Towards a Quality Metric for Dense Light Fields

Chaonan Song

Jun 10, 2018

Abstract

Today I read the last part of Dr. Adhikarla's thesis. In this part Dr. ADhikarla considered several popular image, video, stereo, and multiview quality metrics. In order to use the image quality metrics to obtain the quality of the light field, we applied metrics to the individual light field images, performed experiments on each metric, and compared them at the end. It reveals the quality of different light field reconstruction methods and can directly guide the selection of light field reconstruction techniques.

1. Evaluation of quality metrics

Dr. Adhikarla considered several popular image, video, stereo, and multiview quality indicators. First Dr. Adhikarla briefly described the indicators and then evaluated their performance on our data set. In order to use the image quality metrics to obtain the quality of the light field, Dr. Adhikarla apply metrics to the individual light field images and then average the scores of all the images.

Quality metrics: Although studies have shown that perceptual indicators perform better than absolute differences (AD) [2], because of its important use in image quality assessment, Dr. Adhikarla still consider peak signal-to-noise ratio (PSNR).$SSIM_{2D \times 1D}$ uses 2D¡□1D patch which contains a 2D window extracted from a particular view and a 1D row of pixels that extends from the center of the 2D window in the angular domain (see Figure 1). Dr. Adhikarla applied the metrics to all light-field images without resampling and averaged the scores over all images. Although Dr. Adhikarla experimented with various pooling strategies, he found that the average value performs best. Due to better performance, Adhikarla chose 32and 64-pixel angular window sizes for SSIM2D¡□1D and SSIM3D, respectively. Finally, Adhikarla chose to solve multi-view data and interpret the indicators of interpolation artifacts. The 3DSWIM proposed by Battisti *et al.* [1] first performs shift compensation on the reference image block and the distortion (interpolation) image. These matching blocks undergo a first-order Haar wavelet transform and calculate a histogram of subband-

s corresponding to horizontal details in the block. Finally, the Kolmogorov-Smirnov distances of these histograms are used as metric predictions. The resulting per-pixel error map is then aggregated and converted into a peak SNR measure.
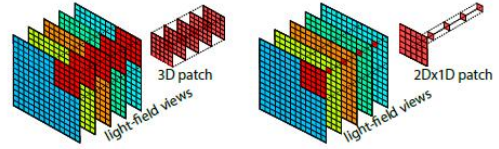

Figure 1. Patches used in our extensions of $SSIM_{2D}$

1.1. Metric performance comparison

It is expected that the quality value predicted by each metric is related to the JOD value, but this relationship may be complex and non-linear. To explain this relationship, we follow a common practice and fit a logical function 1:

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

Where o is the output of the metric. The parameters $a_{1...5}$ are optimized to minimize the given fit measure. Here we report the reduced chi-squared statistic ($\mathcal{X}^2_{red}$) and Pearson correlation coefficient (ρ). $\mathcal{X}^2_{red}$ is computed as a weighted average of the squared differences, in which weights are the inverse of sample variance. $\mathcal{X}^2_{red}$ is calculated as a weighted average of the squared differences, where the weight is the reciprocal of the sample variance. This shows that larger values of the JOD are more uncertain and, therefore, their prediction accuracy may be lower. For fair comparison, Dr. Adhikarla performed seven cross-validations in different scenarios. Adhikarla measure the goodness of fit in two randomly chosen scenarios of cross-validation multiples and average the results at all multiples. The resulting Pearson correlation and $\mathcal{X}^2_{red}$ values are shown in Figure 2. Figure 3 shows the performance of various distortion metrics.

The results show good performance of the 2D image and video quality indicators. Adhikarla observe that angular distortions indirectly translate into differences in spatial patterns, which can explain good performance. Figure 3 shows
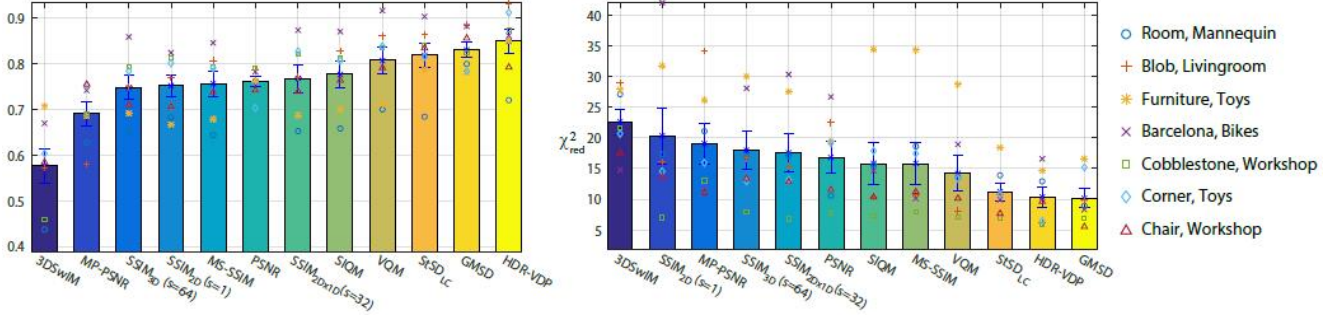
Figure 2. The goodness-of-fit scores for the metrics expressed as Pearson Correlation Coefficient (ρ) and reduced chi-square($\mathcal{X}^2_{red}$) after cross-validation. The results for each cross-validation fold are shown. $\mathcal{X}^2_{red} = 1$ indicates that the goodness of fitbetween the metric predictions and the subjective data is in perfect agreement with the measured subjective variance and $\rho = 1$ indicates perfect positive linear relation between objective scores and JODs. The error bars represent standard error
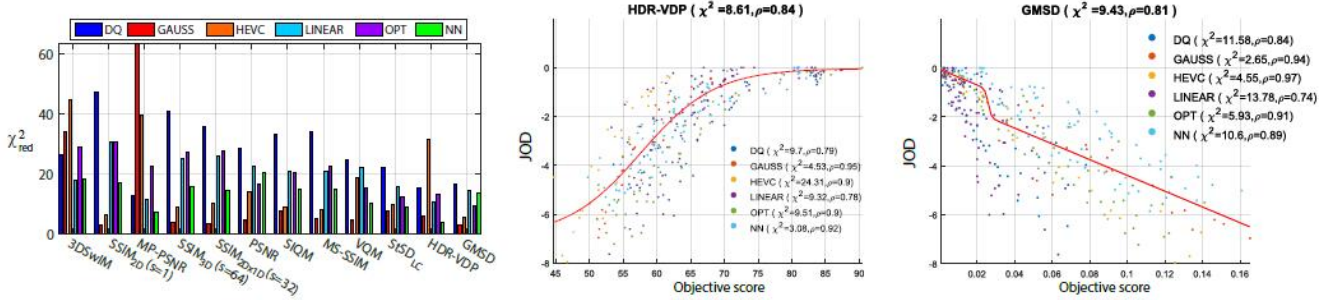

Figure 3. Left: The prediction accuracy per-distortion reported as reduced chi-squared goodness of fit score. Middle and right: $\mathcal{X}^2_{red}$¨Cfit for the metrics HDR-VDP and GMSD over all scenes. The prediction accuracy for individual distortions are shown inside the plots and the overall accuracy is indicated on the top of the plots.

that some indicators are better than others in predicting certain distortion types. For example, HDR-VDP-2 has consistently underestimated the quality of HEVC. Training such indicators for specific types of distortion can significantly improve performance.

1.2. Sparse light-field reference case

Adhikarla provided a high-quality 101 field of view as a reference for quality indicators. In fact, only sparsely sampled light fields are unavailable in most applications. When the sparse light field is used as a reference, a full reference metric is given to compare the two distorted light fields without a perfect reference. This is a task. These metrics are not designed to predict the JOD relative to the perfect reference image, not the JND relative to any other image. For the test light field, Adhikarla considered all light fields with higher distortion levels. For fair comparison, Adhikarla also use the complete 101-view light field as a reference to run metrics on the same subset. The results of these tests are shown in Figure 4 in cyan and blue bars. This indicates that a high quality reference light field must be provided for existing indicators to reliably predict quality. However, if there
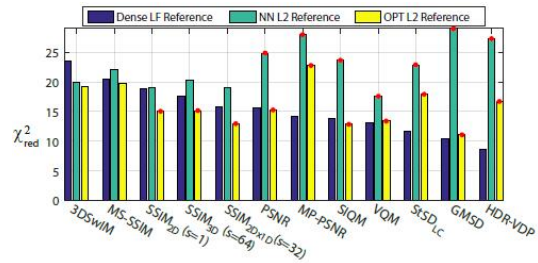

Figure 4. The goodness-of-fit scores for the subset of the dataset when a dense LF is used as a reference (blue), when nearest-neighbour at the 2nd distortion level is a reference (cyan), or when optical flow is used to up-sample the reference LFs. The dots at cyan bars mean that the value is statistically different from the dense LF case and the dots on the yellow bars that the values are statistically different from the NN case. The significance is computed by bootstrapping and running one-tailed test ($\rho = 0.05$).

is no such high-quality reference, can it be approximated? Our subjective data shows that Optical Flow Interpolation (OPT) produces the highest quality results. Therefore, we use OPT to generate a reference 101 perspective light field

from a sparse 21-view field of light and repaint the metric on the subset. The results show that the prediction effect is improved compared to using a sparse light field. This shows that a potential solution to an incomplete reference problem is to use a high quality interpolation method to generate the reference.

2. Conclusions and future work

Adhikarla have established a new 3D dense light field data set and subjective mass scaling of various distortions that occur in light field applications. Different methods in light field processing result in visual artifacts with completely different appearances, such as LINEAR blurring, OPT ghosting, NN image flickering and jumping. Adhikarla's experiments revealed how these different artifacts affect perceived quality. Adhikarla evaluated the potential of existing image, video, stereo, and multiview quality indicators in predicting subjective scores. Adhikarla's observations show that indicators such as HDR-VDP-2, GMSD, STSDLC, and VQM perform quite well when the distorted light field is compared with a dense reference and can be used in applications that require this comparison. This is the case in some applications when dense light fields are not available, and it is not reasonable to use these quality assessment indicators. Adhikarla provide sensory zoom data that can be used to train and validate new light field quality indicators. Adhikarla's results also reveal the quality of different light field reconstruction methods and can directly guide the choice of light field reconstruction techniques.

References

[1] F. Battisti, E. Bosc, M. Carli, P. L. Callet, and S. Perugia. Objective image quality assessment of 3D synthesized views. *Signal Processing: Image Communication*, 30:78-88, 2015.

[2] W. Lin and C. C. J. Kuo. Perceptual visual quality metrics: A survey. *Journal of Visual Communication and Image Representation*, 22:297-312, 2011.