SMART: a Light Field image quality dataset

Pradip Paudyal
Roma TRE University
via Vito Volterra, 62, 00146
Rome, Italy
pradip.paudyal@uniroma3.it

Roger Olsson
Mid Sweden University
Holmgatan 10, 851 70
Sundsvall, Sweden
roger.olsson@miun.se

Mårten Sjöström
Mid Sweden University
Holmgatan 10, 851 70
Sundsvall, Sweden
marten.sjostrom@miun.se

Federica Battisti
Roma TRE University
via Vito Volterra, 62, 00146
Rome - Italy
federica.battisti@uniroma3.it

Marco Carli
Roma TRE University
via Vito Volterra, 62, 00146
Rome - Italy
marco.carli@uniroma3.it

ABSTRACT

In this contribution, the design of a Light Field image dataset is presented. It can be useful for design, testing, and benchmarking Light Field image processing algorithms. As first step, image content selection criteria have been defined based on selected image quality key-attributes, i.e. spatial information, colorfulness, texture key features, depth of field, etc. Next, image scenes have been selected and captured by using the Lytro Illum Light Field camera. Performed analysis shows that the proposed set of images is sufficient for addressing a wide range of attributes relevant for assessing Light Field image quality.

CCS Concepts

•**Information systems** → *Multimedia databases;*

Keywords

Light Field imaging, Plenoptics, Quality of Experience, Content Attributes, SMART Light Field Dataset

1. INTRODUCTION

The Light Field (LF) imaging is considered a next generation imaging technology. The basic concept was first introduced by Lippmann [8] in 1908 as integral photography and improved by many researchers throughout the years [1] [21]. LF imaging is based on a camera recording information about the intensity of light in a scene and the direction in the space of the light rays. Basically, the LF cameras record multiple views of a scene by using a single camera in a single shot, thus reducing the problems related to camera synchronization. This can be achieved thanks to the presence of a micro lens array that allows to record information on the incident light direction at different positions in multiple micro-images. The most appealing feature of the LF cameras is that even a single LF snapshot can provide pictures where focus, exposure, and depth of field can be adjusted after the picture is taken. Therefore, LF cameras open new possibilities in many applications such as photography, astronomy, robotics, medical imaging, and microscopy fields [19].

However, LF imaging system demands very large computational power and presents resolution and image quality issues. The rapidly developing LF technology and consumer/industry/academic interest towards the technology is also pushing the need for the evaluation of the quality of such content. Moreover, the several possible applications of the LF images require the understanding of Quality of Experience (QoE) from different points of view. LF images are subject to several distortions during the acquisition, processing, compression, storage, transmission and reproduction phases. Any of these stages results in a degradation of visual quality, thus pushing the need for evaluation of the quality of such a content. The availability of ground truth information, test image contents and annotated subjective scores, are important and useful tools, needed for training, testing, and benchmark the processing algorithms [33] [7].

In the literature, few LF datasets have been proposed. The main features are reported in Table 1. Stanford LF Archive [30] is widely used; however, the images are captured by using a multi-camera system including gantry, microscope, etc. Nowadays, different LF cameras have been realized, (e.g. Lytro [21], Lytro Illum, and Raytrix [28]), thus allowing the consumers to exploit such a technology. Lytro Illum is the new version of the Lytro plenoptic camera, characterized by increased resolution and processing capabilities, while Raytrix is a so called focused plenoptic camera. As can be noticed, the dataset [30] will not be sufficient to deal with new challenges, perceptual quality evaluation, performance testing for processing algorithms, etc., which arose with the advancement of the LF technology.

Other recently proposed datasets [32] [31] [9] [20] [17] [26] have been designed for specific purposes and the images have been acquired by the Lytro plenoptic camera. In the dataset [29], the Lytro Illum camera has been used. However, most of the images have similar features and motivations behind the particular image content selection have not been reported.

MMSys'16, May 10-13, 2016, Klagenfurt, Austria
© 2016 ACM. ISBN 978-1-4503-4297-1/16/05...$15.00
DOI: http://dx.doi.org/10.1145/2910017.2910623

Table 1: **Most relevant datasets with corresponding features.**

Datasets	Year	Purposes	Features	Acquisition Devices	DM	Remarks
Stanford Light Field Archive [30]	2008	General	More than 20 images	Gantry, LF Microscope, and Camera Array	No	Old
Synthetic Light Field Archive [32]	2013	Compression	more than 17 light field images, includes transparencies, occlusions and reflections, etc.	Camera (Artificial LF)	No	non-natural
Light Field Analysis [31]	2013	Depth Map	7 Blender and 6 Gantry images, does not cover the wide range of natural scenes.	Blender Software and Gantry	Yes	Specific purpose
EPFL Light-Field Image Dataset [29]	2015	General	More than 117 images with different categories: buildings, landscapes, etc.	Lytro Illum	No	Wide Range
LCAV-31 [9]	2014	Object Recognition	More than 31 images	Lytro	No	Specific purpose
Lytro dataset [20]	2015	LF Reconstruction	30 images with indoor, outdoor, motion blur, etc.	Lytro	No	Specific Purpose
Light Field Saliency Dataset (LFSD) [17]	2014	saliency map estimation	more than 100 LF images	Lytro	Yes	Particular for saliency
GUC Light Field Face and Iris Dataset [26]	2016	face and iris Recognition	112 subjects for faces and 55 subjects for eye pattern	Lytro	NO	Particular for biometric

1.1 Motivation

The motivations behind this work are briefly reported in the following:

- From the state of the art survey it results the need of a comprehensive and well defined LF image dataset;
- The carefully selection of LF images is important for the effectiveness of a test dataset. The selected Source Sequences (SRCs) should cover a wide range of content variation, since perceived image quality is significantly influenced by the image content [23];
- In many applications and during pilot-test phases, it is desirable to have a reduced set of SRCs, especially if considering the computational cost of processing LF data.

1.2 Contribution

In this article, a LF image dataset is proposed. The dataset creation methodology, description of LF images, and analysis of LF image content is tailored. In brief:

- the SRCs image content selection criteria is defined,
- a comprehensive LF image quality dataset is proposed and made freely available to the research community [2],
- a spatial information estimation metric is exploited,
- an analysis of the features of the proposed dataset is provided.

The dataset design methodology, adopted in our work, can be used as a guideline for the creation of LF image/video quality datasets. Here, image content is selected based on a defined selection criteria. The dataset has a small number of images but it covers a wide range and dynamics of content related features. The dataset can be used for a variety of applications including design and testing of LF image processing techniques; encoding and refocusing, and LF image QoE assessment.

The rest of the paper is organized as follows: Section 2 briefly describes the adopted methodology (image content selection criteria, number of images, and image acquisition device). Section 3 presents a brief description of the proposed database, while the database analysis results are summarized in Section 4, and finally in Section 5 concluding remarks are drawn.

2. METHODOLOGY

In this section, scene/image contents, required number of SRCs, and acquisition device are discussed.

2.1 Content Selection

Several efforts [24] [14] [33] [6] have been made for image contents classification. In brief, many *low level* image features (i.e., contrast, brightness, edges) can be used resulting in very large number of image content clusters [24]. At the same time, the classification of image content based on *high level* features (i.e., indoor, outdoor) is complex, since high level features can be considered as a combination of low level ones. Therefore, there is no standard procedure for image content definition [14]. In this context, in [33]

Table 2: Considered key quality attributes

Features	Catagory	key importance
Spatial Information (SI)	General	perceptual indicator of spatial information of the scene
Colorfulness (CF)	General	perceptual indicator of naturalness of the images
Texture	General	perceptual indicator of human fixations
Depth of Field (DoF)	LF specific	refocusing applications
Transparency	LF specific	LF camera capability
Reflection	LF specific	LF camera capability

the authors propose spatial and colorfulness information for image content analysis, and a survey of available image quality datasets is presented based on these features. Similarly, image content is explained with the help of color, texture, shape, position, and dominant edges of image objects and regions in [6].

It is useful to underline that the LF camera can record the angle dependent information [3], thus providing information about *depth dependence* and *Lambertian lighting*. Depth dependence implies multiple depth in semitransparent objects, and Lambertian lighting is possible due to the capture of different angular information. Therefore, the inclusion of LF images characterized by transparency and reflection is an important feature in LF image analysis. Furthermore, another LF images key feature is Depth of Field (DoF), generally used in refocusing applications.

Based on widely used image attributes and Human Visual System (HVS), we have selected a set of key features, shown in Table 2, which are categorized into general and LF specific capabilities/attributes. The details about the selected general image quality attributes are reported in Section 4.1.

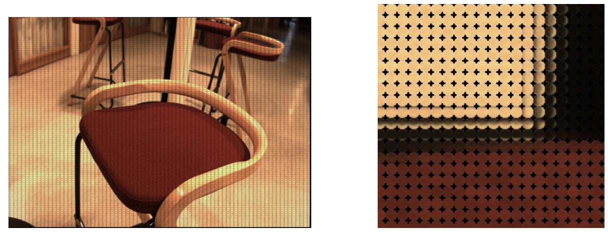
2.2 Dataset cardinality

The number of SRCs needed for a test dataset depends on the required image content diversity and on the possible applications. The proposed dataset will be used for several applications including design and testing of LF image processing techniques and image-related QoE assessment methodologies. To be effective, in quality assessment, the relative quality score in Just Noticeable Differences (JNDs) must be computed based upon data collected from a minimum of ten observers and three scenes [16]. Therefore, the number of SRCs is based on selected key-quality attributes and the number of potential attributes that are covered by a single image. In particular, the number of SRCs is determined based on the key image quality attributes and each attribute must be present in at least three images.

2.3 Image Acquisition Device

In this work, Lytro Illum camera has been used.

Finally, based on the key LF image quality (general and LF specific) attributes image content is selected. The selections of the scenes will help to characterize the impacts for



(a)Sample LF image (b) Close-up

Figure 2: Encoded LF image.

quality attributes [15], and ensure that the observers will examine the wide variety of attributes during the quality evaluation process.

3. DATASET DESCRIPTION

A total number of 15 LF images have been included in the SMART LF dataset. The all-focused 2D views of the images are shown in Figure 1 and the corresponding features are reported in Table 3. One single image can be characterized by many quality attributes; however in the table only the selected key attributes are reported. The images have been selected in such a way that a limited number of images can cover a large number of features and categories. The images are from both indoor and outdoor category, and cover not only the general image content related features but also the LF specific capabilities, reflection, transparency, and DoF variation.

The dataset has the following contents:

Raw LF image content: For each image, the LF data and the relative camera specific calibration data are provided. The calibration data is needed to decode LFR files. Moreover, the depth map, which is extracted by using *Lytro Desktop Software*, information for each images is provided.

Processed LF image content: The recently proposed LF image/video compression methods [4] [18] exploit microlens LF image (as shown in Figure 2) for compression. To create a composite LF image, the microlens images are assembled together and overall image is considered as a single LF image. For this purpose, recorded LF images are decoded by Light Field Toolbox for Matlab [5]. The sampled RGB LF image of size 6510x9390 (with 15x15 micro lens size) is down sampled to 8 bit image and transformed into YUV 4:2:0 format. The available decoded LF images, in both RGB and YUV format, are available in the dataset, and it can be directly used for encoding, rendering and quality evaluation purposes.

4. ANALYSIS AND DISCUSSIONS

In this section, the analysis of the proposed dataset is presented.

4.1 Objective Attributes

In this section, general content related features, colorfulness, spatial distribution, and textural information are considered to analyze the dataset.

Figure 1: All-focused 2D views of the LF images from the SMART LF dataset.

Table 3: Images available in the SMART LF dataset and their brief description with corresponding key features coverage.

S. No.	Name	Description	Key Features	Remarks
(a)	Tile	Tile with background building	Energy, Textures	Outdoor
(b)	Table	Table with sofa	Colorfulness, Correlation	Indoor
(c)	Chair	Chair on the floor	Colorfulness, DoF	Indoor
(d)	Flower	Flower with tile on the floor	SI, texture	Outdoor
(e)	Sky	Sky with natural scenes	Homogeneous, correlation	Outdoor
(f)	Grid	Grid with natural scenes	Depth distribution, grid	Outdoor
(g)	River	Flower and river with reflection of the building	Contrast, DoF	Outdoor
(h)	Building	Building and its reflection on the river	SI, contrast, reflection	Outdoor
(i)	Car	Car roof and building with sky	Homogeneity, DoF	Outdoor
(j)	Stone	Stone on the concrete ground	SI, contrast	Outdoor
(k)	Window	Natural outdoor scene with indoor objects	Energy, Transparency, DoF	Outdoor
(l)	Pilers	Pilers with wood and light	Colorfulness, DoF	Indoor/Outdoor
(m)	Book	Book inside a transparent box	Homogeneous, Transparency	Indoor/Outdoor
(n)	Person	Close-up picture of a person with reflection	Reflection, contrast	Indoor
(o)	White Sky	Natural scene with white sky	Energy, correlation	Outdoor

4.1.1 Colorfulness (CF)

CF is the main perceptual attribute underlying the image perceptual quality and the naturalness of the images. As a perceptual indicator of the variety and intensity of colors in the image, the colorfulness metric is used [11]. CF is computed by

$$M_{CF} = \sigma_{rgyb} + 0.3\mu_{rgyb}, \qquad (1)$$

where, M_{CF} is colorful metric, $\sigma_{rgyb} = \sqrt{\sigma_{rg}^2 + \sigma_{yb}^2}$, $\mu_{rgyb} = \sqrt{\mu_{rg}^2 + \mu_{yb}^2}$, $rg = R - G$, and $yb = 0.5(R+G) - B$, σ is the standard deviation and μ is the mean value. The R, G, and B are the red, green and blue color channels of the image pixels.

4.1.2 Spatial Information (SI)

As a perceptual indicator of the spatial information of the scene, SI is used [22]. The SI filter [12] is proposed by Institute for Telecommunication Science (ITS) to estimate the image spatial information. The filter is similar to the classical Sobel filter, where separate horizontal and vertical filters are applied, then the total edge energy is computed as the Euclidean distance.

In this work, an *SI-based metric* for images is adopted. It is based on the ITU recommended SI metric [13] for videos. The metric is slightly modified by considering only one frame of the video to be applied to images. The luminance, Y, of the image is first filtered by using a *Sobel* filter. The standard deviation over the pixels in each filtered image is then computed as an SI by

$$M_{SI} = \sigma_{space}[Y_{Sobel}], \qquad (2)$$

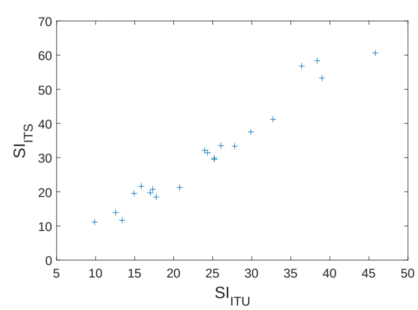


Figure 3: Relationship between the ITU and ITS recommended SI.

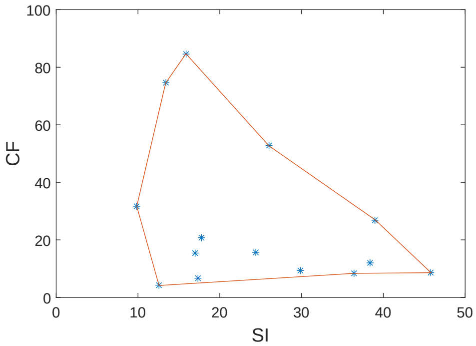


Figure 4: CF and SI distribution.

where, M_{SI} is spatial information metric, σ_{space} is the standard deviation over the pixels, and Y_{Sobel} is the *Sobel* filtered luminance plane of the image.

Both the proposed SI and the SI recommended by ITS were tested for the proposed SMART LF dataset. The results show that, for all-focused 2D images the correlation between two metrics is around 98% (shown in Figure 3).

4.1.3 Texture

Image texture is one of the most important features in image processing. In particular, it has been observed that different image textures attract human fixation with varying degrees [25] [27]. In the literature many features have been used to explain the image texture properties. However, in this work only four key features (contrast, homogeneity, energy, and correlation) are considered. For this purpose, a Gray Level Co-occurrence Matrix (GLCM) [10] is used.

4.2 Dataset Analysis

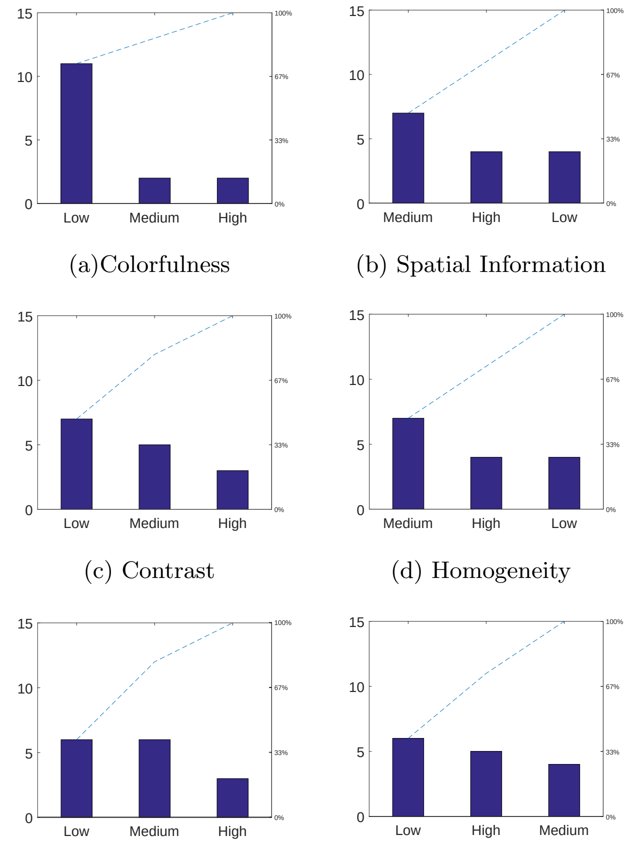
To compute the content descriptors score all in focused 2D view, *thumbnail* image, is used since it covers most of the spatial content related information of the scene.

The combined SI and CF distribution of the proposed dataset is shown in Figure 4; as can be noticed the images cover a wide range of CF and SI.

The image texture is examined by exploiting four key textural related features: contrast, energy, correlation, and homogeneity. For analysis, each general attribute is rated in the range low, medium, and high. The levels have been selected by equally partitioning the space in 3 parts.

The proposed dataset has 15 images and the images are used for analysis. The analysis result is shown in Figure 5. In Figure 5, x-axis shows the range of the quality attributes, while the left y-axis shows the number of images and the right y-axis shows images distribution in percentage. Moreover, the continuous plot is the cumulative sum of the histogram bars. From Figures 4 and 5 we can notice that most of the considered images have low CF. This result indicates that the images are natural; in fact high CF indicates low naturalness of the image [11]. For the remaining general features, Figure 5 shows that the features are well distributed over the full range.

For the analysis of LF specific features analysis, to the best of our knowledge there is no standard objective metric to analyze reflection, transparency, and DoF. Therefore, the images are selected in such a way that they can cover these



(a)Colorfulness (b) Spatial Information

(c) Contrast (d) Homogeneity

(e) Energy (f) Correlation

Figure 5: SMART LF image dataset (y-axis represent the number of images and x-axis represent the levels of each quality attributes).

attributes. As an example, the images *(k) Window* and *(m) Book* are considered to cover the feature transparency, *(g) River* and *(h) Building* are considered to cover the feature reflection, and *(f) Grid* and *(l) Pilers* are considered to cover the feature DoF variation.

5. CONCLUSION AND FUTURE WORK

This article proposes a LF image dataset for the research communities to be used for the design, testing, and benchmarking of LF image processing algorithms and for QoE estimation purposes. In particular, this article provides a brief introduction and analysis of the state-of-the-art LF datasets; the image content selection criteria have been defined for the selection of the content for LF image quality dataset design; the LF image quality dataset made freely available online in [2]. The adopted methodology can be used as a guideline for new image/video quality dataset design. Moreover, the new spatial information estimation metric is exploited for image content analysis.

6. ACKNOWLEDGMENT

Part of this work has been performed in the framework of EU ICT COST Action IC1105, 3D content creation, coding and transmission over future media networks(3D-ConTourNet).

7. REFERENCES

[1] E. H. Adelson and J. Y. A. Wang. Single lens stereo with a plenoptic camera. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (2), 1992.

[2] F. Battisti, M. Carli, R. Olsson, P. Paudyal, and M. Sjostrom. SMART Light Field Image Dataset. *http://www.comlab.uniroma3.it/SMART.html*, 2016.

[3] T. E. Bishop, S. Zanetti, and P. Favaro. Light field superresolution. In *Computational Photography (ICCP), 2009 IEEE International Conference on*, 2009.

[4] C. Conti, L. D. Soares, and P. Nunes. Hevc-based 3d holoscopic video coding using self-similarity compensated prediction. *Signal Processing: Image Communication*, 42:59–78, 2016.

[5] D. G. Dansereau, O. Pizarro, and S. B. Williams. Linear volumetric focus for light field cameras. *ACM Transactions on Graphics (TOG)*, 34(2):15, 2015.

[6] C. Faloutsos, R. Barber, M. Flickner, J. Hafner, W. Niblack, D. Petkovic, and W. Equitz. Efficient and effective querying by image content. *Journal of intelligent information systems*, 3, 1994.

[7] K. Fliegel and C. Timmerer. WG4 databases white paper v1. 5: QUALINET multimedia database enabling QoE evaluations and benchmarking. *Prague/Klagenfurt, Czech Republic/Austria*, 2013.

[8] L. Gabriel. Epreuves reversibles. photographies integrals. *Comptes-Rendus Academie des Sciences*, 146, 1908.

[9] A. Ghasemi, N. Afonso, and M. Vetterli. LCAV-31: a dataset for light field object recognition. In *IS&T/SPIE Electronic Imaging*, 2014.

[10] R. M. Haralick, K. Shanmugam, and I. H. Dinstein. Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*, (6), 1973.

[11] D. Hasler and S. E. Suesstrunk. Measuring colorfulness in natural images. In *Electronic Imaging 2003*, 2003.

[12] ITS. Spatial information (SI) filter. *http://www.its.bldrdoc.gov/resources/*, 2015.

[13] P. ITU-T RECOMMENDATION. Subjective video quality assessment methods for multimedia applications. 1999.

[14] Y. Ke, X. Tang, and F. Jing. The design of high-level features for photo quality assessment. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, 2006.

[15] B. Keelan. *Handbook of image quality: characterization and prediction*. CRC Press, 2002.

[16] B. W. Keelan and H. Urabe. ISO 20462: a psychophysical image quality measurement standard. In *Electronic Imaging*, 2003.

[17] N. Li, J. Ye, Y. Ji, H. Ling, and J. Yu. Saliency detection on light field. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.

[18] Y. Li, M. Sjostrom, R. Olsson, and U. Jennehag. Scalable coding of plenoptic images by using a sparse set and disparities. *Image Processing, IEEE Transactions on*, 25(1):80–91, 2016.

[19] I. Montilla, M. Puga, J. P. Luke, J. G. Marichal-Hernandez, and J. M. Rodriguez-Ramos. Design and laboratory results of a plenoptic objective: From 2D to 3D with a standard camera. *Display Technology, Journal of*, 11(1), 2015.

[20] A. Mousnier and C. Vural, E.and Guillemot. Partial light field tomographic reconstruction from a fixed-camera focal stack. *arXiv preprint arXiv:1503.01903*, 2015.

[21] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan. Light field photography with a hand-held plenoptic camera. *Computer Science Technical Report CSTR*, 2(11), 2005.

[22] P. Paudyal, F. Battisti, and M. Carli. A study on the effects of quality of service parameters on perceived video quality. In *Visual Information Processing (EUVIP), 2014 5th European Workshop on*, pages 1–6. IEEE, 2014.

[23] P. Paudyal, F. Battisti, and M. Carli. Impact of video content and transmission impairments on quality of experience. *Multimedia Tools and Applications*, pages 1–25, 2016.

[24] M. Pedersen, N. Bonnier, J. Y. Hardeberg, and F. Albregtsen. Attributes of a new image quality model for color prints. In *Color and Imaging Conference*, volume 2009, 2009.

[25] C. M. Privitera and L. W. Stark. Algorithms for defining visual regions-of-interest: Comparison with eye fixations. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(9):970–982, 2000.

[26] R. Raghavendra, K. Raja, and C. Busch. Exploring the usefulness of light field camera for biometrics: An empirical study on face and iris recognition. *IEEE Transactions on Information Forensics and Security (TIFS) (Accepted)*, 2016.

[27] U. Rajashekar, L. K. Cormack, and A. C. Bovik. Image features that draw fixations. In *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*, volume 3, pages III–313. IEEE, 2003.

[28] Raytrix. Raytrix light field camera. *http://www.raytrix.de/*, 2015.

[29] M. Rerabek, L. Yuan, L. A. Authier, and T. Ebrahimi. [ISO/IEC JTC 1/SC 29/WG1 contribution] EPFL Light-Field Image Dataset. Technical report, ISO/IEC JTC 1/SC 29/WG1, 2015.

[30] V. Vaish and A. Adams. The (new) stanford light field archive. *http:/lightfield. stanford. edu*, 2008.

[31] S. Wanner, S. Meister, and B. Goldluecke. Datasets and benchmarks for densely sampled 4D light fields. In *Annual Workshop on Vision, Modeling and Visualization: VMV*, 2013.

[32] G. Wetzstein. Synthetic light field archive. *http://web.media.mit.edu/ gordonw/ SyntheticLightFields/*, 2015.

[33] S. Winkler. Analysis of public image and video databases for quality assessment. *Selected Topics in Signal Processing, IEEE Journal of*, 6(6), 2012.