A comprehensive framework for visual quality assessment of light field tensor displays

Irene Viola; Multimedia Signal Processing Group (MMSPG), École Polytechnique Fédérale de Lausanne (EPFL); Lausanne, Switzerland

Keita Takahashi; School of Engineering, Nagoya University; Nagoya, Japan

Toshiaki Fujii; School of Engineering, Nagoya University; Nagoya, Japan

Touradj Ebrahimi; Multimedia Signal Processing Group (MMSPG), École Polytechnique Fédérale de Lausanne (EPFL); Lausanne, Switzerland

Abstract

In recent years, light field technology has attracted the interest of academia and industry, thanks to the possibility of rendering 3D scenes in a more realistic and immersive way. In particular, light field displays have been consistently investigated for their ability to offer a glass-free 3D viewing experience. Among others, tensor displays represent a promising way to render light field contents. However, only a few prototypes of such type of displays have been implemented and are available to the scientific community. As a direct consequence, the visual quality of such displays has not been rigorously investigated. In this paper, we propose a new framework to assess the visual quality of light field tensor displays on conventional 2D screens. The multilayer components of the tensor displays are virtually rendered on a typical 2D monitor through the use of a GUI, and different viewing angles can be accessed by simple mouse interactions. Both single and double stimulus methodologies for subjective quality assessment of light field contents are supported in this framework, while the total time of interaction is recorded for every stimulus. Results obtained in two different laboratory settings demonstrate that the framework can be successfully used to perform subjective quality assessment of different compression solutions for light field tensor displays.

Introduction

Light field photography has recently seen a surge of popularity due to the increased capabilities in acquiring and rendering real-life scenes in a more immersive way. In particular, light field rendering promises to overcome the limitations of stereoscopic representation by allowing for a more seamless transition between multiple point of views, thus giving a more faithful representation of 3D scenes. The recent innovations in the realm of acquisition and compression of light field contents have fueled the need for light field displays on which the data can be natively visualized.

In the past, several multi-view displays have been presented that allow to visualize the scene from several points of views. Among other technologies, parallax barrier displays have been proposed as a glass-free alternative to stereoscopic displays [8]. To allow for multi-view rendering, several technologies have been implemented, such as polarizer [14] or LCD dynamic barriers through viewer tracking [12]. Other approaches include the use of integral imaging [1], or a system of projectors used in combination with a holographic screen to fully render the 3D scene [2]. The Quality of Experience (QoE) associated with such type of light field displays has been recently investigated by Tamboli et

al. [17][18]. Another alternative uses a stack of programmable light-attenuating layers in front of a light-emitting source to provide depth cues without the need of glasses [10][11] [20]. The layer patterns to be displayed in each light-attenuating layers can be obtained from the multi-view light field data through Non-negative Tensor Factorization (NTF) [20]. Recently, a new method has been proposed to generate the layer patterns from a stack of focused images (focal stack), which greatly reduces the number of images that are needed as input for the tensor displays [16]. The method was tested on a prototype 3D display to prove its efficacy [9].

Testing the visual quality of compressed and uncompressed light field contents on native light field display is of extreme importance in future development of both new rendering methods, as well as new compression solutions. However, the limited availability of light field displays hinders the assessment of their visual quality. Moreover, hardware limitations in prototype models considerably lessen the perceptual quality of experience in consuming light field contents. Being able to simulate light field multi-layer rendering in a virtual environment is thus helpful in conducting evaluation of visual quality for light field displays in an ideal scenario.

In this paper, we present a framework to conduct quality assessment of light field contents rendered through a tensor display simulator using 2D screens. Through a GUI the layer patterns composing the multi-layer tensor displays are simulated in a 3D environment. By interacting with the mouse, users can experience the light field from different points of views. A validating experiment is conducted to show the validity of our framework.

Proposed framework

The framework proposed in this paper provides a tool to assess the quality of light field contents rendered through the use of multi-layer tensor displays. It consists of a software application for quality assessment of light field contents that enables visualization from different points of views, while keeping track of both the given ratings and the total time of interaction.

A graphical interface based on the software proposed in [15], simulates the multi-layer structure of light field tensor displays. The layer patterns are given as input to be directly visualized using the interface, along with a file specifying the parameters for the rendering, such as the horizontal and vertical angular resolution (i.e., the number of perspective views) of the input light field, and the number of layers composing the simulated display. The

Figure 1: Example rendering of the input stimuli with the proposed graphical interface, using double stimulus methodology with side-by-side display.

| (a) Bikes | (b) Danger de mort | (c) Stone Pillars Outside | (d) Fountain & Vincent | (e) Friends |

Figure 2: Central perspective view from each content used in the test.

layer patterns are always displayed in their original resolution. By clicking and dragging, users can physically alter the visualization angle of the simulated display on the screen, thus accessing different points of views. The viewing angles are limited by the number of layers and the angular resolutions of the input light field, to ensure only properly rendered points of view are accessible. In particular, denoting V_x and V_y as the number of perspective views of the input light field in the horizontal and vertical dimension, respectively, and L as the number of layers in the simulated display, the maximum viewing angle θ_x and θ_y in the horizontal and vertical direction, respectively, can be defined as such:

$$\theta_x = \arctan \frac{\alpha \left\lfloor \frac{V_x}{2} \right\rfloor}{L}, \tag{1}$$

$$\theta_y = \arctan \frac{\alpha \left\lfloor \frac{V_y}{2} \right\rfloor}{L}. \tag{2}$$

Parameter α depends entirely on the specifications of the 2D monitor used to display the simulation:

$$\alpha = \frac{\sqrt{W^2 + H^2}}{\sqrt{w^2 + h^2}}, \tag{3}$$

in which W and H represent the screen size in meters, while w and h represent the screen resolution in pixels.

The graphical interface has been adapted to be used for subjective quality assessment. In accordance with the ITU-R recommendations [6], both single and double stimulus methodologies can be used for the subjective evaluation. For the double stimulus methodology, both side-by-side and consecutive presentations are available. In the former case, the two stimuli are presented simultaneously on the screen, and any change in viewing angle is rendered in a synchronized way, to allow users to visualize both contents from the same point of view. Conversely, in the consecutive presentation only one stimulus is presented at a time. By using the arrow keys on the keyboard, users can switch between two stimuli. The switching can happen at any viewing angle the user has chosen, thus allowing to compare the two stimuli at any point of view. A mid-grey color has been selected for the environment surrounding the simulated display, in accordance with the ITU-R recommendations [6]. An example rendering from the graphical interface is presented in Figure 1.

Once users are satisfied with their viewing experience of the content, they can score the stimuli using the keyboard. All the scores are saved in an output file. The total time each stimulus was visualized is recorded in a separate file, to be used in analyzing interaction patterns and user behavior [19].

Validating experiment

In this section we will describe the subjective quality experiment we conducted to validate our proposed framework. In particular, we first list the dataset and the coding conditions. We then describe the environment in which the tests were conducted, as well as the employed methodology. Finally, we give an overview

of the statistical analysis conducted on the gathered data.

Dataset and coding conditions

Five light field contents were selected from a publicly available database [13]. The contents were acquired with a Lytro Illum camera and processed using the Light Field Matlab Toolbox [3][4] to obtain a stack of 15×15 perspective image, each having a resolution of 625×434 pixels. Color and gamma corrections were applied on each perspective image for the rendering. To avoid unwanted distortions caused by the lenslet structure of the Lytro Illum camera, only the 9×9 central perspective views were selected for the test. The central perspective view from each content is displayed in Figure 2.

Considering the peculiarities of our rendering system, three viable alternatives for light field compression were employed. The first arranges the 9×9 perspective views in a pseudo-temporal video sequence, which is subsequently encoded. The layer patterns needed for rendering are then created at the receiver side, after decoding the compressed views. The second creates the layer patterns at the encoder side; such layer patterns are arranged in a pseudo-temporal video sequence and compressed. At the receiver side, the decoded layer patterns can be directly rendered without further processing. Finally, the third solution creates a focal stack of refocused images from the perspective views. The focal stack is then arranged in a pseudo-temporal sequence, compressed and transmitted. At the receiver side, the layer patterns are created from the focal stack. For all three solutions the state-of-the-art video encoding standard HEVC was employed for the compression, to ensure a fair comparison.

The layer patterns were created using the software implementation in [15]. To create the focal stack, the Light Field Matlab Toolbox was employed [3][4]. In our validating test, the number of layers was fixed to $L = 3$.

The compression solutions were evaluated at four bit-rates, namely $R1 = 537$ kB, $R2 = 134$ kB, $R3 = 67$ kB, and $R4 = 27$ kB, corresponding to 0.2, 0.05, 0.025 and 0.01 bit per pixel (bpp), respectively. The bpp are computed with respect to the original size of the 9×9 perspective views. The bit-rates were carefully chosen to cover the visual quality space while providing reasonable and fair comparison among the listed compression solutions.

Subjective quality assessment methodology and test environment

For our validating experiment, the Double Stimulus Impairment Scale (DSIS) with side-by-side presentation and 5-point grading scale (*5-Imperceptible, 4-Perceptible but not annoying, 3-Slightly annoying, 2-Annoying, 1-Very annoying*) was selected, according to the ITU-R Recommendation BT.500-13 [6].

Participants were asked to rate the quality of the test stimuli when compared to the uncompressed reference. They were informed beforehand on which side of the screen the reference would be displayed, and its position on the screen was fixed for the duration of the test. In order to accustom the participants with what distortions to expect in the test images, a training session was organized before the experiment. Three training samples, created by compressing one additional content on the test bit-rates, were manually selected by expert viewers.

All the compressed stimuli were shown in one session. Additionally, two hidden references per content were added to the

test: one consisted in the layer patterns generated from the uncompressed stack of perspective views, while the other was created from the uncompressed focal stack. A total of 70 stimuli were evaluated in each session.

The test was conducted in two different settings to test its validity, in the facilities of the École Polytechnique Fédérale de Lausanne (EPFL) and Nagoya University (NU). In EPFL, a laboratory for subjective video quality assessment, which was set up according to ITU-R Recommendation BT.500-13 [6], was used for the test. A 27-inch Apple Display with native resolution of 2560×1440 pixels was used. The monitor settings were adjusted according to the following profile: sRGB Gamut, D65 white point, 120 cd/m^2 brightness, and minimum black level of 0.2 cd/m^2. The controlled lighting system in the room consisted of adjustable neon lamps with 6500 K color temperature against mid-grey background walls. The illumination level measured on the screens was 18 lux. Conforming to requirements in ITU-R Recommendation BT.2022 [5], the distance of the subjects from the monitor was approximately equal to 7 times the height of the displayed content. However, subjects were allowed to move further or get closer to the screen. A total of 20 subjects (10 males and 10 females) participated in the first test, amounting to 20 scores per stimulus. Subjects were between 18 and 35, with a mean age of 23.29 years old. Before starting the test, all subjects were examined for visual acuity and color vision using Snellen and Isihara charts, respectively.

The second test, using the same stimuli, was conducted in NU in a controlled environment, using the same 27-inch Apple Display. However, no calibration on the brightness settings of either the monitor or the room was conducted. A total of 17 subjects (16 males and 1 female) took part in the second test. Subjects were between 18 and 35, with a mean age of 24 years old. Before starting the test, all subjects were examined for visual acuity and color vision using Snellen and Isihara charts, respectively.

Statistical analysis

Outlier detection and removal was performed on the results, independently for each test setting, according to the ITU-R Recommendations [6]. No outlier was detected in either batch of scores. After outlier removal, the mean opinion score (MOS) was computed for each stimulus, independently for each methodology. The corresponding 95% confidence intervals (CIs) were computed assuming a Student's t-distribution.

Following the ITU-T Recommendations [7], several fittings were applied to the MOS values from the two different test settings. In particular, first order and third order fittings were used to compare the MOS values. Absolute prediction error (RMSE), Pearson Correlation Coefficient (PCC), Spearman's Rank Correlation Coefficient (SRCC) and Outlier Ratio (OR) were computed for accuracy, linearity, monotonicity and consistency, respectively.

Results

Figures 3 and 4 show the results of the subjective quality assessment tests conducted in EPFL and NU, respectively, for each content under test. Results corresponding to the two hidden reference are shown in shaded colors. The graphs show that operating the compression on the layer patterns seems to be the preferable solution, as its performance is either statistically equivalent
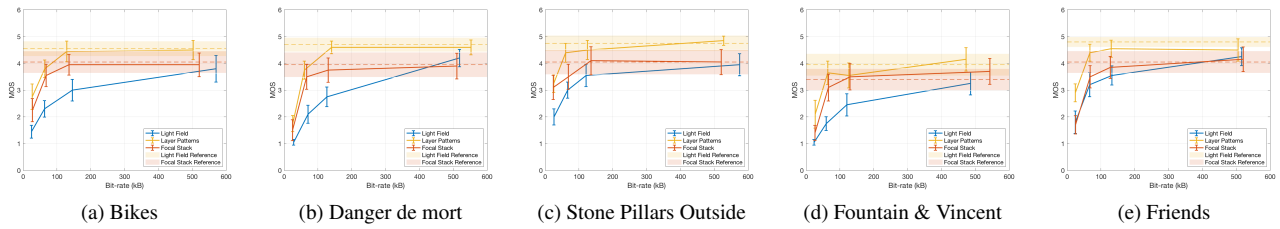
(a) Bikes   (b) Danger de mort   (c) Stone Pillars Outside   (d) Fountain & Vincent   (e) Friends

Figure 3: MOS vs bit-rate with respective CIs, for all contents under test (results obtained in EPFL laboratory).

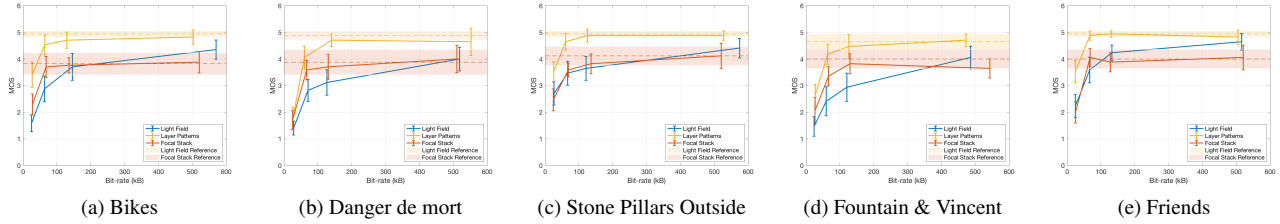(a) Bikes   (b) Danger de mort   (c) Stone Pillars Outside   (d) Fountain & Vincent   (e) Friends
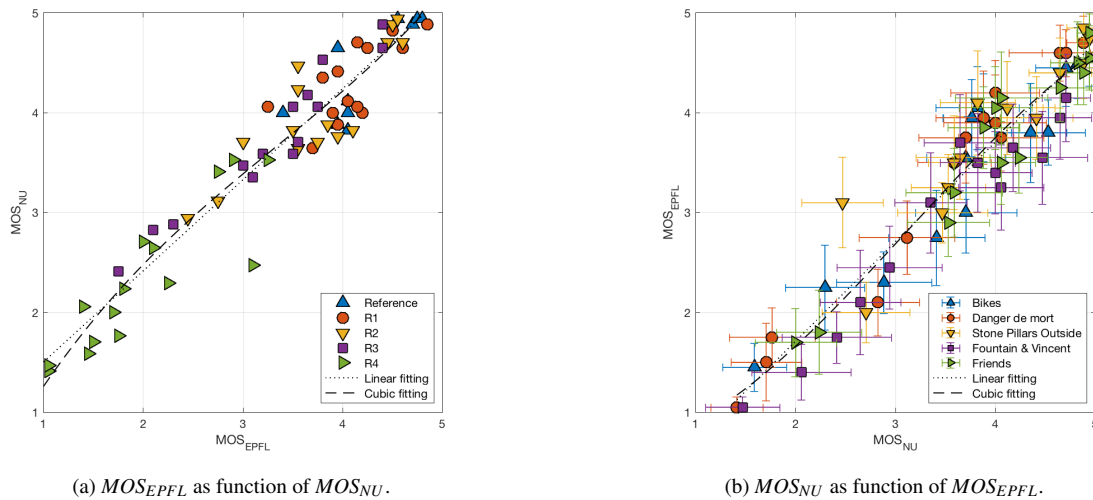
Figure 4: MOS vs bit-rate with respective CIs, for all contents under test (results obtained in NU laboratory).

(a) MOS_{EPFL} as function of MOS_{NU}.

(b) MOS_{NU} as function of MOS_{EPFL}.

Figure 5: Comparison of MOS values obtained in the different test settings, along with linear and cubic fittings. Points are differentiated by compression ratio (a) and by content (b).

or better than the other solutions at all bitrates. It is particularly interesting to see that compressing the focal stack achieves significantly worse results when compared to the compression of layer patterns, despite the improved coding efficiency. The results can be explained by comparing the MOS obtained by the two references: the layer patterns generated from the uncompressed focal stack never reaches transparent quality, and it is always performing worse than the layer patterns generated from the perspective views forming the light field. However, the focal stack compression seems to be working better in some cases when compared to light field compression, at least for low bitrates.

Figure 5 depicts the scatter plot showing the results of the comparison between the MOS scores obtained in the two test settings. In Figure 5 (b), the horizontal and vertical bars represent the CIs corresponding to results obtained in NU and EPFL, respec-

tively. To improve visualization, the points are colored based on compression ratio or content. Linear and cubic fittings are shown for both comparisons. As clearly shown in Figure 5, the scores obtained in the two test settings are strongly correlated. In particular, linear regression performed on the $[\widehat{MOS}_{NU}, MOS_{EPFL}]$ pair reports a slope of 0.9974 and an intercept of -0.2830, which indicates that while a strong correlation can be seen between the scores obtained in the two test settings, ratings obtained in EPFL are consistently lower than their NU counterpart. Such behaviour can be attributed to the difference in population among the two test settings, as well as the influence of the uncontrolled lighting conditions in the NU test with respect to EPFL.

Table 1 reports the results of the performance indexes computed on the data. In particular, the performance indexes are computed for every pair of $\widehat{MOS}_X, MOS_Y$, in which X and Y denote

Table 1: Performance indexes.

	$[\widehat{MOS}_{EPFL}, MOS_{NU}]$				$[\widehat{MOS}_{NU}, MOS_{EPFL}]$			
	PCC	SRCC	RMSE	OR	PCC	SRCC	RMSE	OR
No fitting	0.9542	0.9221	0.4230	4.29%	0.9542	0.9221	0.4230	4.29%
Linear fitting	0.9542	0.9221	0.2904	1.43%	0.9542	0.9221	0.3035	1.43%
Cubic fitting	0.9558	0.9221	0.2856	1.43%	0.9551	0.9221	0.3008	1.43%

the different test settings, and $\widehat{MOS}$ represents the MOS scores obtained after linear and cubic fitting. Results of the performance indexes computed on the MOS pairs confirm the strong correlation between the scores obtained in the two test settings. In particular, cubic regression seems to give the best results among the fittings applied to the MOS pairs, with a PCC index of 0.9558 and 0.9551 for the $[\widehat{MOS}_{EPFL}, MOS_{NU}]$ and $[\widehat{MOS}_{NU}, MOS_{EPFL}]$ pair, respectively.

Conclusions

In this paper we presented a framework for conducting quality assessment of light field contents on simulated multi-layer tensor displays. We present a novel graphical user interface to conduct subjective quality assessment using multi-layer-based rendering on common 2D screens. We then perform a validating experiment in two separate test laboratories, showing the high correlation between the scores obtained in the two test settings. Results show that our proposed framework can be effectively used to perform quality assessment for light field contents.

The software application can be found at the following link: `https://github.com/mmspg/LFDisplaySimulator`. It is free to use, modify or redistribute, according to the MIT license. In case of use of the software for any purpose, publishing or use of any updates and variations based on it, as well as when presenting and publishing results based on the software, a reference to this paper should be provided.

Acknowledgments

References

[1] Jun Arai, Fumio Okano, Masahiro Kawakita, Makoto Okui, Yasuyuki Haino, Makoto Yoshimura, Masato Furuya, and Masahito Sato. Integral three-dimensional television using a 33-megapixel imaging system. *Journal of Display Technology*, 6(10):422–430, 2010.

[2] Tibor Balogh. The Holovizio system. In *Stereoscopic Displays and Virtual Reality Systems XIII*, volume 6055, page 60550U. International Society for Optics and Photonics, 2006.

[3] Donald G. Dansereau, Oscar Pizarro, and Stefan B. Williams. Decoding, calibration and rectification for lenselet-based plenoptic cameras. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2013.

[4] Donald G. Dansereau, Oscar Pizarro, and Stefan B. Williams. Linear volumetric focus for light field cameras. *ACM Transactions on Graphics (TOG)*, 34(2), Feb. 2015.

[5] ITU-R BT.2022. General viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays. International Telecommunication Union, August 2012.

[6] ITU-R BT.500-13. Methodology for the subjective assessment of the quality of television pictures. International Telecommunication Union, January 2012.

[7] ITU-T P.1401. Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models. International Telecommunication Union, July 2012.

[8] Adrian Jacobs, Jonathan Mather, Robert Winlow, David Montgomery, Graham Jones, Morgan Willis, Martin Tillin, Lyndon Hill, Marina Khazova, Heather Stevenson, et al. 2D/3D switchable displays. *Sharp Technical Journal*, pages 15–18, 2003.

[9] Yuto Kobayashi, Shu Kondo, Keita Takahashi, and Toshiaki Fujii. A 3-D display pipeline: Capture, factorize, and display the light field of a real 3-D scene. *ITE Transactions on Media Technology and Applications*, 5(3):88–95, 2017.

[10] Douglas Lanman, Matthew Hirsch, Yunhee Kim, and Ramesh Raskar. Content-adaptive parallax barriers: optimizing dual-layer 3D displays using low-rank light field factorization. In *ACM Transactions on Graphics (TOG)*, volume 29, page 163. ACM, 2010.

[11] Douglas Lanman, Gordon Wetzstein, Matthew Hirsch, Wolfgang Heidrich, and Ramesh Raskar. Beyond parallax barriers: applying formal optimization methods to multilayer automultiscopic displays. In *Stereoscopic Displays and Applications XXIII*, volume 8288, page 82880A. International Society for Optics and Photonics, 2012.

[12] Tom Peterka, Robert L Kooima, Daniel J Sandin, Andrew Johnson, Jason Leigh, and Thomas A DeFanti. Advances in the dynallax solid-state dynamic parallax barrier autostereoscopic visualization display system. *IEEE transactions on visualization and computer graphics*, 14(3):487–499, 2008.

[13] Martin Řeřábek and Touradj Ebrahimi. New light field image dataset. *8th International Conference on Quality of Multimedia Experience (QoMEX)*, 2016.

[14] Kunio Sakamoto and Tsutomu Morii. Multiview 3D display using parallax barrier combined with polarizer. In *Advanced Free-Space Optical Communication Techniques/Applications II and Photonic Components/Architectures for Microwave Systems and Displays*, volume 6399, page 63990R. International Society for Optics and Photonics, 2006.

[15] Keita Takahashi. Light field display project. Available at `http://www.fujii.nuee.nagoya-u.ac.jp/~takahasi/Research/LFDisplay/`.

[16] Keita Takahashi, Yuto Kobayashi, and Toshiaki Fujii. From focal stack to tensor light-field display. *IEEE Transactions on Image Pro-

cessing, 27(9):4571–4584, 2018.

[17] Roopak R Tamboli, Balasubramanyam Appina, Sumohana Channappayya, and Soumya Jana. Super-multiview content with high angular resolution: 3D quality assessment on horizontal-parallax lightfield display. *Signal Processing: Image Communication*, 47:42–55, 2016.

[18] Roopak R Tamboli, Balasubramanyam Appina, Peter A Kara, Maria G Martini, Sumohana S Channappayya, and Soumya Jana. Effect of primitive features of content on perceived quality of light field visualization. In *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–3. IEEE, 2018.

[19] Irene Viola and Touradj Ebrahimi. A new framework for interactive quality assessment with application to light field coding. In *Applications of Digital Image Processing XL*, volume 10396, page 103961F. International Society for Optics and Photonics, 2017.

[20] Gordon Wetzstein, Douglas Lanman, Matthew Hirsch, and Ramesh Raskar. Tensor displays: compressive light field synthesis using multilayer displays with directional backlighting. 2012.

Author Biography

Irene Viola received her B.Sc. in Cinema and Media Engineering from the Polytechnic University of Turin, Italy, in 2013, which was followed by a M.Sc. degree in Computer Engineering from the Polytechnic University of Turin, Italy, in 2015. Since March 2016 she is pursuing a PhD in the Multimedia Signal Processing Group (MMSPG), under the supervision of Professor Touradj Ebrahimi, at the Swiss Federal Institute of Technology in Lausanne (EPFL), Switzerland. Her research interest includes image and video processing, compression and evaluation, with a focus on light field coding.

Keita Takahashi received the B.E., M.S., and Ph.D. degrees in information and communication engineering from The University of Tokyo in 2001, 2003, and 2006, respectively. He was a Project Assistant Professor with The University of Tokyo from 2006 to 2011 and an Assistant Professor with the University of Electro-Communications from 2011 to 2013. He is currently an Associate Professor with the Graduate School of Engineering, Nagoya University, Japan. His research interests include computational photography, image-based rendering, and 3D displays.

Toshiaki Fujii received his B.E., M.E., and Dr.E. degrees in electrical engineering from the University of Tokyo in 1990, 1992, and 1995, respectively. He is currently a Professor at the Graduate School of Engineering, Nagoya University, Japan. His current research interests include multi-dimensional signal processing, multi-view video coding and transmission, and 3D imaging system based on light field acquisition and display. Prof. Fujii is a member of IEEE Signal Processing Society.

Touradj Ebrahimi is currently Professor at EPFL heading its Multimedia Signal Processing Group. He is also the Convener of JPEG standardization Committee. His research interests include still, moving, and 3D image processing and coding, visual information security (rights protection, watermarking, authentication, data integrity, steganography), new media, and human computer interfaces (smart vision, brain computer interface). He is the author or the co-author of more than 200 research publications, and holds 14 patents.