Enhanced prediction for motion estimation in scalable video coding

Citation for published version (APA):
Loomans, M. J. H., Koeleman, C. J., & With, de, P. H. N. (2010). Enhanced prediction for motion estimation in scalable video coding. In *Proceedings of the 17th IEEE International Conference on Image Processing (ICIP 2010), 26-29 September 2010, Hong Kong, Hong Kong* (pp. 1301-1304). Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/ICIP.2010.5650796

DOI:
10.1109/ICIP.2010.5650796

Document status and date:
Published: 01/01/2010

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

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Download date: 17. Nov. 2023

ENHANCED PREDICTION FOR MOTION ESTIMATION IN SCALABLE VIDEO CODING

Marijn J.H. Loomans [a,b] *Cornelis J. Koeleman* [a] *Peter H.N. de With* [b,c]

[a] VDG Security BV
Radonstraat 10-14
2718 TA, Zoetermeer, NL

[b] Eindhoven University of Technology
Den Dolech 2
5612 AZ, Eindhoven, NL

[c] CycloMedia Technology BV
Achterweg 48
4181 AE, Waardenburg, NL

ABSTRACT

In this paper, we present a temporal candidate generation scheme that can be applied to motion estimators in Scalable Video Codecs (SVCs). For bidirectional motion estimation, usually a test is made for each block to determine which motion compensation direction is preferred: forward, bidirectional or backward. Instead of simply using the last computed motion vector field (backward or forward), giving an asymmetry in the estimation, we involve both vector fields to generate a single candidate field for a more stable and improved prediction. This field is generated with the aid of mode decision information of the codec. This single field of motion vector candidates serves two purposes: (1) it initializes the next recursion and (2) it is the foundation for the succeeding scale in the scalable coding. We have implemented this improved candidate system for both HPPS as EPZS motion estimators in a scalable video codec. We have found that it reduces the errors caused by occlusion of moving objects or image boundaries. For EPZS, only a small improvement is observed compared to the simple candidate scheme. However, for HPPS improvements are more significant: when looking at individual levels, motion compensation performance improves by up to 0.84 dB and when implemented in SVC, HPPS slightly outperforms EPZS.

Index Terms— Motion Estimation, Scalable Video Coding, Real-time Systems, Embedded Systems, Parallel Algorithms.

1. INTRODUCTION

Motion estimation is an essential function in state-of-the-art video coding, both in important standards like H.264/AVC [1], as in Scalable Video Coding (SVC), such as the well-known MC-EZBC [2] and the surveillance oriented SVC proposed by the authors [3].

Motion estimators have evolved continuously since their first appearance. Initially, full-search or exhaustive-search motion estimators were proposed, which were improved by utilizing a multi-stage approach. In such recursive approaches, a restricted set of candidates is tested according to a certain pattern, after which the best match is used as the starting point for the next step, which is then tested with the same or another pattern. This process repeats itself for a few iterations, until the optimal vector is found or the maximum amount of iterations is reached. Well-known motion estimators that utilize this multi-stage approach are TSS (Three-Step-Search) [4], ARPS-3 (Advanced Root Pattern Search) [5]. Algorithms like PMVFAST [6] and EPZS (Enhanced Predictive Zonal Search) [7] additionally introduced early-stop criteria that terminate the iterative processing when a certain condition is satisfied, e.g. the error metric drops below a certain threshold.

In sequential software implementations, the Sum of Absolute Differences (SAD) calculation occupies a large part of the complexity. Many of the design decisions in the aforementioned motion estimators are based on this characteristic. However, in current parallel architectures with hardware accelerators, block operations such as the SAD calculation, can be processed significantly faster than in traditional sequential general-purpose processors. As a result, the bottleneck of the motion estimation algorithm shifts from computations to memory bandwidth. Furthermore, in SVCs, the motion estimation processing has to comply with the layering in scalable coding, which is different than in traditional hybrid video coding. Since motion is estimated at various temporal levels, temporal distances of 1, 2, 4 and 8 frames occur. In previous work [8], the HPPS (Highly Parallel Predictive Search) motion estimator satisfied these specific design requirements. HPPS features good mapping on parallel and multi-core architectures, has a fixed computational load and performs well at various temporal levels. However, HPPS had a very simple candidate generation system, which used the last computed motion vector field (backward or forward), giving an asymmetry in the estimation. Therefore, in this paper we propose an enhanced candidate vector generation which (1) generates a single candidate field for a more symmetrical candidate vector framework and (2) utilizes per block mode-decision of the codec for improved prediction.

This paper is organized as follows. Section 2 presents the motion estimation and mode decisions within an $2D+t$ SVC. A summary of the HPPS motion estimation algorithm is presented in Section 3 together with the proposed improvements. Section 4 discusses our results, and we conclude in Section 5.

2. MOTION ESTIMATION IN SVC

In $2D+t$ SVC, many different temporal configurations are possible. Trade-offs can be made between i.e. coding quality, end-to-end delay and memory usage. For more details on this topic, the reader can refer to [3]. For our application, we adopted the Bi-Directional configuration with Low Delay, abbreviated `BDLD`, since it features sufficiently high quality with low end-to-end delay and memory access. In a 4-level temporal configuration, the `BDLD` temporal configuration uses bidirectional prediction at the two lowest levels, and single prediction at the two highest levels, as shown in Figure 1. Bold arrows indicate the motion-compensated lifting steps, in which the Motion Estimation (ME) is situated as well. The figure also shows the various aspects of the motion estimation process in SVC with the bidirectional ME in the top two layers, and the ME over large temporal distances in the bottom.

In SVC, a mode decision is made for each block that is bidirectionally estimated. This mode decision determines if bidirectional motion compensation is beneficial for compression, or introduces artifacts due to one of the two motion vectors being inaccurate. This situation might occur around boundaries of moving objects that oc-

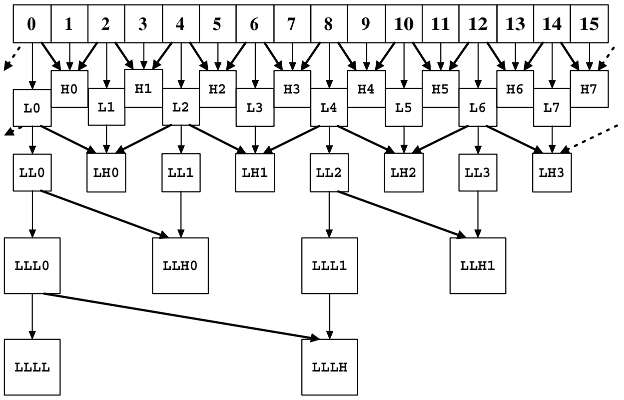


Fig. 1. The BDLD temporal configuration for an SVC with a four-level temporal transform. Bold arrows indicate the lifting steps that include motion estimation and compensation.

clude the background. Three modes are used in our proposed SVC: forward, bidirectional and backward. The SAD of the block is calculated for each of these three modes and the mode with the lowest SAD is chosen.

3. HIGHLY PARALLEL PREDICTIVE SEARCH (HPPS)

HPPS was proposed in [8] to provide a motion estimator suited for SVC while facilitating a smooth mapping on parallel and multi-core architectures. This section will summarize HPPS and explain the Parallelogram-Shaped Scanning (PSS) pattern, the candidate generation and the temporal candidate generation in an SVC, in Sections 3.1, 3.2 and 3.3, respectively.

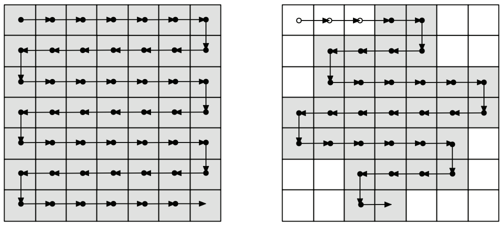
3.1. Parallelogram-Shaped Scanning Pattern

For HPPS we proposed the use of two accelerators that perform common tasks on blocks. First, a block SAD accelerator, that calculates the SAD between two image blocks. Second, a block cyclic rotation accelerator, which rotates the pixels of an image block in left, right and bottom directions.

With these accelerators, a full-search can be implemented easily. Figure 2(a) shows the use of the SAD accelerator (illustrated by the gray blocks) and the rotation accelerator (illustrated by arrows) to perform a 7 × 7 full search. The Parallelogram-Shaped Scanning pattern, called PSS, is visualized in Figure 2(b). The same accelerators are used, however, reducing the amount of SAD calculations by up to 50% without significantly reducing the quality of the found vector. The width of the top and bottom rows of the PSS pattern can be increased so that the time required to fetch the next row from the reference image, is filled with useful SAD calculations.

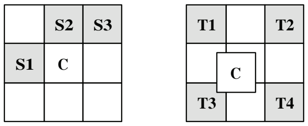
3.2. Candidate generation

HPPS performs a dense search according to the previously discussed PSS pattern around two candidates, one spatial and one temporal. Figure 3 shows how the (a) spatial and (b) temporal candidates are determined by calculating the median of neighboring motion vectors, with C being the current motion vector for which the candidate is calculated.



(a) (b)

Fig. 2. HPPS scanning patterns: (a) full-search and (b) PSS. Arrows illustrate a rotation and gray blocks illustrate the SAD calculation.



(a) (b)

Fig. 3. Motion vector candidates of HPPS: (a) spatial: S1–S3, and (b) temporal: T1–T4. The current motion vector is indicated by C.

3.3. Temporal candidates in SVC

For bidirectional motion estimation, usually a test is made for each block to determine which motion compensation direction is preferred: forward, bidirectional or backward. Instead of simply using the last computed motion vector field as in previous work (backward or forward), giving an asymmetry in the estimation, we involve both vector fields to generate a single candidate field for a more stable and improved prediction.

We build our argumentation starting with the HPPS algorithm. At the lowest level, motion vector fields for the bidirectional estimation are simply reversed, and for higher levels the motion-vector field is multiplied with a factor of two, as is visualized in Figure 4 for a four-level SVC with a `BDLD` temporal configuration.

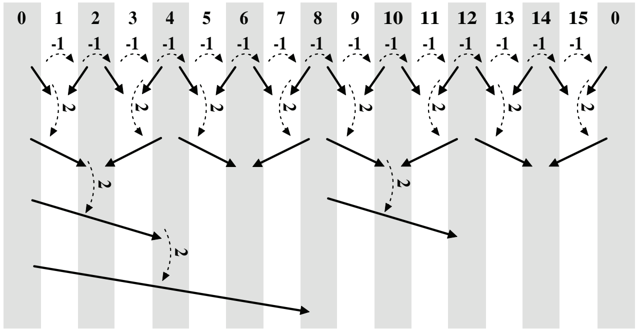


Fig. 4. Temporal candidates in 4-level SVC. Normal arrows indicate ME calculations with candidate vectors represented by the dashed arrows, labeled with the multiplication factor of the motion-vector field.

For moving objects, the uncovered background results in sub-optimal motion vectors which do not correctly represent the motion. For bidirectional ME this uncovered background alternates between forward and backward estimation, and therefore creates erroneous candidates that propagate to the next motion estimation calculation.

The artifacts are mostly visible at the contours of moving objects.

To improve on this simple candidate generation system, we propose to utilize the mode decisions made by the codec, to create a single motion vector field which describes the motion more accurately. Mode decisions are made for each block, describing if the best match is backward, bidirectional or forward. This information is utilized in calculating the temporal candidate as shown in Table 1. From this table it can be seen that the candidate vector is always mapped to the forward direction, ensuring the desired symmetry.

Table 1. Candidate vector calculation based on block mode.

Mode	Candidate vector
Forward	$MV_{cand} = +MV_{fwd}$
Bidirectional	$MV_{cand} = (MV_{fwd} - MV_{bwd})/2$
Backward	$MV_{cand} = -MV_{bwd}$

This combined candidate vector is then propagated in the SVC as shown in Figure 5, where the dashed arrows represent the propagation, and the numbers above these arrows a multiplication factor. The dotted circles indicate groups of (bi)directional motion vectors. Each group only generates one candidate set, and also absorbs one set only. Internally, the received candidates are directly used for the forward motion estimation, and inverted to be used in the backward motion estimation.

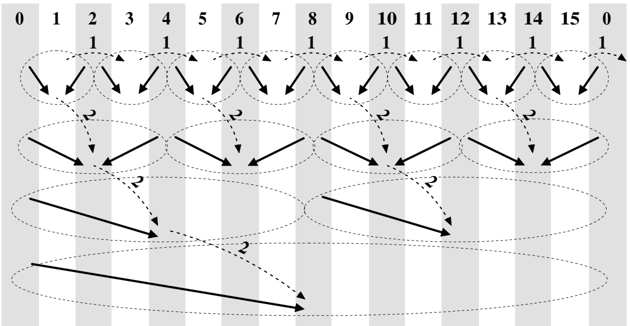


Fig. 5. Enhanced temporal candidates in a 4-level SVC. Normal arrows indicate ME calculations, and circles group (bi)directional vectors. Temporal candidates are represented by the dashed arrows labeled with the multiplication factor of the motion-vector field.

4. EXPERIMENTAL RESULTS

4.1. Various temporal levels

First, we will examine the effectiveness of the proposed ME algorithm with simple and enhanced candidates. We perform PSNR measurements at the 4 levels of the `BDLD` configuration with both HPPS and EPZS motion estimators. Figure 7 shows the results of these measurements for the well-known City sequence, where PSNR is calculated by comparing the original image with the motion-compensated image.

From Figure 7(a), it can be seen that the improvement for the lower levels in HPPS is only marginal. However, at higher levels, Figure 7(b)-(d) show that the improvement is significant. This can also be observed from Table 2, where the mean PSNR for the various configurations is listed. There is no improvement at Level 1, but at Level 3 the improvement is most pronounced at 0.84 dB. It can be noted that the improvement at Level 4 is lower, however, no bidirectional estimation exists at Level 3, so this improvement is fully due to the improved motion vectors at Level 3.

For EPZS, only a very limited improvement can be observed at higher levels from both Figure 7 and Table 2, most likely due to the fact that EPZS always evaluates four temporal candidates, thereby producing a more stable result than a system with only one temporal candidate, such as HPPS.

Table 2. Mean PSNR at various temporal levels for HPPS and EPZS, with simple and enhanced candidates.

	Level 1	Level 2	Level 3	Level 4
Δframes	1	2	4	8
HPPS simple	31.45 dB	30.35 dB	29.04 dB	28.50 dB
HPPS enhanced	31.45 dB	30.55 dB	29.88 dB	28.92 dB
Improvement	0.00 dB	0.20 dB	0.84 dB	0.42 dB

	Level 1	Level 2	Level 3	Level 4
Δframes	1	2	4	8
EPZS simple	31.11 dB	30.25 dB	29.60 dB	28.75 dB
EPZS enhanced	31.07 dB	30.28 dB	29.62 dB	28.77 dB
Improvement	-0.04 dB	0.03 dB	0.02 dB	0.02 dB

4.2. Integrated in the complete SVC

To measure the effectiveness of the enhanced candidates, we have integrated the proposed algorithm in our complete SVC framework. Figure 6 shows the rate-distortion curve for the City sequence using Full Search, HPPS and EPZS motion estimators, with simple and enhanced candidates. In Figure 6(a), the complete rate-distortion curve is given, of which a region around 3 Mbit/s is enlarged in Figure 6(b). This enlarged view shows that the gain for EPZS is small, and more significant for HPPS. With the enhanced predictors, HPPS now even slightly outperforms EPZS.

5. CONCLUSIONS

In this paper, we have enhanced the candidate generation for SVC motion estimators in the following way. Instead of simply using the last computed motion vector field (backward or forward), giving an asymmetry in the estimation, we employ both vector fields to generate a single candidate field for a more stable and improved prediction. In more detail, the information of both vector fields is refined with mode-decision information from the codec to improve the accuracy of the candidates. For each motion block, a test is made within the codec which direction is optimal: forward, bidirectional or backward. This mode decision is then used to generate a single motion vector candidate field. The included symmetry in the construction of the vector field decreases the errors caused by occlusion of moving objects, so that contours of moving objects become less noisy. We have implemented this improved candidate system for both HPPS as EPZS. For EPZS, only a small improvement was observed due to the 4 temporal candidates that are always evaluated. However, for HPPS, improvements are more significant: when looking at individual levels, motion compensation performance improves by up to 0.84 dB and when implemented in the complete SVC, HPPS with enhanced candidates even slightly outperformed EPZS.

6. REFERENCES

[1] T. Wiegand, G.J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the h.264/avc video coding standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, no. 7, pp. 560–576, July 2003.

[2] Peisong Chen and J.W. Woods, "Bidirectional mc-ezbc with lifting implementation," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 14, no. 10, pp. 1183–1194, Oct. 2004.

[3] M.J.H. Loomans, C.J. Koeleman, and P.H.N. de With, "Performance vs. complexity in scalable video coding for embedded surveillance applications," *Visual Communications and Image Processing 2008*, vol. 6822, no. 1, pp. 68220J, 2008.

[4] Reoxiang Li, Bing Zeng, and M.L. Liou, "A new three-step search algorithm for block motion estimation," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 4, no. 4, pp. 438–442, Aug 1994.

[5] G. Ma, K.-K.; Qiu, "Unequal-arm adaptive rood pattern search for fast block-matching motion estimation in the jvt/h.26l," in *Image Processing,. Proceedings International Conference on*, September 2003, vol. 1, pp. I–901–I–904.

[6] A.M. Tourapis, O.C. Au, and M.L. Liou, "Predictive motion vector field adaptive search technique (pmvfast) - enhancing block based motion estimation," in *in the Optimization Model 1.0, in ISO/IEC JTC1/SC29/WG11 MPEG2000/M6194*, 2001, pp. 883–892.

[7] A.M. Tourapis, "Enhanced predictive zonal search for single and multiple frame motion estimation," in *Proceedings of Visual Communications and Image Processing*, 2002, vol. 4671, pp. 1069–1079.

[8] Marijn J. H. Loomans, Cornelis J. Koeleman, and Peter H. N. de With, "Highly-parallelized motion estimation for scalable video coding," in *Image Processing, ICIP. 16th IEEE International Conference on*, 2009.

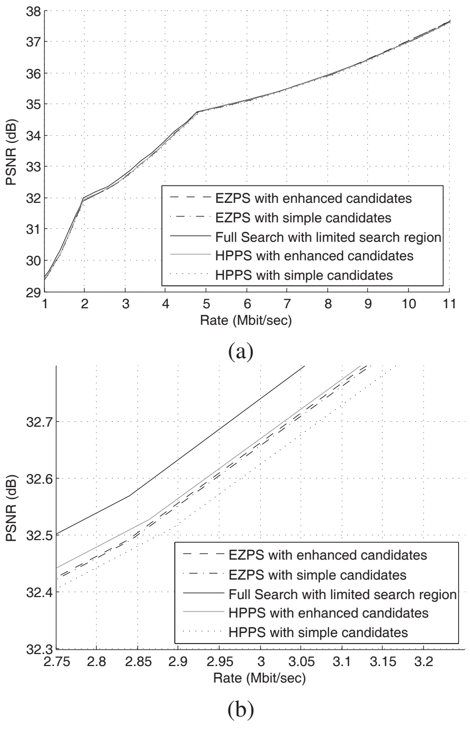


Fig. 6. Rate Distortion curve for the City sequence using the Full-Search (black-solid), HPPS simple and enhanced (black dotted and gray-solid) and EPZS simple and enhanced (black-dash/dotted and black-dashed) with (a) overview and (b) zoom of 3Mbit/s.

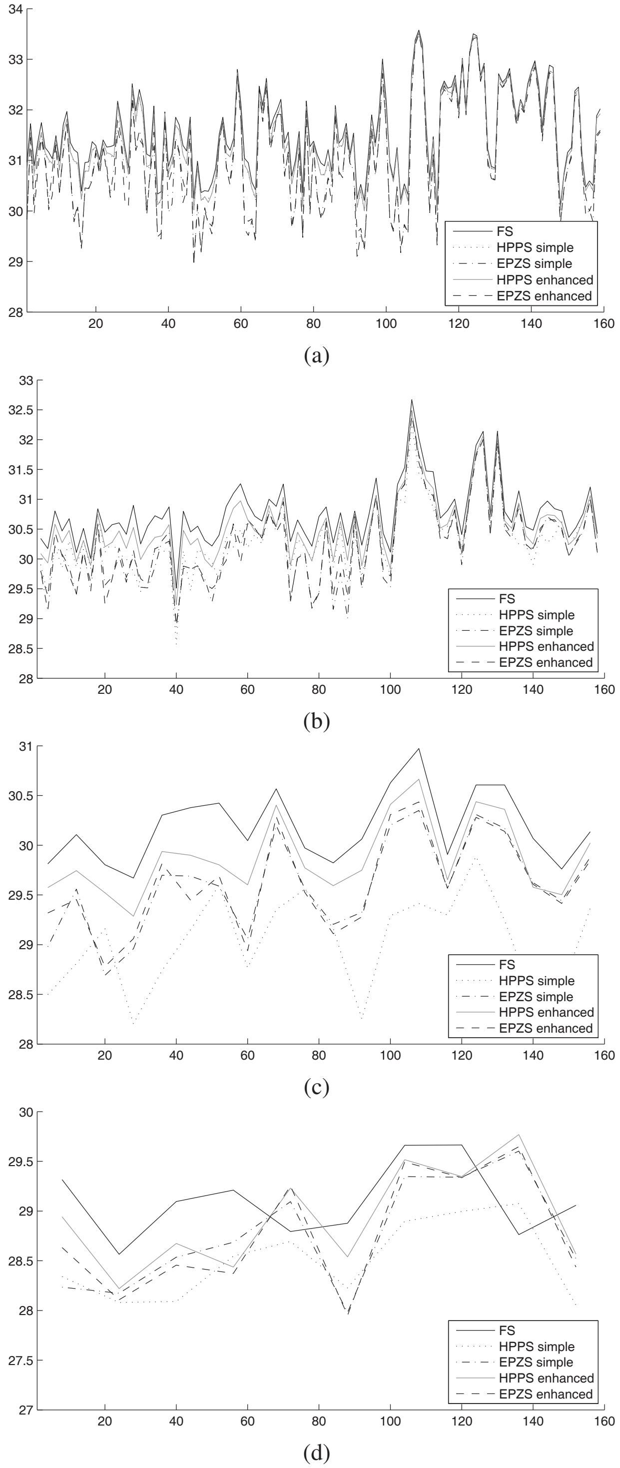


Fig. 7. PSNR measurements per frame for the City sequence using the Full-Search (black-solid), HPPS simple and enhanced (black dotted and gray-solid) and EPZS simple and enhanced (black-dash/dotted and black-dashed) motion estimators for temporal distances of: (a) 1, (b) 2, (c) 4, and (d) 8 frames apart.