OCTAGONAL-AXIS RASTER PATTERN FOR IMPROVED TEST ZONE SEARCH MOTION ESTIMATION

Paulo Goncalves, Marcelo Porto, Bruno Zatt, Luciano Agostini, Guilherme Correa

Video Technology Research Group (ViTech) – Federal University of Pelotas (UFPel), Brazil
{phrgoncalves, porto, zatt, agostini, gcorrea}@inf.ufpel.edu.br

ABSTRACT

Test Zone Search (TZS) is considered the current state-of-the-art fast Motion Estimation algorithm because it presents the best tradeoff between compression efficiency and complexity in comparison to the Full Search strategy. However, it is still one of the most computationally-demanding tools of current video coding standards, such as the High Efficiency Video Coding (HEVC). This paper presents an analysis on the search area opportunities and best match distributions in TZS, which led to the proposal of a novel search pattern in its most complex step, the Raster Search (RS). The new pattern, named Octagonal-Axis Raster Pattern (OARP), allowed an average complexity reduction of 61% in TZS, with a negligible BD-rate increase of 0.0371% in comparison to the original algorithm.

Index Terms—HEVC, Motion Estimation, Test Zone Search, TZS, TZSearch, Complexity Reduction.

1. INTRODUCTION

High Efficiency Video Coding (HEVC) [1] is the most recent video coding standard developed by the Joint Collaborative Team on Video Coding (JCT-VC). HEVC achieves twice the compression of its predecessor, the H.264/AVC standard [2], for the same image quality. However, to reach such compression efficiency, the encoding process can take up to 500% longer than H.264/AVC [3]. This complexity increase of HEVC is mostly due to the larger number of partitions evaluated in the Motion Estimation (ME) process, which is one of the most complex stages in modern video encoders.

The increasing users' demand for high-resolution video sequences led the HEVC standard to allow a flexible frame partitioning structure, which allows both very large (64×64) and very small (4×4) block sizes during the encoding process. Each frame is first divided into a sequence of equal-sized Coding Tree Units (CTUs), which are typically 64×64 blocks. A CTU can be recursively split into smaller blocks, called Coding Units (CUs) in a quadtree structure. These CUs are then further divided into Prediction Units (PUs) and Transform Units (TUs) for prediction and transform purposes, respectively. With this flexible partitioning, ME is evaluated several times with every partitioning combination, consuming about 60-70% of the whole encoding time [4-5]. To simplify the partitioning decision, several authors [6-9] propose different heuristics to prune the quadtree structure, aiming at reducing the complexity of HEVC without decreasing compression efficiency. Even with this pruning, the complexity of ME is still high due to the large number of candidates blocks in reference frames.

The Full Search (FS) algorithm provides pseudo-optimal coding efficiency but incurs in a substantial computational burden because it checks all possibilities within the search area (SA). To speed up ME, the Test Zone Search (TZS) algorithm is used for fast search in most HEVC encoder implementations [10]. TZS employs a four-step strategy based on heuristics that try to predict the best matching, exploiting spatial and temporal correlation in the neighboring areas. However, as this process is still executed for all possible PU sizes in every quadtree possibility [1], ME is still very complex even when TZS is deployed.

Several authors try to reduce the block matching complexity in ME algorithms through the use of low complexity distortion measures, such as the Partial Sum of Absolute Differences (PSAD) [11], or through the use of early-termination schemes [12-13]. However, these strategies usually result in significant compression efficiency losses and turn out to be of difficult implementation in hardware due to lack of regularity.

This paper presents a complexity reduction strategy for TZS based on a novel raster search pattern, called Octagonal-Axis Raster Pattern (OARP). By exploiting the existence of regions within the SA with highest matching probabilities, OARP is able to reduce the TZS complexity by 60.91%, which results in an overall encoding complexity reduction of 21.57% and a negligible Bjontegaard Delta (BD)-rate increase of only 0.0371%. The rest of this paper is organized as follows. Section 2 briefly describes TZS and its complexity. Section 3 presents the statistical analysis of Raster Search in TZS, which led to the proposal of OARP in section 4. Section 5 presents the obtained results and section 6 concludes this paper.

2. TZS OVERVIEW AND COMPLEXITY ANALYSIS

The TZS algorithm consists of four steps: Motion Vector Prediction, First Search, Raster Search, and Refinement, as shown in the flowchart of Fig. 1. Initially, the Motion Vector Prediction, simply referred as Prediction, chooses the best motion vector (MV) out of the median, co-located, left, top and top-right predictors from temporal and spatial neighboring PUs [10]. The predictor that leads to the minimum rate-distortion (RD) cost is chosen as the initial search point for the next step, the First Search.

During First Search, a diamond or square search pattern is traversed, with a stride length varying from 1 to the maximum search range (SR_{max}) in powers of 2. Among all the points tested in the pattern, the best search point is the one that returns the minimum RD cost. First Search step is interrupted if no block with a smaller RD cost is found after the execution of three expansions levels.

The third step is the Raster Search, in which a simple Full Search is performed on a sub-sampled version of the SA. The sub-sampling step is defined by *iRaster*, a constant defined as 5 by default, and it is applied both horizontally and vertically, resulting in a total of 676 candidates blocks in the case of a SA configured as [-64,+64]. Raster Search is only performed if the distance of the SA center to the best block matching found in the second step (First Search) is greater than *iRaster*. Otherwise, it is skipped.

The final step of TZS is the Refinement, which tries to find a better block matching starting a new search around the best position obtained so far. The search pattern used in this step is the same used in First Search, but it updates the center of the SA to the best result obtained in the last iteration. The Refinement step is interrupted if no MV with smaller RD cost is found after two expansion levels.

The complexity analysis of TZS presented in [12] showed that due to its sub-sampling search method within the whole SA, Raster Search is by far the most complex step, reaching 81% of the total TZS execution time on average. The complexity analysis was deepened in this work for different PU sizes, as Fig. 2 shows. In this analysis, the HEVC reference software (HM 16.14) was used under the specifications of the Common Test Conditions (CTC) [14]. Altogether, five high-resolution video sequences were analyzed: *NebutaFestival*, *Traffic, BQTerrace*, *Cactus*, *ParkScene*). All frames were encoded with Quantization Parameters (QPs) 22, 27, 32, and 37, using the Random Access (RA) configuration and the Main Profile [14].

The results in Fig. 2 show that for all PU sizes Raster Search is by far the most complex step of TZS. In the largest PUs (64×64), Raster Search demands 90% of the overall TZS time. This share decreases in smaller blocks, but even in 8×4 and 4×8 PUs Raster Search corresponds to more than 55% of TZS time. Additionally, the analysis in [12] also revealed that despite its high complexity Raster Search is responsible for finding only 0.4% of the best block

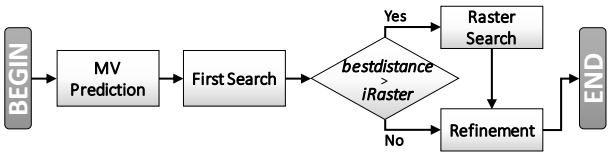

Fig. 1. Flowchart of TZS algorithm

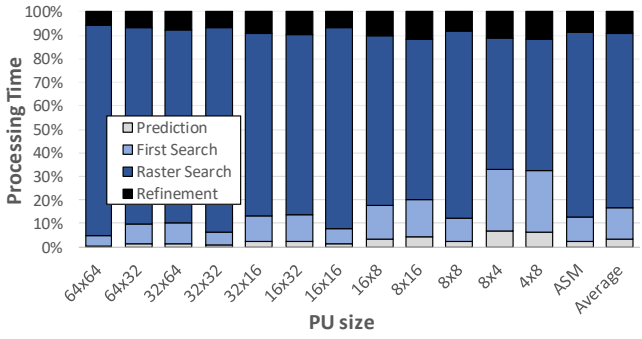

Fig. 2. Distribution of TZS processing time across its four steps for different PU sizes.

matchings in TZS. However, simply eliminating this step from the algorithm would result in prohibitive compression efficiency loss [12]. This way, an analysis of the behavior of Raster Search was performed, aiming at identifying the distribution of the best block matchings returned from this step.

3. BLOCK MATCHING DISTRIBUTION IN RASTER

After analyzing the TZS time distribution for all PU sizes, an analysis of the best block matching occurrence in the SA of Raster Search was performed. Besides the CTC video sequences listed in the previous section, the first 100 frames of four Ultra High Definition (UHD) videos from the Ultra Video Group [15] were analyzed (*Beauty*, *HoneyBee*, *Jockey*, *YachtRide*).

During the execution of Raster Search, all search points are within the SA, also referred as search window, which has a fixed maximum size defined by the Search Range (SR_{max}) parameter. The default value of SR_{max} in HM was used in the experiments, setting the SA to [-256, +256], thus composed of 512×512 positions. Therefore, considering the *iRaster* sub-sampling step, a total of approximately 10,500 positions were tested during each execution of Raster Search. For the distribution analysis presented in this section, the point where the best block matching was found was stored after each execution of Raster Search.

The results were plotted in the heat maps presented in Fig. 3, where the warmest colors represent SA positions with larger occurrence of best block matching and the coolest colors represent positions rarely or never chosen. Fig. 3(a)-(c) present the heat map for three videos sequences (*BQTerrace*, *YachtRide*, and *Cactus*, respectively), whereas Fig. 3(d) is the average obtained for all video sequences

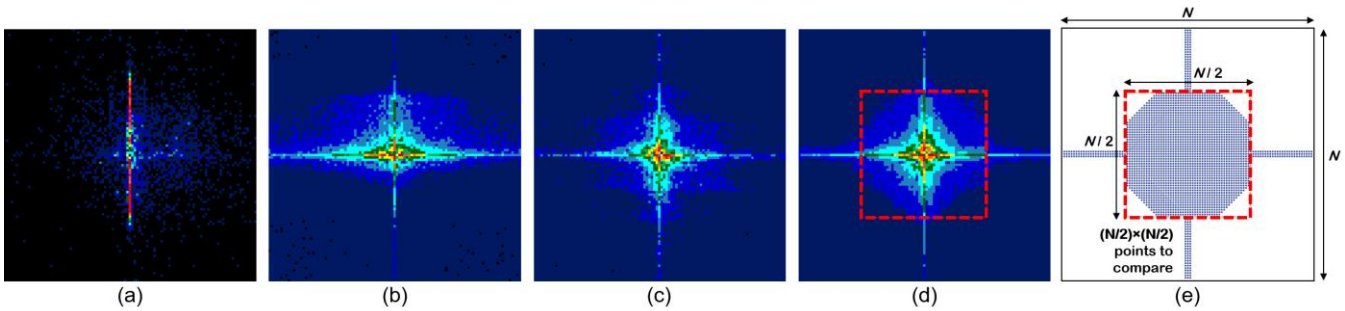

Fig. 3. Block matching distribution within the [-256, +256] search area for (a) *BQTerrace*, (b) *YachtRide*, and (c) *Cactus* sequences; (d) average block matching distribution for the whole set of video sequences tested; (e) proposed Octagonal-Axis Raster Pattern (OARP).

analyzed. *BQTerrace* (Fig. 3(a)) and *YachtRide* (Fig. 3(b)) are corner cases that represent the two most uncommon distributions among all videos analyzed. In Fig. 3(a), a significant concentration of matchings in the central vertical axis of the SA is perceived. This happens due to temporal characteristics of the *BQTerrace* sequence, in which there is a constant camera movement in the upright direction. In Fig. 3(b), the distribution is much more uniform, but the activity above the central horizontal axis is more accentuated than below. This is also due to the *YachtRide* video characteristics, which is mainly composed of a boat floating from left to right, with some upright oscillations. The distribution for *Cactus*, shown in Fig. 3(c), represents a pattern recurrent in most remaining sequences, with a clear concentration around the central point.

The average distribution for the whole set of videos is shown in Fig. 3(d). Notice that besides the high concentration within the inner 256×256 square (in red), a distribution around the central axes is significant, even in areas far from the central point. Still, it is important to notice that even the atypical cases presented in Fig. 3(a)-(b) are covered in the average pattern observed in Fig. 3(d).

4. OCTAGONAL-AXIS RASTER PATTERN

The analysis presented in the previous section shows that the computational effort to perform the search in most of the points during Raster Search is wasted since the majority of compared positions outside the 256×256 square and the axes rarely lead to the best matching. Therefore, it is perceptible that these search points could be removed, with small penalties in coding efficiency.

To exploit the observed characteristics of MV distribution in the search area efficiently, the Octagonal-Axis Raster Pattern (OARP) is proposed to reduce the complexity of Raster Search in TZS. Fig. 3(e) presents the proposed search pattern. The original idea was to perform the search within the square that represents 25% of all the SA, indicated by the analysis presented in section 3. However, it is noticeable in Fig. 3(d) that the four corners of the inner search square rarely lead to best block matchings, while the horizontal and vertical axes are well represented in the heat maps, but not totally covered by the inner square. This way, with the aim of better-exploiting areas with the highest occurrence of best block matching, the search points that compose the four corners were replaced by the exact same amount of points in the axes regions, resulting in the octagonal-plus-axes search area shown in Fig. 3(e).

The idea is generalized in Fig. 3(e) for an original $N\times N$ SA. The OARP strategy reduces it to a $(N/2)\times(N/2)$ SA, which is distributed between the octagon inscribed in the $(N/2)\times(N/2)$ square and the horizontal and vertical axes. Altogether, OARP searches 25% of the original SA, of which 88% are within the octagonal area and 12% refer to the axes. In the specific case of the default SA in HM (512×512), a total of 2,601 positions are tested, from which 2,289 are within the octagon and 312 belong to the axes.

After the definition of OARP, a brief analysis was performed to evaluate how well it covers the best block matchings found by the original Raster Search in TZS. The analysis shows that even performing the search in 25% of the original SA, OARP is able to cover 62.3% of the total best block matchings found by the original Raster Search, considering all sequences analyzed in the previous section. As next section shows, the smaller number of tests performed in OARP allows reducing significantly the TZS complexity, while its large coverage of best matching points still maintains encoding efficiency.

Differently from other fast ME strategies, OARP does not need neighboring cost information for directing the search and is suitable for the Raster Search step of TZS. Besides, its fixed number of search points allows for execution regularity, which is especially important for hardware implementations of ME algorithms.

5. EXPERIMENTAL RESULTS

To evaluate the efficiency of the novel Raster Search pattern proposed for TZS, OARP was implemented in the HEVC reference software (HM-16.14). All simulations have been carried on a workstation with an Ubuntu 14.04.5 OS, running on an Intel Xeon E5-2640v3@2.60GHz processor and with 32 GB of RAM.

The performance of OARP was evaluated for eight high-definition video sequences from classes A1, A2, A and B of the latest versions of the CTC documents [14],[16]. Classes A1 and A2 correspond to UHD sequences, whereas classes A and B correspond to WQXGA and 1080p sequences, respectively. The SR_{max} parameter was set to 256, and the maximum CU size and depth were set to 64×64 and 4, respectively. All sequences were encoded with Quantization Parameter (QP) set to 22, 27, 32, and 37 and the Random Access (RA) configuration, also following the specifications recommended by the CTC.

The HM encoder with OARP replacing Raster Search in TZS was compared to the original HM in terms of total encoding time reduction and TZS-only time reduction. Encoding efficiency was evaluated in terms of Bjøntegaard Delta (BD)-rate [17]. Table 1 presents the obtained results for OARP in terms of time reduction (TR) and encoding efficiency for all test sequences. Notice that the set of videos used in this evaluation are complimentary to those used in the complexity and block matching distribution analyses presented in sections 2 and 3.

Table 1 shows that by using OARP instead of the usual Raster pattern, the processing time of TZS is reduced by 60.91%. When considering the whole encoding process, an encoding time reduction of 21.57% is achieved in HM, with a BD-rate increase of only 0.0371%, which is negligible when taking into consideration the large reductions in processing time. It is important to emphasize that these sequences differ from one another in frame rate, bit depth, spatial resolution and motion/texture content. Thus, some videos benefit better from OARP because they incur in the execution of the Raster Search step more often. In fact, some cases even achieve negative BD-rate values, which means that OARP can perform better than the original TZS. This happens because by narrowing the SA closer to the central SA point, the Refinement step is sometimes executed around a position that would never be chosen in the original Raster Search process.

Table 2 compares OARP with the two best-related works found in the literature for complexity reduction in the TZS algorithm. To allow for a fair comparison, the set of video sequences tested for the comparison in Table 2 is exactly the same as that listed in the two related works. In [18], TZS is accelerated by decreasing the number of search points according to a spiral scan manner instead of a raster scan, whereas in [12] an early-termination strategy based on machine learning is applied after each step that composes the TZS algorithm. However, these reductions are either small compared to the expressive complexity of TZS algorithm [18] or harm the compression efficiency to non-negligible levels [12]. This work, on the other hand, excels when both dimensions are considered. Table 2 shows that OARP achieves a large average complexity reduction, close to that achieved in [12], with a much better compression efficiency, close to the achieved in [18]. Besides, as the strategy proposed in this work is compatible with that in [12], both of them can be combined in order to speed up TZS even more if some encoding efficiency loss can be tolerated.

Table 1. Comparison between OARP and original Raster

Class	Sequence	BD-rate (%)	TZS TR (%)	Total TR (%)
Class A	*PeopleOnStreet*	-0.2519	60.65	18.99
	SteamLocomotive	0.0510	61.40	21.30
Class B	*BasketballDrive*	-0.0291	59.55	20.93
	Kimono	-0.0721	53.19	13.58
Class A1	*Campfireparty*	0.0955	68.16	30.04
	ToddlerFountain	-0.0305	56.96	16.08
Class A2	*CatRobot*	0.0255	61.71	22.88
	DaylightRoad	0.5084	65.66	28.78
Average		**0.0371**	**60.91**	**21.57**

Table 2. Comparison between OARP and related works

Sequences	BD-rate (%)			TZS TR (%)		
	[18]	[12]	OARP	[18]	[12]	OARP
BasketballDrive	-0.10	-	-0.03	24.4	-	59.5
BQTerrace	-0.02	-	0.06	15.0	-	32.9
Cactus	-0.31	-	-0.04	21.6	-	55.7
Kimono	-0.05	0.42	-0.07	29.5	67.7	53.2
ParkScene	0.13	0.35	-0.03	16.2	64.5	42.8
PeopleOnStreet	-	0.65	-0.25	-	52.5	60.6
SteamLocomotive	-	0.33	0.05	-	66.3	61.4
Average*	**-0.09**	**0.49**	**0.037**	**18.5**	**62.5**	**60.91**

* Average results for all video sequences tested in the work

6. CONCLUSIONS

This paper presented a novel search pattern for the Raster Search step of the Test Zone Search (TZS) algorithm, named Octagonal-Axis Raster Pattern (OARP). The strategy was conceived after a deep analysis on the TZS complexity, the search area opportunities and the best block matching distribution in Raster Search, leading to a search area 75% smaller than in the original algorithm. The experimental results showed that OARP is able to reduce the TZS processing time by 61%, with a negligible BD-rate increase of 0.0371%. When implemented in the HEVC reference software, an average total encoding time reduction of 21.6% is achieved. The strategy is compatible with other fast Motion Estimation algorithms that employ Raster Search, and can also be jointly implemented with other complexity reduction strategies.

7. ACKNOWLEDGEMENT

The authors would like to thank the National Council for Scientific and Technological Development (CNPq), the National Council for the Improvement of Higher Education (CAPES) and the Foundation for Research Support of Rio Grande do Sul State (FAPERGS) for the financial support.

8. REFERENCES

[1] G. J. Sullivan, J. Ohm, H. Woo-Jin, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, pp. 1649-1668, 2012.

[2] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 560-576, 2003.

[3] G. Correa, P. Assuncao, L. Agostini, and L. A. da Silva Cruz, "Performance and Computational Complexity Assessment of High Efficiency Video Encoders," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, pp. 1899-1909, 2012.

[4] F. Bossen, B. Bross, K. Sühring, and D. Flynn, "HEVC complexity and implementation analysis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, pp. 1685–1696, 2012.

[5] I. K. Kim, J. Min, T. Lee, W. J. Han, and J. Park, "Block partitioning structure in the HEVC standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, pp. 1679–1706, Dec. 2012.

[6] G. Correa, P. Assuncao, L. Agostini, L. A. da Silva Cruz, "Fast HEVC Encoding Decisions Using Data Mining," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, pp. 660-673, 2015.

[7] C. Seunghyun and K. Munchurl, "Fast CU Splitting and Pruning for Suboptimal CU Partitioning in HEVC Intra Coding," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 23, pp. 1555-1564, 2013.

[8] M. U. K. Khan, M. Shafique, and J. Henkel, "An Adaptive Complexity Reduction Scheme with Fast Prediction Unit Decision for HEVC Intra Encoding," *IEEE International Conference on Image Processing (ICIP)*, pp. 1578-1582, 2013.

[9] Y. Shi, Z. Gao, and X. Zhang, "Early TU Split Termination in HEVC Based on Quasi-Zero-Block," *3rd International Conference on Electric and Electronics*, pp. 450-454, 2013.

[10] ISO/IEC-JCT1/SC29/WG11, "High Efficiency Video Coding (HEVC) Test Model 16 (HM 16) Encoder Description," in Geneva, Switzerland, ed. 17 October 2014.

[11] Cheung, Kwan C., Po, L.M., "Normalized partial distortion search algorithm for block motion estimation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, pp. 417–422, 2000.

[12] P. Goncalves, G. Correa, M. Porto, B. Zatt, L. Agostini, "Multiple Early-Termination Scheme for TZSearch Algorithm based on Data Mining and Decision Trees," *IEEE 19th International Workshop on Multimedia Signal Processing*, 2017.

[13] Pan, Y. Zhang, S. Kwong, X. Wang, and L. Xu, "Early termination for TZSearch in HEVC Motion Estimation," Proc. IEEE Int. Conf. on Acous. Speech and Signal Process. (ICASSP), pp. 1389–1393, 2013.

[14] ISO/IEC-JCT1/SC29/WG11, "Common test conditions and software reference configurations," ed. Stockholm, Sweden, 2012.

[15] Tampere University of Technology. Ultra Video Group, 2016. Available: http://ultravideo.cs.tut.fi/

[16] ISO/IEC-JCT1/SC29/WG11, "Common test conditions and software reference configurations," ed. Geneva, Sweden, 2017.

[17] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," ed. Austin, Texas, 2001.

[18] K.C.R.C., Varma, M.V.P., Kumar, S. Mahapatra, "Search range reduction for uni-prediction and bi-prediction in HEVC," *Journal of Real Time Image Processing*, pp. 1-14, 2016.