

Exploring ERP Distortions to Reduce the Encoding Time of 360 Videos

Iago Storch, Bruno Zatt, Luciano Agostini, Daniel Palomino
Programa de Pós-Graduação em Computação (PPGC), Video Technology Research Group (ViTech)
Universidade Federal de Pelotas (UFPEL)
Pelotas, RS, BRazil
{icstorch, zatt, agostini, dpalomino}@inf.ufpel.edu.br

ABSTRACT

The encoding of 360 videos presents many challenges. These spherical videos require a planar projection stage prior to encoding, and there are no studies on how projection interferes with the encoder decisions. Also, due to the increased resolution, encoding 360 videos is computationally costlier than conventional videos. In this work, we perform an extensive evaluation of how the texture distortions caused by the projection interfere with the encoder's behavior and exploit these interferences to propose multiple fast algorithms to accelerate the encoding. Experimental results show that the proposed algorithms reduce encoding time significantly with minor coding efficiency penalties, while the evaluations expose novel insights into the encoder behavior. The reached results are published in important events and journals.

KEYWORDS

Omnidirectional video, fast video coding, projection distortion, time saving.

1 Introduction

This paper summarizes the dissertation developed by Iago Storch in twelve months and advised by Bruno Zatt, Luciano Agostini, and Daniel Palomino.

Digital videos are ubiquitous in modern society, and in addition to conventional videos, 360 videos are getting popular as well with the popularization of head-mounted displays (HMDs). These videos represent the inner surface of a sphere, and when reproduced in an HMD, the user can freely look around the scene.

Digital videos must be encoded to allow efficient storage and transmission, and for 360 videos, it is not different. However, there is no standard able to perform the encoding in the spherical domain, therefore, 360 videos are projected into a rectangular format and encoded by conventional video coding standards. The most commonly used projection is the equirectangular (ERP), presented in Figure 1, and its main feature is that the closer we are to the north/south poles, the more horizontally stretched are

the textures – similar to the plate carrée projection used in maps, but without the vertical distortion.

Video coding is a task known to be computing intensive – in fact, all multimedia devices come with a dedicated video codec chipset to perform it. If encoding conventional videos is already a challenge, 360 videos push the challenge further: given that only a viewport of the video is watched at a time and the display is usually close to the user's eyes, it is required that the complete 360 video presents an ultra high definition. Although current 360 videos already employ resolutions around 8k, larger resolutions are required for a truly immersive experience.

Following the popularization of 360 videos, the internet traffic of virtual and augmented reality content is expected to grow twelve-fold in the timespan between 2017-2022 [1], showing that the encoding of 360 videos is playing a major role in the multimedia market. In this scenario, the main objectives of this work are to get a deep understanding of how the distortions of ERP 360 videos influence encoder decisions and propose a set of algorithms that leverage this knowledge to reduce the encoding time. The results show that the encoding of 360 videos presents a recurrent behavior different from conventional videos, and the proposed algorithms present time saving and coding efficiency results competitive with related works. A preliminary evaluation of the encoder behavior is published in [2], and an accelerating algorithm targeting a single encoding tool is published in [3]. Finally, a scheme targeting multiple encoding tools (encompassing most of this work) is published in [4].

2 Projection Influence on Encoder Decisions

In modern video coding standards, most of the encoding complexity lies in the interframes prediction (encoding one frame based on previous frames), intraframe prediction (encoding one frame based on its own information), and its auxiliary functions [5]. Also, modern standards allow recursive partitioning: the



Figure 1: Different videos in ERP projection

In: III Concurso de Teses e Dissertações (CTD 2021), Minas Gerais, Brasil. Anais Estendidos do Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia). Porto Alegre: Sociedade Brasileira de Computação, 2021.
©2021 SBC – Sociedade Brasileira de Computação.
ISSN: 2596-1683

video is divided into a grid of regular blocks, and each block is divided into sub-blocks recursively until a maximum depth. In this approach, the prediction is performed for all intermediary blocks, and the same region is encoded multiple times with different block size combinations to find the best tradeoff. Although these properties are common for all modern video coding standards, this work takes the High Efficiency Video Coding (HEVC) standard [6] as a case study. Then, the development of this work is guided by the following evaluations and their conclusions. All encodings are conducted using the HEVC Reference Software Encoder (HM-16.16) [7] besides 360Lib [8] to allow support for 360 videos.

HEVC employs a special interframes prediction mode called SKIP, where a block inherits the motion information from another block [6]. SKIP is mostly used in homogeneous regions, and since the projection creates homogeneous regions in the frame, it may influence the SKIP selections. To evaluate this, the encoder decisions are traced to compute how often SKIP is selected in each block of the frame, and the result is presented in Figure 2 (a). Note that the occurrence rate is superior to 90% in the distorted top/bottom regions, and it decreases towards the center, where there is no distortion. The evaluation of Figure 2 (a) shows that SKIP occurrences increase with projection distortion, especially when considering that conventional videos present a balanced distribution with occurrences similar to undistorted regions of 360 videos. Besides, a second analysis crossed the SKIPS with intermediate encoding information and concluded that this SKIP-distortion relation has a different intensity depending on aspects such as block size and video resolution.

After testing SKIP, integer motion estimation (IME) searches for a block as similar as possible to the current block inside a search range, located in an already encoded frame, and produces a motion vector (MV) representing the movement of the current block. Since the movement in distorted regions is usually warped, the distortion can affect the motion vectors distribution. To assess this, the average and standard deviation of MVs throughout the frame are computed, in addition to computing the distribution of MV angles and assessing the impact of reducing the search range. In conclusion, it was found that the horizontal component of MVs in distorted regions tends to be smaller due to greater horizontal redundancy, and narrowing the search range causes a negligible impact on coding efficiency. On the contrary, the MVs of conventional videos present a similar magnitude and angle irrespective of frame region, and both the vertical and horizontal components present similar importance to coding efficiency.

After IME, fractional motion estimation (FME) uses interpolation to generate samples in fractional positions and tests if using blocks in fractional positions improves the coding efficiency – in HEVC, fractional samples with a precision of $\frac{1}{2}$ and $\frac{1}{4}$ are generated [6]. The horizontal stretching of ERP projection is a product of a similar interpolation, therefore, these successive interpolations may produce samples of redundant value. To verify this, the variance of samples in a horizontal sliding window with dimensions of 4×1 is computed in a row-by-row fashion, and the mean results for each row of 360 videos are presented in Figure 2 (b). Note that closer to the top/bottom regions, the variance is

close to zero, and it increases towards the center. This evaluation shows that performing FME in horizontal direction is more likely to produce redundant samples in the distorted regions. Performing this evaluation in conventional videos produces a flat or random distribution, showing that FME performance is more predictable in 360 videos when compared to conventional videos.

Finally, the recursive partitioning allows selecting block sizes according to motion and texture complexity – during intraframe prediction, however, only texture complexity is considered. In this context, the projection distortion may influence the selection of block sizes. To evaluate this, the selection of block sizes in the rows of the pre-defined grid inside the video is traced, and the distribution of block sizes is presented in Figure 2 (c), where each color represents a block size. This distribution is normalized such that the weight of each block size equals its area. Note that closer to the poles (top/bottom), most of the frame is encoded with larger blocks while smaller blocks are rarely selected, and approaching the center of the frame, smaller blocks are selected more often at the expense of larger blocks. In conventional videos, the distribution for each block size is constant across the frame. In conclusion, it may be unnecessary to evaluate smaller blocks in distorted regions since they are rarely selected.

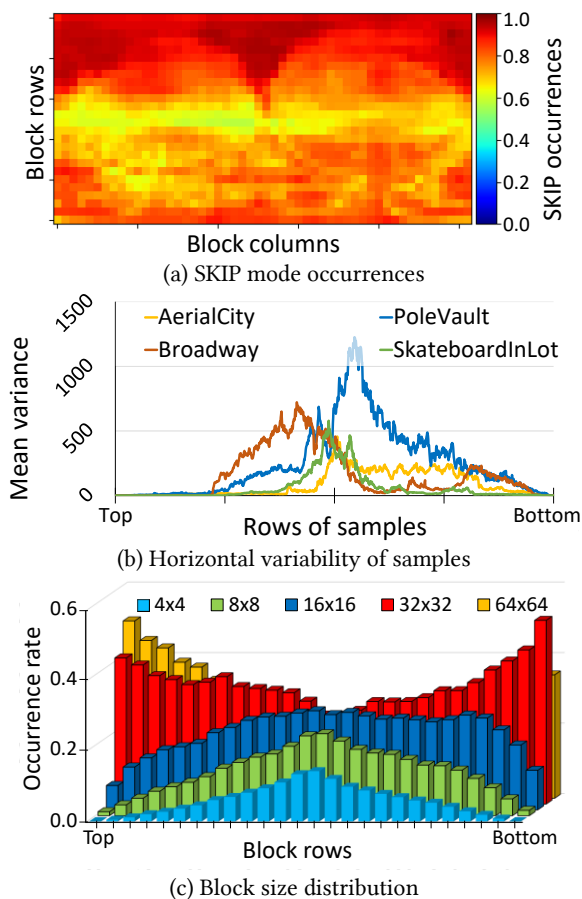


Figure 2: Evaluations performed over ERP 360 videos

3 Proposed Fast Algorithms

Based on the previous analyses, a set of fast algorithms are proposed to reduce the encoding time of 360 videos, namely: *Early SKIP*, *Reduced SR*, *Reduced FME*, and *Reduced Intra Sizes*. The three first algorithms are applied during interframes prediction, and they are positioned in the interframes prediction as presented in Figure 3. *Reduced Intra Sizes* is applied during intraframe prediction, and it does not interact with other algorithms.

Early SKIP is applied directly after testing the SKIP mode. It evaluates the current block variance, size, and position, in addition to the video resolution. In sequence, this information is crossed with a set of pre-computed cumulative distributions of SKIP selection (generated during the evaluations), and based on the result, an early decision is conducted or not. In the case of an early decision, the remaining stages are bypassed, and the current block is encoded with SKIP; otherwise, MERGE (another special mode) is tested and the remaining interframes prediction is performed. The *Early SKIP* algorithm works adaptively to projection distortion intensity, i.e., the more intense is the distortion, the more likely is the early selection of SKIP mode. By bypassing some prediction stages, *Early SKIP* can save significant encoding time.

Reduced SR is applied during IME, and it reduces the search range of Raster Search (the most demanding step of IME) based on distortion intensity. Since ERP does not stretch the textures vertically, the search range height is kept as original. As for the search range width, it is reduced adaptively to distortion intensity: the more distorted is a region, the more narrowed is the search range. By reducing the search range area, the number of blocks evaluated during IME is reduced, accelerating the encoding.

Reduced FME is performed before conducting FME: it evaluates the vertical position of the current block in the frame to estimate the stretching intensity and adjusts the FME precision accordingly. In undistorted regions, it sets FME to use maximum precision in the horizontal direction; in regions of moderate distortion, it sets FME to employ only $\frac{1}{2}$ -pixel precision in the horizontal direction; and in highly-distorted regions, it bypasses FME in the horizontal direction. FME always employs maximum precision in the vertical direction since there is no vertical stretching. By avoiding the interpolation and evaluation of some fractional pixels, the encoding process is accelerated.

Reduced Intra Sizes is distributed in the intraframe prediction and in the partitioning loop of the encoder. During intraframe prediction, it traces what is the best RD-cost for the current block size (RD-cost represents the fitness of a mode/block to encode a

region). In the partitioning loop, *Reduced Intra Sizes* evaluates the contribution of the current block size in the respective frame region (as in Figure 2 (c)) and the RD-cost of the co-located region in a reference frame already encoded. Finally, if the current block size often occurs in the current region, and the RD-costs are close (i.e., the current size is as good as the optimal size was for reference frame), then *Reduced Intra Sizes* disables the intraframe prediction for sub-blocks of the current block. Since this algorithm relies on previous frames, *Reduced Intra Sizes* is disabled once in every three frames to avoid propagating error indefinitely. With this approach, smaller block sizes tend to be evaluated only when they are very likely to be selected, and the encoding is accelerated.

4 Experimental Results

The proposed algorithms are implemented in HM-16.16 [7] with 360Lib-5.0 [8], and the encodings are conducted following the Common Test Conditions for encoding 360 videos [9]. The algorithms are assessed according to time saving and coding efficiency compared to the reference encoder. Time saving is considered the amount of processing time saved using the proposed algorithms in relation to the reference encoder, whereas coding efficiency is measured as BD-BR [10], a metric that represents the bitrate increase for a solution to achieve the same visual quality as the reference. The algorithms for accelerating interframes prediction are evaluated with the *RandomAccess* configuration whereas the algorithms for intraframe prediction are evaluated with the *AllIntra* configuration [9] of HM-16.16.

The average results for each algorithm are presented in Table 1, where *InterTech* is the result of employing all interframes prediction algorithms simultaneously (i.e., *Early Skip*, *Reduced SR*, and *Reduced FME*), and *IntraTech* is the combination of *Reduced Intra Sizes* and a previous work published at [11]. Table I also presents the average results of related works: Liu [12] and Ray [13] present proposals to accelerate the interframes prediction of ERP 360 videos, whereas Wang [14] and Zhang [15] propose to accelerate the intraframe prediction of ERP 360 videos. Liu [12] uses a KNN classifier to discard some unlikely block sizes during interframes prediction; Ray [13] discards unlikely block sizes during interframes prediction and reduces FME precision based on block position; Wang [14] discards some prediction modes and block sizes during intraframe prediction based on block position; and Zhang [15] evaluates the texture of the blocks to evaluate a subset of prediction modes and block sizes. It must be noted that

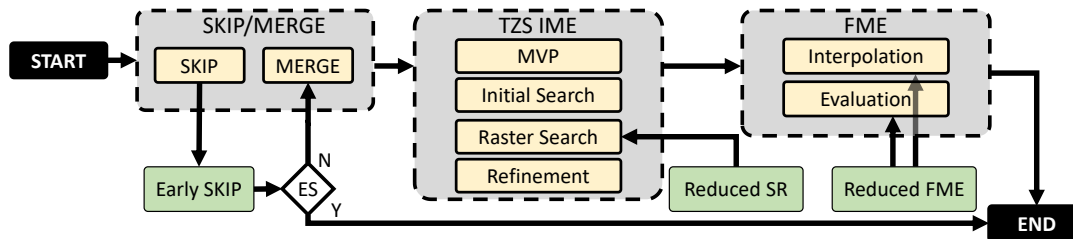


Figure 3: Proposed algorithms for accelerating interframes prediction

Table 1: Experimental results and comparison

Algorithm	Time saving	BD-BR
Early SKIP	13.59%	0.24%
Reduced SR	2.31%	0.00%
Reduced FME	10.20%	0.42%
InterTech	22.84%	0.65%
Liu [12]	32.00%	1.30%
Ray [13]**	15.00%	-0.20%
Reduced Intra Sizes	10.41%	0.37%
IntraTech	23.09%	1.21%
Wang [14]	24.50%	0.24%
Zhang [15]	53.00%	1.30%

** This work uses a different encoder

Ray [13] uses the encoder *Joint Exploration Model* (JEM) [16], an exploratory encoder built on top of HEVC to evaluate new encoding tools for future video coding standards. The works [12] [14][15] use the same HEVC encoder as this work – the HM [7].

The results from Table I show that *InterTech* can reduce encoding time by 22.84%, on average, whereas posing a small average BD-BR of 0.65% (i.e., increasing bitrate in 0.65% for the same visual quality), demonstrating that the proposed algorithms can work together efficiently in reducing the encoding time without harming coding efficiency. When compared to related works, Liu [12] achieves a greater complexity reduction, but the coding efficiency is degraded twofold when compared to *InterTech*. When computing the Time saving per BD-BR ratio of the two algorithms, *InterTech* achieves 35.0 whereas [12] achieves 24.6 – showing that *InterTech* presents a better tradeoff. Ray [13], on the other hand, presents a smaller time saving than *InterTech* whereas improving the coding efficiency by -0.2% BD-BR. However, this work is implemented and tested using the exploratory encoder JEM [16], and it is possible that it would not achieve the same results in case it was tested in a standardized video coding framework such as HEVC.

When the average results for intraframe prediction are evaluated, *Reduced Intra Sizes* reduces the encoding time by 10.41%, on average, whereas posing an average coding efficiency penalty of 0.37% BD-BR. When it is combined with a previous work [11] (*IntraTech* line), the encoding time is reduced by 23.09% with a BD-BR of 1.21%. When compared to related works, it is visible that *IntraTech* is outperformed by [14] and [15]. However, both related works are mainly based on generic complexity reduction techniques that are applied in the same manner for the entire frame, irrespective of projection distortion. In this context, they are likely to achieve similar results if applied to conventional videos, whereas *IntraTech* is specifically designed to exploit the distortions caused by the projection of 360 videos. Finally, since *IntraTech* and [14][15] rely on different properties of the video, it is possible to combine them and improve the overall performance.

5 Conclusion

This work presented a set of evaluations that model the behavior of a video encoder when encoding ERP 360 videos and a

set of algorithms to reduce encoding time. The evaluations showed that the projection distortion interferes in the selection of SKIP mode, on the magnitude of motion vectors, on the redundancy of FME, and on the selection of block sizes. Based on the evaluations, a set of algorithms were designed to accelerate the encoding of ERP 360 videos adaptively to projection distortion. Early SKIP performs an early decision by SKIP mode, Reduced SR reduces the search range of motion estimation, Reduced FME adjusts the precision of FME, and Reduced Intra Sizes avoids the evaluation of some blocks. Experimental evaluations showed that these algorithms present results competitive with related works, and they can be combined with other algorithms to improve the performance. Although many authors propose algorithms to accelerate 360 video coding, most of them rely on empirical approaches and do not provide evidence on how the algorithms are conceived or why they achieve positive results. In this context, the contribution of this work is twofold: provide evidence on the impacts that ERP distortion causes in the encoding, and propose algorithms to exploit such evidence and accelerate the encoding.

ACKNOWLEDGMENTS

The authors thank FAPERGS, CNPq and CAPES (Finance Code 001) for financing this investigation.

REFERENCES

- [1] Cisco visual networking index: forecast and methodology 2017-2022, 2019.
- [2] I. Storch, L. A. da Silva Cruz, L. Agostini, B. Zatt, D. Palomino. 2019. The Impacts of Equirectangular 360-degrees Videos in the Intra-Frame Prediction of HEVC. *Journal of Integrated Circuits and Systems* 14, 1 (2019), 1-10.
- [3] I. Storch, G. Correa, B. Zatt, L. Agostini, D. Palomino. ESA360 - Early SKIP Mode Decision Algorithm for Fast ERP 360 Video Coding. In *2020 28th European Signal Processing Conference (EUSIPCO)*. IEEE, 535-539.
- [4] I. Storch, L. Agostini, B. Zatt, S. Bampi, D. Palomino. FastInter360: A Fast Inter Mode Decision for HEVC 360 Video Coding. In *IEEE Transactions on Circuits and Systems for Video Technology*, (2021), 1-15.
- [5] J. Vanne, M. Viitanen, T. D. Hamalainen, A. Hallapuro. 2012. Comparative rate-distortion-complexity analysis of HEVC and AVC codecs. *IEEE Transactions on Circuits and Systems for Video Technology* 22, 12 (2012), 1885-1898.
- [6] G. Sullivan, J. R. Ohm, W.-J. Han, T. Wiegand. 2012. Overview of the high efficiency video coding (HEVC) standard. In *IEEE Transactions on Circuits and Systems for Video Technology*, 22, 12 (2012), 1649-1668.
- [7] HEVC Test Model, available at: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/. Access in Aug. 2021.
- [8] 360Lib software, available at: https://jvet.hhi.fraunhofer.de/svn/svn_360Lib/. Access on Aug. 2021.
- [9] P. Hanhart, J. Boyce, K. Choi, J.-L. Lin. JVET common test conditions and evaluation procedures for 360 video. JVET meeting document L1012, 1-7, 2018.
- [10] G. Bjontegaard. Calculation of average PSNR differences between RD-curves. VCEG meeting document M33, 1-4, 2001.
- [11] I. Storch, B. Zatt, L. Agostini, G. Correa, L. A. da Silva Cruz, D. Palomino. 2020. Spatially Adaptive Intra Mode Pre-Selection for ERP 360 Video Coding. In *2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2178-2182.
- [12] Z. Liu, P. Song, M. Zhang. 2019. A CU split early termination algorithm based KNN for 360-degree video. In *Data Compression Conference (DCC)*. IEEE, 594-594.
- [13] B. Ray, J. Jung, M. C. Larabi. 2018. A low-complexity video encoder for equirectangular projected 360 video content. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 1723-1727.
- [14] Y. Wang, Y. Li, D. Yang, Z. Chen. 2017. A fast intraprediction algorithm for 360-degree equirectangular panoramic video. In *2017 IEEE Visual Communications and Image Processing (VCIP)*. IEEE, 1-4.
- [15] M. Zhang, X. Dong, Z. Liu, F. Mao, W. Yue. 2019. Fast intra algorithm based on texture characteristics for 360 videos. *EURASIP Journal on Image and Video Processing* 2019, 1 (2019), 1-16.
- [16] J. Chen, E. Alshina, G. Sullivan, J. R. Ohm, J. Boyce. Algorithm description of joint exploration model test model 3. JVET meeting document C1001, 1-30, 2016.