https://doi.org/10.2352/ISSN.2470-1173.2019.11.IPAS-261

Evaluating the effectiveness of image quality metrics in a light field scenario

Giuliano Arru, Federica Battisti, and Marco Carli;
Department of Engineering; Universitá degli Studi Roma Tre, Rome, Italy.

Abstract

In this contribution, an objective metric for quality evaluation of light field images is presented. The method is based on the exploitation of the depth information of a scene, that is captured with high accuracy by the light field imaging system. The depth map is estimated both from the original and impaired light field data. Then, a similarity measure is applied, and a mapping is performed to link the depth distortion with the perceived quality. Experimental test performed by comparing state-of-art metrics with the proposed one, demonstrate the effectiveness of the proposed metric.

Introduction

Light Field (LF) imaging is an emerging technology that allows to capture richer visual information from the surrounding world. Differently from traditional photography, which captures a 2D projection of the light in the scene, in light field there are other information, deriving from the angular domain. In fact, light fields collect radiance from rays in all directions thus demultiplexing the angular information lost in the traditional photography.

On one hand, the higher-dimensional representation of the collected data offers powerful capabilities for scene understanding and its application substantially improves the performance in many fields such as depth sensing, post-capture refocusing, segmentation, video stabilization, or material classification. On the other hand, the high dimensionality of LF opens new challenges in terms of data capture, data compression, content editing, and displaying.

It is useful to notice that LF data is prone to a wide variety of distortions during acquisition, processing, compression, storage, transmission, and reproduction phases. Any of these, may result in a degradation of the data quality. Therefore, especially for applications involving persons, the measure of the perceived quality plays an important role.

Quality can be usually measured in two ways: subjectively and objectively. Subjective methods are based on the judgment given by a set human observers collected during an ad-hoc designed experiment. These systems produce accurate results, however they are time-consuming, expensive, and can not be used in real time applications. Objective methods aim at designing quality measures that can automatically predict perceived image quality, overcoming the limits of subjective quality metrics. However, the available metrics ore often well-performing only for some specific distortions. Furthermore, their results may not be related with the perceived quality. According to the availability of the original data or of some information about it, the objective metrics can be classified into Full-Reference, Reduced-Reference, and No-Reference.

In this contribution, we highlight the connection between depth information and human perception in case of LF data and, by studying the impact of LF data distortion on the estimated depth, we define a Reduced-Reference objective metric for LF image. In more details, the metric is based on the exploitation of the depth information of a scene, that is captured with high accuracy by the LF imaging system. The depth map is estimated both from the original and impaired LF data. Then, a similarity measure is applied, and a mapping is performed to link the depth distortion with the perceived quality.

The rest of the paper is organized as follows: in Section *Quality assessment of Light Field images* the LF is defined and a literature survey of existing methods for quality assessment of LF images is reported. The proposed metric is described in Section *Proposed method* while, in Section *Experiments and Results*, the test performed for evaluating its performances are reported. Finally, in Section *Concluding remarks* the conclusions are drawn.

Quality assessment of Light Field images

The LF may be represented through the plenoptic function [17], that is a multidimensional function which describes the set of light rays traveling in every direction through every point in 3D space, from a geometric optics perspective.
To collect this information, the light rays, at every possible location (x,y,z), from every possible direction of arrival (θ,ϕ), at every light wavelength γ and, at every time t, should be measured. In more details, the plenoptic function is a 7D function defined as follows:

$$L(x,y,z,\theta,\phi,\gamma,t) \tag{1}$$

Under specific assumptions the complexity associated with the sampling of the plenoptic function can be reduced. As a first step, only the luminance component for still images can be considered, thus reducing the measured function to be monochromatic and time-invariant.

Then, the simplified model proposed by Levoy and Hanrahn [1] and Gortler et al. [2] is adopted. The authors assume the light field to be measured in free space. Under this hypothesis, the light ray radiance remains constant along a straight line thus obtaining the 4D function.

For practical applications, the generally adopted model for representing the 4D light field function, relies on the parameterization of the light rays by the coordinates of their intersections with two planes placed at arbitrary positions. Let us denote with (u,v) and (s,t) the coordinates system for the first and second plane, respectively. An oriented light ray defined in the system first intersects the *uv* plane at coordinate (u,v) and then intersects the *st* plane at coordinate (s,t). Thus, the plenoptic function becomes:

$$L(u,v,s,t) \tag{2}$$

reducing the dimensions from 7 to 4 dimensions, parametrized by four coordinates.

In literature, the problem of measuring the perceived quality for 2D images, video, and sparse multiview content, has been largely investigated. Only few attempts have been performed for evaluating the quality of LF images.

In [18] and [19], the PSNR metric has been used for evaluating the performance of LF encoding methods. However, as well known for *classical* 2D images, this metric is not well correlated with the perceived quality. In [20], Benoit et al. propose an objective quality metric for stereoscopic images based on the comparison of reference and distorted disparity maps. A survey on the existing LF quality assessment metrics is presented in [6] from Adhikarla et al. The authors propose an interactive light field viewing setup for the subjective evaluation of angular consistency. Furthermore, they extend the SSIM metric [3] to a 3D context for light field-specific angular assessment, and evaluate the performance of existing quality assessment metrics.

Proposed method

The proposed metric exploits the capacity of LF to capture the depth information and relates the distortions of the processed depth map to the perceived quality. In more details, the 4D LF representation essentially contains multiple views of the same scene, thus allowing a depth map estimation for each view pair. The possibility to exploit these features from every sub-aperture image pair leads to expands the disparity space to a continuous space [10], making depth estimation more robust and precise.

In literature the link between depth map information and visual attention or saliency model has been investigated in [21] proving its connection with the perceived quality. In [22], the authors show that depth quality is an essential aspect of perceived quality for 3D stereoscopic images. In [23], Banitalebi et alt. show that in 3D video the perceived quality is directly correlated with depth map quality. The high correlation existing between the perceived quality and depth map in the case of 3D images, pushed us to investigate the role of depth map in case of LF images, and the relation eventually existing between LF depth map and quality of experience. Our metric is based on the hypothesis that the measure of distortion on depth map is highly correlated to the LF image objective quality. That is:

$$Q_{LF} = f(dis) \tag{3}$$

where Q is the perceived quality, f is a mapping function between metric and stimuli, and dis is a measure of distortion on depth map.

In Figure 1 the disparity map extracted from a light field scene affected by the neigh-bor interpolation (NN) distortion with increasing distortion is presented. It is possible to notice the correlation between the depth map degradation and the amount of the distortions.

The quality score is computed in five different steps:

1. Disparity map estimation from the reference (original) LF $DisMap_{ref}$;
2. Disparity map estimation from the distorted LF $DisMap_{dis}$;
3. Estimation of the distortion level of the depth map comparing the reference ($DisMap_{ref}$) and distorted ($DisMap_{dis}$) disparity maps, by using a similarity function.

$$dis = similarity(DisMap_{ref}, DisMap_{dis}), \tag{4}$$

4. Selection and application of the pooling strategy;
5. Using a mapping model to estimate the perceptual quality of distorted LF;

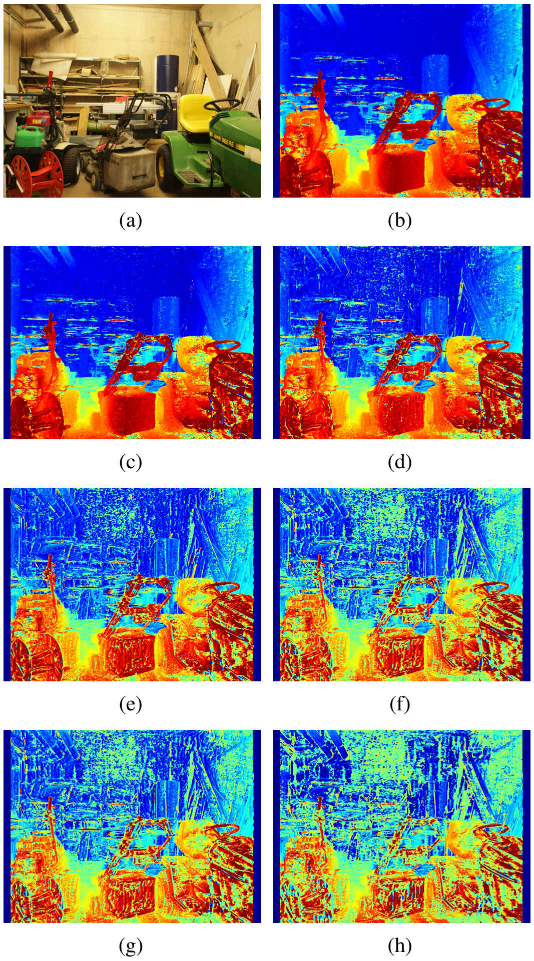

(a) (b) (c) (d) (e) (f) (g) (h)

Figure 1: Impact of neighbor interpolation (NN) distortion on the estimated depth map. 1(a) Central View Reference LF. 1(b) Depth map Reference LF. 1(c) Depth map from LF with distortion NN with severity 4. 1(d) Depth map from LF with distortion NN with severity 7. 1(e) Depth map from LF with distortion NN with severity 10. 1(f) Depth map from LF with distortion NN with severity 14. 1(h) Depth map from LF with distortion NN with severity 24.

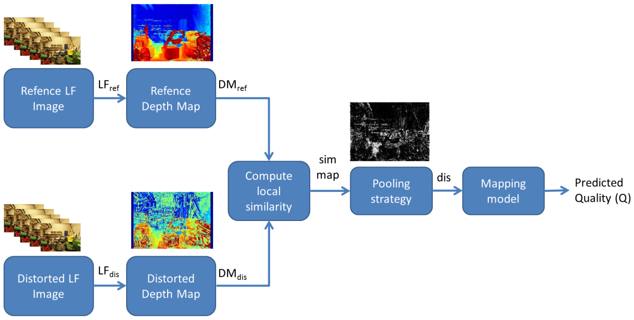


Figure 2: Block diagram of the proposed metric.

It is worth to notice that the extraction of a gray scale disparity/depth map is an operation of dimensional reduction. In fact, a single disparity map has a limited size with respect to the LF content and also to a single view. Therefore, the proposed metric can be considered a reduced reference metric, since it uses only disparity/depth map information of reference and distorted LFs for estimating the quality. In this work, a multi-resolution approach is used to compute the disparity map [4]. Multiple views of scene

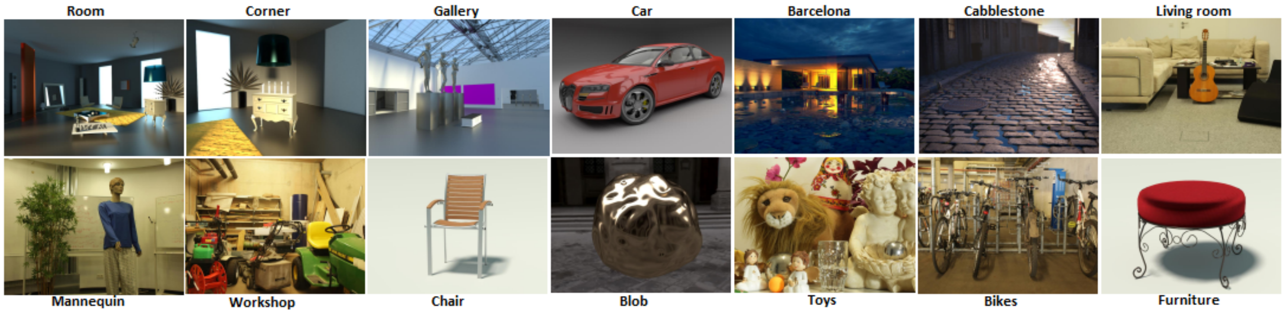


Figure 3: Central view images of all light fields in the dataset.

are used for estimating the depth map. In brief, a log-likelihood functional of depth of field for a given pair of sub-aperture views (center view and others view) is defined as the conditional joint probability of the view given the depth of field divided by arbitrary function that does not depend on depth of field. The depth map is expressed as a Maximum Likelihood estimate of the depth of the functional, and a weighted median filter is used for rendering the estimated map. It is useful to underline that the adopted depth estimation method, even if resulting in one of the most accurate depth estimation methods, presents ambiguity in the selection of the maxima of the likelihood functional in flat, uniform areas. Therefore, lower performance are possible on flat regions without textures.

To select the similarity function several tests were performed. Mean Square Error (MSE), Peak Signal to Noise Ratio (PSNR), and Structural Similarity index (SSIM) [3] have been tested. SSIM includes three components: luminance, contrast, and structure. It is based on the assumption that the Human Vision System extracts information from image textures. Higher correlation with MOS has been obtained by using the SSIM. This result can be explained by considering that the structural information is an important feature for the depth map. Thus the distortion on depth maps by measuring SSIM between reference (DM_{ref}) and distorted disparity map (DM_{dis});

$$Dis = SSIM(DM_{ref}, DM_{dis}) \tag{5}$$

$$SSIM(DM_{ref}, DM_{dis}) = \frac{(2\mu_x\mu_y + c1)(2\sigma_x y) + c2}{(\mu_x^2 + \mu_y^2 + c1)(\sigma_x^2 + \sigma_y^2 + c2)} \tag{6}$$

where $c1$ and $c2$ are two normalization factors. μ_x, μ_y, σ_x, and σ_y are the mean and standard deviation and $\sigma_x y$ is the covariance of DM_{ref} and DM_{dis} respectively.

The output of structure similarity is a dissimilarity map. To combine local quality results in a single quality score, different pooling strategies are available in literature [5] (i.e., Minkowski, Local Quality, Average, or Saliency). Based on performed tests, in the proposed method the average strategy has been adopted. The mean is a special case of Minkowsky pooling strategy where $p = 1$:

$$M = \sum_{i=1}^{N} m_i^p \tag{7}$$

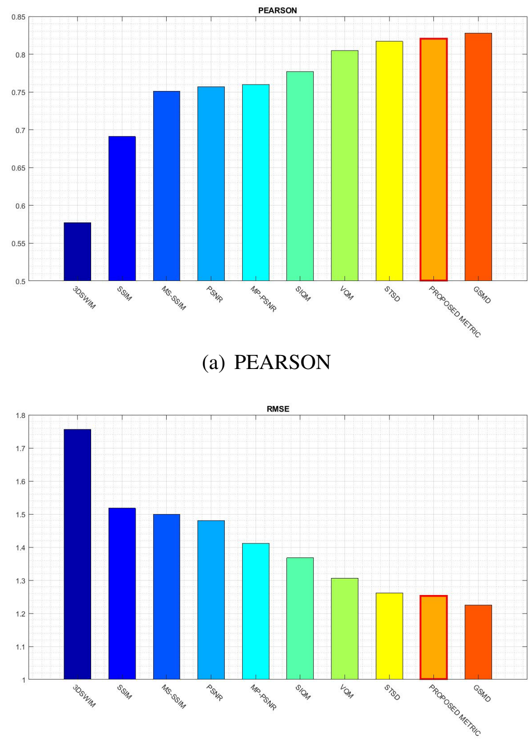
N is the number of samples in the quality/distortion map, p is the Minkowsky power and M is the global result obtained combining the local scores.

In literature many studies have been performed to understand the relationship between human perception and physical stimuli. The relation between metric and stimuli can be complex and not linear. To account this relation a mapping model is needed to estimate the perceptual quality [7]. In this work, the logistic function with five parameters (i.e., a logistic function with an added linear term, constrained to be monotonic), has been used. In more details:

$$q(o) = a_1\left\{\frac{1}{2} - \frac{1}{1+exp[a_2(o-a_3)]}\right\} + a_4 o + a_5 \tag{8}$$

where o is the output of the metric. The parameters $a_{1...5}$ are optimized to minimize q given the goodness-of-fit measure.

Experiments and Results



(a) PEARSON

(b) RMSE

Figure 4: Comparative analysis of the performances of the proposed metric. Above: Pearson Correlation Coefficient (Pearson). Below: Root Mean Squared Error (RMSE).

To validate the proposed metric a dataset designed for the light field metric evaluation is used [6]. In this dataset there are nine synthetic and five real word scenes 3. They span a large variety of different conditions, for example daylight/night, outdoor/indoor etc. All the light fields are of identical spatial and angular resolution (960x720x101x1 pixels). Four different distortions with six severity levels were applied to every scene. To all the synthetic scenes were applied nearest neighbor interpolation (NN), nearest linear interpolation (LINEAR), image warping using optical flow estimation (OPT), and quantized depth maps (DQ). For all real-world scenes, it is used nearest neighbor interpolation (NN), image warping using optical flow estimation (OPT), Gaussian blur in angular domain (GAUSS) and, 3D extension of HEVC encoder (HEVC). Including original light fields, the dataset consists of 350 different light fields. The MOS for each image is available.

For further investigation on the performances of the proposed metric, a comparison with the results achieved by existing state-of-the-art metrics on the same dataset has been performed [6]. In more details, the following metrics have been considered:

- Image-based: Structure Similarity (SSIM) [3], Peak Signal to Noise Ratio (PSNR), Multi Scale Structure Similarity (MS-SSIM) [9], Gradient Magnitude Similarity Deviation (GMSD) [12];
- Video-based: Video Quality Metric (VQM) [11];
- Multiview-based: 3DSWIM [14], MP-PSNR [13];
- Stereoscopic image quality metric: SIQM [15] that is based on the concept of cyclopean image where, it is averaged scores obtained from all stereo pairs shown in our experiment.
- Stereoscopic video quality metric: *STSD* [16]

The performances of the proposed metric have been tested versus the benchmark metrics by evaluating the Pearson correlation and the Root Mean Squared Error (RMSE).

The goodness-of-fit scores were computed after the logistic function fitting. For a fair comparison, a seven-fold cross-validation was used where the whole dataset was divided based on the scenes. Each fold was constructed by a testing set corresponding to two scenes, while the others were used for training.

The experiments show that the proposed metric is one of the best performing and confirms the hypothesis that depth map gets a good approximation of the perceived quality. In Fig. 4 the obtained results for the different performances indexes are shown. The bars indicate the scores that are averaged after testing across different cross-validation folds. If we consider the two different indexes (PEARSON, RMSE), the proposed metric performs better than all metrics except GMSD.

The proposed metric, even if is not the best performing for some LF images, well matches the subjective scores as shown by the Pearson correlation. This result confirms that for LF images, the depth information can be used for estimating the impact of distortion on the perceived quality. It is useful to underline that the depth map, extracted from the original image, can be lossless compressed and used as a reduce reference of the original signal.

In Table 1, the performance analysis of the proposed metric with respect to the specific images in the dataset is shown. As can be noticed, the content does not severely influence the metric performances. Lower performances are obtained for the scenes: Room and Blob. These images are characterized by uniform depth areas in low illumination.

In Table 2, the performance analysis of the proposed metric with respect to the considered distortions is shown. Overall, the metrics well matches the human judgment. Lower performances are obtained for the distortions: image warping using optical flow estimation (OPT) and quantized depth maps (DQ).

As a last test, the performances of proposed metric in case of computer generated images have been considered. In fact, advances in image synthesis techniques allow us to simulate the distribution of light energy in a scene with great precision. Unfortunately, this does not ensure that the display end image will have a high fidelity visual appearance. Therefore, it is useful to test the metric for computer generated images. The results are shown in Table 3. As can be noticed, the metric well performs in both cases.

Scene	RMSE	Pearson
Barcelona	1.05	0.87
Bikes	1.13	0.87
Blob	1.38	0.76
Car	1.25	0.81
Chair	1.31	0.84
Cobblestone	1.46	0.81
Corner	1.22	0.78
Furniture	0.76	0.94
Gallery	0.93	0.88
Living room	0.83	0.96
Mannequin	1.19	0.91
Room	1.83	0.58
Toys	1.34	0.86
Workshop	1.17	0.80

Table 1: Performance analysis of the proposed metric with respect to the images in the dataset.

Distortion	RMSE	Pearson
DQ	1.355	0.65
Gauss	1.10	0.87
HEVC	1.40	0.89
Linear	1.21	0.84
NN	0.89	0.90
OPT	1.45	0.60

Table 2: Distortion and content based analysis

Scene type	RMSE	Pearson
Real	1.28	0.76
Synthetic	1.14	0.85

Table 3: Performances of the proposed metric averaged for real and computer generated images.

Concluding remarks

In this contribution a Reduced Reference metric for assessing the quality of LF images is presented. The proposed metric exploits the capacity of LF to capture the depth information. The metric is tested with a robust methodology and compared with the state-of-art techniques. The experiments show that the effectiveness of the proposed metric and confirm the hypothesis that depth map is strictly connected to the perceived quality, even for LF images. The reduced reference information can be used in broadcasting scenario. On going work is devoted for coping with flat and uniform depth areas.

Acknowledgement

The authors would like to thank Michele Brizzi for his support in selection and tuning the disparity map algorithm.

References

[1] Marc Levoy and Pat Hanrahan. 1996. Light field rendering. In Proceedings of the 23rd annual conference on Computer graphics and interactive techniques (SIGGRAPH '96). ACM, New York, NY, USA, 31-42. DOI=http://dx.doi.org/10.1145/237170.237199

[2] S.J. Gortler et al., The Lumigraph, Proc. ACM Siggraph, ACM Press, 1996, pp. 43-54.

[3] Zhou, W., A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. "Image Qualifty Assessment: From Error Visibility to Structural Similarity." IEEE Transactions on Image Processing. Vol. 13, Issue 4, April 2004, pp. 600612.

[4] A. Neri, M. Carli and F. Battisti, "A multi-resolution approach to depth field estimation in dense image arrays," 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, 2015, pp. 3358-3362. doi: 10.1109/ICIP.2015.7351426

[5] Z. Wang and X. Shang, "Spatial Pooling Strategies for Perceptual Image Quality Assessment," 2006 International Conference on Image Processing, Atlanta, GA, 2006, pp. 2945-2948. doi: 10.1109/ICIP.2006.313136

[6] V. K. Adhikarla et al., "Towards a Quality Metric for Dense Light Fields," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 3720-3729. doi: 10.1109/CVPR.2017.396

[7] H. R. Sheikh, M. F. Sabir and A. C. Bovik, "A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms," in IEEE Transactions on Image Processing, vol. 15, no. 11, pp. 3440-3451, Nov. 2006.

[8] Bovik, A.C.: Mean squared error: love it or leave it? - A new look at signal fidelity measures. IEEE Sig. Process. Mag. 26, 98-117

[9] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multiscale structural similarity for image quality assessment. In Signals, Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on, volume 2, pages 13981402 Vol.2, Nov 2003. 6

[10] S. Wanner and B. Goldluecke, Variational light field analysis for disparity estimation and super-resolution, IEEE TPAMI, vol. 36, no. 3, pp. 606619, 2014.

[11] M. H. Pinson and S. Wolf. A new standardized method for objectively measuring video quality. IEEE Transactions on Broadcasting, 50(3):312322, 2004. 2, 6

[12] W. Xue, L. Zhang, X. Mou, and A. C. Bovik. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. IEEE Transactions on Image Processing, 23(2):684695, Feb 2014. 6

[13] D. Sandic-Stankovic, D. Kukolj, and P. Le Callet. Multiscale synthesized view assessment based on morphological pyramids. Journal of Electrical Engineering, 67(1):311, 2016. 2, 6

[14] F. Battisti, E. Bosc, M. Carli, P. L. Callet, and S. Perugia. Objective image quality assessment of 3D synthesized views. Signal Processing: Image Communication, 30(C):7888, 2015. 2, 6

[15] M.-J. Chen, C.-C. Su, D.-K. Kwon, L. K. Cormack, and A. C. Bovik. Full-reference quality assessment of stereopairs accounting for rivalry. Signal Processing: Image Communication, 28(9):11431155, 2013. 6

[16] V. D. Silva, H. K. Arachchi, E. Ekmekcioglu, and A. Kondoz. Toward an impairment metric for stereoscopic video: A full-reference video quality metric to assess compressed stereoscopic video. IEEE Transactions on Image Processing,22(9):33923404, Sept 2013. 6

[17] H. Adelson, Edward and R. Bergen, James, The plenoptic function and the elements of early vision, Computational Models of Visual Processing, 1991.

[18] A. Vieira, H. Duarte, C. Perra, L. Tavora and P. Assuncao, "Data formats for high efficiency coding of Lytro-Illum light fields," 2015 International Conference on Image Processing Theory, Tools and Applications (IPTA), Orleans, 2015, pp. 494-497.

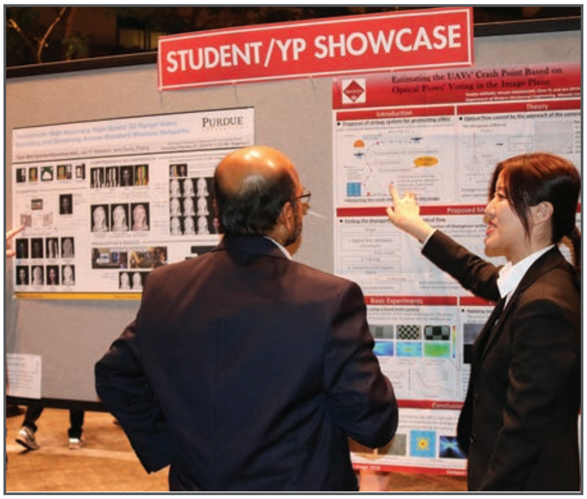
[19] Y. Li, M. Sjstrm, R. Olsson and U. Jennehag, "Scalable Coding of Plenoptic Images by Using a Sparse Set and Disparities," in IEEE Transactions on Image Processing, vol. 25, no. 1, pp. 80-91, Jan. 2016.

[20] Alexandre Benoit, Patrick Le Callet, Patrizio Campisi, Romain Cousseau. Using disparity for quality assessment of stereoscopic images. IEEE International Conference on Image Processing, ICIP 2008, Oct 2008, San Diego, United States. 2008.

[21] H. Liu and I. Heynderickx, "Visual Attention in Objective Image Quality Assessment: Based on Eye-Tracking Data," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 21, no. 7, pp. 971-982, July 2011.

[22] J. Wang, S. Wang, K. Ma and Z. Wang, "Perceptual Depth Quality in Distorted Stereoscopic Images," in IEEE Transactions on Image Processing, vol. 26, no. 3, pp. 1202-1215, March 2017.

[23] A. Banitalebi-Dehkordi, M. T. Pourazad and P. Nasiopoulos, "A study on the relationship between depth map quality and the overall 3D video quality of experience," 2013 3DTV Vision Beyond Depth (3DTV-CON), Aberdeen, 2013, pp. 1-4.


STUDENT/YP SHOWCASE
PURDUE

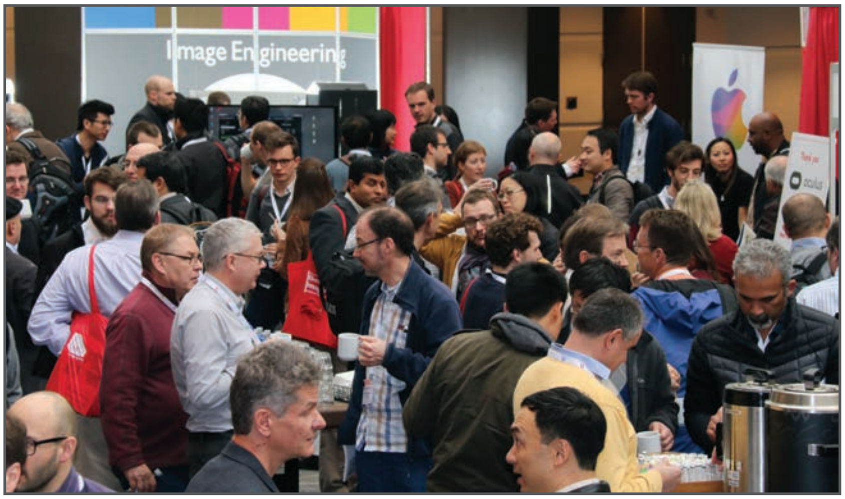


Image Engineering

AR is a fundamental and transformative interface and platform.
20+ years from my initial working demo to today's AR displays
Optimistic that it won't take another 20 to reach ubiquitous consumer AR!
My first working AR prototype
Today's platforms
Ubiquitous consumer AR