Depth-projected Determination for Adaptive Search Range in Motion Estimation for HEVC

Tsz-Kwan Lee, Yui-Lam Chan, and Wan-Chi Siu
Center for Signal Processing,
Department of Electronic and Information Engineering, The Hong Kong Polytechnic University
Email: eie.glorylee@connect.polyu.hk, enylchan@polyu.edu.hk, enwcsiu@polyu.edu.hk

***Abstract*—High Efficient Video Coding (HEVC) improves coding efficiency but suffers from high computational complexity due to its quad-tree partitioning structure in motion estimation (ME). In recent development of 3D video technology, depth map from the 3D video provides an intimation of the objects' distance from the projected screen in a 3D scene, which inspires the exploration in adaptive search range determination for complexity reduction in HEVC. The proposed algorithm exploits the high temporal correlation between the depth map and the motion in texture. By utilizing this correlation and the potential impact of 3D-to-2D projection, a depth/motion relationship is built for a tailor-made search range with a depth-projected scale factor to skip unnecessary search points in ME. Besides, the proposed ASR algorithm can work well with other fast ME algorithms with up to 53% of average coding time reduction whereas the coding efficiency can be maintained.**

I. Introduction

The latest HEVC video coding standard has achieved the coding gain of bitrate reduction of 50% at similar perceptual quality compared to H.264 [1]. However, high computational complexity is induced since a flexible quad-tree partitioning structure is adopted where a recursive split of a coding unit (CU) is conducted for every cycle of ME [2]. With quad-tree partitioning, inter prediction consumes about 60-70% of the whole encoding time [3]. Therefore, fast ME search approaches are used to expedite the inter prediction process. Directional search was suggested applying various search patterns to reduce search points within a fixed search range [4]. Nevertheless, arbitrary search patterns are not preferable for hardware implementation due to their irregular data flow [5]. In this circumstance, full search with an adaptive search range (ASR) can provide both search point reduction and regular data flow. Some ASR algorithms proposed in H.264 determine the search range of a block by motion characteristics of its neighbors. Two most recent algorithms extended to support HEVC are Maximum Likelihood Estimation Laplace Distribution Algorithm (MLELD) [6], and Linear Adaptive Model for Adaptive Search Range Algorithm (LAMASR) [7]. MLELD [6] models the motion vector (MV) differences of the previous frame by the zero-mean Laplace distribution where the parameters are solved by maximum likelihood estimation (MLE) to set the final ASR. LAMASR [7] adopts a linear ASR model with a fixed scale factor, including an overdetermined equation system. The system is solved by parameters of prediction unit (PU) size, MV difference and MV predictors.

To the best of our knowledge, only one work in [8] tried to reveal the usage of depth information for fast mode decision rather than ASR determination in H.264. We are the first one raising the depth information to design an efficient ASR algorithm for HEVC. In our previous work [9], a technique of depth intensity mapping for ASR was proposed. Different from our previous work, this paper further proposes a depth/motion relationship map (DMRMap) for retrieval of ASR in individual x and y directions. Furthermore, a depth projection aided scale factor is proposed based on the geometric correlation between 3D scene and 2D image plane. This paper starts with exploring the high temporal correlation between depth maps and the motion in texture in Section II. In Section III, the proposed DMRMap is introduced. By utilizing the DMRMap, the retrieval of ASR for the encoding block is then presented. The ASR adjustment due to the impact of the 3D-to-2D plane projection is discussed in Section III. Simulation results are provided in Section IV. Section V concludes this paper.

II. Temporal Correlation between Depth Map and Motion in Texture Streams

Texture plus depth is one of the data representation formats including both the color texture and the depth map of a scene [10]. The color texture stream captures luminance and chrominance information of every pixel while the depth map records the distances of the objects from the camera [9]. Thus, the depth map being a piece of additional information objectively reflects distance between objects with their dynamic movements between frames. This motivates us to utilize depth map to expedite the texture video coding process in HEVC.

Fig. 1 plots the maximum amplitude of MVs of color texture along average depth intensity values of all blocks in two consecutive frames. Fig. 1(a) and Fig. 1(b) reveal that the maximum horizontal MV (in x direction) amplitude distribution could be very similar in consecutive frames. Fig. 1(c) and Fig. 1(d) also show the same high temporal correlation in the vertical MV (in y direction). Fig. 1 demonstrates that the depth information of an object not only represents the physical object position but also exhibits the motion activities of the object between frames. By this temporal correlation between the depth map and motion in texture, the proposed algorithm can determined a tailor-made search range for each block.

III. The Proposed DMRmap-based ASR Algorithm and Depth-projected Scale Factor

Revealed by Section II, a depth/motion relationship map (DMRMap) is proposed to reflect the probable range of movements for any object such that motion activities of objects could be predicted by depth maps. The ASR is then set

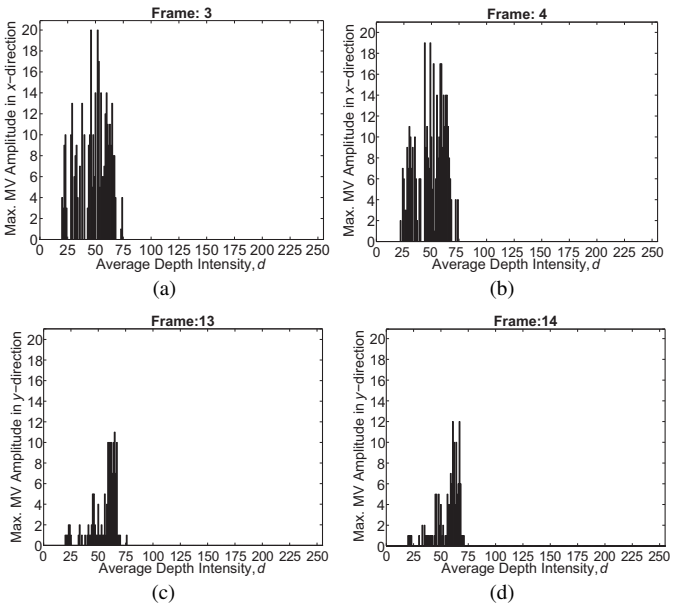

Fig. 1. The maximum MV amplitude of color texture against average depth intensity values between consecutive frames, x-component for (a) Frame 3 and (b) Frame 4, and y-component for (c) Frame 13 and (d) Frame 14 of "Lovebird".

according to the proposed DMRMap, followed by an adjustment with a depth projection aided scale factor where the impact of 3D-to-2D projection by motion activity is considered. Unnecessary search points within the pre-defined search range can be removed accordingly for x and y directions.

A. DMRMap Construction and Motion Vector Extraction

The DMRMap is established after the first inter-frame has gone through a conventional full rate-distortion optimization (RDO) inter coding and once the MVs of all blocks in the reference frame are obtained. Let d be the average depth intensity value of a block, where $0 \leq d \leq 255$. It is noted that depth maps are always estimated using stereo matching methods [11], which induces slight variation or noise of depth values within the same object. To tolerate the noise, d is quantized uniformly by a quantization factor Q=8 into $\hat{d}$, where $0 \leq \hat{d} \leq \lceil 255/Q \rceil$. Note that $\lceil\ \rceil$ is the ceiling function.

By the quantized average block intensity, DMRMap comprises mapping sets, $S^{\hat{d}}$ containing blocks with the particular $\hat{d}$. Examples of the mapped candidates with $\hat{d}$=2 and $\hat{d}$=8 are marked on the reference frame as depicted in Fig. 2. They are linked with the corresponding current block with $\hat{d}$=2 and $\hat{d}$=8, respectively. To extract motion activities from the DMRMap,

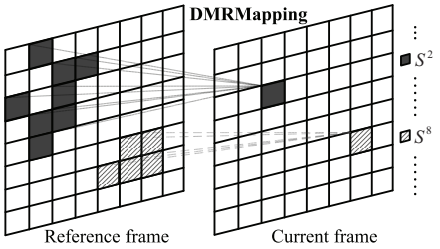

Fig. 2. Illustration of DMRMap establishment between consecutive frames.

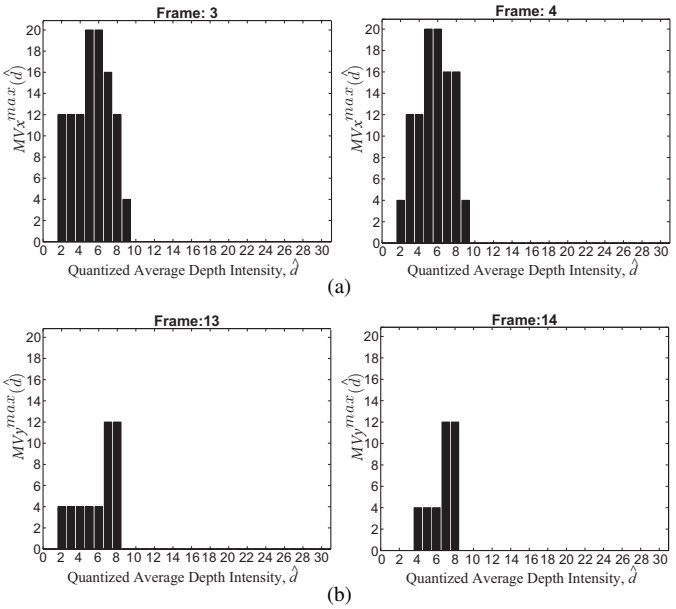

Fig. 3. The largest motion vector (a) $MVx^{\max}(\hat{d})$ by quantized depth in the x direction, and (b) $MVy^{\max}(\hat{d})$ in the y direction from a pair of consecutive frames for "Lovebird".

$S_x^{\hat{d}}$ and $S_y^{\hat{d}}$ are the sets of MVs obtained by the mapped candidates in the x and y direction, respectively. The largest MV, $(MVx^{\max}(\hat{d}), MVy^{\max}(\hat{d}))$, in the x and y directions are

$$MVx^{\max}(\hat{d}) = \max(S_x^{\hat{d}}) \tag{1}$$

and

$$MVy^{\max}(\hat{d}) = \max(S_y^{\hat{d}}), \tag{2}$$

where $\max(S)$ gives the maximum value of the set S. $MVx^{\max}(\hat{d})$ and $MVy^{\max}(\hat{d})$ are the descriptors of the DMRMap, the largest motion in both x and y directions can be determined for the given $\hat{d}$. Two pairs of DMRMaps in the x and y directions constructed from consecutive frames for "Lovebird" are illustrated in Fig. 3(a) and Fig. 3(b), respectively, in which Q is set to 8. They record $MVx^{\max}(\hat{d})$ and $MVy^{\max}(\hat{d})$ for each $\hat{d}$, where $\hat{d}$ is from 0 to 31. Fig. 3 reveals that the motion distributions of both directions are in high similarity between consecutive frame pairs against $\hat{d}$.

B. Adaptive Search Range Decision based on DMRMap

By utilizing the high temporal correlation of DMRMaps, the ASR of which search ranges are respectively denoted as $Rx(B_t^n)$ and $Ry(B_t^n)$ in the x and y directions, is determined for block n being encoded in frame t. From (1) and (2), $Rx(B_t^n)$ and $Ry(B_t^n)$ can respectively be computed as

$$Rx(B_t^n) = MVx^{\max}(QDepth(B_t^n)) \tag{3}$$

and

$$Ry(B_t^n) = MVy^{\max}(QDepth(B_t^n)). \tag{4}$$

Here, $QDepth(B_t^n)$ is the quantized average depth intensity value for B_t^n. If $QDepth(B_t^n)$ is empty in the DMRMap, the search ranges of B_t^n in both x and y directions are set to 64. It is the default search range of the main profile in HEVC, which is always larger than or equal to the proposed ASR.

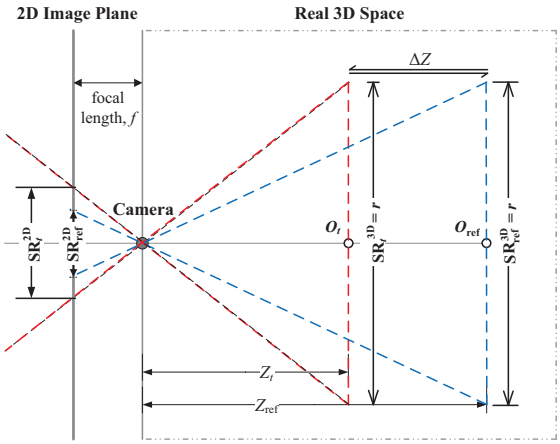

Fig. 4. Geometric relationship between depth of object and motion activity on the 2D image plane.

C. The Proposed Depth-projected Scale Factor

An object moving towards and away from the camera, or zoom effect from the camera changes the distance between the object being captured and the camera along time. This motion activity along the camera axis (z-axis) has the potential to weaken the DMRMap temporal correlation. Taking these distance/depth changes and the impact of 3D-to-2D projection on the image plane into consideration, a scale factor of B_t^n, $\rho(B_t^n)$, is applied to the proposed ASR in

$$Rx^{\rho}(B_t^n) = \rho(B_t^n) \times MVx^{\max}(QDepth(B_t^n)) \quad (5)$$

and

$$Ry^{\rho}(B_t^n) = \rho(B_t^n) \times MVy^{\max}(QDepth(B_t^n)). \quad (6)$$

In (5) and (6), $\rho(B_t^n)$ is underlying on the motion parallax stating that, given the same horizontal or vertical motions of objects in the 3D space, objects that are closer to the camera move faster on the 2D image plane than the objects that are farther. In other words, the degree of the projected displacement of an object on the image plane is always related to its distance from the camera. This situation is illustrated in Fig. 4 with an example of the geometric relationship when an object moving towards the camera and how its displacement on the 2D image plane varies. Let O_{ref} denote an initial reference position of an object. Its actual distance from the camera is Z_{ref}, and its search range implied by its motion range is assumed to be SR_{ref}^{3D} in the 3D space. When an object moves from O_{ref} to O_t between time $t-1$ and time t such that the object moving towards the camera, and now $Z_t < Z_{\text{ref}}$ since O_t is closer to the camera than O_{ref}. The dotted lines in Fig. 4 indicate the trajectory of the 3D-to-2D projections through the camera lens onto the image plane. Therefore, the projection amplitudes show $SR_{\text{ref}}^{2D} < SR_t^{2D}$ on the image plane provided that O_{ref} and O_t have the same search range in the 3D space (i.e., $SR_{\text{ref}}^{3D} = SR_t^{3D} = r$). Consequently, depth intensity changes can be used to determine a deviation of the projection ratios and form $\rho(B_t^n)$ for scaling the proposed ASR due to the motion along the z-axis by the triangular similarity as

$$\rho(B_t^n) = \frac{SR_t^{2D}}{SR_{\text{ref}}^{2D}} = \frac{\frac{r \cdot f}{Z_t}}{\frac{r \cdot f}{Z_{\text{ref}}}} = \frac{Z_{\text{ref}}}{Z_t}. \quad (7)$$

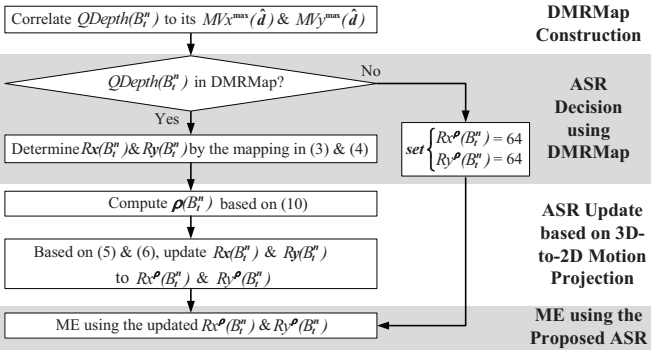

Fig. 5. Flowchart of the proposed depth projected ASR algorithm.

In (7), f is the focal length of the camera, and r is the search range amplitude of the object in the 3D space. The actual distances, Z_t and Z_{ref}, in the 3D space can be computed from the average depth intensity values without quantization in the depth maps, $Depth(B_t^n)$ and $Depth(B_{\text{ref}}^n)$, for B_t^n in the current frame and its co-located block B_{ref}^n in the reference frame, respectively, as

$$Z_t = 1/\left[\frac{Depth(B_t^n)}{255} \times \left(\frac{1}{Z_{\text{near}}} - \frac{1}{Z_{\text{far}}}\right) + \frac{1}{Z_{\text{far}}}\right] \quad (8)$$

and

$$Z_{\text{ref}} = 1/\left[\frac{Depth(B_{\text{ref}}^n)}{255} \times \left(\frac{1}{Z_{\text{near}}} - \frac{1}{Z_{\text{far}}}\right) + \frac{1}{Z_{\text{far}}}\right], \quad (9)$$

where Z_{near} and Z_{far} are, respectively, the smallest and the largest actual distances captured by the camera, which are freely available in the camera configure files of the test sequences recommended by the ISO/IEC and ITU-T JCT-3V group. Putting (8) and (9) into (7), $\rho(B_t^n)$ is finally expressed as

$$\rho(B_t^n) = \frac{Depth(B_t^n)(Z_{\text{far}} - Z_{\text{near}}) + 255 \times Z_{\text{near}}}{Depth(B_{\text{ref}}^n)(Z_{\text{far}} - Z_{\text{near}}) + 255 \times Z_{\text{near}}}, \quad (10)$$

and it can be summarized as

$$\begin{cases} \text{case 1}: \rho(B_t^n) > 1, & Depth(B_t^n) > Depth(B_{\text{ref}}^n) \\ \text{case 2}: \rho(B_t^n) = 1, & Depth(B_t^n) = Depth(B_{\text{ref}}^n) \\ \text{case 3}: \rho(B_t^n) < 1, & Depth(B_t^n) < Depth(B_{\text{ref}}^n) \end{cases}, \quad (11)$$

where case 1 represents a scenario that the object moving towards the camera, case 2 represents the object without z-axis motion, and case 3 represents the object moving away from the camera. The flowchart of the proposed DMRMap-based ASR algorithm with the depth-projected scale factor for encoding a block in HEVC is provided in Fig. 5.

IV. Simulation Results

The proposed depth-projected scaling DMRMap-based ASR algorithm (SDMRMap) and algorithms in literatures for comparisons were implemented respectively in the conventional full-search (FS) and fast Test Zone Search (TZS) in HEVC model, HM14.0 [12]. Bjontegaard (BD) measurement [13] in terms of BD-rate (%) and BD-PSNR (dB), and Δtime (%) representing coding time change in percentage were evaluated. The coding time includes the computational cost for all CU quad-tree levels. Positive and negative values denote increments and decrements, respectively.

TABLE I. BD AND CODING TIME CHANGES OF THE PROPOSED FS+SDMRMAP AND LITERATURES AGAINST FS

Sequences (720p & 1080p)	FS+MLELD			FS+LAMASR			Proposed FS+SDMRMap		
	Δtime (%)	BD-PSNR (dB)	BD-rate (%)	Δtime (%)	BD-PSNR (dB)	BD-rate (%)	Δtime (%)	BD-PSNR (dB)	BD-rate (%)
Balloons	-63.17	0.00	+0.07	-74.01	0.00	-0.06	-89.63	-0.01	+0.26
Kendo	-80.74	-0.01	+0.23	-89.99	-0.01	+0.25	-94.01	-0.01	+0.26
Lovebird	-87.87	0.00	+0.12	-94.96	0.00	+0.06	-99.46	0.00	+0.08
Newspaper	-73.84	-0.02	+0.47	-90.94	-0.01	+0.23	-97.23	-0.01	+0.31
Poznan_Street	-42.87	-0.01	+0.51	-69.47	-0.02	+0.74	-97.15	0.00	+0.17
Poznan_Hall2	-41.25	-0.01	+0.28	-40.37	-0.01	+0.22	-86.72	-0.01	+0.41
Undo_Dancer	-36.78	-0.01	+0.39	-34.74	-0.02	+0.69	-91.74	-0.01	+0.30
GT_Fly	-27.32	0.00	+0.03	-36.41	0.00	+0.14	-95.47	-0.01	+0.27
Average:	**-56.73**	**-0.01**	**+0.26**	**-66.36**	**-0.01**	**+0.28**	**-93.93**	**-0.01**	**+0.26**

TABLE II. BD AND CODING TIME CHANGES OF THE PROPOSED TZS+SDMRMAP AND LITERATURES AGAINST TZS

Sequences (720p & 1080p)	TZS+MLELD			TZS+LAMASR			Proposed TZS+SDMRMap		
	Δtime (%)	BD-PSNR (dB)	BD-rate (%)	Δtime (%)	BD-PSNR (dB)	BD-rate (%)	Δtime (%)	BD-PSNR (dB)	BD-rate (%)
Balloons	-15.48	-0.01	+0.14	-14.24	0.00	+0.02	-48.92	0.00	-0.01
Kendo	-17.46	0.00	+0.12	-25.92	-0.01	+0.30	-50.14	0.00	+0.13
Lovebird	-29.30	-0.01	+0.26	-37.90	0.00	+0.02	-63.06	0.00	+0.10
Newspaper	-26.98	-0.02	+0.45	-22.34	0.00	+0.13	-54.50	-0.01	+0.24
Poznan_Street	-22.98	-0.01	+0.30	-21.36	-0.01	+0.64	-56.63	-0.01	+0.23
Poznan_Hall2	-39.54	0.00	+0.04	-31.37	0.00	0.00	-54.25	0.00	+0.14
Undo_Dancer	-12.82	-0.01	+0.15	-5.98	-0.02	+0.58	-55.56	-0.01	+0.19
GT_Fly	-20.29	0.00	+0.01	-30.14	0.00	+0.09	-42.61	0.00	+0.12
Average:	**-23.11**	**-0.01**	**+0.18**	**-23.66**	**-0.01**	**+0.22**	**-53.21**	**0.00**	**+0.14**

A. Results of Applying SDMRMap to FS

Table I lists the coding performance and Δtime of our proposed SDMRMap against FS, denoted as FS+SDMRMap. The proposed FS+SDMRMap can averagely save 93.93% of coding time over FS since it utilizes the high temporal correlation of motions revealed by depth intensity mapping. While significant coding time reduction can be achieved, FS+SDMRMap attains negligible loss on BD-PSNR by 0.01dB and a trivial increment of 0.26% in BD-rate as compared to FS. From Table I, the proposed FS+SDMRMap can save more computational time by about 37%, and 27%, respectively, as compared with the algorithms in literatures, FS+MLELD [6], and FS+LAMASR [7]. The reason is that FS+SDMRMap considers the search range in the x and y directions separately for tracing the true MVs. But, FS+MLELD, and FS+LAMASR consider the search range in both directions jointly, which provide more than enough search range dimension. Furthermore, FS+SDMRMap utilizes an adaptive scale factor for ASR adjustment. But, FS+LAMASR simply multiplies a fixed scale factor to its resultant ASR.

B. Results of Applying SDMRMap to Fast TZS

Table II shows the BD measurement and the coding time change of TZS+MLELD [6], TZS+LAMASR [7], and the proposed TZS+SDMRMap, compared to TZS. To have fair comparison since TZS is only suited for a squared search window, the proposed ASR by TZS+SDMRMap was computed as $\max(Rx^{\rho}(B_t^n), Ry^{\rho}(B_t^n))$ for the sake of simplicity. TZS+SDMRMap could save 42.61% to 63.06% coding time for various sequences. Meanwhile, almost no loss in terms of BD-PSNR and BD-rate (averagely 0.14% increment) was found. The above results indicate that the proposed TZS+SDMRMap is well compatible with the fast search strategy in HEVC. On average, TZS+SDMRMap attains a better BD performance and reduce more coding time by 29% compared to TZS+MLELD and TZS+LAMASR.

V. CONCLUSION

The proposed ASR algorithm for HEVC reduces the computational complexity of ME. It proposes a depth/motion relationship map (DMRMap) which builds the linkage on the same object among consecutive frames which reflects the probable range of movements. Based on the DMRMap, a depth intensity mapping is contrived to form an ASR for ME. Furthermore, a depth-projected scale factor for the ASR adjustment has been proposed to comply with the impact of 3D-to-2D projection during depth intensity changes. The proposed ASR algorithm performs well with FS and fast ME algorithms such as TZS in HEVC for complexity reduction. Simulation results demonstrated that the proposed algorithm is able to reduce up to 53% of average coding time in TZS while the coding efficiency can be maintained.

ACKNOWLEDGMENT

Tsz-Kwan Lee would like to thank the Center for Signal Processing, Dept. of EIE, The Hong Kong Polytechnic University, the University for the research studentship, and the Research Grants Council of the HKSAR, China for a grant (Grant No. PolyU 5119/12E).

REFERENCES

[1] J. Vanne, M. Viitanen, T. Hmlinen, and A. Hallapuro, "Comparative rate-distortion-complexity analysis of HEVC and AVC video codecs," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1885–1898, Dec 2012.

[2] B. Bross, P. Helle, S. Oudin, T. Nguyen, D. Marpe, H. Schwarz, and T. Wiegand, "Quadtree structures and improved techniques for motion representation and entropy coding in HEVC," in *Proc. IEEE Int. Conf. on Consumer Electronics, (ICCE-Berlin)*, Sept 2012, pp. 26–30.

[3] F. Bossen, B. Bross, K. Suhring, and D. Flynn, "HEVC complexity and implementation analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1685–1696, Dec 2012.

[4] C. M. Kuo, Y. H. Kuan, C. H. Hsieh, and Y. H. Lee, "A novel prediction-based directional asymmetric search algorithm for fast block-matching motion estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 6, pp. 893–899, June 2009.

[5] W. Dai, O. Au, S. Li, L. Sun, and R. Zou, "Adaptive search range algorithm based on cauchy distribution," in *Proc. IEEE Int. Visual Commun. and Image Process. (VCIP 2012)*, Nov 2012, pp. 1–5.

[6] L. Jia, O. C. Au, C. ying Tsu, Y. Shi, R. Ma, and H. Zhang, "A diamond search windowbased adaptive search range algorithm," in *Proc. IEEE Int. Conf. Multimedia and Expo Workshops*, July 2013, pp. 1–4.

[7] L. Du, Z. Liu, T. Ikenaga, and D. Wang, "Linear adaptive search range model for uni-prediction and motion analysis for bi-prediction in HEVC," in *Proc. IEEE Int. Conf. on Image Process. (ICIP 2014)*, Oct 2014, pp. 3671–3675.

[8] Y. H. Lin and J. L. Wu, "A depth information based fast mode decision algorithm for color plus depth-map 3d videos," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 542–550, June 2011.

[9] T. K. Lee, Y. L. Chan, and W. C. Siu, "Depth-based adaptive search range algorithm for motion estimation in HEVC," in *Proc. Int. Conf. Digital Signal Process. (DSP2014)*, Aug 2014, pp. 919–923.

[10] K. Muller, P. Merkle, and T. Wiegand, "3-D video representation using depth maps," *IEEE Proc.*, vol. 99, no. 4, pp. 643–656, April 2011.

[11] J. Jiao, R. Wang, W. Wang, S. Dong, Z. Wang, and W. Gao, "Local stereo matching with improved matching cost and disparity refinement," *IEEE MultiMedia*, vol. 21, no. 4, pp. 16–27, Oct 2014.

[12] K. McCann, B. Bross, W. J. Han, I. K. Kim, K. Sugimoto, and G. J. Sullivan, "High efficiency video coding (HEVC) test model 14 (HM14) encoder description," JCT-VC and ISO/IEC, San Jose, US, March 2014.

[13] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," VCEG-M33, March 2001.