RAW PLENOPTIC VIDEO CODING UNDER HEXAGONAL LATTICE RESOLUTION OF MOTION VECTORS

Thuc Nguyen Huu, Vinh Duong Van, Jonghoon Yim, and Byeungwoo Jeon

Department of Electrical and Computer Engineering, Sungkyunkwan University, Korea

ABSTRACT

In raw plenoptic video, the optimal motion searching points mostly follow the hexagonal structure of micro-images. Based on this understanding, we propose a new motion vector resolution, namely the *hexagonal lattice* (HL) resolution which reflects micro-image structure. The HL resolution can be efficiently represented by HL basis. A study in this paper shows that motion vectors are highly concentrated at hexagonal lattice points, leading to use of the proposed resolution in the context of video compression. In this regard, we demonstrate the compression benefit brought by estimating motion vectors at HL resolution in the VVC codec.

Index Terms— plenoptic video, lenslet video, motion vector resolution, hexagonal lattice resolution, video coding.

1. INTRODUCTION

Plenoptic camera [1] [2] records the 4D plenoptic function [3] (or 4D light field [4]) by inserting an additional microlens array (MLA) between main lens and image sensor as in Fig. 1. The sensor output of plenoptic camera is known as the raw plenoptic image since it yet requires additional pre-processing steps to generate final plenoptic function. The plenoptic image consists of many micro-images (drawn as circles in Fig. 1) which are distributed following a particular structure. Here, the structure of micro-images depends on MLA layout which can be either in rectangular or hexagonal. To increase the fill-factor of a sensor, the MLA often follows the hexagonal layout, leading to hexagonal structure of micro-images.

The raw plenoptic data, especially for raw plenoptic video, calls for compression due to its extremely high spatial resolution (it could be up to 7728 x 5368 for sensor output of Lytro camera). In this regard, motion estimation and compensation of plenoptic video are essential enabling techniques to effectively reduce the temporal redundancy. Unlike conventional video, plenoptic video has a unique characteristic to utilize in motion estimation, i.e., its optimum searching points mostly (around 80%) follow the micro-images structure [5][6]. This interesting property is illustrated in Fig. 2(b) which plots the motion vectors (MVs) of raw plenoptic video estimated by full search. The observation that motion vectors are, in general, sparsely located in plenoptic video motivated development of fast search methods [5][6] which can speed up the motion search while maintaining the compression performance. Noting that previous works [5][6] only paid attention to fast motion search algorithms, we suggest that the special motion property is not fully utilized yet, and propose a new motion vector resolution, namely the hexagonal lattice (HL) resolution.

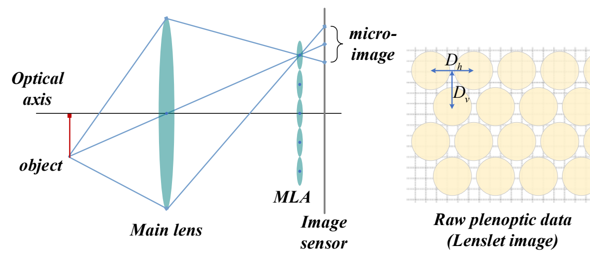
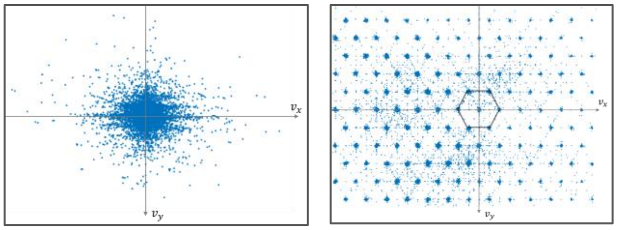


Fig. 1. Optical structure of plenoptic camera and its sensor output

In this paper, we study using the HL resolution of the plenoptic video since it may work efficiently with the conventional MV resolutions in video coding standards (e.g., H.264/AVC, H.265/HEVC, H.266/VVC, etc.) considering its better rate distortion performance than the conventional MV resolutions for most cases. We focus on the sensor output of Lytro camera [1] (standard plenoptic camera) and the research for Raytrix camera [2] is left for future work.

The paper is organized as follows. In the next section, we present fundamental background of motion vector and motion vector resolution in video coding standards. Following, in Section 3, we propose the HL resolution which is a new MV resolution for raw plenoptic video. A comprehensive study of HL resolution and its usage are also presented. In Section 4, we give experimental results and next we conclude the paper.

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (NRF-2020R1A2C2007673).



(a) (b)

Fig. 2. Distribution of motion vectors in (a) conventional video; and (b) raw plenoptic video data. In (b), a hexagonal shape is drawn at origin to highlight the unique property of motion vectors. The motion vectors are found by full search using VVC encoder [14] under low delay configuration.

2. MOTION VECTOR RESOLUTION AND OUR MOTIVATION

To exploit the correlation in video data existing in temporal domain, video coding techniques perform motion compensated-prediction:

$$I_{cur}(\mathbf{x}) - I_{ref}(\mathbf{x}+\mathbf{v}), \tag{1}$$

where $\mathbf{x} \in \mathbb{Z}^2$ is an integer pixel location, and $\mathbf{v} = (v_x, v_y) \in \mathbb{R}^2$ is a motion vector defining pixel displacement between current picture I_{cur} and reference picture I_{ref}. It is worth noting in this paper that we normalize unit of all measurements as pixel (*pel*) size. When the continuous motion is represented in a finite precision, a higher MV precision gives better prediction quality, but in turn, consumes more bit overhead to signal the information. Therefore, choosing a proper MV precision (or resolution) is a fundamental issue in terms of trade-off between rate and distortion.

A motion vector $\mathbf{v}$ is said to be at resolution $\Delta \in \mathbb{R}^+$ if it is represented by an integer multiple of Δ:

$$\mathbf{v} = \mathbf{n}\Delta \qquad \text{where} \quad \mathbf{n} \in \mathbb{Z}^2 \tag{2}$$

Most common resolutions of Δ are quarter-pel (1/4-pel), half-pel (1/2-pel), integer-pel (1-pel), all of which have a form $\Delta = 2^{-k}$ $(k \in \mathbb{Z})$. When the resolution Δ is non-integer pel, it is usually called as sub-pel resolution. H.264/AVC and H.265/HEVC use a fixed ¼-pel MV resolution [7][8]. This design choice not only has considered the RD performance, but also the complexity of motion estimation and interpolation filtering process. In estimating a motion vector at a sub-pel resolution, in general, interpolation is executed, which requires additional computational processing [9]. The high processing power of modern CPUs and GPUs have made it possible to have higher MV resolutions. The AV1 compression allows motion resolution up to 1/8-pel [10] while the VVC adopts even higher resolution, that is, 1/16-pel for control points MV in affine motion mode [11].

In Fig. 2, we illustrate MVs obtained from conventional video and raw plenoptic video, which points out a key difference between them: for conventional video, MVs are densely distributed along the origin as in Fig. 2(a) whereas sparsely distributed in a hexagonal pattern in case of raw plenoptic video as in Fig. 2(b). It clearly hints that current MV resolutions are well-fitted for conventional videos, however they might not be necessarily the best for raw plenoptic video where the sparsity of MVs dominates. It motivates us to define a new MV resolution specifically for raw plenoptic video.

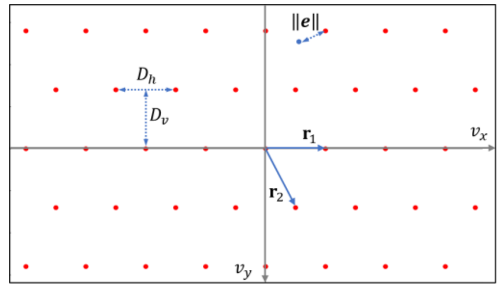


Fig. 3. The concept of HL resolution for motion vector resolution/precision. Here red dots correspond to motion vectors at HL locations which is formed by two basis vectors $\mathbf{r}_1, \mathbf{r}_2$.

3. PROPOSED HEXAGONAL LATTICE RESOLUTION

3.1. Definition

In raw plenoptic video, a motion vector $\mathbf{v} \in \mathbb{R}^2$ *is said to be at hexagonal lattice (HL) resolution if* $\mathbf{v}$ *is representable by a HL basis:*

$$\mathbf{v} = \mathbf{A}\mathbf{n}: \quad \mathbf{n} \in \mathbb{Z}^2, \tag{3}$$

where $\mathbf{A} \in \mathbb{R}^{2\times 2}$ *is a generator matrix consisting of two basis vectors of HL.* Although the choice of basis vectors is not unique, for the sake of simplicity, we adopt a set of basis vectors $\{\mathbf{r}_1, \mathbf{r}_2\}$ in Fig. 3, which can be denoted as follows:

$$\mathbf{r}_1 = \begin{pmatrix} D_h \\ 0 \end{pmatrix}; \mathbf{r}_2 = \begin{pmatrix} D_h/2 \\ D_v \end{pmatrix}, \tag{4}$$

where D_h, D_v are micro-images distance shown in Fig. 1. These distances are known by camera manufactures through their optical configurations. In short, the matrix $\mathbf{A}$ can be explicitly specified: $\mathbf{A} = (\mathbf{r}_1 \quad \mathbf{r}_2)$.

3.2. Study on HL locations

For raw plenoptic video, motion vectors are quite expected to be highly concentrated at HL locations. We verify this statement by setting up a statistical analysis of motion vectors in raw lenslet video. To obtain sufficient MV accuracy, we encode raw plenoptic video under VVC low delay configuration [12] with full search motion estimation (the motion cost involving rate is ignored during the RD check, and search range is set 128). Here, the MV resolution is set to ¼-pel and the affine motion mode [10] is turned off. After the encoding, we obtain a ground truth of MV data and their locations are shown in Fig. 2(b). Interestingly, the MVs are seen to be located mostly at HL points. To measure the degree

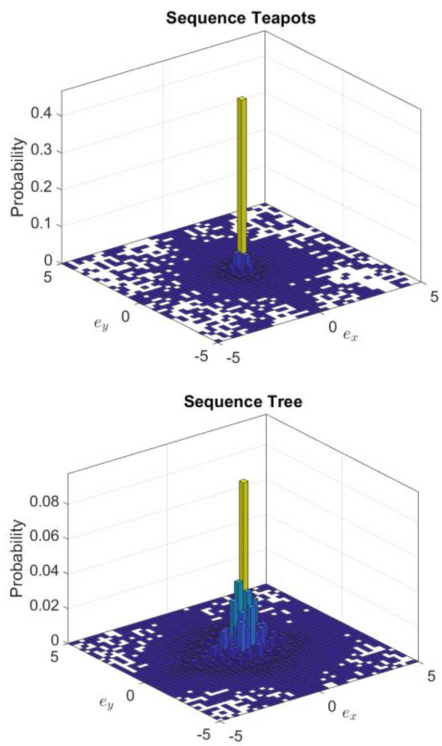


Fig. 4. Probability distribution of error vector $\mathbf{e} = (e_x, e_y)$ for test sequence Teapots and Tree.

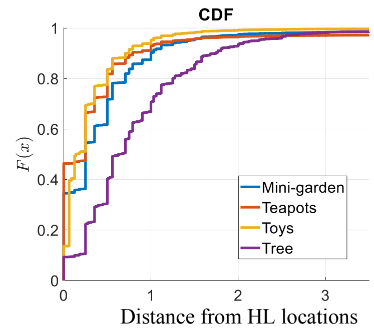


Fig. 5. Cumulative distribution function (CDF) of error magnitude $\|\mathbf{e}\|$ for various test sequences.

of concentration, we treat the observed motion vector $\mathbf{v}_{obs}$ as a linear measurement:

$$\mathbf{v}_{obs} = \mathbf{A}\mathbf{n} + \mathbf{e}, \tag{5}$$

where $\mathbf{A} = (\mathbf{r}_1 \quad \mathbf{r}_2)$ is the generator matrix for HL, and $\mathbf{e} = (e_x, e_y)$ is an error vector. A smaller value of $\|\mathbf{e}\|$ indicates a closer MV to HL position.

We plot the probability distribution of the error vector $\mathbf{e}$ and cumulative distribution function (CDF) of the error magnitude $\|\mathbf{e}\|$ in Fig. 4 and Fig. 5, respectively. As can be seen from Fig. 4, the error vector $\mathbf{e}$ is mostly located near zero with very high probability. Regarding CDF shown in Fig. 5, $\Pr(\|\boldsymbol{e}\| \leq 0.5)$ is respectively 0.6, 0.7, 0.75 for test sequences Mini-garden, Teapots, Toys, Tree. These CDF values are even higher if we consider $\Pr(\|\boldsymbol{e}\| \leq 1.0)$ which are 0.87, 0.9, 0.95, 0.65, respectively. It means that one can mostly find a motion vector within a radius of single pixel distance surrounding HL locations. Therefore, it can be safely stated that MVs are very concentrated at HL locations.

3.3. The potential uses of HL resolution

A good MV resolution should be the one which consumes less number of bits to represent the MV data. A low resolution might not be able to represent actual motion in a scene accurate enough, thus reducing the performance of motion compensated prediction. However, for raw plenoptic video, because of the high concentration at HL locations, it could be better to replace conventional MV resolution with our proposed HL resolution.

One suggested way of introducing HL resolution into video codec can be described briefly as follows. Firstly, a plenoptic video encoder sends a quantized motion vector by using a *lattice quantizer* [13]:

$$\mathbf{n}_{sent} = [\mathbf{A}^{-1}.\mathbf{v}], \tag{6}$$

where $[.]$ denotes an element-wise rounding operator. Secondly, a reconstructed motion vector $\mathbf{v}_{rec}$ can be found by using an inverse quantization process:

$$\mathbf{v}_{rec} = \mathbf{A}[\mathbf{A}^{-1}.\mathbf{v}] = \mathbf{A}\mathbf{n}_{sent}$$

Once the motion vectors are reconstructed, it is possible to obtain the motion compensated prediction similar to Eq. (1):

$$I_{cur}(\mathbf{x}) \leftarrow I_{ref}(\mathbf{x} + \mathbf{v}_{rec}). \tag{8}$$

4. EXPERIMENTAL RESULTS

4.1. Test condition

To evaluate performance, we implement the proposed HL resolution on top of the VVC codec (VVenC v1.2.0 [14]) under low-delay configuration as in Table I with four raw plenoptic video test sequences (see Fig. 6) [15]. In this dataset, the micro-images distances are $D_h = 16$ pixels and $D_v = 14$ pixels. Each test sequence is compressed four times by using quantization parameters of {24, 30, 36, 42}. In our experiments, we compare the BD-Rate (%) performance of our proposed method (HL resolution) with those of three conventional MV resolutions: int-pel, half-pel, quarter-pel.

Table I. Summary of test condition

Software	• VVenC v1.2.0 [14] • Low delay configurations • NumThreads = 8
Affine mode [10]	Off
AMVR mode [10]	Off
Motion search	Full search with search range = 128
Quantization parameters	24, 30, 36, 42

Fig. 6. The dataset for the experiment.

4.2. Test results

In Table II, we evaluate the compression performance of our HL resolution by comparing BDBR of the proposed method with three conventional MV resolutions as anchor individually. A slightly mixed coding performance can be seen from the results. Except the sequence Tree, the HL resolution is observed to greatly outperform the int-pel and the half-pel resolutions by average of -1.83% and -3.10% respectively while it achieves a slightly worse performance to quarter-pel resolution. Breaking down to individual test sequence, we observe a good performance gain of sequence Teapots (ranging from -4.0% to -7.6%), and performance loss of sequence Tree (ranging from +0.8% to +5.4%). It indicates that HL resolution might not always perfectly represent the actual motion in sequence Tree. The reason can be seen from Fig. 5 where CDF curve of sequence Tree is the lowest one, among four test sequences. It tells us that a considerable amount of MVs is far away from HL location, thus they cannot be represented effectively in HL resolution. Therefore, we want to address this issue in the future work.

Table II. BD-Rate (%) of the proposed HL resolution with three conventional MV resolutions as anchor respectively

Anchor / Sequences	*Int-pel*	*Half -pel*	*Quarter -pel*
Mini-garden	-2.5%	-3.9%	+1.8%
Teapots	-5.7%	-7.6%	-4.0%
Toys	-1.6%	-1.7%	-0.5%
Tree	+2.5%	+0.8%	+5.4%
Avg.	-1.83%	-3.10%	+0.68%

5. CONCLUSION

In this paper, we have presented the HL resolution which is a new MV resolution for raw plenoptic video coding. The HL resolution is seen to achieve better rate-distortion performance than the int-pel, the half-pel resolution. As a future work, the HL resolution can be combined with conventional ones to build an adaptive scheme of MV resolution, which might further improve the coding performance.

6. REFERENCES

[1] R. Ng, "Light field photography with a hand-held plenoptic camera," Technical report, Stanford University, 2005.

[2] A. Lumsdaine and T. Georgiev, "The focused plenoptic camera," 2009 IEEE Int. Conf. on Computational Photography (ICCP), 2009.

[3] E. Adelson and J. Bergen, The plenoptic function and the elements of early vision. Cambridge, Mass.: Vision and Modeling Group, Media Laboratory, Massachusetts Institute of Technology, 1991.

[4] M. Levoy and P. Hanrahan, "Light field rendering," Proceedings of the 23rd annual conference on Computer graphics and interactive techniques - SIGGRAPH '96, 1996.

[5] L. Li and X. Jin, "Macropixel-constrained Collocated Position Search for Plenoptic Video Coding," 2019 IEEE Visual Communications and Image Processing (VCIP), 2019.

[6] T. N. Huu, V. Van Duong and B. Jeon, "Fast and Efficient Microlens-Based Motion Search for Plenoptic Video Coding," 2021 IEEE Int. Conf. on Image Processing (ICIP), 2021.

[7] T. Wiegand, G. J. Sullivan, G. Bjontegaard and A. Luthra, "Overview of the H.264/AVC video coding standard," in IEEE Trans. on Circuits and Systems for Video Tech., vol. 13, no. 7, pp. 560-576, July 2003.

[8] G. J. Sullivan, J. Ohm, W. Han and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," in IEEE Trans. on Circuits and Systems for Video Tech., vol. 22, no. 12, pp. 1649-1668, Dec. 2012.

[9] H. Lv, R. Wang, X. Xie, H. Jia and W. Gao, "A comparison of fractional-pel interpolation filters in HEVC and H.264/AVC," 2012 Visual Communications and Image Processing, 2012, pp. 1-6.

[10] Y. Chen et al., "An Overview of Core Coding Tools in the AV1 Video Codec," 2018 Picture Coding Symp. (PCS), 2018, pp. 41-45.

[11] B. Bross et al., "Overview of the Versatile Video Coding (VVC) Standard and its Applications," in IEEE Trans. on Circuits and Systems for Video Tech., vol. 31, no. 10, pp. 3736-3764, Oct. 2021.

[12] F. Bossen, J. Boyce, X. Li, V. Seregin and K. Sühring, "VTM Common Test Conditions and Software Reference Configurations for SDR Video", document JVET-T2010 of JVET, Oct 2020.

[13] Gibson, Jerry D., and Khalid Sayood. "Lattice quantization." Advances in electronics and electron physics 72 (1988): 259-330.

[14] Fraunhofer Versatile Video Encoder software (VVenC v1.2.0), https://github.com/fraunhoferhhi/vvenc.

[15] M. Teratani, X. Jin, L. Li, G. Lafruit, and L. Yu, "Exploration experiments and common test conditions for dense light fields," ISO/IEC JTC1/SC29/WG11, Marrakesh, Morocco, N18173, 2019.