The following publication T. -K. Lee, Y. -L. Chan and W. -C. Siu, "Adaptive Search Range for HEVC Motion Estimation Based on Depth Information," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 27, no. 10, pp. 2216-2230, Oct. 2017 is available at https://doi.org/10.1109/TCSVT.2016.2583979.

# Adaptive Search Range for HEVC Motion Estimation based on Depth Information

Tsz-Kwan Lee, *Student Member, IEEE*, Yui-Lam Chan, *Member, IEEE*, and Wan-Chi Siu, *Fellow, IEEE*

***Abstract*—High Efficient Video Coding (HEVC) achieves twofold coding efficiency improvement compared to its predecessor H.264/MPEG-4 AVC. However, it suffers from high computational complexity due to its quad-tree structure in motion estimation. This paper exposes the usage of depth maps in the multiview video plus depth format for relieving the computational burden. The depth map provides an intimation of the objects' distance from the projected screen in a 3D scene, which is explored in adaptive search range determination in this paper. The proposed algorithm exploits the high temporal correlation between the depth map and the motion in texture. By utilizing this correlation, a depth/motion relationship map is built for a mapping process. For each block, this forms a tailor-made search range with a motion-aware asymmetric shape to skip unnecessary search points in motion estimation. The obtained search range can be further adjusted by taking the influence of 3D-to-2D projection into consideration. Simulation results reveal that, compared to the full search approach, the proposed algorithm can reduce the complexity by 93% on average whereas the coding efficiency can be maintained. Besides, the proposed search range determination can work well with other fast search motion estimation algorithms in the literature.**

***Index Terms*—Adaptive search range, high efficiency video coding, motion estimation, multiview video plus depth, video coding.**

## I. Introduction

THE latest High Efficiency Video Coding (HEVC) standard is targeted for efficient compression of high resolution (720p and 1080p) and 3D videos [1]. Compared with the H.264/MPEG-4 AVC standard, HEVC can reduce the bit rate by almost 50% with the similar perceptual video quality [2], [3]. HEVC adopts the same block-based hybrid video coding scheme [4], [5] used in the prior video compression standards. The achievement in coding gain results mainly from its more flexible block partition mechanism at the cost of high computational complexity [6]–[8]. In the encoding process as shown in Fig. 1, each picture is divided into coding tree units (CTUs), which is the base unit in HEVC [9]–[11]. The size of a CTU can be chosen as 64×64, 32×32, 16×16 or 8×8. A CTU is composed of a luma coding tree block (CTB), two chroma CTBs, and the associated syntax elements. The luma and chroma CTBs can be further partitioned into smaller blocks using a quad-tree structure. The leaves of the CTBs are specified as coding blocks (CBs). One luma CB, and its corresponding two chroma CBs, together with the syntax elements form a coding unit (CU). The CU shares the identical prediction mode (intra, inter, skip, or merge), and it acts as the root for a prediction unit (PU) partitioning structure. Fig. 1 lists out all possible PU modes. The PU is composed of prediction blocks (PBs) where the same prediction process is applied for its luma and chroma PBs. In the PU partitioning structure of HEVC, each luma/chroma CB can be further partitioned into one, two, or four rectangular shaped PBs. In HEVC, it adopts square motion partitions, symmetric motion partitions, and asymmetric motion partitions [12], as shown in Fig. 1. It means that every CU undergoes motion predictions by various types of PU partitions. With this flexible block partitioning mechanism, inter prediction consumes about 60-70% of the whole encoding time [13], [14].

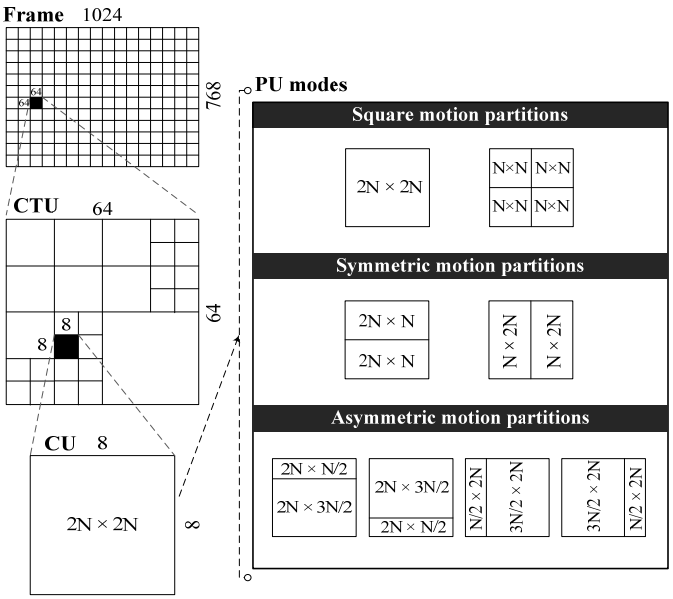



Fig. 1. Flexible block partitioning in HEVC: CTUs, CUs, and PUs.

Manuscript received September 17, 2015; revised February 23, 2016; accepted June 11, 2016. Date of publication XX XX, 2016; date of current version June 21, 2016. This work was supported by the Center for Signal Processing, Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, the research studentship provided by the University, and a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Grant No. PolyU 5119/12E).

The authors are with the Center for Signal Processing, Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong (e-mail: eie.glorylee@connect.polyu.hk; enylchan@polyu.edu.hk; enwcsiu@polyu.edu.hk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier XX.XXXX/TCSVT.XXXX.XXXXXXX

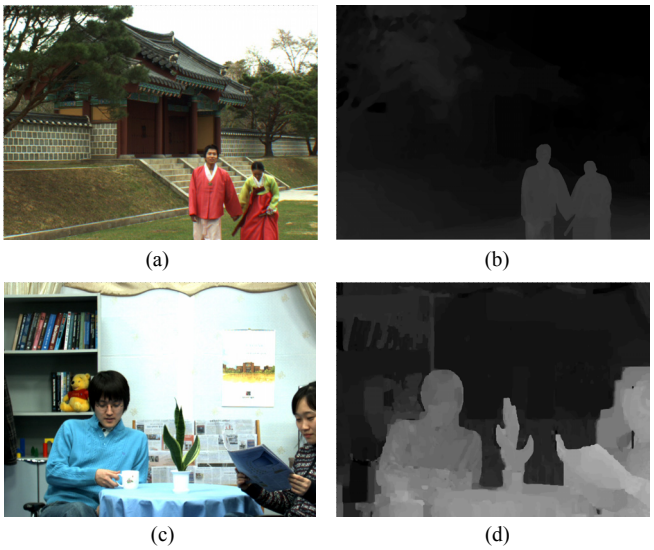



Fig. 2. Color texture and its associated depth map for a frame. (a) Color texture in "Lovebird1". (b) Depth map in "Lovebird1". (c) Color texture in "Newspaper". (d) Depth map in "Newspaper".

Recently, many researchers have devoted their efforts to expedite the inter prediction using fast mode decision or early mode termination in HEVC [15]–[20]. Early termination based on various coding information was suggested in [15]–[18]. For instance, zero coded block [15], [16] and the selection of SKIP mode [17], [18] are employed to trigger early termination of CU size decision. The works in [19], [20] further exploited the spatio-temporal analysis, motion homogeneity, and RD cost to determine the condition of early termination. These methods focus on reducing the computational complexity of selecting the best CU and PU, which are also highly related to the motion estimation (ME) algorithms. Fast ME algorithms always restrict the number of search locations. Test Zone Search (TZS) is one of the popular methods implemented in the HEVC test model (HM) [21], [22]. TZS starts with a diamond or square search pattern with different stride lengths of 1, 2, 4, 8, 16, 32, and 64 to locate an initial search point. This initial search point is taken as the center search point for the possible raster search and refinement. In [23], a rotating hexagonal grid with alternate horizontal/vertical hexagonal patterns was suggested for TZS to locate the global minima with early termination. However, the multiple initial search point decision is still a major burden on TZS [24]. In addition to TZS, other search strategies [25]-[27] and directional search were suggested in [28], [29]. These works focus on applying specific search patterns to reduce search points within a fixed search range. Nevertheless, various search patterns are not preferable for hardware implementation due to their irregular data flow [30]-[33]. In this circumstance, full search with an adaptive search range (ASR) can provide both search point reduction and regular data flow. Besides, ASR can also be applied to various search patterns in software implementation to further reduce the number of search points. In [34], [35], the search range is modeled by the Cauchy distribution and Laplace distribution, which exhibit good results in terms of quality and complexity in H.264. Other ASR algorithms proposed in H.264 correlates the search range of the current block with the motion characteristics of its neighbors. Examples of these motion characteristics include motion vector predictors [36], [37], sum of absolute difference [38], and motion activities [39] from neighboring blocks, motion vector differences in previous frames [40], etc. Recently, these concepts of ASR have been directly extended to support the flexible block partition in HEVC [30], [41]–[46]. But there are only a few of them trying to adopt the new features provided in bitstreams for ASR. For example, the ASR algorithm given in [46] uses the correlation between views in 3D-HEVC for ASR adjustment. However, the disparity among views might reduce the correlation between the search range of the view being coded and the motion vectors of its neighboring views. In [47], the authors tried to reveal the usage of depth information for fast mode decision. This algorithm makes the fast decision on selecting SKIP, inter-mode, and intra-mode in H.264 coding only. To the best of our knowledge, there is no existing work of ASR considering the depth information of 3D videos, which has gained great attention recently. In this paper, the depth information brings new room for designing an efficient ASR algorithm in HEVC.

The rest of this paper is organized as follows. Section II illustrates the motivation of using depth maps for search range determination. Section III exploits the temporal correlation between the depth map and motion in texture. The correlation makes the development of a new ASR algorithm possible for speeding up the ME process in HEVC. The proposed idea of linkage between depth maps and motions for the purpose of search range adjustment is then introduced in Section III. First, we describe the construction of a depth/motion relationship map (DMRMap) based on the correlation between the depth map and the motion in texture. Second, by making use of the DMRMap, the retrieval of ASR for the block being encoded is presented. Furthermore, the final adjustment of the search range due to the influence of the 3D space to the 2D image plane projection is discussed in Section IV. The entire proposed depth-based ASR algorithm based on the construction of DMRMap and the search range adjustment due to 3D-to-2D projection is conveyed in Section V. Simulation results of the proposed algorithm are provided in Section VI. Finally, Section VII concludes this work, with some idea for future work.

## II. Motivation of using Depth Maps in Search Range Adjustment of Motion Estimation

In recent years, we have witnessed the rapid development of 3D video technology. Among various 3D video representations, the multiview video plus depth (MVD) [48] is emerging as the most flexible format. The MVD includes both the color texture and the depth map of the captured scene. Two color textures, "Lovebird1" and "Newspaper", and their associated depth maps are shown in Fig. 2. In Fig. 2(a) and Fig. 2(c), the objects in front are the couples walking towards the camera and two teenagers sitting together, respectively. These color textures capture the luminance and chrominance

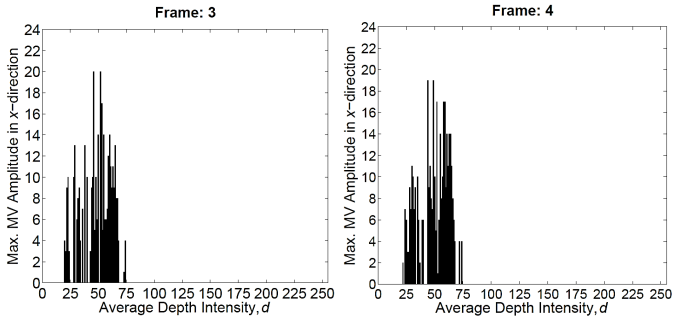



Fig. 3. The maximum amplitude of *x*-component motion vectors *MV* in quarter pixel of color texture for various average depth intensity values between consecutive frames, (a) Frame 3 and (b) Frame 4 of "Lovebird1".

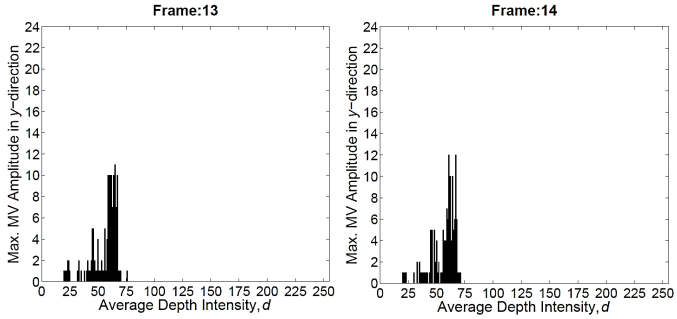



Fig. 4. The maximum amplitude of *y*-component motion vectors *MV* in quarter pixel of color texture for various average depth intensity values between consecutive frames, (a) Frame 3 and (b) Frame 4 of "Lovebird1".

information of pixels in the scenes while the depth maps in Fig. 2(b) and Fig. 2(d) record the distances of the objects associated with every pixel of the color textures. The lighter of the grayscale in the depth map, the closer of the object to the viewer. At the decoder side, depth maps can be used to synthesize arbitrary numbers of extra views through Depth-Image-Based Rendering (DIBR) techniques [48], [49]. The depth map for each frame is a piece of additional information to be encoded in the state-of-the-art 3D video encoder.

In [47], the authors utilized depth information for fast mode decision. By using depth information, a video scene is divided into near, middle, and far regions. Various mode candidates of a macroblock are chosen according to the classified region the macroblock belongs to. Instead of fast mode decision, our previous work in [50] has initially designed an efficient ASR algorithm in HEVC by depth information, but only limits to two particular CU sizes with symmetric search range. In this paper, we further extend our work in [50] to all block sizes with an asymmetric search range and a flexible scale factor depending on depth information. This paper proposes a depth information based ASR algorithm to adjust the search range to speed up ME. Depth maps are able to provide the additional intimation on areas/pixels belonging to objects in the same distance. In Fig. 2(b) and Fig. 2(d), the distinguished objects can be obviously figured out by the depth map since different video objects should have the distinct distance in the scene. In other words, same object, which shares the similar motion activities, can easily be identified by the depth maps. Once these objects are located, different search ranges can be applied to different objects in order to expedite ME.

## III. Temporal Correlation between Depth Map and Motion in Texture Streams

Fig. 3(a) and 3(b) plots the maximum amplitude of motion vectors of color texture in the *x*-direction for various average depth intensity values of all blocks in two consecutive frames while Fig. 4(a) and 4(b) shows the maximum amplitude of motion vectors of color texture in the *y*-direction for various average depth intensity. From these graphs, it can be seen that they have very similar distribution. It is because the depth information of an object not only represents the physical object position but also exhibits the motion activities of the object itself on each frame. It reflects that blocks with similar average depth intensity value will usually have similar motion vectors over a period of time. By making use of this temporal correlation between the depth map and motion in texture, we establish depth and motion relationship for each frame, and it is referred to as a depth/motion relationship map (DMRMap).

With the aid of the relationship map, motion activities of objects between consecutive frames could be roughly predicted by depth maps. In this paper, the search range is adopted according to the proposed DMRMap. Therefore, unnecessary search points within the pre-defined search range can be removed.

### *A. DMRMap Construction in Reference Frame*

This paper proposes a framework to obtain and maintain the depth/motion relationship map (DMRMap) of a reference frame, which can be used to determine the search range of the current frame. The DMRMap captures the relationship between motion activity and average depth intensity of all blocks in a reference frame. The proposed algorithm should start with any frame other than the first inter-frame because the reference frame of the first inter-frame is intra-coded, and no motion vectors from this reference frame can be obtained. Therefore, the first inter-frame will go through a conventional full rate-distortion optimization (RDO) inter coding. Once the motion vectors of all blocks in the frame are obtained, its DMRMap is constructed for ME of the next frame. Let $d$ be the average depth intensity values of a block in the reference frame, where $0 \le d \le 255$. It is noted that depth maps are always estimated using stereo matching methods [51], which induces slight variation or noise of depth values within the same object. To tolerate the variation of depth values in an object, $d$ is divided into an appropriate number of ranges, each containing many similar values of $d$. To do so, $d$ is quantized uniformly by a quantization factor $Q$ into $\hat{d}$, where $0 \le \hat{d} \le \lceil 255/Q \rceil$. Note that $\lceil\ \rceil$ is the ceiling function. The DMRMap relates the largest motion vectors to all possible values of $\hat{d}$ in the reference frame. Assume that $S_x^{\hat{d}}$ and $S_y^{\hat{d}}$ are the sets of motion vectors in the *x*- and *y*-direction, respectively, with the blocks in which their quantized average depth intensity value is $\hat{d}$. The largest motion vector, $\left(mv_x^{max}(\hat{d}), mv_y^{max}(\hat{d})\right)$, in

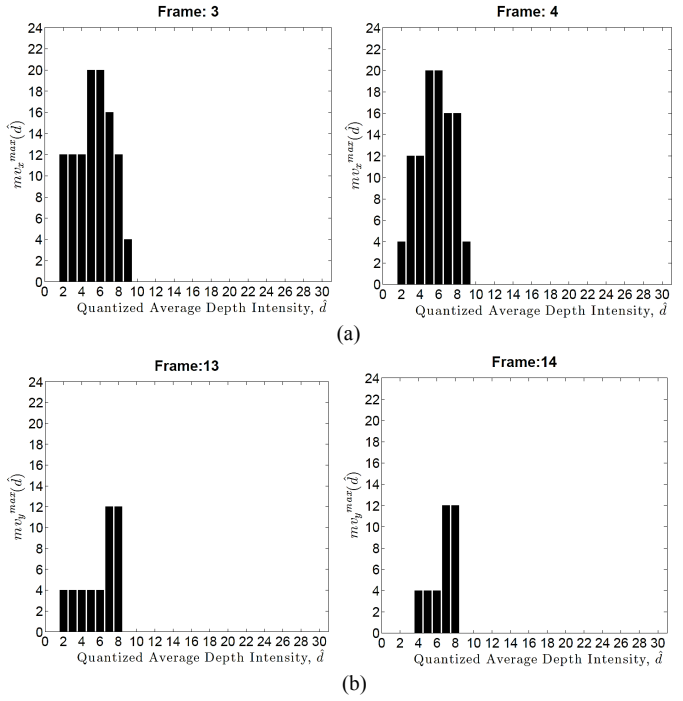



Fig. 5. The largest motion vector (a) $mv_x^{max}(\hat{d})$ in the *x*-direction, and (b) $mv_y^{max}(\hat{d})$ in the *y*-direction from a pair of consecutive frames for "Lovebird1".

the *x*- and *y*-directions are the maximum values in $S_x^{\hat{d}}$ and $S_y^{\hat{d}}$ as follows

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

and

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

where max(*S*) gives the maximum value of the set *S*. Actually, $mv_x^{max}(\hat{d})$ and $mv_y^{max}(\hat{d})$ can be used to describe the DMRMap, which constructs the relationship between the largest motion vectors and $\hat{d}$. In other words, the largest motion vector in both of the *x*- and *y*-directions can be determined for the given $\hat{d}$.

The DMRMap will be updated frame by frame. Two relationship maps in the *x*- and *y*-directions constructed from a pair of consecutive frames for "Lovebird1" are illustrated in Fig. 5(a) and 5(b), respectively, in which *Q* is set to 8. They record $mv_x^{max}(\hat{d})$ and $mv_y^{max}(\hat{d})$ for each $\hat{d}$ within the frame, where $\hat{d}$ is from 0 to 31. The value of *Q* depends on the level of noise in the depth map. The more the noise in the depth map, the larger the value of *Q* is used to absorb depth variation in the same object. However, a large *Q* results in affecting the precision of DMRMap, which will have a detailed discussion in Section VI. Setting *Q* to 8 is always appropriate for the quality of most depth maps recommended by the ISO/IEC and ITU-T JCT-3V group with the reasonably good DMRMap. From Fig. 5(a) and 5(b) for the DMRMap, it can be observed that the distributions are very similar to each other for consecutive frames in both of the horizontal and vertical movements.

### *B. Adaptive Search Range Decision based on Mapping Process using DMRMap*

By utilizing the temporal correlation of DMRMaps between two consecutive frames, the mapping process from the average depth intensity for the $i^{th}$ block being encoded in frame *t*, $B_t^i$, to its search range, denoted as $SR_x(B_t^i)$ and $SR_y(B_t^i)$ in the *x*- and *y*-directions, respectively, is conducted. The mapping is based on the DMRMap in the reference frame as defined in (1) and (2). Let $QDepth(B_t^i)$ be the average depth intensity values after quantization for $B_t^i$. From (1) and (2), $SR_x(B_t^i)$ and $SR_y(B_t^i)$ can then be computed as

$$SR_x(B_t^i) = mv_x^{\max}(QDepth(B_t^i)) \tag{3}$$

and

$$SR_y(B_t^i) = mv_y^{\max}(QDepth(B_t^i)) \tag{4}$$

This mapping process is to correlate the temporal information by the average depth intensity value of the block being encoded to those blocks in the reference frame. Since depth maps indicate the location of an object in the video scene from the image plane, the average depth intensity value could therefore be a criterion for distinguishing various objects with different distances in a video scene. Based on this observation, it is likely that the blocks belonging to one particular video object across consecutive frames have consistent motion associated with the similar average depth intensity values. Once the average depth intensity value of the current block is matched to those in the same intensity value of the reference frame, ASR decision can be made from the DMRMap in the reference frame. It is noted that, if $QDepth(B_t^i)$ is empty in the DMRMap of the reference frame, the search ranges of $B_t^i$ in both *x*- and *y*-directions are set to 64. It is the default search range of the main profile in HEVC, which is larger or equal to the values obtained in (3) and (4). After all motion vectors of the frame being encoded are determined, the DMRMap is updated for both *x*- and *y*-directions for the next frame.

## IV. Influence of 3D-to-2D Projection on Motion Activity on 2D Image Plane

The working principle of the mapping process in (3) and (4) is based on the very strong temporal correlation of DMRMaps between the current and reference frames. However, 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 between frames. This motion activity along the camera axis (*z*-axis) has the potential to weaken the degree of this correlation, which reduces the prediction accuracy of the search range in (3) and (4). Taking this into consideration, a scale factor for the search range of $B_t^i$, $\rho(B_t^i)$, is added to offer extra flexibility in the determination of the search range in (3) and (4). The search range prediction is then scaled as follow

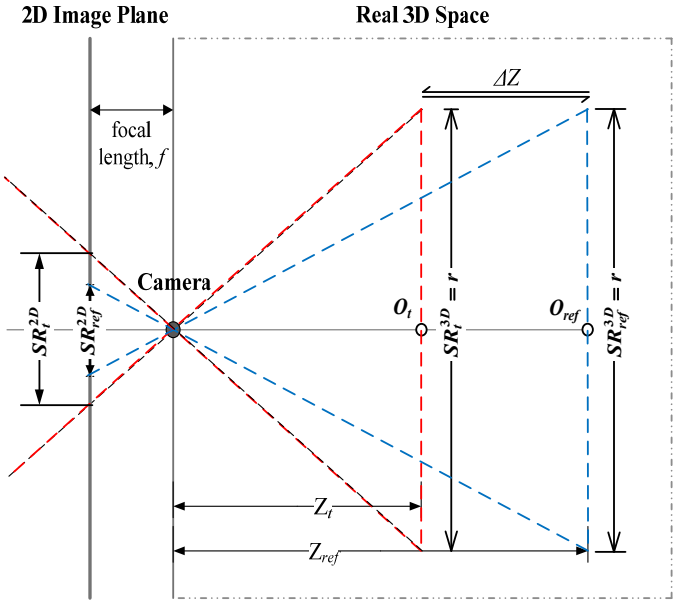



Fig. 6. Geometric relationship between depth of object and motion activity on the 2D image plane.

$$SR_x^{\rho}(B_t^i) = \rho(B_t^i) \times mv_x^{\max}(QDepth(B_t^i)) \tag{5}$$

and

$$SR_y^{\rho}(B_t^i) = \rho(B_t^i) \times mv_y^{\max}(QDepth(B_t^i)) \tag{6}$$

Note that $\rho(B_t^i) = 1$ is the case for the scene without motion along the $z$-axis. In this case, (5) and (6) are equal to (3) and (4), respectively. The change in depth intensity between frames actually reflects the degree of $z$-axis motion, which in turn gives a good estimation of $\rho(B_t^i)$. The way to determine $\rho(B_t^i)$ is underlying on the motion parallax. It states 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 2D image plane is always influenced by how close it is located to the camera in the 3D space. When the object is closer to the camera, projected displacement on the 2D image plane is larger. This situation is illustrated in Fig. 6. In this figure, an example of the geometric relationship between the depth information of an object moving towards the camera and its displacement variation on the 2D image plane is depicted. Let $O_{ref}$ denote a 3D position of an object in the 3D space. It is noted that the subscript *ref* represents the reference position in the following discussion. The actual distance value between $O_{ref}$ and the camera is $Z_{ref}$, and its search range is assumed to be $SR_{ref}^{3D}$ in the 3D space. Assume that $O_{ref}$ moves to $O_t$ with the actual distance value of $Z_t$ at time $t$. In this case of the object moving towards the camera, $Z_t$ is smaller than $Z_{ref}$ since $O_t$ is closer to the camera than $O_{ref}$, as shown in Fig. 6. Similarly, the search range of $O_t$ is $SR_t^{3D}$. In Fig. 6, $SR_{ref}^{2D}$ and $SR_t^{2D}$ are therefore the projections of $SR_{ref}^{3D}$ and $SR_t^{3D}$ on the 2D image plane, respectively. For the same search range for $O_{ref}$ and $O_t$ in the 3D space (i.e., $SR_{ref}^{3D} = SR_t^{3D} = r$), $SR_{ref}^{2D}$ is smaller than $SR_t^{2D}$ on the 2D image plane after projection, as illustrated in Fig. 6. Consequently, this phenomenon can be used to determine $\rho(B_t^i)$. In (5) and (6), $\rho(B_t^i)$ is a factor to scale the search range projected on the 2D image plane at time $t$ due to the motion along the $z$-axis, which is the ratio of $SR_t^{2D}$ to $SR_{ref}^{2D}$ defined by

$$\rho(B_t^i) = \frac{SR_t^{2D}}{SR_{ref}^{2D}} \tag{7}$$

The dotted lines in Fig. 6 indicate the trajectory of the projections through the camera lens onto the 2D image plane. Using triangular similarity, the relationship between $SR_{ref}^{2D}$ and $SR_t^{2D}$ can be correlated to the actual distances, $Z_{ref}$ and $Z_t$ in the 3D space as

$$\frac{SR_{ref}^{2D}}{f} = \frac{r}{Z_{ref}} \tag{8}$$

and

$$\frac{SR_t^{2D}}{f} = \frac{r}{Z_t} \tag{9}$$

where $f$ is the focal length of the camera, and $r$ is the search range of the object in the 3D space. By combining (7), (8) and (9), $\rho(B_t^i)$ can be formulated as

$$\rho(B_t^i) = \frac{Z_{ref}}{Z_t} = 1 + \frac{\Delta Z}{Z_t} \tag{10}$$

where $\Delta Z$ is the change in the actual distance between $O_{ref}$ and $O_t$ in the 3D space due to the $z$-axis motion of the object, as shown in Fig. 6. The positive ΔZ means the object moving towards the camera since physically $Z_t < Z_{ref}$ while the negative ΔZ means the object moving away from the camera due to $Z_t > Z_{ref}$. In addition, there is no z-axis motion when ΔZ is equal to zero.

In (10), it introduces the scale factor based on the changes in the actual distance between the current and reference blocks. 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, $Depth(B_t^i)$ and $Depth(B_{ref}^i)$, for $B_t^i$ in the current frame and its co-located block $B_{ref}^i$ in the reference frame, respectively, as

$$Z_t = 1 / \left[ \frac{Depth(B_t^i)}{255} \cdot \left( \frac{1}{Z_{near}} - \frac{1}{Z_{far}} \right) + \frac{1}{Z_{far}} \right] \tag{11}$$

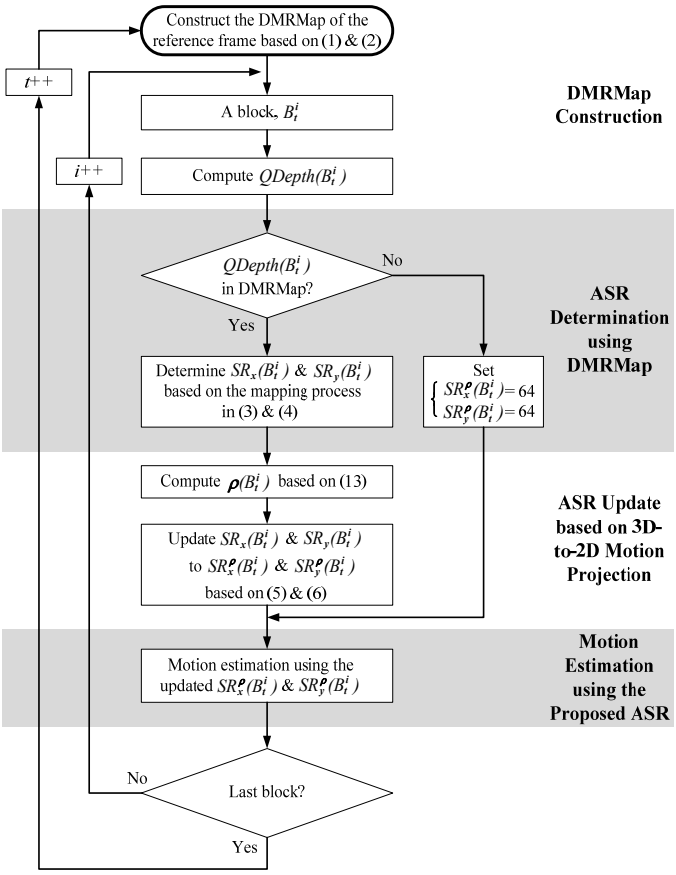



Fig. 7. Flowchart of the proposed DMRMap-based ASR algorithm.

TABLE I
VALUES OF $Z_{NEAR}$ AND $Z_{FAR}$ IN VARIOUS SEQUENCES

| Sequences | $Z_{near}$ (cm) | $Z_{far}$ (cm) |
|---|---|---|
| Balloons | 448.251214 | 11206.280350 |
| Kendo | 448.251214 | 11206.280350 |
| Lovebird1 | -2228.745812 | -156012.206815 |
| Newspaper | -2715.181648 | -9050.605493 |
| Poznan_Street | -34.506386 | -2760.510889 |
| Poznan_Hall2 | -23.394160 | -172.531931 |
| Undo_Dancer | 2289 | 213500 |
| GT_Fly | changes every frame between a range of $\{Z_{near}, Z_{far}\}$ ={3156.3, 1000000000} | |

and

$$Z_{ref} = 1 \Big/ \left[ \frac{Depth(B_{ref}^i)}{255} \cdot \left( \frac{1}{Z_{near}} - \frac{1}{Z_{far}} \right) + \frac{1}{Z_{far}} \right] \quad (12)$$

where $Z_{near}$ and $Z_{far}$ are the smallest and largest actual distance among all points captured by the camera, which are recorded in the camera configure files of the test sequences recommended by the ISO/IEC and ITU-T JCT-3V group. Their values of $Z_{near}$ and $Z_{far}$ are listed in Table I in which positive or negative values denote the viewing direction of the camera. It is noted that the values of $Z_{near}$ and $Z_{far}$ are signaled with the 3D videos for a correct geometric displacement in synthesized intermediate views [49] at decoder side. By putting (11) and (12) into (10), $\rho(B_t^i)$ is expressed as

$$\rho(B_t^i) = \frac{Depth(B_t^i)\left(Z_{far} - Z_{near}\right) + 255 Z_{near}}{Depth(B_{ref}^i)\left(Z_{far} - Z_{near}\right) + 255 Z_{near}} \quad (13)$$

and it can be summarized as

$$\begin{cases} \text{case 1:} & \rho(B_t^i) > 1, \ Depth(B_t^i) > Depth(B_{ref}^i) \\ \text{case 2:} & \rho(B_t^i) = 1, \ Depth(B_t^i) = Depth(B_{ref}^i) \\ \text{case 3:} & \rho(B_t^i) < 1, \ Depth(B_t^i) < Depth(B_{ref}^i) \end{cases} \quad (14)$$

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.

## V. PROPOSED DMRMAP-BASED ADAPTIVE SEARCH RANGE (ASR) ALGORITHM

Fig. 7 shows the flowchart of the proposed DMRMap-based ASR algorithm to encode a frame in HEVC. The proposed ASR algorithm has three new features: (a) DMRMap construction of the reference frame; (b) ASR determination using DMRMap; and (c) ASR update based on 3D-to-2D motion projection. Combining these three techniques, the proposed DMRMap-based ASR algorithm can be applied to the block $B_t^i$ being encoded as follows.

Step (i) Construct the DMRMap of the reference frame according to (1) and (2).
Step (ii) Compute $QDepth(B_t^i)$ for the mapping process.
Step (iii) If $QDepth(B_t^i)$ is available in DMRMap, go to Step (iv); otherwise, go to Step (vii)
Step (iv) Obtain the horizontal $SR_x(B_t^i)$, and the vertical $SR_y(B_t^i)$ based on the mapping processing in (3) and (4), respectively.
Step (v) Determine the scale factor $\rho(B_t^i)$ as (13).
Step (vi) Update the horizontal $SR_x(B_t^i)$ to $SR_x^\rho(B_t^i)$ and the vertical $SR_y(B_t^i)$ to $SR_y^\rho(B_t^i)$ according to (5) and (6), respectively by $\rho(B_t^i)$. Go to Step (viii).
Step (vii) Set both of the horizontal $SR_x^\rho(B_t^i)$ and the vertical $SR_y^\rho(B_t^i)$ to 64.
Step (viii) Perform motion estimation using $SR_x^\rho(B_t^i)$ and $SR_y^\rho(B_t^i)$.

## VI. SIMULATION RESULTS

To evaluate the performance of the DMRMap-based ASR algorithm, the techniques proposed in Section III and Section IV have been integrated into the HM 14.0 reference software [52], and tested under the low-delay P configuration specified in the common test condition [53] of the HEVC standardization in which the main profile of HEVC was used. An I-frame was allowed in the first frame only, and the rest were encoded as P-frames. All CU-level of 64×64, 32×32, 16×16, and 8×8 were enabled. For PU and TU, a full quad-tree structure was utilized. All tested algorithms were evaluated

TABLE II
BJONTEGAARD (BD) MEASUREMENT AND CODING TIME CHANGE OF FS+LSMF, FS+MLELD, FS+LAMASR, FS+DMRMAP, AND FS+DMRMAP+SCALING FOR ASR AGAINST FS IN HEVC

| Seq. | QP | FS | | FS+LSMF [43] | | | FS+MLELD [40] | | | FS+LAMASR [44] | | | FS+DMRMap | | | FS+DMRMap+Scaling | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | PSNR (dB) | Bitrate (kbps) | Δtime (%) | BD-PSNR (ΔdB) | BD-rate (%) | Δtime (%) | BD-PSNR (ΔdB) | BD-rate (%) | Δtime (%) | BD-PSNR (ΔdB) | BD-rate (%) | Δtime (%) | BD-PSNR (ΔdB) | BD-rate (%) | Δtime (%) | BD-PSNR (ΔdB) | BD-rate (%) |
| *720p* | | | | | | | | | | | | | | | | | | |
| Balloons | 37 | 38.38 | 335.94 | -40.54 | +0.01 | +0.13 | -63.17 | 0.00 | +0.07 | -74.01 | 0.00 | -0.06 | -93.43 | -0.01 | +0.18 | -89.63 | -0.01 | +0.26 |
| | 32 | 41.24 | 593.35 | | | | | | | | | | | | | | | |
| | 27 | 43.56 | 1171.64 | | | | | | | | | | | | | | | |
| | 22 | 45.46 | 3134.29 | | | | | | | | | | | | | | | |
| Kendo | 37 | 39.66 | 372.48 | -53.53 | 0.00 | +0.07 | -80.74 | -0.01 | +0.23 | -89.99 | -0.01 | +0.25 | -93.01 | -0.01 | +0.30 | -94.01 | -0.01 | +0.26 |
| | 32 | 42.27 | 654.66 | | | | | | | | | | | | | | | |
| | 27 | 44.44 | 1236.53 | | | | | | | | | | | | | | | |
| | 22 | 46.29 | 2900.38 | | | | | | | | | | | | | | | |
| Lovebird1 | 37 | 34.32 | 164.72 | -71.10 | -0.01 | +0.35 | -87.87 | 0.00 | +0.12 | -94.96 | 0.00 | +0.06 | -99.42 | 0.00 | +0.11 | -99.46 | 0.00 | +0.08 |
| | 32 | 37.22 | 353.62 | | | | | | | | | | | | | | | |
| | 27 | 40.29 | 830.90 | | | | | | | | | | | | | | | |
| | 22 | 43.64 | 2069.84 | | | | | | | | | | | | | | | |
| Newspaper | 37 | 35.71 | 242.51 | -54.77 | -0.02 | +0.45 | -73.84 | -0.02 | +0.47 | -90.94 | -0.01 | +0.23 | -95.82 | -0.01 | +0.39 | -97.23 | -0.01 | +0.31 |
| | 32 | 38.45 | 451.86 | | | | | | | | | | | | | | | |
| | 27 | 41.08 | 960.40 | | | | | | | | | | | | | | | |
| | 22 | 43.81 | 2711.91 | | | | | | | | | | | | | | | |
| *1080p* | | | | | | | | | | | | | | | | | | |
| Poznan_Street | 37 | 35.10 | 461.62 | -39.99 | -0.01 | +0.55 | -42.87 | -0.01 | +0.51 | -69.47 | -0.02 | +0.74 | -96.38 | -0.01 | +0.48 | -97.15 | 0.00 | +0.17 |
| | 32 | 37.48 | 1142.23 | | | | | | | | | | | | | | | |
| | 27 | 39.94 | 4223.58 | | | | | | | | | | | | | | | |
| | 22 | 43.24 | 24872.19 | | | | | | | | | | | | | | | |
| Poznan_Hall2 | 37 | 39.59 | 237.06 | -22.98 | 0.00 | +0.05 | -41.25 | -0.01 | +0.28 | -40.37 | -0.01 | +0.22 | -85.24 | -0.01 | +0.50 | -86.72 | -0.01 | +0.41 |
| | 32 | 41.09 | 501.60 | | | | | | | | | | | | | | | |
| | 27 | 42.25 | 1568.65 | | | | | | | | | | | | | | | |
| | 22 | 44.24 | 13711.65 | | | | | | | | | | | | | | | |
| Undo_Dancer | 37 | 33.01 | 1525.83 | -25.14 | -0.01 | +0.42 | -36.78 | -0.01 | +0.39 | -34.74 | -0.02 | +0.69 | -94.59 | -0.01 | +0.29 | -91.74 | -0.01 | +0.30 |
| | 32 | 35.76 | 4303.43 | | | | | | | | | | | | | | | |
| | 27 | 39.02 | 11217.01 | | | | | | | | | | | | | | | |
| | 22 | 42.69 | 25191.12 | | | | | | | | | | | | | | | |
| GT_Fly | 37 | 35.82 | 1166.74 | -25.41 | 0.00 | 0.00 | -27.32 | 0.00 | +0.03 | -36.41 | 0.00 | +0.14 | -96.56 | -0.01 | +0.21 | -95.47 | -0.01 | +0.27 |
| | 32 | 38.44 | 3023.78 | | | | | | | | | | | | | | | |
| | 27 | 41.10 | 7445.85 | | | | | | | | | | | | | | | |
| | 22 | 43.85 | 16927.17 | | | | | | | | | | | | | | | |
| **Average:** | | | | **-41.68** | **-0.01** | **+0.25** | **-56.73** | **-0.01** | **+0.26** | **-66.36** | **-0.01** | **+0.28** | **-94.31** | **-0.01** | **+0.31** | **-93.93** | **-0.01** | **+0.26** |

with four QPs of 22, 27, 32, and 37 using eight test sequences with two resolutions of 1280×720 and 1920×1088.

Two sets of experiments were performed to evaluate the overall efficiency of applying our proposed DMRMap-based ASR algorithm to various ME search strategies. First, the proposed DMRMap algorithm with and without the scale factor $\rho(B_t^i)$ in (13) have been incorporated into the conventional full-search (FS) in order to provide ASR for ME, and let us call them FS+DMRMap+Scaling and FS+DMRMap, respectively. $Q$ was set to 8 in both of FS+DMRMap+Scaling and FS+DMRMap. Three most recent ASR algorithms have also been implemented for comparisons, and they are referred to as FS+MLELD [40], FS+LSMF [43], and FS+LAMASR [44], respectively. FS+MLELD [40] models the motion vector differences of the previous frame by the zero-mean Laplace distribution where the parameters are solved by maximum likelihood estimation (MLE) for a motion estimator. The estimated distribution model is then used to set the final ASR. FS+LSMF [43] classifies the current block either in a small motion frame (SMF) or a large motion frame (LMF) by the distribution of motion vector differencs in the previous frame. Consequenctly, it differentiates the current block into two sub-classes of different degrees of motion activity by the average motion vector difference of the co-located CTU. Larger ASR is assigned in high motion activity blocks and vice versa. FS+LAMASR [44] adopts a linear adaptive search range model including an overdetermined equation system. The parameters in the system can be solved in considerations on PU size, motion vector difference and motion vector predictors. The ASR is then finalized with a fixed scale factor. Second, we demonstrate the performance of the proposed DMRMap applied to the Test Zone Search (TZS), named as TZS+DMRMap+Scaling. TZS was employed in the H.264 joint scalable video model (JSVM) [13], and TZS is also the only fast method adopted in the HEVC reference software [21], [22]. It can prove that our

TABLE III
SEARCH RANGE DIMENSION AND AVERAGE NUMBER OF SEARCH POINTS PER CU OF FS, FS+LSMF, FS+MLELD, FS+LAMASR, AND FS+DMRMAP+SCALING

| | Dimension (*Dn*, integer pixel) | | | | | | Average number of search points per CU (*SrPt*) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Sequences | FS | FS+ LSMF [43] | FS+ MLELD [40] | FS+ LAMASR [44] | FS+ DMRMap+ Scaling | | FS | FS+ LSMF [43] | FS+ MLELD [40] | FS+ LAMASR [44] | FS+ DMRMap+ Scaling |
| ***720p*** | $Dn_x = Dn_y$ | $Dn_x = Dn_y$ | | | $Dn_x$ | $Dn_y$ | ***2-D search window*** | | | | |
| Balloons | 64 | 47 | 38 | 32 | 23 | 14 | 16641 | 9025 | 5929 | 4225 | 1363 |
| Kendo | 64 | 42 | 27 | 20 | 13 | 18 | 16641 | 7225 | 3025 | 1681 | 999 |
| Lovebird1 | 64 | 34 | 21 | 14 | 3 | 6 | 16641 | 4761 | 1849 | 841 | 91 |
| Newspaper | 64 | 41 | 32 | 19 | 15 | 8 | 16641 | 6889 | 4225 | 1521 | 527 |
| ***1080p*** | $Dn_x = Dn_y$ | $Dn_x = Dn_y$ | | | $Dn_x$ | $Dn_y$ | ***2-D search window*** | | | | |
| Poznan_Street | 64 | 48 | 45 | 35 | 14 | 10 | 16641 | 9409 | 8281 | 5041 | 609 |
| Poznan_Hall2 | 64 | 55 | 48 | 50 | 28 | 22 | 16641 | 12321 | 9409 | 10201 | 2565 |
| Undo_Dancer | 64 | 54 | 50 | 52 | 11 | 25 | 16641 | 11881 | 10201 | 11025 | 1173 |
| GT_Fly | 64 | 54 | 53 | 51 | 17 | 11 | 16641 | 11881 | 11449 | 10609 | 805 |
| Average | 64 | 47 | 39 | 34 | 16 | 14 | 16641 | 9174 | 6796 | 5643 | 1017 |

proposed DMRMap-based algorithm is compatible to a fast ME algorithm.

In all simulations, Bjontegaard (BD) measurement [54] in terms of BD-rate (%) and BD-PSNR (ΔdB) were used to measure the average coding efficiency of various algorithms, and Δtime (%) represents coding time change in percentage as compared with the benchmarking algorithms. Positive and negative values denote increments and decrements, respectively. Note that the coding time includes the computational cost for all CU quad-tree levels. The test platform used for simulations was a 64-bit MS Windows 8.1 OS running on an Intel Core i7-4770 CPU of 3.4 GHz and 16.0 GB RAM.

### *A. Results of Applying DMRMap to FS*

The full-search (FS) algorithm gives the best and optimal rate distortion (RD) performance in block-based ME since it searches all points inside the predefined search range. The objective of the proposed DMRMap-based ASR algorithms is to provide a suitable and reasonable search range for both vertical and horizontal directions per block for ME in HEVC. As a result, unnecessary search points can be skipped such that better resource utilization in ME can be achieved.

Table II lists the BD measurement and Δtime of our proposed FS+DMRMap and FS+DMRMap+Scaling against FS for the eight depth-enhanced sequences. FS undergoes its fixed search range of 64 pixels, which means that 16641 search points are used for each block in ME. From Table II, FS+DMRMap and FS+DMRMap+Scaling can averagely save 94.31% and 93.93% of coding time over FS, respectively. The search ranges obtained by FS+DMRMap and FS+DMRMap+Scaling are always smaller than that of FS since they utilize the high temporal correlation of motions revealed by depth intensity mapping. A significant time reduction of around 99% by FS+DMRMap and FS+DMRMap+Scaling can be observed at "Lovebird1" sequence. This can be explained by the fact that "Lovebird1" consists of large portion of slow movement so that the proposed techniques can offer remarkable reduction in search range in the mapping process. As a result, both FS+DMRMap and FS+DMRMap+Scaling only consume about 1% encoding time of FS. While significant coding time reduction can be achieved, the coding efficiency of the proposed FS+DMRMap and FS+DMRMap+Scaling can be maintained as compared to FS. From the results of Table II, FS+DMRMap obtains negligible loss on BD-PSNR by 0.01dB as compared to FS while only 0.31% of BD-rate is raised. With the help of the proposed scale factor $\rho(B_t^i)$ on search range due to the 3D-to-2D projection, FS+DMRMap+Scaling also attains negligible loss on BD-PSNR by 0.01dB as compared to FS. At the same time, it only costs an increment of 0.26% in BD-rate.

Table II further lists out the results of FS+LSMF [43], FS+MLELD [40], and FS+LAMASR [44]. It can be observed that FS+LSMF [43], FS+MLELD [40], and FS+LAMASR [44] reduce the computational complexity by averagely 41.68%, 56.73%, and 66.36%, respectively while the proposed FS+DMRMap+Scaling reduces the complexity by 93.93%. The proposed FS+DMRMap+Scaling can save more computational time by about 52%, 37%, and 27%, respectively, as compared with FS+LSMF [43], FS+MLELD [40] and FS+LAMASR [44]. Meanwhile, these algorithms obtain very similar BD-rate deterioration. The reason is that FS+DMRMap+Scaling considers the search range in the *x*- and *y*-directions separately for tracing the true motion vectors. Furthermore, FS+DMRMap+Scaling utilizes an adaptive scale factor for ASR adjustment. On the other hand, all the algorithms in [40], [43], and [44] consider the search range in the *x*- and *y*-directions jointly. In addition, FS+LAMASR [44] simply multiplies a fixed scale factor to the sum of the amplitude for ASR.

From the results of Table II, it can be found that the gain in computational time for 1080p sequences is less significant compared to that of 720p sequences in FS+LSMF [43], FS+MLELD [40], and FS+LAMASR [44]. This can be explained by the fact that FS+LSMF [43] and FS+MLELD [40] adopt the motion vector difference distribution of the previous frame to determine the search range, and FS+LAMASR [44] uses the sum of amplitude differences

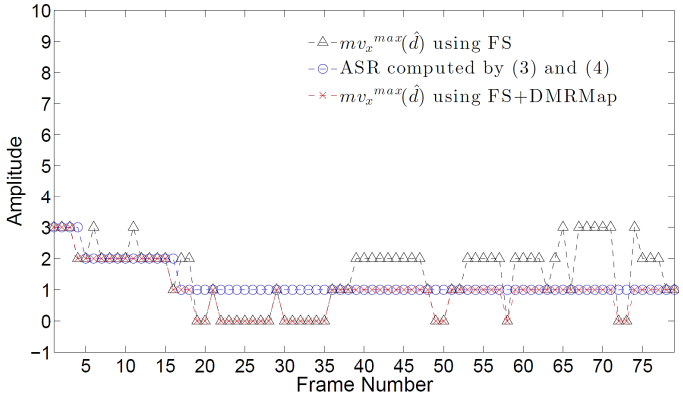



Fig. 8. The maximum absolute amplitude of moiton vectors, $mv_x^{max}(\hat{d})$ using FS and FS+DMRMap, and ASR with $\hat{d}$ = 7 along frames for color texture of "Lovebird1".

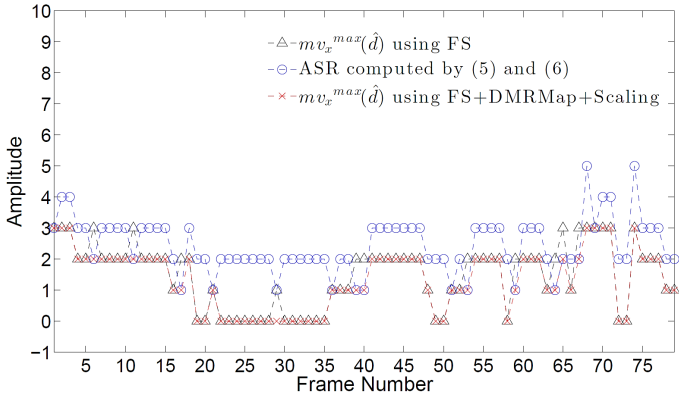



Fig. 10. The maximum absolute amplitude of moiton vectors, $mv_x^{max}(\hat{d})$ using FS and FS+DMRMap+Scaling, and ASR with $\hat{d}$ = 7 along frames for color texture of "Lovebird1".

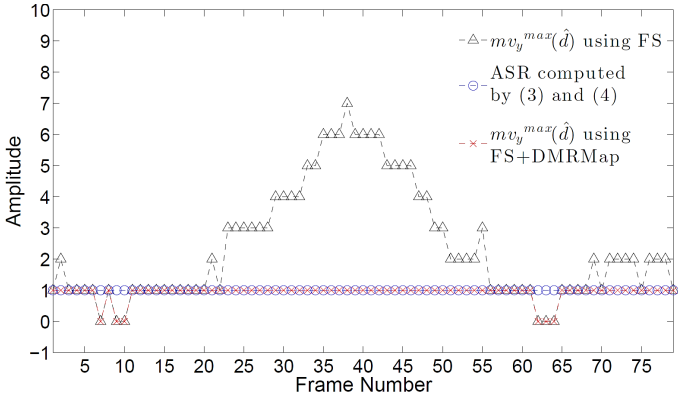



Fig. 9. The maximum absolute amplitude of moiton vectors, $mv_y^{max}(\hat{d})$ using FS and FS+DMRMap, and ASR with $\hat{d}$ = 14 along frames for color texture of "Newspaper".

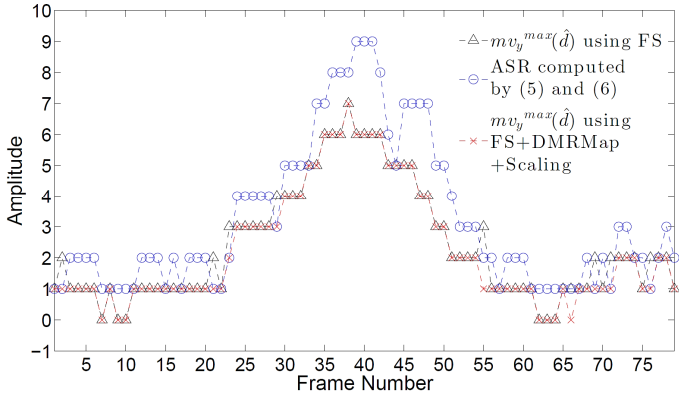



Fig. 11. The maximum absolute amplitude of moiton vectors, $mv_y^{max}(\hat{d})$ using FS and FS+DMRMap+Scaling, and ASR with $\hat{d}$ = 14 along frames for color texture of "Newspaper".

among all motion vector predictors of the current block to initialize the adaptive search range. In general, the motion vector differences among blocks are used as the hint to guide the search range determination for these three ASR algorithms. However, motion activities in 1080p test sequences are always richer than those in 720p test sequences [55]. In other words, motion vector differences between blocks are more likely to have abrupt change such that FS+LSMF [43], FS+MLELD [40] and FS+LAMASR [44] will have a larger search range in 1080p sequences (i.e. less time reduction as a result). On the contrary, our proposed FS+DMRMap and FS+DMRMap+Scaling can obtain the consistent gain in computational time for both 1080p and 720p sequences, as shown in Table II, since they only use the depth map for search range determination, which is insensitive to the video resolution.

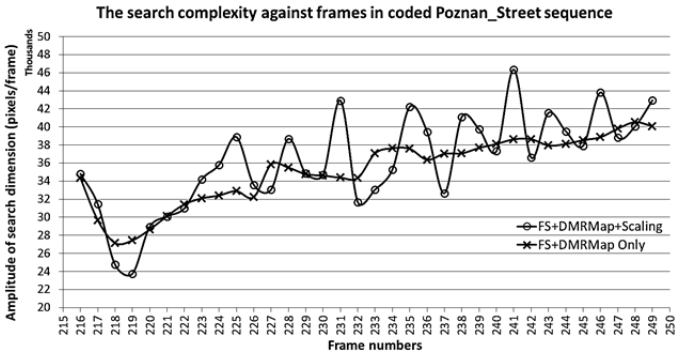
Table III further compares the average sizes of the search range in the *x*- and *y*-directions of FS, FS+LSMF [43], FS+MLELD [40], FS+LAMASR [44], and the proposed FS+DMRMap+Scaling. This table records the *x*- and *y*-dimensions of the search range ($Dn_x$ and $Dn_y$, respectively) and the average number of search points *(SrPt)* per block for the five tested algorithms. For FS, FS+LSMF [43], FS+MLELD [40], and FS+LAMASR [44], $Dn_x$ is equal to $Dn_y$, and they are 64, 47, 39, and 34 on average, respectively, as shown in Table III. It implies that all FS, FS+LSMF [43], FS+MLELD [40], and FS+LAMASR [44] obtain a search window with aspect ratio of 1. For FS+DMRMap+Scaling, the search ranges in the *x*- and *y*-directions are computed independently. Along the sequences, the aspect ratio of the search window is no longer equal to 1, and it depends on the motion characteristic of the sequence. The proposed FS+DMRMap+Scaling therefore can adopt the search window with various aspect ratios for well fitting the true motion. As a result, the average number of search points for each CU is computed as (15) and listed in Table III.

$$SrPt = (2 \times Dn_x + 1) \cdot (2 \times Dn_y + 1) \tag{15}$$

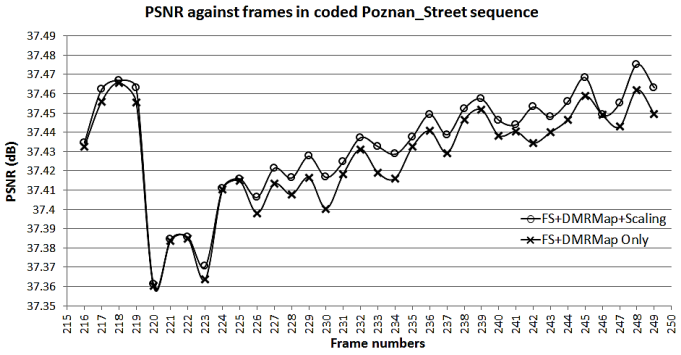
In (15), *SrPt* is defined as the number of search points in a search window based on $Dn_x$ and $Dn_y$. Finally, Table III shows that FS+LAMASR only requires around one third of search points per CU compared to FS whereas the proposed FS+DMRMap+Scaling only occupies averagely less than one tenth of search points for compared to FS.

### *B. Gains of Scaling Technique on DMRMap*

From the results in Table II, we can see that FS+DMRMap+Scaling obtains a slight decrease in BD-rate compared with FS+DMRMap. The gain is contributed from the scale factor $\rho(B_t^i)$ in (13) that can adjust the final search range based on *z*-axis motion. Fig. 8 and Fig. 9 exemplify the




(a)




(b)

Fig. 12. Performance of FS+DMRMap+Scaling over FS+DMRMap (from frame 216 to frame 249) in "Poznan_Street". (a) Search compleixty in term of amplitude of search dimensions. (b) Resultant PSNR.

inefficiency in FS+DMRMap. In these figures, $mv_x^{max}(\hat{d})$ and $mv_y^{max}(\hat{d})$ represent the largest motion vectors in the $x$- and $y$-directions at the quantized depth value $\hat{d}$, respectively, obtained by FS and FS+DMRMap. Fig. 8 displays $mv_x^{max}(\hat{d})$ at $\hat{d}$=7 in "Lovebird1" from frame 1 to frame 80. In most of the time, $mv_x^{max}(\hat{d})$ of FS+DMRMap is the same as that of FS. However, there are some discrepancies in a number of frames. It can be observed that $mv_x^{max}(\hat{d})$ of FS+DMRMap cannot follow the increase in $mv_x^{max}(\hat{d})$ of FS. This is because adaptive search range (ASR) decision of FS+DMRMap makes use of depth/motion relationship in the reference frame. It implies that the ASR at particular $\hat{d}$ of the current block cannot be larger than $mv_x^{max}(\hat{d})$ of the reference frame.

The ASR decision based on the DMRMap of the reference frame is also plotted in Fig. 8 (the blue curve marked with circle dots). It is clearly shown that the resultant ASR is non increasing along frames due to the use of the DMRMap in the reference frame. This situation is more obvious in Fig. 9 where $mv_y^{max}(\hat{d})$ at $\hat{d}$=14 in "Newspaper" is shown. For instance, starting from frame 20 in "Newspaper", there is motion of an object along the $z$-axis, which results in reducing the temporal correlation of DMRMaps between the current and reference frames, as discussed in Section IV. As a consequence, FS+DMRMap may not catch the actual motions for the moving object, and may lead to RD deterioration. By contrast, FS+DMRMap+Scaling utilizes $\rho(B_t^i)$ to provide

(a) (b) (c)

Fig. 13. Sample texture frames of "Poznan_Street". (a) Frame 216. (b) Frame 232. (c) Frame 248.

(a) (b) (c)

Fig. 14. Sample depth frames of "Poznan_Street". (a) Frame 216. (b) Frame 232. (c) Frame 248.

additional flexibility in ASR decision for fitting the maximum texture motion. Fig. 10 and Fig. 11 illustrate how $\rho(B_t^i)$ can contribute the prediction accuracy of ASR. From these figures, it can be seen that FS+DMRMap+Scaling is able to catch up with $mv_x^{max}(\hat{d})$ and $mv_y^{max}(\hat{d})$ of FS along frames. It is due to the reason that the derivation of $\rho(B_t^i)$ from (13) complies with the influence of 3D-to-2D projection such that the ASR can be enlarged or diminished accordingly. In other words, $\rho(B_t^i)$ allows ASR to rebound to a larger value, as depicted in Fig. 10 and Fig. 11.

The advantage shown in Fig. 10 and Fig. 11 of FS+DMRMap+Scaling cannot be fully depicted in the results of Table II as the phenomenon in Fig. 8 and Fig. 9 only happens in a very short period of most sequences. However, both of the BD-rate and BD-PSNR in Table II measure the whole sequence in which the gain of FS+DMRMap+Scaling as compared with FS+DMRMap might be averaged out. To demonstrate this benefit of FS+DMRMap+Scaling, Fig. 12 further shows the performance of FS+DMRMap+Scaling over FS+DMRMap in very short time period. Fig. 12(a) shows the variation of the average search complexity depending on the search range size for all $\hat{d}$ with respect to the frame number from 216 to 249 in "Poznan_Street". During this period, "Poznan_Street" contains a car moving forward along the $z$-axis and a man walking away to the background, as shown in Fig. 13. Fig. 14 also shows the corresponding depth maps that exhibit remarkable changes in the moving object. Those changes in depth maps can be detected by FS+DMRMap+Scaling. In contrast to FS+DMRMap, the search range obtained by FS+DMRMap+Scaling is then enlarged or diminished accordingly, as shown in Fig. 12(a). This mechanism allows FS+DMRMap+Scaling to successfully provide a more adaptive search range for ME. Furthermore, PSNR results for coding "Poznan_Street" are plotted against the same series of frames in Fig. 12(b). From the results, FS+DMRMap+Scaling achieves a better quality of coded frames over FS+DMRMap. The observed PSNR gains by FS+DMRMap+Scaling verify that the proposed scaling scheme can provide a proper adjustment of the search range. In conclusion, FS+DMRMap+Scaling can balance the search range prediction accuracy, the complexity of ME, and the

TABLE IV
BJONTEGAARD (BD) MEASUREMENT AND CODING TIME CHANGE OF TZS+LSMF, TZS+MLELD, TZS+LAMASR AND TZS+DMRMAP+SCALING FOR ASR AGAINST TZS IN HEVC

| Seq. | QP | TZS | | TZS+LSMF [43] | | | TZS+MLELD [40] | | | TZS+LAMASR [44] | | | TZS+DMRMap+Scaling | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | PSNR (dB) | Bitrate (kbps) | Δtime (%) | BD-PSNR (ΔdB) | BD-rate (%) | Δtime (%) | BD-PSNR (ΔdB) | BD-rate (%) | Δtime (%) | BD-PSNR (ΔdB) | BD-rate (%) | Δtime (%) | BD-PSNR (ΔdB) | BD-rate (%) |
| ***720p*** | | | | | | | | | | | | | | | |
| Balloons | 37 | 38.36 | 337.13 | -12.07 | -0.01 | +0.11 | -15.48 | -0.01 | +0.14 | -14.24 | 0.00 | +0.02 | -48.92 | 0.00 | -0.01 |
| | 32 | 41.24 | 594.51 | | | | | | | | | | | | |
| | 27 | 43.56 | 1171.83 | | | | | | | | | | | | |
| | 22 | 45.46 | 3131.69 | | | | | | | | | | | | |
| Kendo | 37 | 39.64 | 372.91 | -16.90 | 0.00 | +0.03 | -17.46 | 0.00 | +0.12 | -25.92 | -0.01 | +0.30 | -50.14 | 0.00 | +0.13 |
| | 32 | 42.27 | 655.30 | | | | | | | | | | | | |
| | 27 | 44.44 | 1238.99 | | | | | | | | | | | | |
| | 22 | 46.29 | 2901.65 | | | | | | | | | | | | |
| Lovebird1 | 37 | 34.32 | 164.99 | -19.75 | -0.01 | +0.14 | -29.30 | -0.01 | +0.26 | -37.90 | 0.00 | +0.02 | -63.06 | 0.00 | +0.10 |
| | 32 | 37.22 | 353.52 | | | | | | | | | | | | |
| | 27 | 40.30 | 831.54 | | | | | | | | | | | | |
| | 22 | 43.64 | 2073.03 | | | | | | | | | | | | |
| Newspaper | 37 | 35.71 | 242.02 | -21.80 | -0.02 | +0.34 | -26.98 | -0.02 | +0.45 | -22.34 | 0.00 | +0.13 | -54.50 | -0.01 | +0.24 |
| | 32 | 38.44 | 452.42 | | | | | | | | | | | | |
| | 27 | 41.09 | 960.82 | | | | | | | | | | | | |
| | 22 | 43.80 | 2710.81 | | | | | | | | | | | | |
| ***1080p*** | | | | | | | | | | | | | | | |
| Poznan_Street | 37 | 35.09 | 462.86 | -14.89 | -0.01 | +0.16 | -22.98 | -0.01 | +0.30 | -21.36 | -0.01 | +0.64 | -56.63 | -0.01 | +0.23 |
| | 32 | 37.47 | 1144.32 | | | | | | | | | | | | |
| | 27 | 39.94 | 4223.44 | | | | | | | | | | | | |
| | 22 | 43.24 | 24864.72 | | | | | | | | | | | | |
| Poznan_Hall2 | 37 | 39.56 | 238.55 | -28.10 | 0.00 | +0.22 | -39.54 | 0.00 | +0.04 | -31.37 | 0.00 | 0.00 | -54.25 | 0.00 | +0.14 |
| | 32 | 41.07 | 502.38 | | | | | | | | | | | | |
| | 27 | 42.24 | 1562.84 | | | | | | | | | | | | |
| | 22 | 44.24 | 13723.17 | | | | | | | | | | | | |
| Undo_Dancer | 37 | 32.99 | 1532.64 | -5.70 | -0.01 | +0.17 | -12.82 | -0.01 | +0.15 | -5.98 | -0.02 | +0.58 | -55.56 | -0.01 | +0.19 |
| | 32 | 35.75 | 4327.77 | | | | | | | | | | | | |
| | 27 | 39.01 | 11262.33 | | | | | | | | | | | | |
| | 22 | 42.68 | 25270.98 | | | | | | | | | | | | |
| GT_Fly | 37 | 35.78 | 1169.20 | -26.96 | 0.00 | +0.02 | -20.29 | 0.00 | +0.01 | -30.14 | 0.00 | +0.09 | -42.61 | 0.00 | +0.12 |
| | 32 | 38.41 | 3040.38 | | | | | | | | | | | | |
| | 27 | 41.08 | 7481.00 | | | | | | | | | | | | |
| | 22 | 43.84 | 16976.31 | | | | | | | | | | | | |
| **Average:** | | | | **-18.27** | **-0.01** | **+0.15** | **-23.11** | **-0.01** | **+0.18** | **-23.66** | **-0.01** | **+0.22** | **-53.21** | **0.00** | **+0.14** |

PSNR performance with the help of the scale factor $\rho(B_t^i)$.

### C. Results of Applying DMRMap to Fast TZS

Section VI-A and Section VI-B demonstrate that the proposed DMRMap-based ASR algorithm with the scaling factor is successful in FS for complexity reduction. It is worth noting that our proposed algorithm for search range determination is not only applied to FS, it is also compatible with other fast search algorithms in HEVC. To validate this, our proposed search range determination has been also used by the fast TZS in HEVC, named as TZS+DMRMap+Scaling. TZS only searches points on the vertexes of the blocks and the diamond patterns with various sizes inside the fixed search range. Instead of using fixed search range in TZS, an ASR is determined by TZS+DMRMap+Scaling. It is noted that the search strategy of TZS is only suited for a squared search window. However, the DMRMap-based ASR algorithm can handle the horizontal and vertical search ranges separately. For the sake of simplicity, based on (5) and (6), the search range of the squared search window is then computed by $\max(SR_x^\rho(B_t^i), SR_y^\rho(B_t^i))$.

Table IV shows the BD measurement and the coding time change of the proposed TZS+DMRMap+Scaling compared to TZS. As far as TZS+DMRMap+Scaling concerned, 42.61% to 63.06% coding time can be saved in various sequences. Meanwhile, the coding efficiency almost has no loss in terms of BD-PSNR and BD-rate (0.14% increment). The above result indicates that the proposed DMRMap-based ASR

algorithm is well compatible with the fast search strategy in HEVC and provides up to around 53.21% time saving on average with only negligible loss in BD measurements. Besides, Table IV also shows the results when LSMF in [43], MLELD in [40], and LAMASR in [44] are used in TZS for ASR determination as denoted by TZS+LSMF [43], TZS+MLELD [40], and TZS+LAMASR [44], respectively. On average, TZS+DMRMap+Scaling attains a better BD performance comprising 0.01dB BD-PSNR gain and a range from 0.01% to 0.08% BD-rate decrement compared to others. As shown in Table IV, TZS+DMRMap+Scaling can reduce more coding time from 29% to 34%. It shows that an accurate ASR determination by TZS+DMRMap+Scaling is very crucial in the fast search ME process. The above experimental results demonstrate the proposed ASR scheme based on the temporal correlation of depth/motion relationship maps and the 3D-to-2D projection can figure out a more precise range for ME. As a result, motion vectors are obtained quickly.

### *D. Influence of Q on DMRMap Accuracy*

In the following, we discuss the influence of the quantization factor $Q$ on the performance of the proposed DMRMap-based algorithm. As mentioned in Section III-A, $Q$ is used to absorb depth variation in DMRMap construction due to the noise of a depth map. Fig. 15 illustrates two DMRMaps using different $Q$. For the example in Fig. 15(b) where $Q$ = 16, group $A$ is exactly equivalent to group $A_1$ and group $A_2$ ($Q$ = 8) in Fig. 15(a) with the same largest motion vector. It implies that the DMRMaps using $Q$ = 8 and $Q$ = 16 will not affect the accuracy of the mapping processing. In contrast, group $B_1$ and group $B_2$ in Fig. 15(a) of $Q$ = 8 associate with different largest motion vectors while they are combined to group $B$ in Fig. 15(b) of $Q$ = 16. It means that a large search range is required for large $Q$, but has a chance to achieve better BD-rate in this scenario. It is also the tradeoff between the computational complexity and BD performance of the proposed FS+DMRMap+Scaling. The evidence can be seen in Table V where the performances in terms of the BD measurement and the coding time change for various $Q$ are shown. As expected, the complexity reduction increases as $Q$ decreases for all sequences. Nevertheless, it only shows little variation for nearly all sequences, except "Balloons" and "Undo_Dancer".

It is interesting to note that the depth map of "Undo_Dancer", as shown in Fig. 16(a), is different from most of other sequences. Its depth map is computer generated sequence using 3D models and its depth map is ground truth without noise. Besides, the dancer contains diverse motion activities in different parts of his body, as shown in Fig. 16(b) to Fig. 16(e). However, these different parts of his body have very close depth values. The quantization process in the construction of DMRMap might merge parts with different motions of the hand, leg and head into one group if $Q$ is large, which leads to the increase in the computational complexity of "Undo_Dancer", as shown in Table V. It also happens in "Balloons" where the balloons in the foreground have similar depth values, but diverse motions. In conclusion, for sequences having noiseless depth maps and complicated motion with similar depth value, it is beneficent to adopt small $Q$ for the DMRMap construction.

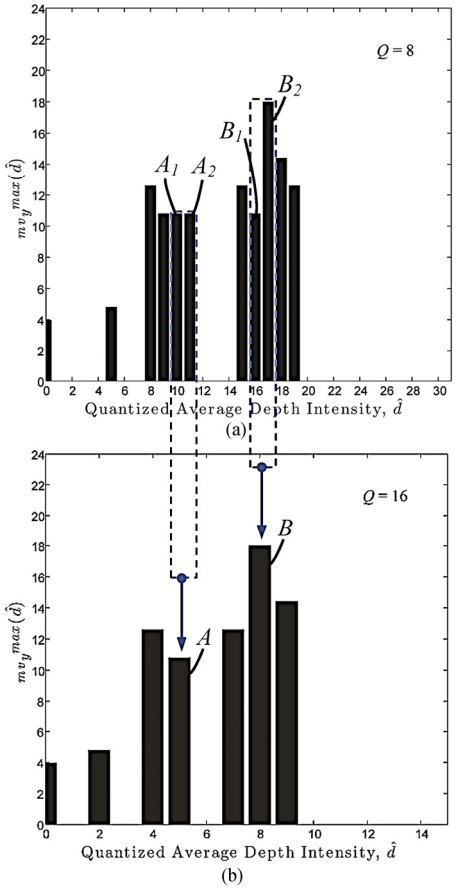



Fig. 15. DMRMaps with various $Q$, (a) $Q$ = 8, and (b) $Q$ = 16, in "Undo_Dancer".

## VII. Conclusion and Future Work

In this paper, we have proposed an efficient ASR algorithm for HEVC to reduce the computational complexity of ME by exploiting the temporal correlation between the depth map and motion in texture. The new depth/motion relationship map (DMRMap) is then established, and is interpreted to control the ASR for each block. DMRMap builds the linkage on the same object among consecutive frames which reflects the probable range of movements for the object. Based on this, a depth intensity mapping is contrived to form an asymmetric search range for ME. It results in reducing unnecessary search points in ME. Furthermore, the impact of the depth intensity variations of the block in 3D-to-2D projection on ASR has been analyzed. By taking this into account, a scale factor has been proposed to comply with the impact of 3D-to-2D projection. The proposed DMRMap could be jointly worked with FS and other fast search algorithms such as TZS in HEVC for complexity reduction. Simulation results demonstrated that the proposed DMRMap-based ASR algorithm is able to reduce up to 53% of average coding time among various sequences in fast ME algorithms. In the meantime, the coding efficiency can be maintained compared to FS and TZS in terms of the BD measurement.

TABLE V
BJONTEGAARD (BD) MEASUREMENT AND CODING TIME CHANGE OF FS+DMRMAP+SCALING WITH VARIOUS *Q* AGAINST FS IN HEVC

| Sequences | FS+DMRMap+Scaling | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | ***Q* = 4** | | | ***Q* = 8** | | | ***Q* = 16** | | |
| | Δtime (%) | BD-PSNR (ΔdB) | BD-rate (%) | Δtime (%) | BD-PSNR (ΔdB) | BD-rate (%) | Δtime (%) | BD-PSNR (ΔdB) | BD-rate (%) |
| ***720p*** | | | | | | | | | |
| **Balloons** | -93.43 | -0.01 | +0.34 | -89.63 | -0.01 | +0.26 | -88.57 | -0.01 | +0.14 |
| **Kendo** | -95.09 | -0.01 | +0.34 | -94.01 | -0.01 | +0.26 | -92.71 | -0.01 | +0.16 |
| **Lovebird1** | -99.51 | 0.00 | 0.00 | -99.46 | 0.00 | +0.08 | -99.42 | 0.00 | +0.01 |
| **Newspaper** | -97.85 | -0.01 | +0.38 | -97.23 | -0.01 | +0.31 | -95.82 | -0.01 | +0.32 |
| ***1080p*** | | | | | | | | | |
| **Poznan_Street** | -98.07 | -0.01 | +0.25 | -97.15 | 0.00 | +0.17 | -96.08 | 0.00 | +0.17 |
| **Poznan_Hall2** | -87.42 | -0.01 | +0.44 | -86.72 | -0.01 | +0.41 | -85.32 | 0.00 | +0.24 |
| **Undo_Dancer** | -95.04 | -0.02 | +0.42 | -91.74 | -0.01 | +0.30 | -87.75 | -0.01 | +0.27 |
| **GT_Fly** | -97.19 | -0.01 | +0.34 | -95.47 | -0.01 | +0.27 | -93.07 | -0.01 | +0.19 |
| **Average:** | **-95.45** | **-0.01** | **+0.31** | **-93.93** | **-0.01** | **+0.26** | **-92.34** | **-0.01** | **+0.19** |

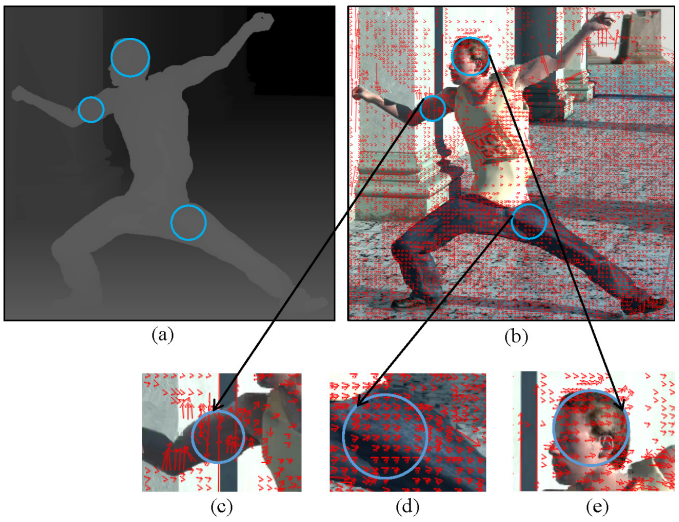



Fig. 16. (a) Depth map, and (b) the corresponding texture of "Undo_Dancer". Magnified regions with similar depth intensity values in parts of (c) hand, (d) leg, and (e) head with different amplitudes of motion vectors.

In the future, some effective and efficient ways to determine the quantization factor in DMRMap construction will be studied. As mentioned in Section VI-D, it depends on the noise level of depth maps, the amount of complicated motion activities in a same object, the diversity of motion activities in different objects with similar depth values, etc. A challenging research topic is to generalize our DMRMap under different kinds of depth maps by a sequence-dependent quantization factor. This could be a point for our immediate future work.

## REFERENCES

[1] K. Ugur, K. Andersson, A. Fuldseth, G. Bjontegaard, L. P. Endresen, J. Lainema, A. Hallapuro, J. Ridge, D. Rusanovskyy, C. Zhang, A. Norkin, C. Priddle, T. Rusert, J. Samuelsson, R. Sjoberg, and Z. Wu, "High performance, low complexity video coding and the emerging HEVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 12, pp. 1688–1697, Dec. 2010.

[2] J. R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the coding efficiency of video coding standards - including high efficiency video coding (HEVC)," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1669–1684, Dec. 2012.

[3] J. Vanne, M. Viitanen, T. D. Hämäläinen, 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.

[4] G. J. Sullivan, J. M. Boyce, Y. Chen, J. R. Ohm, C. A. Segall, and A. Vetro, "Standardized extensions of high efficiency video coding (HEVC)," *IEEE J. Sel. Top. Sign. Proces.*, vol. 7, no. 6, pp. 1001–1016, Dec. 2013.

[5] G. J. Sullivan, J. R. Ohm, W. J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[6] W. Han, J. Min, I. Kim, E. Alshina, A. Alshin, T. Lee, J. Chen, V. Seregin, S. Lee, Y. Homg, M. Cheon, N. Shlyakhov, K. McCann, T. Davies, and J. Park, "Improved video compression efficiency through flexible unit representation and corresponding extension of coding tools," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 12, pp. 1709–1720, Dec. 2010.

[7] G. Corrêa, P. Assunção, L. Agostini, and L. A. da Silva Cruz, M, "Performance and computational complexity assessment of high-efficiency video encoders," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1899–1909, Dec. 2012.

[8] C. E. Rhee, K. Lee, T. S. Kim, and H. J. Lee, "A survey of fast mode decision algorithms for inter-prediction and their applications to high efficiency video coding," *IEEE Trans. Consum. Electron.*, vol. 58, no. 4, pp. 1375–1383, Nov. 2012.

[9] M. Karczewicz, P. Chen, R. Joshi, X. Wang, W. J. Chien, R. Panchal, Y. Reznik, M. Coban, and I. S. Chong, "A hybrid video coder based on extended block sizes, improved interpolation and flexible motion representation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 12, pp. 1698–1708, Dec. 2010.

[10] J. L. Lin, Y. W. Chen, Y. W. Huang, and S. M. Lei, "Motion vector coding in the HEVC standard," *IEEE J. Sel. Top. Sign. Proces.*, vol. 7, no. 6, pp. 957–968, Dec. 2013.

[11] D. Marpe, H. Schwarz, S. Bosse, B. Bross, P. Helle, T. Hinz, H. Kirchhoffer, H. Lakshman, T. Nguyen, S. Oudin, M. Siekmann, K. Suhring, M. Winken, and T. Wiegand, "Video compression using nested quadtree structure, leaf merging, and improved techniques for motion representation and entropy coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 12, pp. 1676–1687, Dec. 2010.

[12] J. Vanne, M. Viitanen, and T. D. Hämäläinen, "Efficient mode decision schemes for HEVC inter prediction," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 9, pp. 1579–1593, Sept. 2014.

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

[14] I. K. Kim, J. Min, T. Lee, W. J. Han, and J. Park, "Block partitioning structure in the HEVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1679–1706, Dec. 2012.

[15] R. H. Gweon and Y. L. Lee, *Early termination of CU Encoding to Reduce HEVC Complexity, Joint Collaborative Team on Video Coding*

*(JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Document JCTVC-F045,* Torin, Italy, Jul. 2011.

[16] K. Lee, H. Lee, J. Kim, and Y. Choi, "A novel algorithm for zero block detection in high efficiency video coding," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 6, pp. 1124-1134, Dec. 2013.

[17] J. Kim, J. Yang, K. Won, and B. Jeon, "Early determination of mode decision for HEVC," in *Proc. Picture Coding Symp. (PCS 2012)*, May 2012, pp. 449–452.

[18] K. Choi, S. H. Park, and E. S. Jang, *Coding Tree Pruning Based CU Early Termination, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Document JCTVC-F092,* Turin, Italy, Jul. 2011.

[19] L. Shen, Z. Liu, X. Zhang, W. Zhao, and Z. Zhang, "An effective CU size decision method for HEVC encoders," *IEEE Trans. Multimedia,* vol. 15, no. 2, pp. 465-470, Feb. 2013.

[20] J. H. Lee, C. S. Park, B. G. Kim, "Novel fast PU decision algorithm for the HEVC video standard," in *Proc. IEEE Int. Conf. on Image Process. (ICIP 2013)*, Sept. 2013, pp. 1982–1985.

[21] 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, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Document JCTVC-P1002*, San Jose, US, Mar. 2014.

[22] F. Bossen, D. Flynn, and K. Sühring, HM software manual, *Document JCTVC-Software Manual, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11*, Apr. 2014.

[23] N. Purnachand, L. N. Alves, and A. Navzrro, "Fast motion estimation algorithm for HEVC," in *Proc. IEEE Int. Conf. on Consumer Electronics (ICCE 2012)*, Sept. 2012, pp. 34-37.

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

[25] Y. L. Chan, K. C. Hui, and W. C. Siu, "Adaptive partial distortion search algorithm for block motion estimation," *J. Visual Commun. and Image Represent.*, vol. 15, no. 4, pp. 489– 506, Dec. 2004.

[26] K. C. Hui, W. C. Siu, and Y. L. Chan, "Fast motion estimation of arbitrarily shaped video objects in MPEG-4," *J. Signal Process: Image Commun.*, vol. 18, no. 1, pp. 33– 50, Jan. 2003.

[27] Y. L. Chan and W. C. Siu, "On block motion estimation using an novel search strategy for an improved adaptive pixel decimation," *J. Visual Commun. and Image Represent.*, vol. 9, no. 2, pp. 139– 159, Jun. 1998.

[28] 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. 893–898, Jun. 2009.

[29] S. H. Yang, J. Z. Jiang, and J. J. Yang, "Fast motion estimation for HEVC with directional search," *Electronics Letters*, vol. 50, no. 9, pp. 673-675, Apr. 2014.

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

[31] Y. L. Chan and W. C. Siu, "New adaptive pixel decimation for block motion vector estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 1, pp. 113–118, Feb. 1996.

[32] Y. L. Chan, H. W. Lam, and W. C. Siu, "A block motion vector estimation using pattern based pixel decimation," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS 1997),* Jun. 1997, pp. 1153–1156.

[33] Y. L. Chan and W. C. Siu, "A new block motion vector estimation using adaptive pixel decimation," in *Proc. IEEE Int. Conf. on Acous. Speech and Signal Process. (ICASSP 1995),* May. 1995, pp. 2257–2260.

[34] C. C. Lou, S. W. Lee, C. C. J. Kuo, "Adaptive motion search range prediction for video encoding," *IEEE Trans. Circuits Syst. Video Technol.,* vol. 20, no. 12, pp. 1903–1908, Dec. 2010.

[35] Y. H. Ko, H. S. Kang, and S. W. Lee, "Adaptive search range motion estimation using neighboring motion vector difference," *IEEE Trans. Consumer Electronics,* vol. 57, no. 2, pp. 726-730, May 2011.

[36] S. Na, and C. M. Kyung, "Activity-based motion estimation scheme for H.264 scalable video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 11, pp. 1475–1485, Nov. 2010.

[37] Y. Liu, J. Wang, and F. Fu, "Adaptive search range adjustment and multiframe selection algorithm for motion estimation in H.264/AVC," *J. Electron. Imaging.*, vol. 22, no. 2, pp. 023031– 023031-8, Jun. 2013.

[38] S. Lee, "Fast motion estimation based on adaptive search range adjustment and matching error prediction," *IEEE Trans. Consumer Electronics,* vol. 55, no. 2, pp. 805-811, May 2009.

[39] Z. Chen, Q. Liu, T. Ikenaga, and S. Goto, "A motion vector difference based self-incremental adaptive search range algorithm for variable block size motion estimation," in *Proc. IEEE Int. Conf. on Image Process. (ICIP 2008)*, Oct. 2008, pp. 1988–1991.

[40] L. Jia, O. C. Au, C. Tsu, Y. Shi, R. Ma, and H. Zhang, "A diamond search window based adaptive search range algorithm," in *Proc. IEEE Int. Conf. Multimedia and Expo Workshops (ICMEW 2013)*, Jul. 2013. pp.1–4.

[41] T. K. Lee, Y. L. Chan, and W. C. Siu, "Adaptive search range by neighbouring depth intensity weighted sum for HEVC texture coding," *Electronics Letters*, vol. 52, no. 2, pp. 1018-1020, Jun. 2016.

[42] Z. P. Deng, Y. L. Chan, K. B. Jia, C. H. Fu, and W. C. Siu, "Fast motion and disparity estimation with adaptive search range adjustment in stereoscopic video coding," *IEEE Trans. Broadcast.*, vol. 58, no. 1, pp. 24-33, Mar. 2012.

[43] S. Kim, D. K. Lee, C. B. Sohn, and S. J. Oh, "Fast motion estimation for HEVC with adaptive search range decision on CPU and GPU," in *Proc. IEEE China Summit & Int. Conf. on Signal and Information Process. (ChinaSIP 2014)*, Jul. 2014, pp. 349–353.

[44] 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.

[45] W. D. Chien, K. Y. Liao, and J. F. Yang, "Enhanced AMVP mechanism based adaptive motion search range decision algorithm for fast HEVC coding," in *Proc. IEEE Int. Conf. on Image Process. (ICIP 2014)*, Oct. 2014, pp. 3696–3699.

[46] H. R. Tohidypour, M. T. Pourazad, P. Nasiopoulos, and V. Leung, "A content adaptive complexity reduction scheme for HEVC-based 3D video coding," in *Proc. 2013 18th International Conference on Digital Signal Processing (DSP 2013),* Jul. 2013, pp. 1–5.

[47] Y. H. Lin, and J. Ling, "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, Jun. 2011.

[48] K. Müller, P. Merkle, and T. Wiegand, "3-D video representation using depth maps," *Proceedings of the IEEE*, vol. 99, no. 4, pp. 643-656, Apr. 2011.

[49] L. Zhang and W. Tam, "Stereoscopic image generation based on depth image for 3D TV," *IEEE Trans. Broadcast.*, vol. 51, no. 2, pp. 191-199, Jun. 2005.

[50] 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. on Digital Signal Process. (DSP 2014)*, Aug. 2014, pp. 919–923.

[51] 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.-Dec. 2014.

[52] JCT-VC, 2014. HM14.0. HEVC software repository (main at HHI) [online].Available:https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/branches/HM-14.0-dev/.

[53] F. Bossen, Common HM test conditions and software reference configuration, *JCTVC-L1100*, Jan. 2013.

[54] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," *VCEG-M33*, Mar. 2001.

[55] Q. Zhang, M. Chen, X. Huang, N. Li, and Y. Gan, "Low-complexity depth map compression in HEVC-based 3D video coding," *EURASIP J. Image Video Process.*, vol. 2015, no. 2, pp1-14.

**Tsz-Kwan Lee** (S'11) received the BSc (Honors) degree and the M.Phil. degree from Department of Electronic and Information Engineering of The Hong Kong Polytechnic University in 2008 and 2011, respectively. She is currently pursuing the Ph.D. degree at the Center for Signal Processing under the supervision of Dr. Y. L. Chan and Prof. W. C. Siu in the same Department and University. During her Ph.D. studies, she was awarded Li Po Chun Charitable Trust Fund Scholarship in 2013. Her research interests include signal processing, video compression, 3D and multi-view video coding, and future video coding standards.

**Yui-Lam Chan** (S'94–A'97–M'00) received the BEng with a First Class Honours degree and the Ph.D degree from the Hong Kong Polytechnic University in 1993 and 1997, respectively. Dr Chan joined the Hong Kong Polytechnic University in 1997, and is now an Associate Professor in the Department of Electronic and Information Engineering. Dr Chan is actively involved in professional activities. In particular, Dr Chan serves as an Associate Editor of IEEE Transactions on Image Processing. He was the Secretary of the 2010 IEEE International Conference on Image Processing (ICIP2010). He was also the Special Sessions Co-Chairs and the Publicity Co-Chairs of 2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC 2015), the Technical Program Co-Chairs of 2014 International Conference on Digital Signal Processing (DSP2014), etc. Besides, Dr Chan has published over 100 research papers in various international journals and conferences. His research and technical interests include multimedia technologies, signal processing, image and video compression, video streaming, video transcoding, video conferencing, digital TV/HDTV, 3DTV/3DV, multi-view video coding, future video coding standards, error-resilient coding, and digital VCR.

**Wan-Chi Siu** (S'77-M'77-SM'90-F'12) received the PhD degree from Imperial College, London, in 1984, and is Life-Fellow of the IEEE. He joined Hong Kong Polytechnic University as a Lecturer in 1980 and has been Chair Professor since 1992. He was Head of Department (EIE) and subsequently Dean of Engineering Faculty between 1994 and 2002. Professor Siu is an expert in digital signal processing, fast algorithms, video coding, 3D videos, pattern recognition and visual surveillance. He has published over 500 research papers (over 200 of which appeared in international journals), and has 8 recent patents. Prof. Siu was an independent non-executive director of a listed video surveillance company in Hong Kong for over 15 years. His works are well received by peers with high citations, and have been ported into hi-tech industrial uses.

Prof. Siu was a Vice President of the IEEE Signal Processing (SP) Society, Chairman of Conference Board and a core member of the Board of Governors (2012–2014). Recently, he has also been elected as the President-Elect (2015-2016) of the Asia-Pacific Signal and Information Processing Association (APSIPA). Professor Siu is now Subject Editor (2015-17, in charge of Image Processing) of Electronics Letters, Associate Editor (2015-2017) of IEEE Transactions on Circuits & Systems for Video Technology, editorial board member of Journal of VLSI Signal Processing Systems for Signal, Image, Video Technology, in addition to other journals. He was also Guest Editor and Associate Editor of IEEE Transactions on Image Processing, IEEE Transactions on Circuits and Systems, Pattern Recognition, and the EURASIP Journal on Applied Signal Processing. He is a co-editor of the book, 'Multimedia Information Retrieval and Management', Springer Berlin Heidelberg 2003, and the book 'Trends in Digital Signal Processing: a Festschrift in Honour of Tony Constantinides', Pan Stanford Publishing, 2016. He is a very popular lecturing staff member within the University, while outside the University he has been a keynote speaker of over 12 international/national conferences in the recent 10 years. He received many awards, such as Distinguished Presenter Award, the Best Teacher Award, the Best Paper Award and IEEE Third Millennium Medal. He was a chief organizer of several prestigious IEEE Society sponsored flagship international conferences, including ICIP2010 (General Chair), ICASSP2003 (General Chair) and ISCAS1997 (TPC Chair). Home Page: http://www.eie.polyu.edu.hk/~wcsiu/mypage.htm