

# Performance comparison of video encoders in light field image compression

Hadi Amirpour<sup>1</sup>, Antonio M. G. Pinheiro<sup>1</sup>, Manuela Pereira<sup>1</sup>, and Mohammad Ghanbari<sup>2,3</sup>

<sup>1</sup>Instituto de Telecomunicações and Universidade da Beira Interior, Covilhã, Portugal

<sup>2</sup>School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran

<sup>3</sup>School of Computer Science and Electronic Engineering, University of Essex, Colchester, UK

## Abstract

Efficient compression plays a significant role in Light Field imaging technology because of the huge amount of data needed for their representation. Video encoders using different strategies are commonly used for Light Field image compression. In this paper, different video encoder implementations including HM, VTM, x265, xvc, VP9, and AV1 are analysed and compared in terms of coding efficiency, and encoder/decoder time-complexity. Light field images are compressed as pseudo-videos.

## Introduction

Light field technology is a promising representation for 3D imaging. It enables some post-processing tasks like refocusing, synthesising new views and depth estimation [1]. These features are enabled at the cost of significant data increase which makes compression a major task to enable its practical use. Independently of acquisition with a camera multi-array or a lenslet camera, a captured light field image of a scene can be represented by a set of multi-views of a scene captured from different viewpoints.

Light field images are characterized with four dimensions  $LF(u, v, s, t)$  [2].  $(s, t)$  are used to select the view, and  $(u, v)$  represents the intensity of pixel  $(u, v)$  in view  $(s, t)$ . One example of a light field image is shown in Figure 1.

Video encoders can be exploited to compress these images using different scan orders. Due to the similarity of different views, exploiting inter-coding capabilities of video encoders provides an efficient way of compressing the views of a light field image. In this regard, views are ordered as video frames resulting in a so-called pseudo-video. Then, the pseudo-video is used as an input to a video encoder to be compressed. The frames are inversely converted to the light field views using the initial order. The procedure is shown in Figure 2.

Several video encoders were used to compress the pseudo-videos that were result from light field images. H.264 or MPEG-4 Part 10, Advanced Video Coding (MPEG-4 AVC) [3] has been used in [4] to compress pseudo-videos using spiral scan order. High Efficiency Video Coding (HEVC) [5] is widely used to compress light field pseudo-videos. In [6], raw lenslet images are tiled as a pseudo-temporal sequence and HEVC is used to compress the pseudo-video. A circular scan order is used to form pseudo sequences in [7]. Then, HEVC is deployed to compress the pseudo-sequences. Ordering based on perceptual quality has been proposed in [8] which uses HEVC to encode the pseudo-sequence. Scanning order in [9] is based on utilizing spatially closer sub-aperture images and HEVC is used as the video encoder. [10] is an extension of [9] which uses the reference soft-

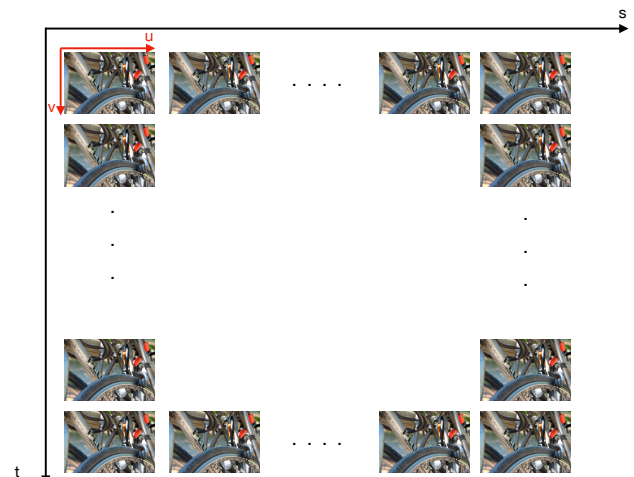


Figure 1: Light field images can be represented by the different views captured from one scene.

ware for the new Versatile Video Coding (VVC) using the VVC Test Model (VTM). Joint Exploration Model (JEM) is another video encoder that is used in some methods to encode the light field views pseudo-videos. In [11], JEM is utilized to compress pseudo-sequence in which views have a symmetrical 2D hierarchical structure. A hybrid scan order has been proposed in [12] which uses JEM as an encoder. Coding efficiency of video encoders might have a significant impact on compressing the light field views pseudo-videos. With the increasing number of video encoders selecting an appropriate encoder is a challenging task. Consequently, encoding light field pseudo-sequences as a specific case of a video using various state-of-the-art encoders is studied in this work.

Video coding standards have been mainly developed by ITU-T and ISO/IEC standard organizations. H.261 and H.263 were defined by ITU-T while MPEG-1 and MPEG-4 Visual were defined by ISO/IEC. H.262/MPEG-2 and H.264/MPEG-4 AVC were defined jointly by these two standard organizations. In 2010, a collaboration team between the ISO/IEC MPEG and ITU-T VCEG was established to develop a new video compression standard with the aim of 50% bitrate reduction compared to H.264/AVC. This joint team was named Joint Collaborative Team on Video Coding (JCT-VC) and their effort was finalized in 2013 and named HEVC. This standard was named MPEG-H Part 2 (ISO/IEC 23008-2) in ISO/IEC and ITU-T Recommendation H.265 in ITU-T.

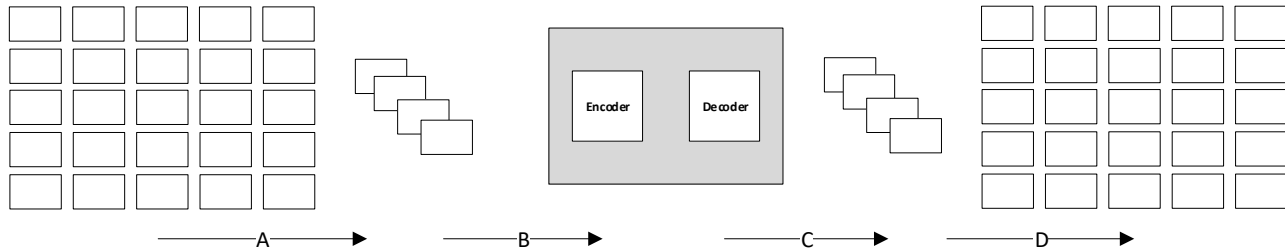


Figure 2: Workflow for pseudo-sequence based light field image compression methods. A) Views are converted to a pseudo-video based on a predefined scan order. B) Pseudo-video is encoded. C) Pseudo-video is decoded. D) Decoded Pseudo-video is ordered to the views based on initial scan order.

As a next-generation video compression standard and successor of HEVC, the Joint Video Exploration Team (JVET) which joint experts from MPEG and ITU was established to develop the Versatile Video Codec (VVC). It aims to reach at least 30% bitrate saving over HEVC with support of up to 16K video and 360° degrees videos.

The paper is organized in the following orders. Selected state-of-the-art encoders and their settings are presented in the next section. Section 3 provides details about implementation setup and dataset. Simulation results are given in section 4 and section 5 concludes the paper.

In addition to standardization activities some companies had developed their own codecs like, e.g. VP8 which was developed mainly for web applications. VP8 that initially was developed by On2 Technologies<sup>®</sup>, was purchased by Google<sup>®</sup>. VP8 evolution resulted in the VP9 codec in 2013, and was aiming for better compression efficiency than HEVC and 50% bitrate saving over VP8. In 2015, the Alliance for Open Media<sup>®</sup> was established between the main players on the multimedia consumption and services market (like Google<sup>®</sup>, Microsoft<sup>®</sup>, Intel<sup>®</sup> or Netflix<sup>®</sup>, between others) to develop a royalty free video format as an alternative to licensed formats like HEVC. Their effort led to a new video codec AV1 with a stable version released in 2018. AV1 is based on VP9 with additional tools aiming to gain better coding efficiency than other state-of-the-art codecs. The performance of the different codecs has been compared in [13, 14, 15, 16, 17]. In this paper, their performance is evaluated for coding of light field pseudo-video sequences.

## Selected Encoder Softwares

### HM software

HEVC Test Model (HM) software is selected as the main implementation of HEVC standard. Particularly, its latest version HM-16.20<sup>1</sup> was selected for the simulations. Three different configurations namely Low delay P, Low delay B, and random access were selected. None of the parameters were changed unless *IntraPeriod*, which is set to -1 for random access configuration file like LowDelay configuration files. This is because in the light field pseudo-videos, scene cut is not expected and frames are similar. Based on HEVC common test condition (CTC), four fixed QPs were selected for the I-frames namely 22, 27, 32, and 37 for the random access configuration (HM-RA) and for the Low delay B (HM-LB) and Low delay P (HM-LP) configurations, QPs obtained experimentally to fit bitrates of HM-RA.

<sup>1</sup><https://hevc.hhi.fraunhofer.de/>

### VTM software

VVC Test Model (VTM) software has been selected as representative of VVC codec. VTM version 4.0.1<sup>2</sup> as the latest version of the VTM software has been used for the simulations. Similar to HM, three different configurations namely Low delay P, Low delay B, and random access were selected. Based on JVET CTC and similar to HEVC CTC, four fixed QPs were selected for the I-frames namely 22, 27, 32, and 37. *IntraPeriod* is set to -1 for the random access (VTM-RA) configuration file. For VTM-LB and VTM-LP, QPs were obtained experimentally.

### VP9 software

WebM Project VP9 Encoder v1.8.0-173-g986b2bef7<sup>3</sup> is selected as the latest version of the VP9. It is very important to have as much as possible similar coding configurations compared with HM and VTM softwares. To have a fixed QP setting like HM and VTM, fixed quality mode *--end-usage* is set to 3. Key frame placement (*disable-kf*) is disabled similar to setting *IntraPeriod* to -1 in HM and VTM. To have a similar Group Of Pictures (GOP) in random access configurations in HM and VTM which is 16, minimum (*min-gf-interval*) and maximum (*max-gf-interval*) golden frame intervals are set to 16. Selected settings for the VP9 encoder have been summarized in Table 1.

Table 1: Selected settings for VP9 encoder

```
--best --cpu-used=0 --codec=vp9 --aq-mode=0
--end-usage=3 --disable-kf --tune=psnr
--min-gf-interval=16 --max-gf-interval=16
--fps=<FrameRate> -w <Width> -h <Height >
--cq-level=<QP>
```

### AV1 software

AOMedia Project AV1 Encoder 1.0.0-1001-gf5c9213e7<sup>4</sup> has been selected as the latest version of AV1. AV1 setting is selected to be similar to VP9, as summarized in Table 1. The only changes are the codec definition that was changed to AV1 (*--codec=av1*) and (*--best*) quality deadline was removed.

### xvc software

The xvc codec is a software-defined video compression format, with first version released on 2017. The xvc codec is mainly

<sup>2</sup><https://jvet.hhi.fraunhofer.de/>

<sup>3</sup><https://www.webmproject.org/vp9/>

<sup>4</sup><https://aomedia.google.com/aom/>



Figure 3: Central views of the selected light field images.

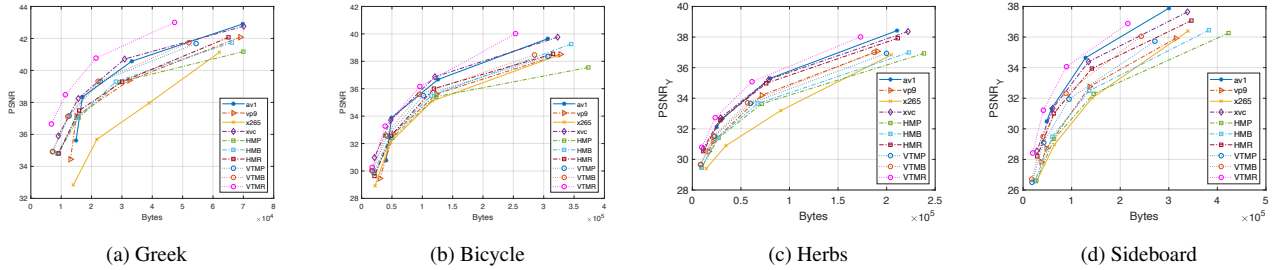


Figure 4: PSNR vs. Bytes.

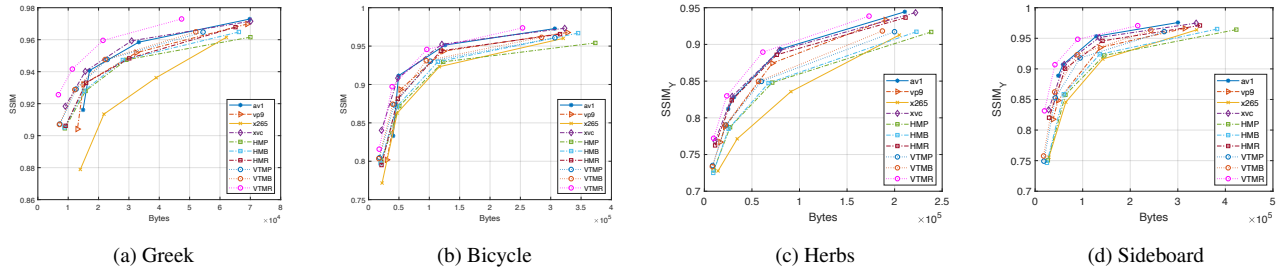


Figure 5: SSIM vs. Bytes.

based on MPEG standards like HEVC, but including some extra tools [18].

The second version of *xvc*<sup>5</sup> is used for the simulations. To set GOP to 16 like other selected codecs, *-sub-gop-length* is set to 16. To satisfy *IntraPeriod=-1* like other codecs, *-max-keypic-distance* is set to more than number of total frames (default 640 was used).

### x265 software

An open-source and commercial version of HEVC is known as *x.265* software. HEVC encoder version *3.0\_RC + 10 - 672ce0547e97* is selected for the simulations. Keyframe interval or GOP length is set to -1 to satisfy *IntraPeriod* parameter of the other software's.

Table 2: Selected settings for *x.265* encoder

```
--input-res <W x H> --preset placebo
--tune psnr --keyint -1
--fps <FrameRate> --qp=<QP>
```

<sup>5</sup><https://github.com/divideon/xvc>

## Implementation setup and Dataset

Four challenging light field images, Greek, Bicycle, Herbs, and Table from 4D Light Field Dataset [19] were selected to be encoded with the selected codecs based on expert review. They are  $9 \times 9$  views with  $512 \times 512$  pixels resolution. The central view of the selected images are shown in Figure 3. The light field image views are firstly converted to 8-bit YCbCr color format and then are downsampled to 4:2:0 chroma subsampling. The views are ordered using a serpentine scan to define a pseudo-video.

## Simulation results

Selected QPs for the encoders are listed in Table 3.

Table 3: Selected QPs for the different encoders.

	VTM-LP	VTM-LB	HM-LP	HM-LB	AV1	VP9	xvc	x265
QP1	25	25	25	25	38	37	25	28
QP2	31	31	31	31	51	51	30	35
QP3	36	36	36	36	62	62	35	40
QP4	41	41	41	41	63	63	40	44

Two objective metrics, PSNR and SSIM [20], were selected to evaluate the performance of the different encoders, based in the conclusions of [21]. Rate-Distortion curves for selected images

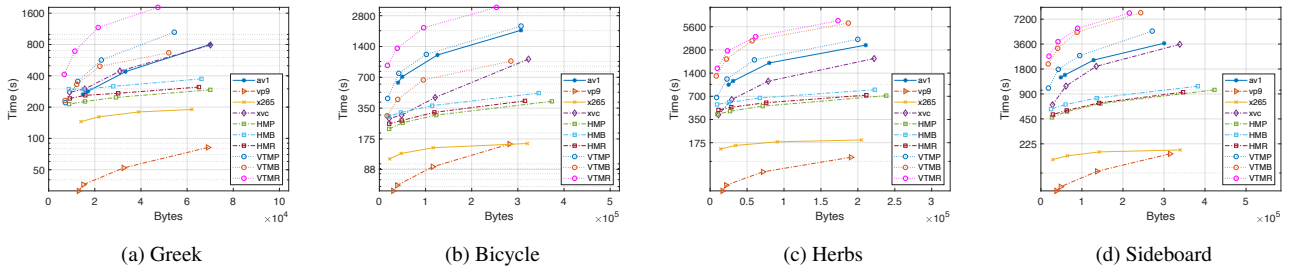


Figure 6: Encoding time-complexity.

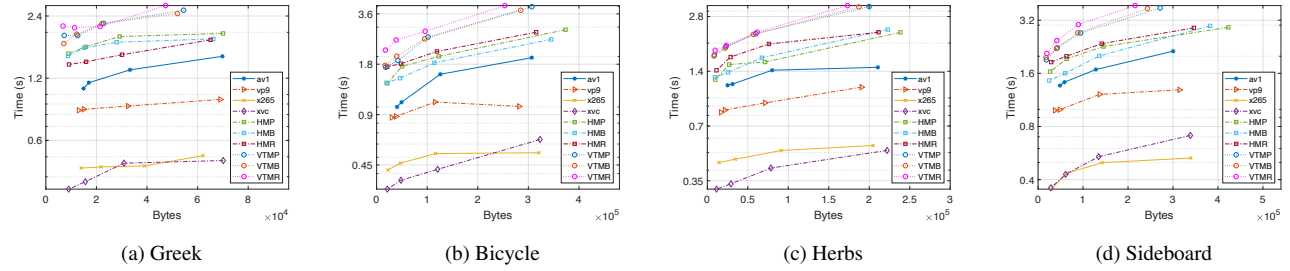


Figure 7: Decoding time-complexity.

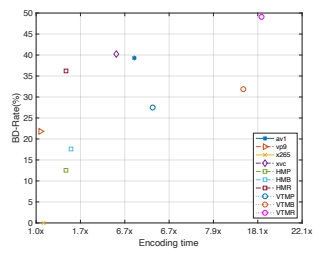


Figure 8: BD-Rate versus encoding time over x265

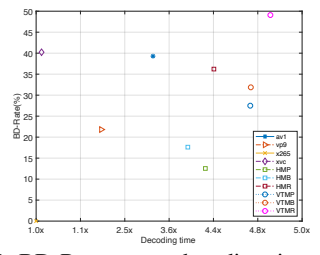


Figure 9: BD-Rate versus decoding time over xvc

are illustrated in Figure 4 and Figure 5 using PSNR and SSIM, respectively. As the x265 has the lowest performance, BD-BR and BD-PSNR Bjontegaard metrics were calculated against x265 using PSNR. Results are given in Table 4 and Table 5.

Although using the most recent compression tools might lead to better compression efficiency, they also increase the time-complexity. Encoding a light field image becomes a very challenging task especially for real-time applications. Encoder time-complexity might have less importance for streaming applications. However, both real-time and streaming applications demand low decoding complexity. Encoder and decoder time-complexities for both encoders and decoders were computed and the results are illustrated in Figure 6 and Figure 7, respectively. As x265 does not provide a decoder implementation, FFMPEG was

Table 4: BD-BR metric using PSNR.

	Greek	Bicycle	Herbs	Sideboard	Average
AV1	-49.1709	-45.9679	-52.5433	-48.7632	-39.2891
VP9	-33.0129	-19.7724	-39.7812	-16.4915	-21.8116
xvc	-53.3568	-45.9976	-56.0594	-45.7097	-40.2247
HMP	-30.3892	-1.7754	-28.5981	-1.9125	-12.5335
HMB	-32.8183	-10.1144	-32.8794	-12.2543	-17.6133
HM	-50.3435	-39.731	-53.6149	-37.371	-36.2121
VTMP	-41.8141	-23.9882	-41.0971	-30.5625	-27.4924
VTMB	-44.0025	-30.3989	-44.5448	-40.3917	-31.8676
VTM	-63.8174	-56.6733	-64.0256	-60.909	-49.0851

Table 5: BD-PSNR metric using PSNR.

	Greek	Bicycle	Herbs	Sideboard	Average
AV1	2.7535	2.038	2.2062	2.8139	1.9623
VP9	1.6704	0.69738	1.3821	0.71955	0.89388
xvc	2.8931	1.8962	2.1353	2.328	1.8505
HMP	1.4121	0.05382	0.85754	0.11665	0.48803
HMB	1.5431	0.32619	1.0124	0.5201	0.68035
HM	2.6713	1.5527	1.9788	1.775	1.5956
VTMP	2.0686	0.84081	1.3319	1.3672	1.1217
VTMB	2.2326	1.1241	1.5119	2.0098	1.3757
VTM	3.851	2.5925	2.6823	3.6445	2.5541

used to decode the x265 compressed light field images. Fig. 8 and Fig. 9 show BD-rate as a function of encoding and decoding time, respectively. It is important to observe that xvc reveals a high compression performance, while it keeps a low complexity on both coding and decoding.

### Conclusion

In this paper, the performance of the various state-of-the-art encoders for compressing light field images was compared. VTM, HM, x265, xvc, AV1 and VP9 were selected for the com-

parison. In terms of coding efficiency, VTM outperforms other encoders. However, its encoding time-complexity is the highest among encoders. Decoding time-complexity of VTM and HM is much higher than others making their applicability challenging.

## Acknowledgements

This research was funded by the Portuguese FCT-Fundação para a Ciência e Tecnologia and co-funded by FEDER-PT2020 partnership agreement under the project PTDC/EEI-PRO/2849/2014 - POCI-01-0145-FEDER-016693, under the project UIDB/EEA/50008/2020, PLive X-0017-LX-20, and by operation Centro-01-0145-FEDER-000019 - C4 - Centro de Competencias em Cloud Computing, cofinanced by the European Regional Development Fund (ERDF) through the Programa Operacional Regional do Centro (Centro 2020), in the scope of the Sistema de Apoio a Investigação Científica e Tecnológica - Programas Integrados de IC&DT.

## References

- [1] G. Wu, B. Masia, A. Jarabo, Y. Zhang, L. Wang, Q. Dai, T. Chai, and Y. Liu, "Light field image processing: An overview," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 7, pp. 926–954, Oct 2017.
- [2] Hadi Amirpour, Antonio Pinheiro, Manuela Pereira, and Mohammad Ghanbari, "Analysis of motion vectors and parallel computing in pseudo-sequence based light field image compression methods," in *Proc. SPIE 10752, Applications of Digital Image Processing XLII, 107520C*, 2018.
- [3] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the h.264/avc video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, July 2003.
- [4] F. Dai, J. Zhang, Y. Ma, and Y. Zhang, "Lenselet image compression scheme based on subaperture images streaming," in *2015 IEEE International Conference on Image Processing (ICIP)*, Sep. 2015, pp. 4733–4737.
- [5] Gary J. Sullivan, Jens-Rainer Ohm, Woo-Jin Han, and Thomas Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Cir. and Sys. for Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.
- [6] C. Perra and P. Assuncao, "High efficiency coding of light field images based on tiling and pseudo-temporal data arrangement," in *2016 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, July 2016, pp. 1–4.
- [7] H. P. Hariharan, T. Lange, and T. Herfet, "Low complexity light field compression based on pseudo-temporal circular sequencing," in *2017 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, June 2017, pp. 1–5.
- [8] Y. Y. Liu, C. Zhu, and M. Mao, "Light field image compression based on quality aware pseudo-temporal sequence," *Electronics Letters*, vol. 54, no. 8, pp. 500–501, 2018.
- [9] H. Amirpour, M. Pereira, and A. Pinheiro, "High efficient snake order pseudo-sequence based light field image compression," in *2018 Data Compression Conference*, March 2018, pp. 397–397.
- [10] Hadi Amirpour, Antonio M. G. Pinheiro, Manuela Pereira, and Mohammad Ghanbari, "Fast and efficient lenslet image compression," *CoRR*, vol. abs/1901.11396, 2019.
- [11] D. Liu, L. Wang, L. Li, Zhiwei Xiong, Feng Wu, and Wenjun Zeng, "Pseudo-sequence-based light field image compression," in *2016 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, July 2016.
- [12] S. Zhao, Z. Chen, K. Yang, and H. Huang, "Light field image coding with hybrid scan order," in *2016 Visual Communications and Image Processing (VCIP)*, Nov 2016, pp. 1–4.
- [13] T. Laude, Y. G. Adhisantoso, J. Voges, M. Munderloh, and J. Ostermann, "A comparison of jem and av1 with hevc: Coding tools, coding efficiency and complexity," in *2018 Picture Coding Symposium (PCS)*, June 2018, pp. 36–40.
- [14] Dan Grois, Tung Nguyen, and Detlev Marpe, "Performance comparison of av1, jem, vp9, and hevc encoders," 2018.
- [15] M. A. Layek, N. Q. Thai, M. A. Hossain, N. T. Thu, L. P. Tuyen, A. Talukder, T. Chung, and E. Huh, "Performance analysis of h.264, h.265, vp9 and av1 video encoders," in *2017 19th Asia-Pacific Network Operations and Management Symposium (APNOMS)*, Sep. 2017, pp. 322–325.
- [16] D. Grois, T. Nguyen, and D. Marpe, "Coding efficiency comparison of av1/vp9, h.265/mpeg-hevc, and h.264/mpeg-avc encoders," in *2016 Picture Coding Symposium (PCS)*, Dec 2016, pp. 1–5.
- [17] P. Akyazi and T. Ebrahimi, "Comparison of compression efficiency between hevc/h.265, vp9 and av1 based on subjective quality assessments," in *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, May 2018, pp. 1–6.
- [18] Jonatan Samuelsson and Per Hermansson, "Image compression with xvc," in *CVPR Workshops*, 2018.
- [19] Katrin Honauer, Ol Johannsen, Daniel Kondermann, and Bastian Goldluecke, "A dataset and evaluation methodology for depth estimation on 4d-light fields," in *Computer Vision – ACCV 2016*, Shang-Hong Lai, Vincent Lepetit, Ko Nishino, and Yoichi Sato, Eds. 2017, pp. 19–34, Springer International Publishing.
- [20] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004.
- [21] H. Amirpour, A. M. G. Pinheiro, M. Pereira, and M. Ghanbari, "Reliability of the most common objective metrics for light field quality assessment," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 2402–2406.

## Author Biography

Hadi Amirpour received his BSc degrees in electrical and biomedical engineering from Amirkabir University of Technology and IAU-South Tehran Branch, respectively. He pursued his MSc in electrical engineering at the K. N. Toosi University of Technology between 2011-2013. Currently, his research interests are on image processing and compression, video processing and compression, emerging 3D imaging technology and medical image analysis.

Antonio Pinheiro (M'99, SM'15) is Professor at U.B.I. (Universidade da Beira Interior), Covilha, Portugal and a researcher at I.T. (Instituto de Telecomunicacoes), Portugal. He received the BSc (Licenciatura) from I.S.T., Lisbon in 1988 and the PhD in Electronic Systems Engineering from University of Essex in 2002. Currently, his research interests are on Multimedia Quality Evaluation, Emerging 3D Imaging Technologies, and Medical Image Analysis. He the JPEG Communication Subgroup chair and senior member of IEEE.

Manuela Pereira received the 5-year B. S. degree in Mathematics and Computer Science in 1994 and the M. Sc. degree in Computational Mathematics in 1999, both from the University of Minho, Portugal. She

*received the Ph. D. degree in Signal and Image Processing in 2004 from the University of Nice Sophia Antipolis, France. Currently she is Associate Professor of the University of Beira Interior, Portugal and the Head of the Multimedia Signal Processing Group.*

*Mohammad Ghanbari (M'78, SM'97, F'01, LF14) is a professor of video networking at the University of Tehran, Iran and Emeritus Professor at the university of Essex, United Kingdom. He is best known for his pioneering work on two-layer video coding for ATM networks (which earned him an IEEE Fellowship in 2001). Prof. Ghanbari has authored or co-authored more than 700 journal and conference papers. On January 2014 he was honored to Life Fellow of IEEE.*

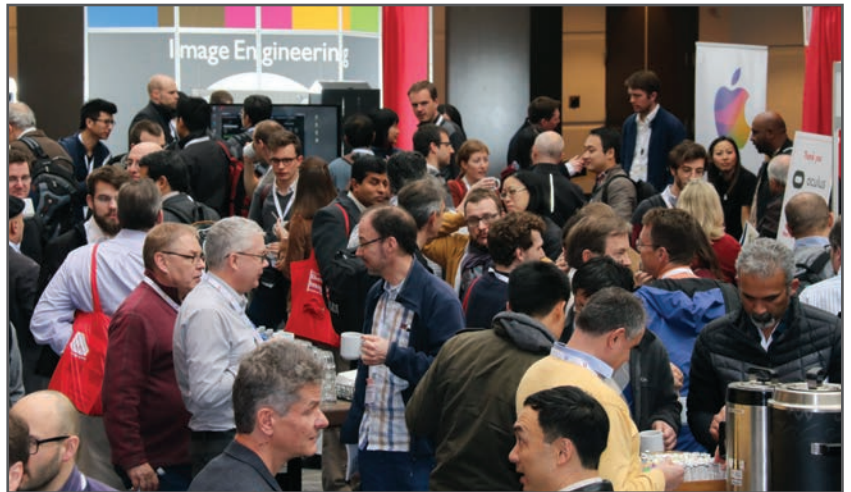
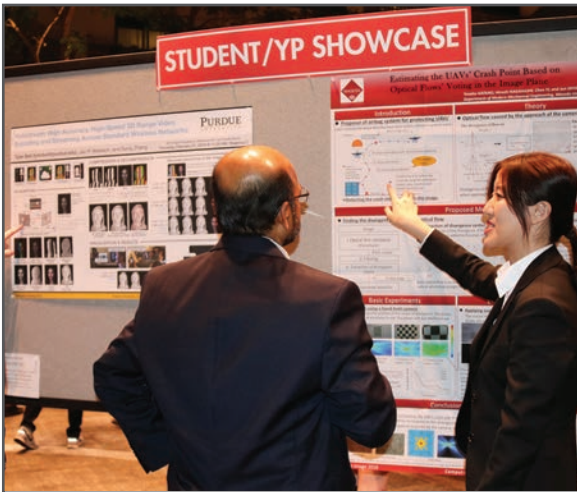
**JOIN US AT THE NEXT EI!**

IS&T International Symposium on

# Electronic Imaging

SCIENCE AND TECHNOLOGY

*Imaging across applications . . . Where industry and academia meet!*



- **SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •**
- **INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •**

[www.electronicimaging.org](http://www.electronicimaging.org)

