the computational effort over a large range making our system feasible for low-cost mobile MPEG systems.

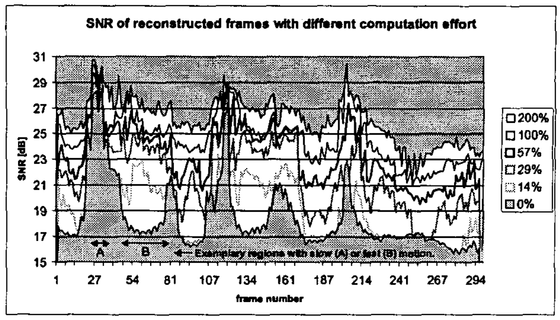

Figure 1. SNR of reconstructed B-frames with different computation effort, P-frames are not shown for the sake of clearity ($N = 16$, $M = 4$). The percentage shows the reduction of the amount of computations by omitting the computation of vector fields in the precomputation stage.

The ME technique has been split into a precomputation stage and an approximation stage. Optionally, a refinement stage can be added to come to the quality of a conventional MPEG encoder (or even outperform it). In the precomputation stage, we used a simple recursive block matcher to find rather good motion estimates because the frames are processed in time-consecutive order. In the approximation stage, vector fields are scaled and added or substracted (thus having multi-temporal references), which is less complex than performing advanced vector searches. The computation of e.g. the backward motion estimation can be omitted to save computational effort and memory bandwidth usage.

Our proposal allows scaling of the computational effort by a factor of 14, resulting in a global variation of 7 dB in picture quality. Furthermore, our system slightly outperforms a 128×128 full-search motion estimation in quality at a lower complexity.

REFERENCES

1. C. Hentschel et al., "Scalable Algorithms for Media Processing," *IEEE Int. Conf. on Image Processing (ICIP 2001)*, vol. 3, pp. 342–345, Oct. 2001.
2. J. Tham et al., "A Novel Unrestricted Center-Biased Diamond Search Algorithm for Block Motion Estimation," *IEEE Trans. Circuits and Systems for Video Technol.*, vol. 8, pp. 369–377, Aug. 1998.
3. P. de With, "A simple recursive motion estimation technique for compression of HDTV signals," *IEEE Int. Conf. Image Proc. and its Applic. (IPA'92)*, pp. 417–420, 1992.
4. G. de Haan et al., "True-Motion Estimation with 3-D Recursive Search Block Matching," *IEEE Trans. Circuits and Systems for Video Technol.*, vol. 3, pp. 368–379, Oct. 1993.
5. F.S. Rovati et al., "An Innovative, High Quality and Search Window Independent Motion Estimation Algorithm and Architecture for MPEG-2 Encoding," *IEEE Trans. Consum. Electron.*, vol. 46, pp. 697–705, Aug. 2000.