A SURVEY ON MOTION ESTIMATION TECHNIQUES

Jagrati Talreja and Dr. Manoj Kumar Gupta

Abstract

This article is about comparison between traditional and newer standards of motion estimation (ME) techniques and algorithms such as High Efficiency Video Coding (HEVC) and Versatile Video Coding (VVC). In this technique, we determine the motion vectors (MV), which provide modification from one image to alternative in many video sequences. Therefore, motion estimation (ME) process can be performed where motion is present. The image data in an image sequence that is approximately the same during motion. For understanding the image data laying-off in image sequences, the ME is required. The ME is applicable for many fields such as video enhancement, compression, stereo correspondence, object tracking and determination of optical flow. Many specific methods have been proposed related to one or more of these applications. To compare the similarity of two pixels or two patches of pixels ME algorithms either operate directly in the image domain or in finding the matching metrics of motion in that domain. This review article mainly focuses on various ME methods such as HEVC and VCC techniques.

Key words: Motion estimation, algorithms, versatile video coding, motion vector.

Introduction

Presently, video content continuously increased on the internet with technological advancement, globally. People can access video content from their handy devices due to technological advancements in 3G/4G mobile. The circulation of video data increased ~70% of total video content on the internet in the next few years, increasing the demands of newer techniques mainly compression techniques (video and image compression [1]. The moving picture expert group (MPEG), made a team in the collaboration of ITUT video coding expert group (VCEG) and ISO/IEC) on video coding (JCT-VC) in April 2010 that focus on the development of newer video encoding namely HEVC/H.265 and MPEG-H. This research group mainly focuses on the considerable developments in coding competence and reduction in the intricacy for transportable devices compared with that of H.264/AVC [2], predominantly for high resolution (HR) video sequences. The first version of HEVC was completed in 2013 that accomplished 50% gain in video coding over H.264/AVC [3]. Nonetheless, the relatively high computational cost for this coding remains a concern. Therefore, there is needed to be newer coding technology that combats issues associated with HEVC. In this context, VVC has potential to fulfil such issues associated with HEVC.

The VVC technology was established by the joint video exploration team (JVET) of ITU-T. The video coding was well optimized and enhanced (H.264/AVC and H.265/HEVC) by experts' group (VCEG) and ISO/IECMPEG. The HEVC is a consistent video compression standard but still there is a requirement of newer competent video coding system for HR and experts in JVET have been actively conducting much research for developing the newer techniques in video compression other than HEVC. The VVC introduces a testimonial software prototype namely VVC Test Model (VTM) [4], and VTM2 [5] was established at the 11th JVET meeting, with the addition of a group of HEVC coding elements with new coding features. The basic framework of VVC consists of partition of block, prediction (inter-and intra-prediction), transform, loop filter and entropy coding that are like the HEVC. Interpretation mainly focusses to achieve identical block in the reference frames that diminished the sequential redundancy, is an indispensable element in video coding that based on two important tools for inter prediction, (1) motion compensation (MC), (2) motion estimation (ME). These tools aided advantages to observed accurate correlation between successive frames for final coding performance. For higher accuracy, newer technologies including matching block based ME and MC have been executed in H.264/AVC and H.265/HEVC. This review articles summarize various ME technologies, drawbacks of existing technologies, and prospects for long-term solutions.

Motion Compensation (MC)

The ME is the process that estimate object velocity or transformation one object (image) to another object (image) onto a 2D-image plane. The motion vectors (MV) might be related to the entire image or component (rectangular and arbitrary blocks). The MV represents various models including translational, which might be estimated the motion of video (rotation, translation and zoom in 3D). The MV applied to image for the prediction of transforming images in moving camera, the entire process named as motion compensation (MC). The combination of these two processes (ME and MC) is one of the main parts of compression that is effectively used in MPEG-1, MPEG-2, and MPEG-4 etc.

Motion Estimation (ME)

Motion Estimation process is extensively owned in numerous applications such as compression (images and video), medical imaging, computer vision, and video processing. Motion is an imperative resource of information in video processing. Motion occurs because of various factors like camera and apparent motion. ME process aided an advantage to recuperate all information by analyzing image. This information is basic for understating the video and tracking the object. Moreover, effective and precise estimation of motion is necessary in all domains. Therefore, with the help of ME process various domains including computer vision, and sequence analysis (image) might be accurate and efficient model.

Recently, ME process has been widely used in various applications mainly robotics, navigation in vehicles, video surveillance, analysis of motion (human), and restoration of video. Additionally, accurate motion is required for operations like frames per second conversion in video analytics. However, relatively higher cost, and slow process remains a concern. Therefore, there is need to be fast, inexpensive algorithm for ME process.

Motion Estimation in HEVC

To predict motion between video frames, HEVC employs many motion estimation techniques. The following are the most often used motion estimation techniques in HEVC:

Integer Pel Motion Estimation (IME): This is the fundamental motion estimation technique in which motion vectors are only searched at integer pixel positions, resulting in full-pixel accuracy.

Fractional Pel Motion Estimation (FME): This technique refines motion vectors by looking at fractional pixel positions, resulting in sub-pixel precision. It increases video quality by reducing motion compensating for errors.

Adaptive Motion Vector Resolution (AMVR): HEVC can choose the resolution of the motion vectors (full-pixel or sub-pixel) based on the motion characteristics of the video block.

Advanced Motion Vector Prediction (AMVP): AMVP reduces the number of bits necessary to transmit motion vectors by leveraging temporal correlations and predicting motion vectors that have already been encoded.

Merge Mode: This technique allows for the combination of motion information from surrounding blocks in order to increase forecast accuracy and reduce the amount of motion vectors.

Motion Estimation in VVC:

VVC, the successor to HEVC, expands on the motion estimation techniques used in HEVC while also introducing certain improvements for improved compression efficiency. VVC employs the following motion estimating techniques:

Fractional Pel Motion estimate with Extended Search Ranges: To improve motion estimate accuracy, VVC employs a wider range of fractional pixel search positions.

Hierarchical Search: VVC offers a hierarchical search structure to improve motion estimation efficiency by performing coarse search first and then refining the search at finer levels.

Bi prediction: By allowing vector motion from both prior and upcoming frames, VVC enhances bidirectional prediction, leading to more precise predictions.

Adaptive Motion Vector Resolution (AMVR) with Extended Precision: VVC, like HEVC, employs AMVR, but with increased precision for sub-pixel motion vectors, allowing for more accurate motion correction.

Enhanced Motion Vector Prediction (EMVP): VVC uses more advanced algorithms to improve the efficiency of motion vector prediction.

High Efficiency Video Coding (HEVC)

High Efficiency Video Coding (HEVC) is a newly developed video coding that enhances compression ability. HEVC is also referred to as H.265, and MPEGH Part 2, developed as a replacement to the broadly utilized Advanced Video Coding (AVC, H.264, and MPEG-4 Part 10). HEVC provides approximately 50% more compression with a similar bit rate and quality of video compared with that of the AVC. The AVC (8-bit) has higher reliability and easily integrated into all supporting hardware with up to 8192 × 4320 resolution. HEVC easily rivals free AV1-coding for consistency by the group Net-VC of the Internet Engineering Task Force (IETF). HEVC uses integer DCT and DST changes using different size of blocks ranges (4 × 4 to 32 × 32), whereas The Integer Discrete Cosine Transform (DCT) is used by AVC with a variety of block sizes (4 × 4 to 8 × 8). Based on HEVC, another format called High Efficiency Image Format (HEIF) also exists. The next most prevalent video coding method after AVC is HECV, which is effectively employed by about 43% of video creators.

HEVC Algorithm *TZS Algorithm*

Video coding requires very precise techniques like motion estimation to reduce temporal dismissal. The HEVC test model consists of two different ME algorithms: First is Full Search (FS) which deals with the eminence of predicted picture however when it comes to stimulation it deals with various complications. Second, ME algorithms are faster than FS. In this article TZS has been discussed. ME algorithm of Most of the reference application utilizes Test Zone search (TZS) [6] in demand of speed and excellence when comparing with first one. TZS utilizes 3 different search patterns viz., (1) diamond (2) square and (3) raster. While its explanation can be given by five methods, (1) Prediction of Motion Vector, (2) Initial Grid Search, (3) Raster Search, (4) Star Refinement, and (5) Two Point Search.

Prediction of Motion Vector

Prediction of motion vector in TZS algorithm utilizes midpoint predictor in three directions like left, up and right. This motion initialized at zero distance from predictors for further consecutive steps.

Initial Grid Search

Grid search in TZS utilizes diamond or square pattern and initialized at distance from starting point equals to 1 to 64. In this search it is mandatory to fix the distance double of initial step. Usually, default diamond search is operated which depends on the distance it stops either three steps or almost after finding excellent similar candidate or it will reach at the edge of the search window. After this the destination of search depends upon the total distance of the best matched if it is found in this step or need to follow further consecutive steps. Which includes two-point search, namely (1) raster, and (2) refinement. To eradicate sequential redundancy of mage sequences ME is a very indispensable procedure in video coding. The HEVC test model offers mainly two kinds of ME algorithms: (1) Full Search (FS) and (2) Test Zone Search (TZS). FS is superior based on quality of predicted image, but its computational complication is high whereas TZS is faster than FS. ME algorithm of TZS is used in the reference software [6] that provides superior quality and speed as compared with the FS-ME. TZS-ME algorithm used mainly three types of searches shapes: (1) diamond, (2) square and (3) raster shape.

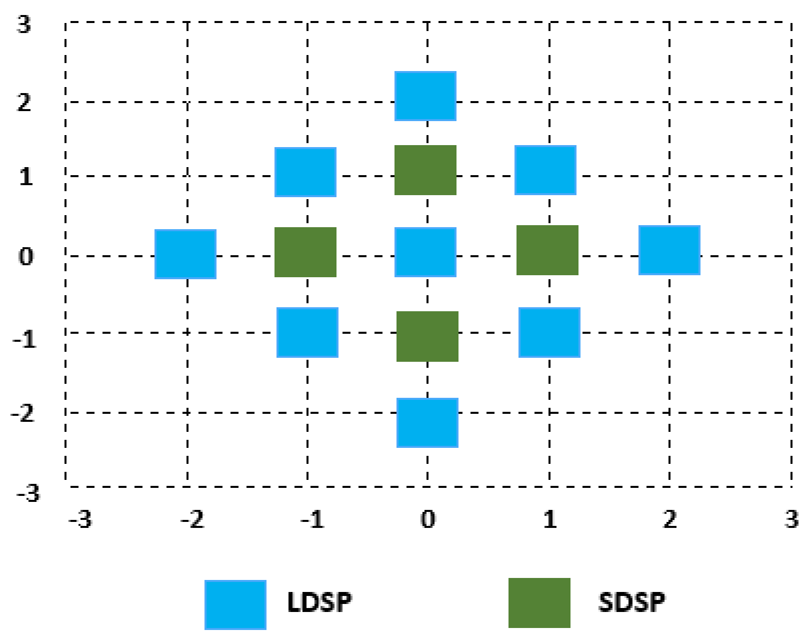

Fig. 1. Diamond Search

Raster Search

Raster search is generally operated when the usual space of the superlative match values is more than the iRaster values. Usually, it deals with FS on the down-sample version of search window. A predefined number is used which is a sampling factors set before the finalization of code for the search window in iRaster during raster scan.

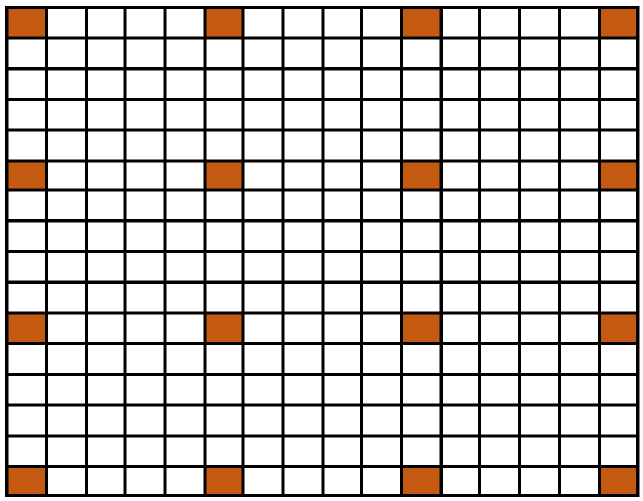

Fig. 2. Raster Search with iRaster equal to 1.

Star Refinement

This type of search utilizes diamond search that is used in initial grid shape. It is recognized as the final grid search as well. It's a 2-shape search specialty initial grid shape. The most common aspect of this search is that can only utilize distance greater than one. Unlike initial grid search, it normally finishes at either two-step after suitable matched values achieve the search window boundary.

Two Point Search

Two-point search applies when excellent similar candidate distance in first step is ~1. At point search will only be allowed to operate when initial results of this step are the ending finest matched values in the search window. Low power operated devices do not usually come under the category of TZ search ME algorithm as this search utilizes different types of shapes and high frequency for accessing the memory capability. Several literatures report the modification such as Prachanda *et al.* reports the modification in shape of TZS ME algorithm from diamond to hexagonal rotating shape [9]. In their modification they utilize threshold for early termination. Unfortunately, due to the pattern following limitation of rotating hexagonal shape this modification has not received any attention. Another report by Yang et al, [7] demonstrates the directional search ME algorithm. In their modification, the author utilizes 2 different search paths viz cross search and square search having distance range of four and one respectively for various search pattern in ME algorithm. Their search model has an advantage over the available ME algorithm is that it saves time of 60% average in integral pictures with almost negligible distortion. The shortcoming of this model was that this modification was not applicable for trapping the movement (directional) in video sequence.

Diamond grid search (DGS) Algorithm

DGS algorithm utilizes two different steps (1) Motion vector prediction and (2) Initial grid (IG) search. These initial steps have resemblance with TZS algorithm this algorithm utilizes uses two different shapes: (1) Small diamond (SD)- this shape require five different search points and (2) Large diamond (LD), this LD search require nine different search points.

Motion Vector Prediction

Motion vector prediction is basically the first step of DGS algorithm. The method helps in predicting the origin distortion and position and helps to select the best starting point.

Initial Grid Search

Both search (TZS and DGS) is almost similar, and the only difference exists between both is that it utilizes diamond search pattern. It will select based on the distance from the origin whether it will be a point search or star refinement search for finding the best possible value.

Star Refinement

Star refinement search is possibly operated depending on the difference of best probable values in previous step is >1.

Like TZS, it also utilizes diamond search pattern. It has an extended limit of distance up to four which is different from TZS. Next step will be followed by or not will depend on possibility of finding best possible values.

Two Point Search

Two-point search is only possible to execute when the distance of best values is almost ~1. As mentioned in TZ search, this search is possible to give the best final values which tend to match with the search window.

A basic design of Test Zone Search Algorithm is given in the above figure 3 which illustrates the flow of motion vector throughout the algorithm for motion estimation. Initially all the motion vectors are fed to the prediction unit which gives the best x and y coordinates and distance between the motion vectors which is either minimum or zero. In the next step it passes through the Initial Grid Search pattern. In the third step it goes from raster Search and finally the best result is obtained from the final step which is Star Refinement.

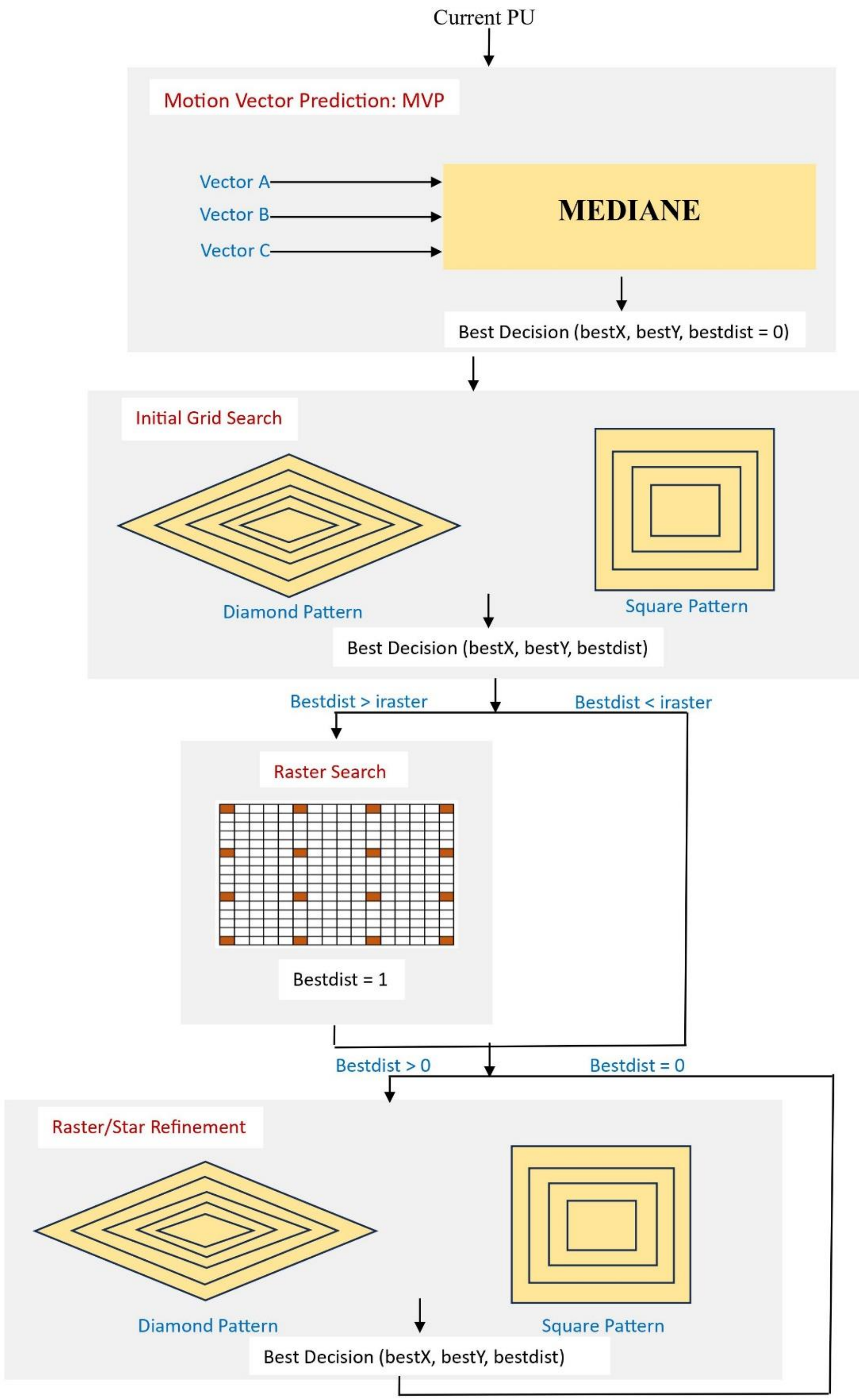

Fig. 3. Flowchart of TZS algorithm

Versatile Video Coding (VVC)

The VVC (MPEG-I Part-3) is a well-known merit for video compression by the Joint Video. After optimization H.264/AVC and H.265/HEVC, VVC is being consistent by the JVET of ITU-T - VCEG and ISO/IECMPEG. Currently, it is known as ITU-T H.266 and Future Video Coding (FVC). This is going to replace HEVC.

VVC Algorithm

The fundamental structure of VVC is like HEVC that consists of the partition of block, inter-and intra-prediction blocks, entropy coding, and loop filter. To reduce sequential dismissal, a required component of video coding, the inter-prediction plan seeks to obtain an identical block in the reference panel. Usually, ME and MC are important tools for inter-prediction. Numerous algorithms are used to calculate ME and MC.

Block-matching algorithm

This is a process that establishes indistinguishable macroblocks in video sequence frames for the assessment of motion. In order to produce appropriate objects on the next frame, the essential concept of ME is the matching of patterns to entities with their surroundings in frames of video series change inside the context. This approach may be used to determine progressive dismissal in video sequences by enhancing the interframe video encoding capability by referencing the unimportantly various sections of a macro-block to describe its contents. To calculate a translation motion vector (MV) with a rectangular block shape is used in Conventional ME (CME). With the variation in size of coding unit (CU) in motion estimation of 64×64 pixels. The HEVC standard accepts a QT structure. In scope of a prediction unit (PU) of the CU might be further divided into eight parts [8]. The MTT structures [9] have been recently introduced for accomplishing more flexible partitioning structures for expectation and realizing even greater compression efficacy. MTT is a tree like framework, in MTT one block divides into three different blocks, (1) QT, (2) BT, and (3) TT from the root. CTU divides into 3 trees. The starting point for a CTU can only be QT framework. Only the BT and TT, however, have a transferable archetype, which limits the further subdivision of the BT or TT to sub-BT or sub-TT.

Affine Motion Estimation (AME) and Motion Compensation (MC) in VVC

Usually, an Affine Motion Compensation (AMC) is executed that supports mainly two models in VVC, namely. (1) 4- parameter and (2) 6-parameter motion models. The Joint Video Exploration Team first developed Joint Exploration Model (JEM) which utilizes AMC coding tool. In terms of motion models, AMC differs from other existing models in that previous models use translational motion models; however, AMC uses three different motion models, namely translation, rotation, and zooming. AMC can operate under four parameter affine models which works between current block and reference block employ under JEM. AME might visage the motion more precisely (as it deals with three different motion) than other existing models, if a camera rotates/zooms to record a video. Li et al [10], developed a four parameter AM model for video coding. The study showed that the four parameters AM model increase from AME is around 10%. The study describes the significant coding gain obtained from AME [11] in the recent JEM executions. VTM may create motion vectors for AME using any of the two affine models that are principally dependent on the control point parameters, namely the 4-parameter and 6-parameter affine models [9]. A minimum of two or three vectors are compulsory to produce an affine transformed block. In order to define the location in a block, an affine model must take vectors in a 2D rectangular coordinate block. In this regard, mv at a particular location (x, y) can be expressed as (mvx, mvy) inside block to be expected, where mvx denotes a location on the abscissa , and mvy demonstrates a location on the ordinate. CME can also predict motion in two ways, (1) Unidirectional, and (2) bi-directional prediction in VTM. The control point motion vectors (CPMVs) express the motion vector field (MVF) of a coding unit in two categories: (a) two CP_s (4-parameter) or (b) three CP_s (6- parameter). Direction representation defines CP_0, CP_1 and CP_2 as the upper-left, upper-right and lower-left corners, respectively.

MV at sample position (x, y) in a CU for 4-parameter AM model is defined as:

$$mvx(x,y) = \left((mvx_1 - mvx_0) \times \left(\frac{x}{W}\right)\right) - \left((mvy_1 - mvy_0) \times \left(\frac{y}{W}\right)\right) + mvx_0 \quad \ldots\ldots\ldots (1)$$

$$mvy(x,y) = \left((mvy_1 - mvxy) \times \left(\frac{x}{W}\right)\right) - \left((mvx_1 - mvx_0) \times \left(\frac{y}{W}\right)\right) + mvy_0 \quad \ldots\ldots\ldots (2)$$

MV at sample position (x, y) in a CU for 6-parameter AM model is defined as:

$$mvy(x,y) = \left((mvx_1 - mvx_0) \times \left(\frac{x}{W}\right)\right) - \left((mvy_2 - mvy_0) \times \left(\frac{y}{W}\right)\right) + mvx_0 \quad \dots\dots\dots.(3)$$

$$mvy(x,y) = \left((mvy_1 - mvxy) \times \left(\frac{x}{W}\right)\right) - \left((mvx_2 - mvx_0) \times \left(\frac{y}{W}\right)\right) + mvy_0 \quad \dots\dots\dots.(4)$$

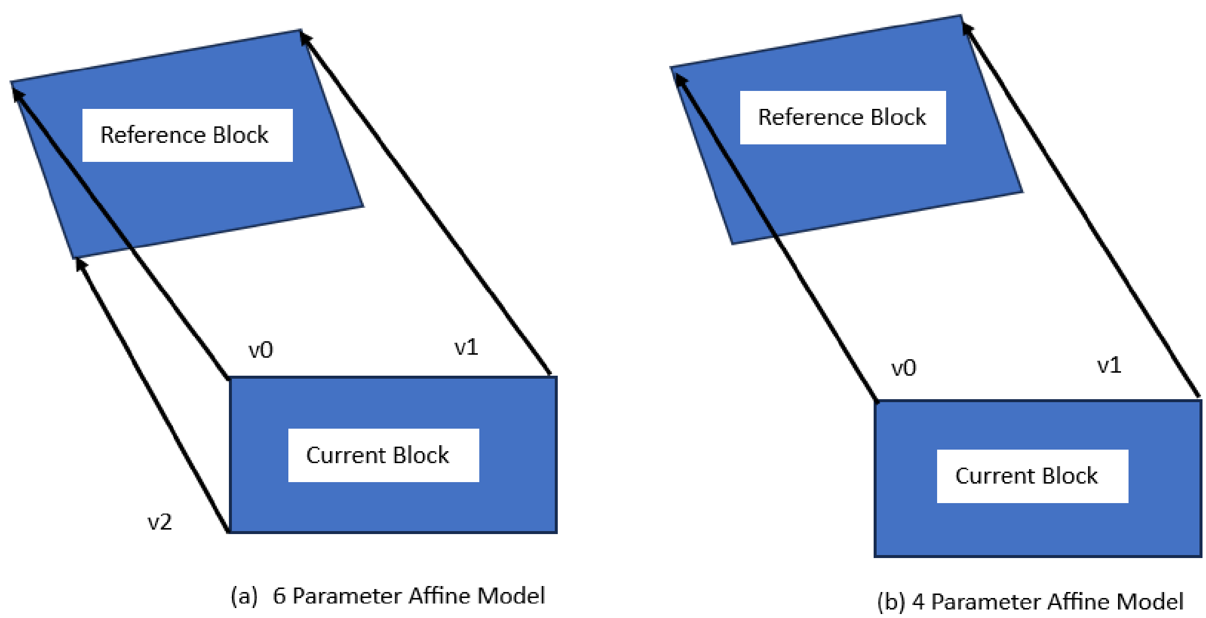

Fig.4 (a) six-parameter and (b) four-parameter affine model.

Unidirectional prediction or bi prediction can be applied to any AME process which entails both the related reference frames of an affine archetype as shown in Fig: 4. Therefore, escalating the encoding intricacy of VTM. When determining the variety of mandatory context sets for each ME operation, the AME technique required two iterations of the CME process. The VTM provides a level-based assessment system for quick AME coding because AME is so complex. Therefore, it is suggested to employ an acceptable level to decide whether AME is required.

Fast Affine Motion Estimation and Compensation

The ME and MC stages are typically the most difficult aspects of CU encoding. The intricacy increases with the variety of context sets utilized for ME, which also provide better accuracy in the result and additionally help in coding improvements, and conversely. Real-world videos have a diversity of motions that construct more and more computational and memory.

complexity. The complexity associated with memory might be incredulous for quick encoders. The encoder of VTM applies numerous approaches in ME [12]. Therefore, HM is also valuable for decisive whether the VTM might be effortlessly augmented. Fast Motion Estimation might be categorized under three approaches: (1) simple search pattern, (2) by adjusting the search parameters, and (3) premature termination. These two preceding approaches have been examined for decades. For example, Zhu et al 2000 [13], developed novel two-diamond search algorithm and situated in HM and VTM for rapid identification of blocks matching ME. The data suggested that the developed algorithm was efficiently used for the matching of identical blocks. Yang et al 2014 [14], develop a HEVC incorporated with directional search-patterns for rapid identification of ME. Another study [16] developed rotating hexagonal pattern. Nalluri and co-workers [15] developed HECV adaptive search for identification of ME. The data suggested that with the help of adaptive search minimizes the complexity of HM coding; thereby providing realistic results in lesser search range (i.e., 64) when compared to VTM. ME's complexity can be decreased further by limiting the total of searches (context sets). Pan et al. offered alternative methodologies [17] and created a low-complexity HEVC encoder in which a baseline choice method is employed to lessen the coding intricacy of ME. For quick ME coding, the number of preparatory reference frames is considered. Wang et al 2013 [18], developed multi-references frame ME for HEVC. In which adjoining reference frames with high resemblance possess have high probability to be chosen [18]. The coding competence and computational intricacy considerably vary that depending on reference frames are in use [19, 20]. Therefore, reference frames must be cautiously preferred in the contexture of other prediction tools. A premature end strategy is an effectual method that terminates superfluous ME operations (either for unidirectional or bidirectional prediction) in each block. The encoding time then significantly reduces if such ME process is skipped and results in loss of quality, thereby trusting high accurateness on the decision progression for the early cessation approach. The recursive partition of blocks in HEVC coding offers a sturdy association of data, allowing the ME process to be completed with great precision.

Pan et al. 2016 [21], for example, devised an instant ME method based on a minimized-complexity H.265/HEVC converter. The PU information stored in the QT structure is retained. Another study extended this work with minor modifications of motion search in [22]. The study suggested that there was no loss of coding. Rhee et al 2013, studied the prediction direction for HEVC. The bi prediction ME process may end prematurely if PU content in the QT structure is present. In the meantime, PU segregating does no longer exists in VTM those early cessation approaches can't be addressed directly to the MTT architecture. Therefore, a novel statistical study of the MTT is expected to be helpful for achieving low complexity in VTM encoders. In this context, VTM 3.0 was a novel strategy to be used for the state-of-the art algorithm on light weight AME in VVC. In 2018, the primitive experimental design (VTM 1.0) was introduced, and since then various research [24] which are being applied to the test model achieving good performance n VVC. These trials were applied to search for an improved stability between coding proficiency and stimulation complexity. For very specific, instead of deriving the parameters again [25], Zhou et al, suggested to reprocess affine motion parameters of neighboring blocks. Another report by Zhang et al. [25], recommended VTM 3.0 to avoid the complexity of VTM 2.0 as that overall memory band width decreases using this process and to evade the small block's attribute compilation operation.

To further enhance the process of Affine Motion Estimation Sagrilo *et al.* in [26] introduce a learning-based technique which can aid the capturing of complicated correlations in motion data and produce more efficient or accurate forecasts. In [27], Duarte *et al.* used machine learning to accelerate the motion estimation in AME. Hong *et al* in [28] further introduced enhanced edge detection techniques as part of the motion estimation process. Edge detection aids in identifying visual boundaries and conspicuous features, which might be critical for correct motion estimations. Currently many attention-based designs are being used to improve motion estimation in video coding. In [29] an attention based bi prediction network in introduced that enables the model to focus on specific parts of the input data that are more relevant for the task.

Discussion

In this article we have presented a thorough survey on various motion estimation techniques used in HEVC and VVC. Various algorithms and their applications along with advantages and disadvantages have been explained which have been widely used for ME in video coding techniques. Some effective strategies for increasing the pace of ME such as AME are introduced by utilizing learning-based techniques. This breakthrough has the potential to improve video compression efficiency and real-time applications. Further study could investigate incorporating machine learning into other elements of video coding in order to continue pushing the limits of performance and speed. This unique solution addresses video compression and transmission difficulties by leveraging attention mechanisms and bi prediction to improve efficiency and quality. The network's use in 5G contexts demonstrates its ability to transform real-time video streaming and broadcasting. As 5G technology evolves, the attention-based bi-prediction network provides a timely and significant answer for optimizing multimedia experiences across high-speed networks. The study also tackles a crucial feature of video coding at the Coding Tree Unit (CTU) extent. The algorithm's emphasis on hardware efficiency indicates its applicability for real-time video processing applications. The hardware-friendly CTU-level technique contributes to optimizing video compression and transmission via VVC-encoded streams as video data becomes more demanding in modern multimedia systems.

Conclusion

This review summarized the software model VTM for VVC which offers superiority in terms of performance when compared to that of HEVC. The affine extrapolation contributes considerably to compression performance by taking a superior variation of motions found in nature among all newly adopted technologies for VTM. Additionally, the AME complication is a tie-up for minimal complication encoding implementations. The studies suggested that VVC technology might be next-generation technology for the detection of motion.

REFERENCES

[1] M. Sinangil, V. Sze, M. Zhou, and A. Chandrakasan, “Cost and coding efficient motion estimation design considerations for high efficiency video coding (hevc) standard,” IEEE J. Selected Topics Signal Process.,, vol. 7, no. 6, pp. 1017–1028, Dec 2013.

[2] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, “Overview of the h.264/avc video coding standard,” IEEE Trans. Circuits Syst. Video Technol., vol. 13, no. 7, pp. 560–576, 2003.

[3] J. Ohm, G. 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. VideoTe chnol.,, vol. 22, no. 12, pp. 1669–1684, Dec 2012.

[4]. Chen, J.; Alshina, E. Algorithm Description for Versatile Video Coding and Test Model 1 (VTM 1); JVET-J1002; JVET: San Diego, CA, USA, 2018.

[5]. Chen, J.; Ye, Y.; Kim, S. Algorithm Description for Versatile Video Coding and Test Model 2 (VTM 2); JVET-K1002; JVET: Ljubljana, Slovenia, 2018.

[6] N. Purnachand, L. Alves, and A. Navarro, “Fast motion estimation algorithm for hevc,” in IEEE Int. Conf. Consumer Electronics - Berlin (ICCE-Berlin), 2012, Sept 2012, pp. 34– 37.

[7] S.-H. Yang, J.-Z. Jiang, and H.-J. Yang, “Fast motion estimation for hevc with directional search,” Electronics Letters, vol. 50, no. 9, pp. 673–675, April 2014.

[8] 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.

[9] J. Chen, Y. Ye, and S. H. Kim, Algorithm Description for Versatile Video Coding and Test Model 3 (VTM 3), Joint Video Experts Team (JVET), document ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JVETL1002, Oct. 2018

[10] L. Li, H. Li, D. Liu, Z. Li, H. Yang, S. Lin, H. Chen, and F. Wu, ‘‘An efficient four-parameter affine motion model for video coding,’’ IEEE Trans. Circuits Syst. Video Technol., vol. 28, no. 8, pp. 1934–1948, Aug. 2018

[11] K. Zhang, Y.-W. Chen, L. Zhang, W.- J. Chien, and M. Karczewicz, ‘‘An improved framework of affine motion compensation in video coding,’’ IEEE Trans. Image Process., vol. 28, no. 3, pp. 1456–1469, Mar. 2019

[12] C. Rosewarne, B. Bross, M. Naccari, K. Sharman, and G. Sullivan, High Efficiency Video Coding (HEVC) Test Model 16 (HM 16) Improved Encoder Description Update 9, Joint Collaborative Team on Video Coding (JCT-VC), document ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JCTVC-AB1002, Jul. 2017

[13] S. Zhu and K.-K. Ma, ‘‘A new diamond search algorithm for fast block matching motion estimation,’’ IEEE Trans.
Image Process., vol. 9, no. 2, pp. 287–290, Feb. 2000.

[14] S.-H. Yang, J.-Z. Jiang, and H.-J. Yang, ‘‘Fast motion estimation for HEVC with directional search,’’ Electron. Lett., vol. 50, no. 9, pp. 673–675, Apr. 2014

[15] 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. ICIP, Paris, France, Oct. 2014, pp. 3696–3699.

[16] P. Nalluri, L. N. Alves, and A. Navarro, ‘‘Complexity reduction methods for fast motion estimation in HEVC,’’ Signal Process.-Image Commun., vol. 39, pp. 280–292, Nov. 2015.

[17] Z. Pan, P. Jin, J. Lei, Y. Zhang, X. Sun, and S. Kwong, ‘‘Fast reference frame selection based on content similarity for low complexity HEVC encoder,’’ J. Vis. Commun. Image Represent., vol. 40, pp. 516–524, Oct. 2016.

[18] S. Wang, S. Ma, S. Wang, D. Zhao, and W. Gao, ‘‘Fast multi reference frame motion estimation for high efficiency video coding,’’ in Proc. ICIP, Melbourne, VIC, Australia, Sep. 2013, pp. 2005–2009.

[19] D. M. M. Rahaman and M. Paul, ‘‘Virtual view synthesis for free viewpoint video and multiview video compression using Gaussian mixture modelling,’’ IEEE Trans. Image Process., vol. 27, no. 3, pp. 1190–1201,Mar.2018.

[20] M. Paul, ‘‘Efficient multiview video coding using 3-D coding and saliencybased bit allocation,’’ IEEE Trans. Broadcast., vol. 64, no. 2, pp. 235–246, Jun. 2018.

[21] Z. Pan, J. Lei, Y. Zhang, X. Sun, and S. Kwong, ‘‘Fast motion estimation based on content property for low-complexity H.265/HEVC encoder,’’ IEEE Trans. Broadcast., vol. 62, no. 3, pp. 675–684, Sep.2016.

[22] S. Park and E. S. Jang, ‘‘Fast motion estimation based on content property for low-complexity H.265/HEVC Encoder,’’ IEEE Trans. Broadcast., vol. 63, no. 4, pp. 740–742, Dec. 2017.

[23] C. E. Rhee and H. J. Lee, ‘‘Early decision of prediction direction with hierarchical correlation for HEVC compression,’’ IEICE Trans. Inf. Syst., vol. 96, no. 4, pp. 972–975, Apr. 2013.

[24] M. Zhou, CE4: Test Results of CE4.1.11 on Line Buffer Reduction for Affine Mode, Joint Video Experts Team (JVET), document ITUT SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JVET-L0045,Oct.2018.

[25] K. Zhang, L. Zhang, H. Liu, Y. Wang, P. Zhao, D. Hong, CE4: Affine Prediction with 4ˆŮ4 Sub-Blocks for Chroma Components (Test 4.1.16), Joint Video Experts Team (JVET), document ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JVETL0265, Oct. 2018.

[26] Sagrilo, F., Loose, M., Viana, R., Sanchez, G., Corrêa, G. and Agostini, L., 2023, May. Learning-Based Fast VVC Affine Motion Estimation. In *2023 IEEE International Symposium on Circuits and Systems (ISCAS)* (pp. 1-5). IEEE.

[27] Duarte, A., Gonçalves, P., Agostini, L., Zatt, B., Correa, G., Porto, M. and Palomino, D., 2022, May. Fast affine motion estimation for vvc using machine-learning-based early search termination. In *2022 IEEE International Symposium on Circuits and Systems (ISCAS)* (pp. 1-5). IEEE.

[28] Hong, J., Dong, Z., Zhang, X., Song, N. and Cao, P., 2023. A Fast Gradient Iterative Affine Motion Estimation Algorithm Based on Edge Detection for Versatile Video Coding.

[29] Choi, Y.J., Lee, Y.W., Kim, J., Jeong, S.Y., Choi, J.S. and Kim, B.G., 2023. Attention-Based Bi-Prediction Network for Versatile Video Coding (VVC) over 5G Network. *Sensors*, *23*(5), p.2631.