Journal of Imaging Science and Technology® 59(3): 030404-1–030404-13, 2015.
© Society for Imaging Science and Technology 2015

Measuring the Relative Image Contrast of Projection Displays

Ping Zhao▲, **Marius Pedersen**▲, **and Jon Yngve Hardeberg**▲
Gjøvik University College, Gjøvik, Norway
E-mail: marius.pedersen@hig.no

Jean-Baptiste Thomas▲
Université de Bourgogne, Dijon, France

Abstract. *Projection displays, compared to other modern display technologies, have many unique advantages. However, the image quality assessment of projection displays has not been well studied so far. In this paper, we propose an objective approach to measure the relative contrast of projection displays based on the pictures taken with a calibrated digital camera in a dark room where the projector is the only light source. A set of carefully selected natural images is modified to generate multiple levels of image contrast. In order to enhance the validity, reliability, and robustness of our research, we performed the experiments in similar viewing conditions at two separate geographical locations with different projection displays. In each location, we had a group of observers to give perceptual ratings. Further, we adopted state-of-art contrast measures to evaluate the relative contrast of the acquired images. The experimental results suggest that the Michelson contrast measure performs the worst, as expected, while other global contrast measures perform relatively better, but they have less correlation with the perceptual ratings than local contrast measures. The local contrast measures perform better than global contrast measures for all test images, but all contrast measures failed on the test images with low luminance or dominant colors and without texture areas. In addition, the high correlations between the experimental results for the two projections displays indicate that our proposed assessment approach is valid, reliable, and consistent.*
© 2015 Society for Imaging Science and Technology.
[DOI: 10.2352/J.ImagingSci.Technol.2015.59.3.030404]

INTRODUCTION

Flat-panel display technologies in the liquid crystal display (LCD) family have dominated the consumer market for many years, and this is especially true for desktop/laptop monitors, mobile phone screens, televisions, and many large outdoor information displays. The strongest appeal for consumers on displays is perhaps the ability to share information and collaborate with teammates quickly, easily, and conveniently. Projection displays, compared to other display technologies, have unique advantages in terms of portability, flexibility for deployment, and large screens to visualize information for a target audience. Recently, there has been an increasingly general interest to embed mini projectors into portable imaging devices such as smart phones or handheld video recorders, so that the pictures/videos can be reviewed and shared with a crowd of people in the field right after they are recorded.[1,2] In more advanced scenarios, multiple projections can be tiled up to produce a single large perceptual seamless image which visualizes information to target audiences, and they will enjoy a fully immersive visual experience.[3,4] In this context, image quality assessment of projection displays has gradually become an essential topic in both academic research and industrial commercial communities. The goal of image quality assessment is not limited to the establishment of a unified approach to evaluate the quality of image reproductions, but also in defining a systematic way to continuously improve the perceptual image quality within a closed work flow.

Image quality can be characterized and interpreted based on a set of image quality attributes which are terms of human perceptions of lightness, contrast, colors, sharpness, and artifacts (including noises).[5] Physical properties such as screen dimension, display resolution, and refreshing rate have impacts on the perceived image quality, but in a typical work flow of image quality assessment they can be assumed to be constants, since they are independent from image content and normally do not vary over time. In this paper, we only focus on the image quality attributes which are content independent. According to the existing literature, there have been many attempts to characterize displays such as cathode-ray tube (CRT)[6,7] and LCD[6–8] desktop/laptop monitors. The characterization of projection displays has a similar approach. Previous characterizations of projection displays primarily focused on black level estimation,[9] display uniformity,[10–12] and colorimetry,[11,13] but limited attention has been paid to measuring the contrast of projected images on the screen. More specifically, the measured contrast of an image has been shown to be of a significant impact on the visual experience.[14–16]

Experiments measuring the contrast of projection displays have largely been conducted based on absolute acquisitions with a radiometer or a spectrometer, etc.[11,17,18] These devices are well designed to produce accurate measurements, but they are expensive and require professional training; in a typical projection environment where it is common to have a low light condition, it takes a long time to collect a large number of measurements at discrete sample spots. Using a

▲ IS&T Members.

Received Mar. 22, 2015; accepted for publication June 1, 2015; published online June 22, 2015. Associate Editor: Maria V. Ortiz Segovia.
1062-3701/2015/59(3)/030404/13/$25.00

camera as a relative acquisition device has the advantage of recording all displayed pixels in one shot.[19] Once we have the captured images, we can process them with image quality measures to predict the actual image contrast and correlate these results with perceptual ratings. So, camera based acquisition can be a fast alternative approach to measure the relative contrast of projection displays at low cost.

This paper presents a study on the measurement of relative contrast of projection displays based on acquisitions with a digital single-lens reflex (DSLR) camera. The main goal of this work is to evaluate state-of-art contrast measures based on their correlations with subjective ratings. The results of the evaluation can be used to improve the design of image quality measures, and they can also to be extended in the development and enhancement of general image reproduction technologies.

This paper is organized as follows. First, in the next section, we introduce the background of image contrast and the state-of-art of contrast measurements. Then, in the third section, a full description of the experimental environment, setup, and experimental procedure is given. The results and discussions on the interaction between measured contrast and perceptual contrast are presented in the fourth section. Finally, in the fifth section, conclusions are drawn based on the data analysis.

CONTRAST MEASURES

The contrast measures for images can be broadly classified into two categories with respect to their measurements at either the global or local level. The global contrast measures determine the contrast at each pixel or a few representative pixels of the input image; so the contrast operator is applied individually to each component without involving its neighborhoods. However, at the local level of contrast measurement, the neighborhoods are involved possibly by following a hierarchical structure.

Global Contrast Measures

It is important to have a clear understanding of what image contrast is before we start to measure it. However, giving a comprehensive definition of perceptual contrast can be difficult, because it depends on how subjective the observers are, how the observers are related to the observation task, and how much experience the observers turn out to have.[20] Due to these difficulties, the early research on perceptual contrast confined itself to controlled viewing conditions with limited types of visual stimuli. The studies began with measuring the contrast of a periodic pattern such as a sinusoidal grating with a simple formula at a global level.

The most commonly used global contrast measure is defined with the Michelson formula

$$C^M = (L_{\max} - L_{\min})/(L_{\max} + L_{\min}),$$

where $L_{\max}$ and $L_{\min}$ stand for the maximum and minimum luminance values, respectively.[21] In a similar fashion, the contrast can also be defined with the Weber fraction

$$C^W = \Delta L/L_b,$$

where $\Delta L = (L_{\max} - L_{\min})/2$ and L_b stands for the luminance of a uniform background around the stimulus.[22] King-Smith and Kulikowski[23] defined contrast as

$$C^K = (L_{\max} - L_a)/L_a,$$

where L_a stands for the average luminance of the visual stimulus, while Burkhardt et al.[24] replaced L_a with the average luminance of the stimulus background. Among these measures, Michelson contrast is the most widely incorporated as a performance reference against others. It is obvious that the measures assume that extreme luminance values dominate the contrast of the whole image.

Pavel et al.[25] proposed a root-mean-square (RMS) measure

$$C^{\mathrm{RMS}} = \sqrt{\sum_{i=1}^{n}(x_i - x')^2/(n-1)},$$

where x_i stands for the normalized luminance value at the ith pixel, x' stands for the mean of x_i, and n stands for the number of pixels.[25] With respect to the formula definition, it is clear that this measure ignores the spatial frequency of image content and spatial distribution of contrast in that image. Pedersen et al.[26] proposed a LAB variance measure

$$C^{\mathrm{LAB}} = \sqrt[3]{\mathrm{std}^2(L) * \mathrm{std}^2(a) * \mathrm{std}^2(b)},$$

where L, a, and b define the coordinate of each pixel in the perceptually uniform CIELAB color space. This measure accounts both luminance and chromatic channels; however, the equal weighting for each channel is inconsistent with the known fact that luminance has stronger impact on the perceived contrast than that of chrominance.[27–29]

For image quality assessment of displays in industry, the international standards TCO Certified Display 6.0[30] and SPWG Notebook Panel Specification 3.8[31] both recommend contrast defined as

$$C^{\mathrm{TCO}} = L_{\max}/L_{\min}.$$

The Information Display Measurements Standard 1.03[32] follows a similar fashion, but further classifies contrast measurements into multiple categories as signal contrast, sequential contrast, starfield contrast, and corner-box contrast by taking the spatial information into account. Part 307 of ISO standard 241 defines contrast as

$$C^{\mathrm{ISO}} = (L_{\max} + L_D + L_S)/(L_{\min} + L_D + L_S),$$

where L_D and L_S stand for the luminance component reflected from diffuse illumination and the luminance component specularly reflected from large aperture sources of illumination, respectively.[33] The contrast definitions in the international standards mentioned above were originally designed to verify the display performance; however, the contrast of actual displayed images is not a part of their concerns.

In summary, the existing global contrast measures are largely inheritances or variations of Michelson contrast to determine the contrast of displays. The contrast definitions above account only the extreme or average luminance values, and they are confined to specific viewing conditions with gray patches or periodic patterns such as sinusoidal gratings; as a result, their use in natural images might be inappropriate.

Local Contrast Measures

For contrast measurement, the local nature of contrast changes across an image and spatial frequency content are related and should be considered together.[14] Local contrast measures divide the images into many subimages, possibly at multiple hierarchical levels, which may be overlapped with each other; the contrast is defined by taking pixel neighborhoods or specific local features into account. Depending on the division granularity, the contrast can be determined at a pixel level with respect to the luminance and/or chrominance information in a certain color space.

Boccignone et al.[34] followed the Weber–Fechner law to replace the subject luminance with the luminance $I(x, y, t)$ of pixel (x, y) at instant t and to replace the background luminance with the average luminance $I_b(x, y, t)$ in the surrounding area of pixel (x, y) at instant t. The instant t is changed by an iterative application of the anisotropic diffusion equation; so the most optimal local contrast for a pixel (x, y) is determined as

$$C^{\mathrm{MWF}}(x, y) = \max_{t\in[t_{\mathrm{inf}}, t_{\mathrm{sup}}]} \ln[I(x, y, t)/I_b(x, y, t)].$$

A. J. Ahumada[35] applied two rounds of low-pass filters F_a and F_b to the input image I and generated two filtered images:

$$I_a(x, y) = I(x, y) * F_a(x, y),$$
$$I_b(x, y) = I_a(x, y) * F_b(x, y).$$

Then the local contrast for each pixel (x, y) is defined as

$$C(x, y) = I_a(x, y)/I_b(x, y) - 1,$$

and eventually the final local contrast is calculated as

$$E(x, y) = C^2(x, y) * F_e(x, y),$$

where F_e is another low-pass filter. Despite the luminance information, the chrominance components in images contribute to the measured contrast as well. Matkovic et al.[36] introduced a global contrast factor (GCF) method to compute the local contrast by averaging the differences between spatially filtered super pixels, and then the global contrast is determined as the mean of local contrast with respect to weighting factors that are estimated based on a psychophysical experiment.

Peli et al.[14] proposed calculating the contrast separately at each pixel of an image to address the variation of contrast across the whole image. In this case, multiple band limited versions of the original image are obtained by applying a radically symmetric band-pass filter in the frequency domain. Then the contrast for each limited band is defined as the ratio between the filtered image and its local luminance mean image. Tadmor and Tolhurst[37] proposed a modified contrast measurement for natural scenes based on the conventional difference of Gaussian (DOG) receptive field model. They proposed a contrast measurement scheme as

$$C^{\mathrm{MDOG}} = [R_c(x, y) - R_s(x, y)]/[R_c(x, y) + R_s(x, y)]$$

at a pixel (x, y) in the image, where

$$R_c(x, y) = \sum_{i=x-3r_c}^{x+3r_c} \sum_{j=y-3r_c}^{y+3r_c} \mathrm{Center}\,(i - x, j - y)$$

and

$$R_s(x, y) = \sum_{i=x-3r_s}^{x+3r_s} \sum_{j=y-3r_s}^{y+3r_s} \mathrm{Surround}\,(i - x, j - y)$$

stand for the center and surrounding components of the receptive field, respectively, with

$$\mathrm{Center}\,(x, y) = \exp[-(x/r_c)^2 - (y/r_c)^2]$$
$$\mathrm{Surround}\,(x, y) = 0.85(r_c/r_s)^2 \exp[-(x/r_s)^2 - (y/r_s)^2],$$

where r_c and r_s stand for the radius of the center and the surroundings of the receptive field, respectively. Eventually, the global contrast is calculated as the mean of the local contrast measurements at many randomized locations in the image.

Rizzi et al.[38] proposed a contrast measure RAMMG that subsamples the input image in order to generate multiple pyramid images in the CIELAB space with a nearest neighborhood algorithm. Then the local contrast is calculated by summing up the absolute differences between one pixel and its surrounding pixels in every channel and at every pyramid level. The local contrast values from the same channel are normalized and finally weighted. The final global contrast is the mean of outputs from all levels:

$$C^{\mathrm{RAMMG}} = \frac{1}{N_L} \sum_{i=1}^{N_L} \sum_{j=1}^{3} \sum_{k=1}^{N_p} W_j C_k,$$

where N_L stands for the number of pyramid levels, N_p stands for number of pixels in each pyramid image, C_k stands for the local contrast for each pixel and its surroundings, and W_j stands for the weighting factor which needs to be determined for each CIELAB channel. Inspired by the RAMMG measure, Simone et al.[16] proposed a measure RSC which employs the DOG formula. They did not merely recombine the mean of averaged local contrast from each pyramid level in the lightness channel, but also in the chromatic channel.

In summary, the existing local contrast measures account for luminance and chrominance components as well as the frequency component in the input images to determine the local contrast at various granularities. The global contrast is eventually determined by pooling local contrast values. In recent years, there has been a general increasing interest in incorporating low-level neuron science knowledge to improve contrast models further.

EXPERIMENTAL SETUP AND PROCEDURE

In order to enhance the validity, reliability, and robustness of our research, we performed the experiments under the same viewing conditions but at two separate geographical locations with two different projection displays and one group of observers at each location. In this case, we had two separate experimental sessions in total.

Experimental Setup

The first experimental session was conducted at the University of Burgundy in France with 10 observers (6 males and 4 females, age from 24 to 33), and we used a portable three-chip LCD projector, a Mitsubishi XL9 (1024 × 768) to display images on the screen. The second experimental session was conducted at the Gjøvik University College in Norway with 17 observers (14 males and 3 females, age from 25 to 53), and we used another three-chip LCD projector, a SONY APL-AW15 (1280 × 768). All observers were confirmed to have neither myopia vision nor color deficiency. Both the Mitsubishi and SONY projectors are three-chip LCD based, and they represent one dominant projector category in the current consumer market. However, the Mitsubishi projector has a more powerful bulb and it appears to be much brighter than the SONY projector in the default settings. Consequently, the Mitsubishi projector suffers from a stronger light leaking problem. The Mitsubishi projector appears to be optimized for document presentation automatically so that the color of the displayed pictures appear to be more bluish. The two projectors are widely used by people for meetings and presentations on a daily basis; their device status is "natural" so that the corresponding perceived contrast is close to what we expect to experience in real practice. Other aspects of the experimental sessions were exactly the same. The projector was placed on a flat table in front of the projection screen at a distance of 3 m away (Figure 1(a)). In our experiments, we were simulating a typical home-theater-like environment. All observers sat in a dark room at an equal distance from the screen. The viewing distance, projection area, and visual angles were all fixed. It is possible for observers to sit closer to or get further away from the screen in practice, but in that case the experimental environment is totally different and the underlying research should be extended to consider additional factors (non-uniform sunlight illumination in a daylight meeting room, for example). In this experiment, since all observers were confirmed to have no myopia or color deficiency difficulty, the visual acuity for them was approximately the same. In the objective experiments, the camera was replaced by the observers (Fig. 1(b)). The principal projection axis is pointed at and is perpendicular to the screen center. On the screen, the projection size was approximately 2 × 1.5 m. The projector was connected to a controlling laptop with a VGA cable. In order to minimize the influence of projector temporally stability, the projector lamp was warmed up at least one

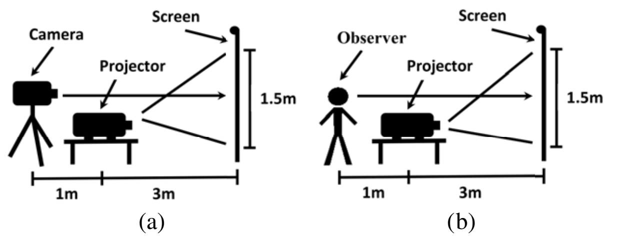


Figure 1. The experimental setup for both experimental sessions. The projectors were placed at a distance of 3 m away from the screen. The camera and observers were located at the same position, which was about 4 m away from the screen. The experiments were conducted in a home-theater-like environment which was typical for projection systems. (a) The setup for the camera and the projectors. (b) The setup for the observers and the projectors.

hour in advance. All settings related to projector brightness, contrast, and color enhancements were switched off to make sure that the input image was projected as it was. In this case, we assumed that the projection displays were uniform in terms of both their luminance and chromatic nature.

For all experiments, we used the same camera to capture the projections. We used a Nikon D610 DSLR camera with imaging resolution of 6016 × 4016 and with a VR 18–100 mm f/3.5–5.6G (VR off) lens to capture the images. We set the camera on a tripod and placed the camera approximately at the height of the observers. The pictures were always taken remotely with software installed on the controlling laptop without physically touching the camera. We selected 10 test images (Figure 2) from the Colour Lab Image Database: Image Quality[39] with respect to their image content (800 × 800 in pixels), so we can cover as many features as we may have in the natural images. The features are, for example highlight/lowlight components, wide range/dominant colors, and large smooth/texture areas. We normalized the RGB values of all pixels in the test images and transformed them in each color channel simultaneously with the formula

$$S_i = (C_i - m) * (j + 6)/6 + m,$$

where S_i stands for the scaled RGB value for the ith pixel in the distorted image, j is an integer scaling factor for contrast distortion in the range $[-3, 3]$, C_i stands for the normalized input RGB value for the ith pixel in the input image, and m stands for the mean of all C_i in the same color channel, so we obtained seven distortion levels for each test image (Figure 3). Any overscaled RGB values (either larger than 1 or smaller than 0) were clipped.

Since the main goal of this research was to evaluate the performance of contrast measures, we only needed to produce multiple levels of contrast distortions, and certain perceptual contrast distances were expected between consecutive distortion levels. A linear contrast tuning is sufficient for achieving the goal without introducing brightness differences between the distorted images, so the variances of perceived contrast due to the perceptual adaptation of luminance are minimized. It is possible to tune the contrast with respect to other types of curves like sigmoid curves; however, the tuning is not expected to significantly

Figure 2. The thumbnails of the 10 test images. We generated 7 levels of contrast distortions of each test image, so there are 70 distorted images in total for each observers to evaluate. The test images were carefully selected to cover many features such as highlight/lowlight components, wide range/dominant colors, and large smooth/texture areas. (a) 1st test image, (b) 2nd test image, (c) 3rd test image, (d) 4th test image, (e) 5th test image, (f) 6th test image, (g) 7th test image, (h) 8th test image, (i) 9th test image, (j) 10th test image.

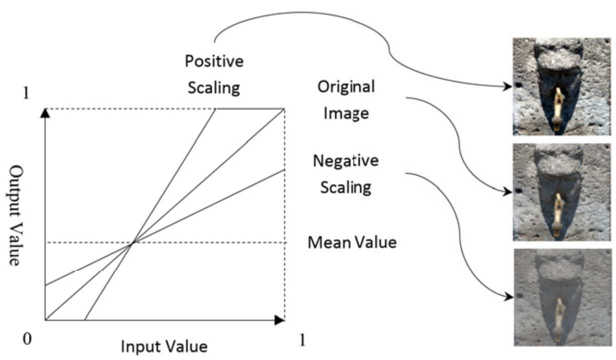


Figure 3. The linear transformation function. Positive scaling represents an enhancement of the actual image contrast, while negative scaling represents a decrement. After the scaling, the mean luminance remains the same. All overscaled values were clipped.

affect the rank order of measured contrast, which is an important aspect of determining correlation coefficients.

Experimental Procedure

Subjective Experiment

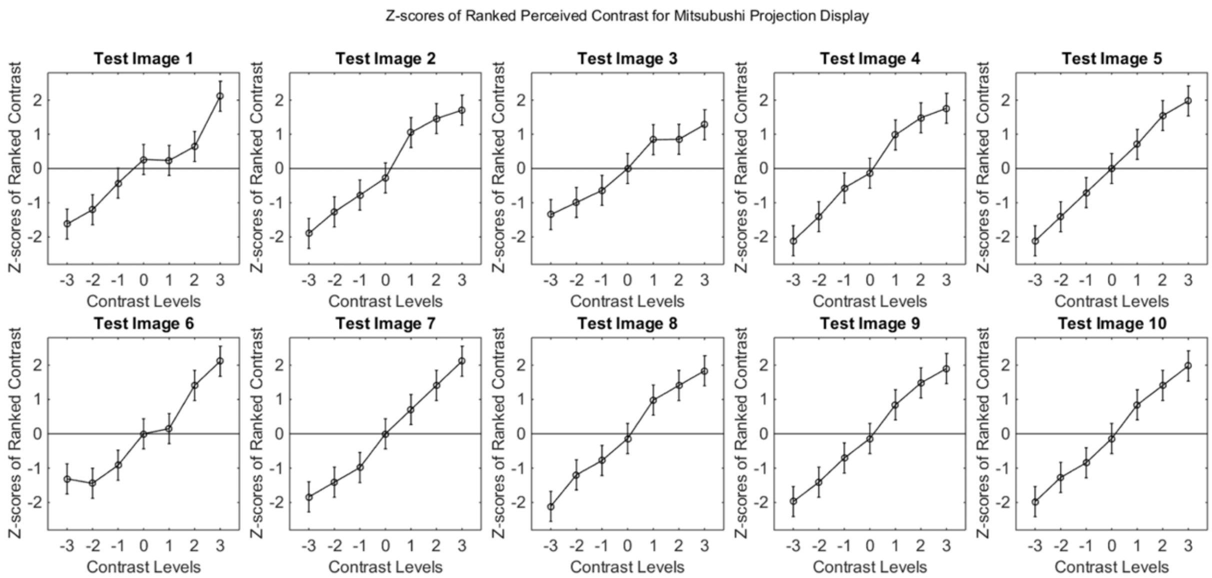
The subjective experiment was conducted by using the software QuickEval,[40] which is an interactive software running on the controlling laptop for psychometric scaling experiments. The software interacts with users within a standard web browser. All observers operate directly on the laptop and they are experiencing exactly the same stimulates in identical viewing conditions. In short, the experiment is performed locally in a controlled manner and it is very different from many typical web-based perceptual experiments.[41,42] Based on this system, each observer is required to perform two tasks. In the first assignment, we display each group of distorted images in randomized order (corresponding to the same input image) on the projection screen at the same time (Figure 4). The observers need to rank this group of distorted images in a descending order

Figure 4. A screenshot of the software QuickEval, which is an interactive software running on the controlling laptop for psychometric scaling experiments. The software interacts with users within a standard web browser. Each observer is required to rank the displayed images with respect to either perceived or preferred contrast. The images at the bottom are distortion thumbnails, and the two larger windows on the top are used to display images of interest in their original size.

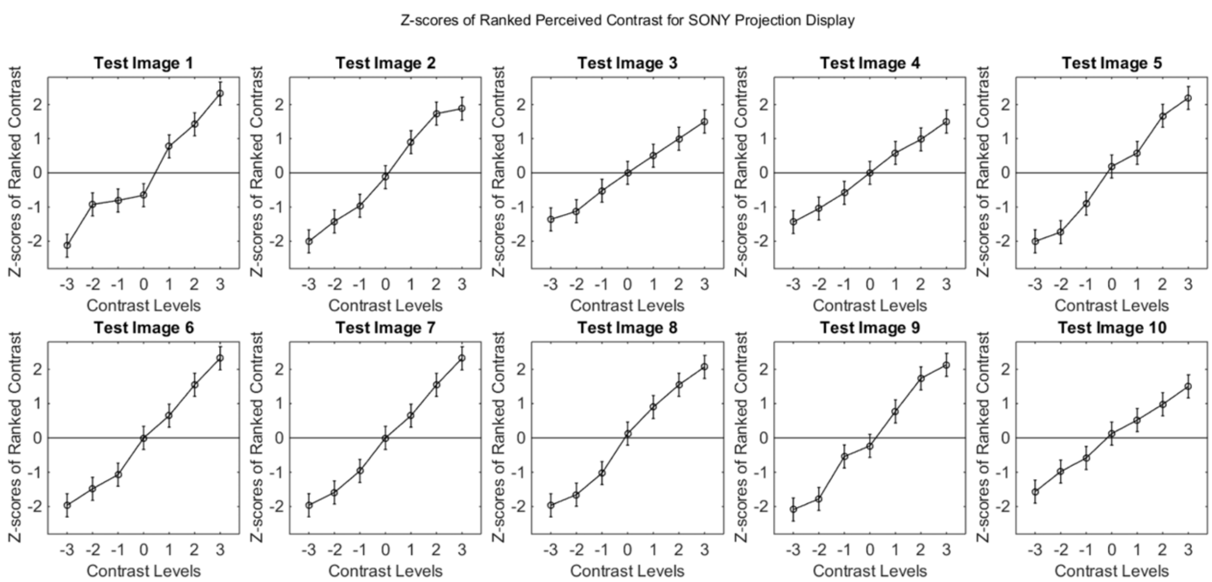
based on their perceived contrast; so the images with higher contrast should be ranked to the left, while the rest with lower contrast will be ranked to the right. In the second assignment, we display the images in the same way but we require the observers to rank each group of distorted images in descending order with respect to their own preference of contrast. The images with the preferred contrast should be ranked to the left, while the rest with less preferred contrast should be ranked to the right. The two windows on the top of the software are used to display observers' selected images at their original size. The software automatically records the ranking results and exports them as a rating matrix in the final report.

Objective Experiment

For the objective experiment part, we used a camera as an acquisition device and further processed these captured images with all types of contrast measures. We set the camera



(a)

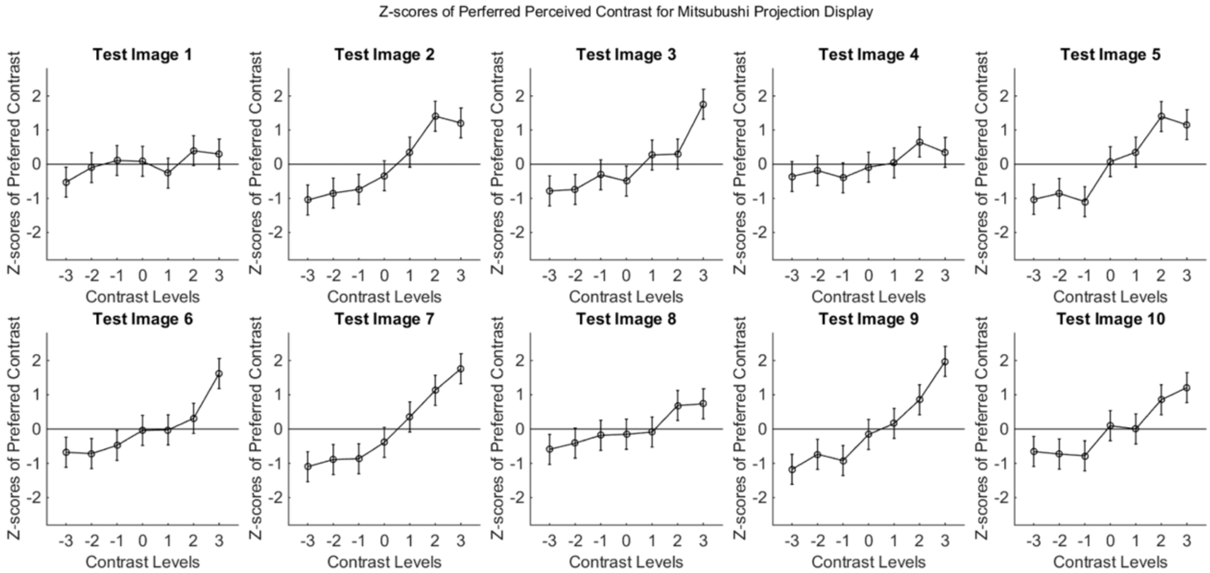


(b)

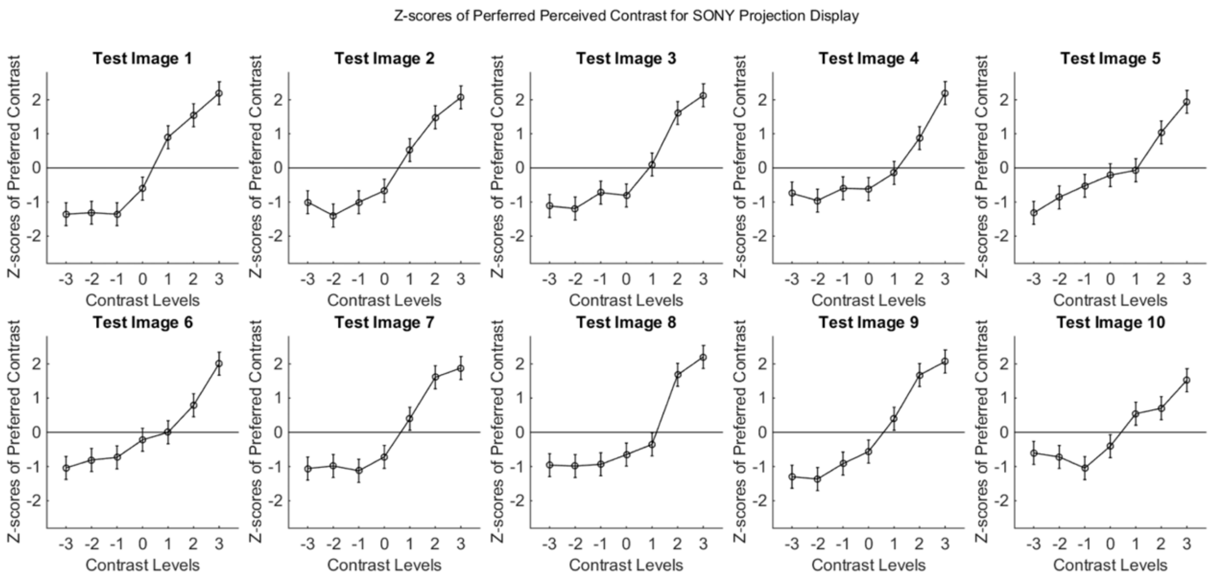
Figure 5. The Z-scores of ranked perceived contrast. The X labels stand for contrast distortion levels in the range $[-3, 3]$. The number 0 stands for the original image. All plots have identical Y value ranges. The circles stand for each Z-score of mean ratings for the distorted image, and the vertical bars indicate the 95% confidence interval[46] of Z-scores as $1.96 \times (1/\sqrt{N})$, where N stands for the number of observers. (a) Mitsubishi XL9 projection display. (b) SONY APL-AW15 projection display.

up with ISO 100 to minimize the camera sensor noise, and performed a standard MTF test[43] in order to acknowledge that the best aperture for the underlying camera and lens was $f/7.1$. We set the shutter speed at a certain value at the initial state and we took several pictures of the peak white projection and observed their histograms. We adjusted the shutter speed setting iteratively to make sure that no camera sensor was either underexposed or overexposed. Since we capture images in raw format, we can apply the spot white balance algorithm to determine the linear scaling factors of the RGB channels respectively. Then we apply these factors to linearly scale all subsequent pictures we take in order to correct the captured luminance. Captured pictures are known to have a vignetting effect, namely an undesirable gradual intensity fall off from the image center to its external limits. We incorporated the method proposed in our previous research[44] to correct camera vignetting based on the captures of a hazy sky which is closely uniform in gray.

In the experiment, we used the following image quality measures to evaluate the image contrast: Michelson contrast,[21] RMS,[25] Lab variance,[26] RAMMG,[38] RSC,[16] and GCF.[36] The Michelson contrast measure was selected because it is representative of global contrast measurement and it is typically used as a reference for contrast measurement in research. RMS and LAB variance measures are selected because they are representative of measurements which



(a)



(b)

Figure 6. The Z-scores of preferred perceived contrast. The X labels stand for contrast distortion levels in the range $[-3, 3]$. The number 0 stands for the original image. All plots have identical Y value ranges. The circles stand for each Z-score of mean ratings for the distorted image, and the vertical bars indicate the 95% confidence interval of Z-scores as $1.96 \times (1/\sqrt{N})$, where N stands for the number of observers. (a) Mitsubishi XL9 projection display. (b) SONY APL-AW15 projection display.

account on statistics; however the RMS measure works only on luminance, while the LAB variance measure further take colors into account in the perceptual uniform CIELAB color space. RAMMG and RSC measures are representative of the measures incorporating low-level visual system models. The GCF measure addresses the problem from the spatial frequency perspective.

EXPERIMENTAL RESULTS

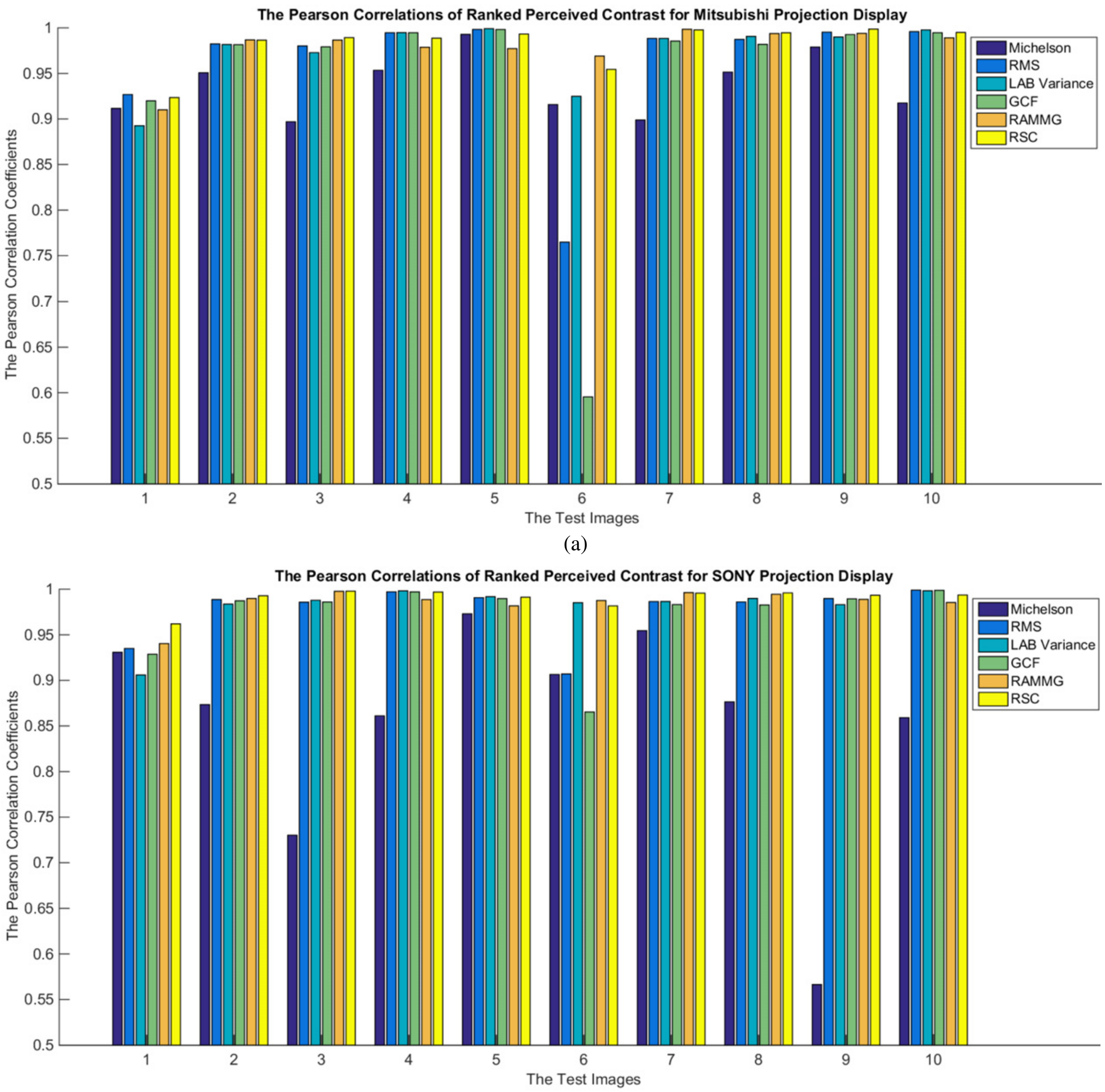
We collected raw subjective ratings, scaled them, and calculated the Z-scores;[4] meanwhile, we processed them with selected image quality measures in order to evaluate the image contrast.

Subjective Results

We collected the subjective ratings for the ranked perceived contrast and preferred perceived contrast, respectively. All collected raw ratings were scaled in order to calculate their Z-scores.

Ranked Perceived Contrast

The Z-scores of ranked perceived contrast for the projectors are shown in Figure 5. It is clear that the rank of perceived contrast has a closely linear relationship with the actual rank of modified contrast. Since the Z-score values in all plots are monotonically increasing, the relationship between perceived contrast and the actual image contrast is almost linear for all types of images.



(a)

(b)

Figure 7. The Pearson correlations between the mean Z-scores of ranked perceived contrast and the measurement scores for the ten selected test images for (a) the Mitsubishi projection display and (b) the SONY projection display. The Y values are limited to between 0.5 and 1.

Preferred Perceived Contrast

The Z-scores of preferred perceived contrast for the projection displays are shown in Figure 6. The general tendency of the Z-scores of preferred contrast no longer follows a linear relationship with the actual image contrast. This observation suggests that the observers tend to rank all distortions into two groups: either relatively less preferred (contrast level -3 to -1) or more preferred perceived contrast (contrast level 1 to 3). In the group of less preferred contrast, since the confidence intervals of Z-scores are largely overlapped, the perceived contrasts have no significant difference, while in the group of more preferred contrast, the confidence intervals are less overlapped. This suggests that the majority of observers prefer the enhanced contrast even though the luminance has been overscaled. In some cases, for both projectors, the contrast level 0, which stands for the original image, is neither preferred nor not preferred because it is very close to the center line for all test images. The preferred perceived contrast values for the two projectors are obviously different.

Objective Results

The evaluations of the objective contrast measures are presented first for the ranked contrast, and then for the preferred perceived contrast for the two projection displays.

Ranked Perceived Contrast

We applied the measures to all modified images to calculate the objective scores, and to determine the Pearson correlation

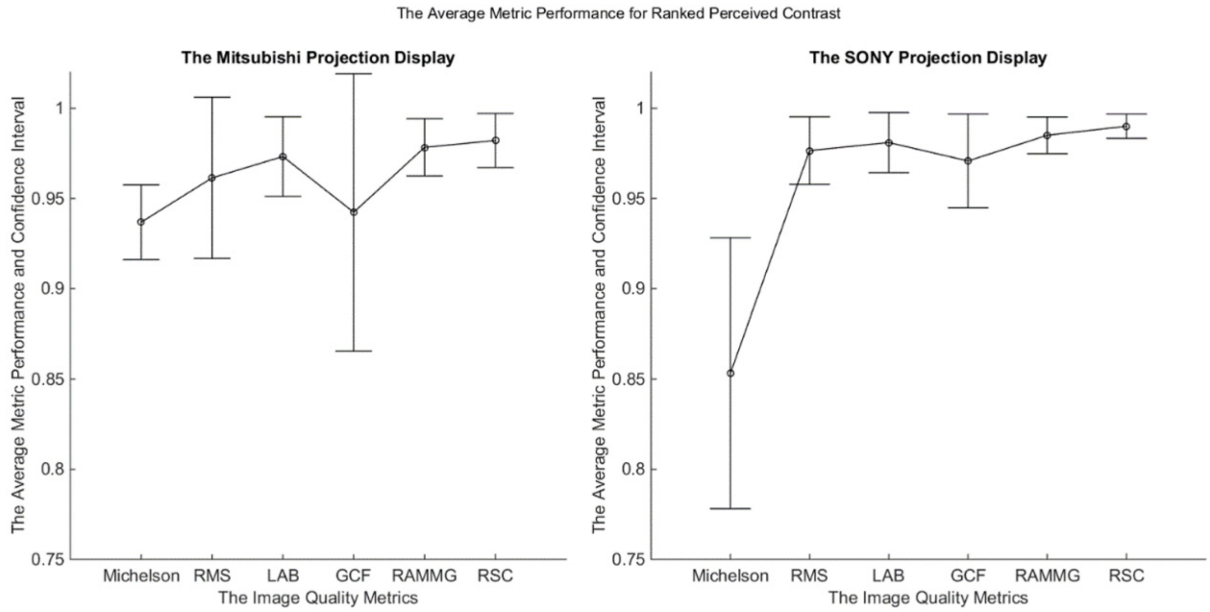


Figure 8. The average performance for ranked perceived contrast with respect to distribution of their Pearson correlation coefficients. The circles indicate the mean of correlation coefficients which are calculated on a per image basis. The bars stand for the 95% confidence interval.

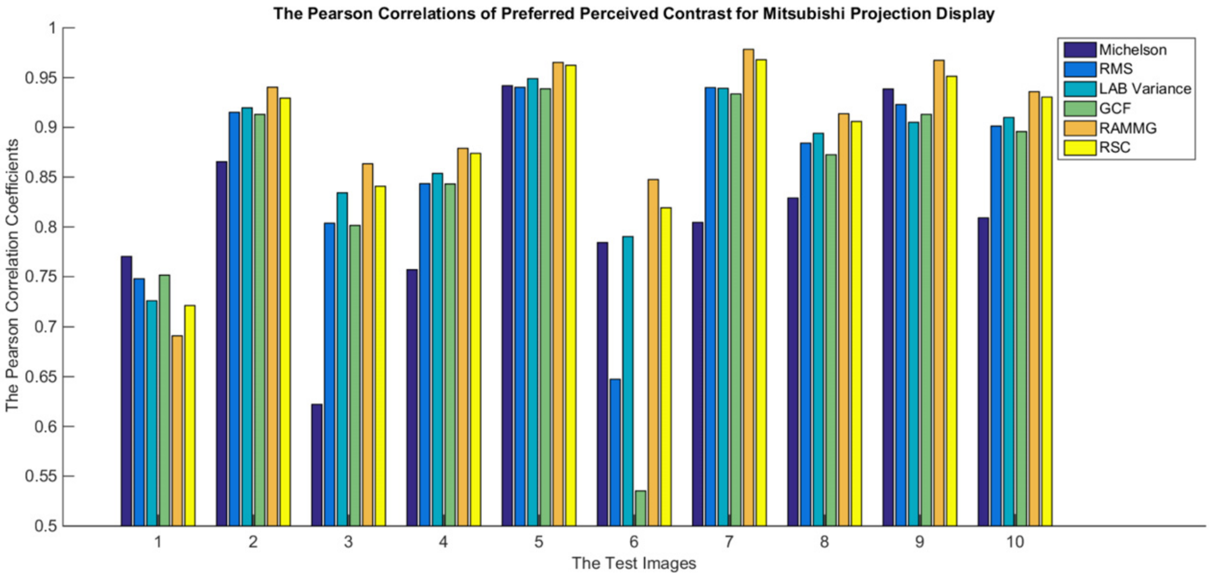
coefficients between the objective scores and the mean Z-scores of both ranked perceived contrast (Figure 7). Based on the observation, it is clear that, for the Mitsubishi projection display, most contrast measures produce high correlation coefficients above 0.85, except that the RMS and GCF measures produce low coefficients on test image 6. However, the observation cannot be obtained from the correlation results for the SONY projector. For the SONY projection display, the Michelson contrast measure performs relatively worse than the other contrast measures, and this is especially true for test image 9. Other contrast measures have very similar performance for both Mitsubishi and SONY projection displays on test images 2, 3, 4, 5, 7, 8, 9, and 10, but not on test images 1 and 6. It is not very clear which measure has the best performance in general. In this case, we generated the box plots of the Pearson correlation coefficients over all test images for both projection displays (Figure 8). It is clear that the Michelson contrast measure performs worse than other contrast measures, not merely because it has a low average correlation value around 0.85, but also its 95% confidence interval is much larger. For the Mitsubishi projection display, the contrast measure GCF performs badly with respect to its confidence interval as well. Although the Mitsubishi and SONY projection displays are supposed to produce different contrast on the screens, the mean of correlation coefficients over all test images follow a very similar general tendency. Based on the observation on the variance of confidence intervals, the RSC contrast measure produces the most stable outcome regardless of the actual image content.

Preferred Perceived Contrast

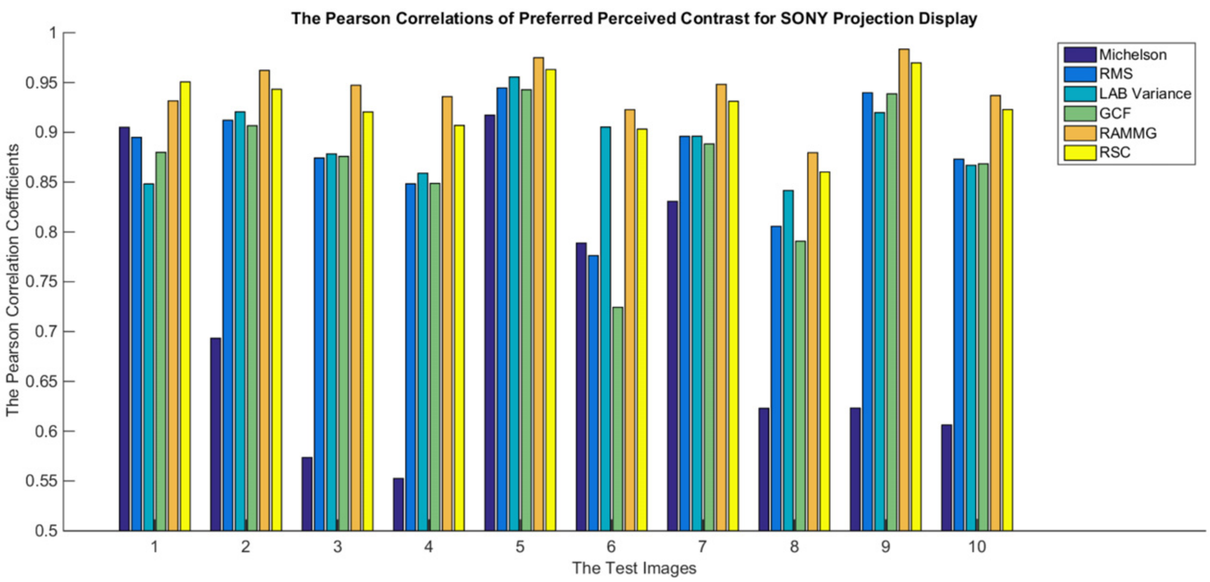
For the preferred perceived contrast, we followed a similar approach to calculate the Pearson correlation coefficients for all contrast measures on all test images (Figure 9). It is clear that the Michelson contrast measure performs the worst for both ranked and preferred contrast. In addition, the RMS and GCF measures both perform relatively worse for test image 6 for the two projection displays as well. For the preferred contrast of both Mitsubishi and SONY projection displays, the RAMMG and RSC still have the highest correlations; however, the correlation from the RAMMG contrast measure is slightly higher than that for the RSC contrast measure. This observation is different from the one for ranked perceived contrast. The rank order between RMS, LAB, GCF, RAMMG, and RSC contrast measures is largely preserved for test images 2, 3, 4, 5, 7, 8, 9 and 10, but not for test images 1 and 6. This observation can be obtained from the ranked perceived contrast for both projection displays as well, but not from the preferred contrast for the SONY projection display. By looking at the average overall contrast measurement performance shown in Figure 10, the general tendency of the average Pearson correlation over all test images is almost the same as the one obtained from the preferred contrast.

Overall Results

In Figs. 8 and 10, we showed the average performance of the contrast measures over all test images for each projection display. In this case, we calculated the Pearson correlation coefficients not on a per image basis but we did the calculation over all test images, so we could observe how the metrics performed regardless of the image content (Figure 11). In Fig. 11 we can see that, for the Mitsubishi projection display, the mean correlation coefficients are almost identical, and the 95% confidence intervals are largely overlapped for both ranked and preferred perceived contrast. This indicates that for the Mitsubishi projection display the contrast measurements have almost the same



(a)



(b)

Figure 9. The Pearson correlations between the mean Z-scores of preferred perceived contrast and the measurement scores for the ten selected test images for (a) the Mitsubishi projection display and (b) the SONY projection display.

performance. However, for the SONY projection display, the Michelson contrast measure performs relatively worse. For both projection displays, the ranked and preferred perceived contrasts share a similar general tendency.

We also calculated the Pearson correlations between the average performances over all contrast measurements with respect to their types of contrast versus the types of projection displays. The results are shown in Table I. Considering that the Michelson contrast measure produces low correlation coefficients and large variances for most test images for all projection displays, we removed the Michelson contrast measure and recalculated the data; the results are shown in Table II.

On looking at the data in Table I, it is clear that the average performances of ranked and preferred perceived contrast measurements have high correlations; they are all above 0.9. However, there is no evidence to indicate that there is any good relationship for ranked or preferred contrast between one projection display and another, since their correlation values are all below 0.85. After taking the Michelson contrast measure away, the low coefficients presented in

Table I. The Pearson correlations between the average correlation coefficients.

Contrast/Projector	Ranked, Mit.	Ranked, SONY	Preferred, Mit.	Preferred, SONY
Ranked, Mit.	1	0.7431	0.9629	0.8396
Ranked, SONY	0.7431	1	0.8155	0.9668
Preferred, Mit.	0.9629	0.8155	1	0.9248
Preferred, SONY	0.8396	0.9668	0.9248	1

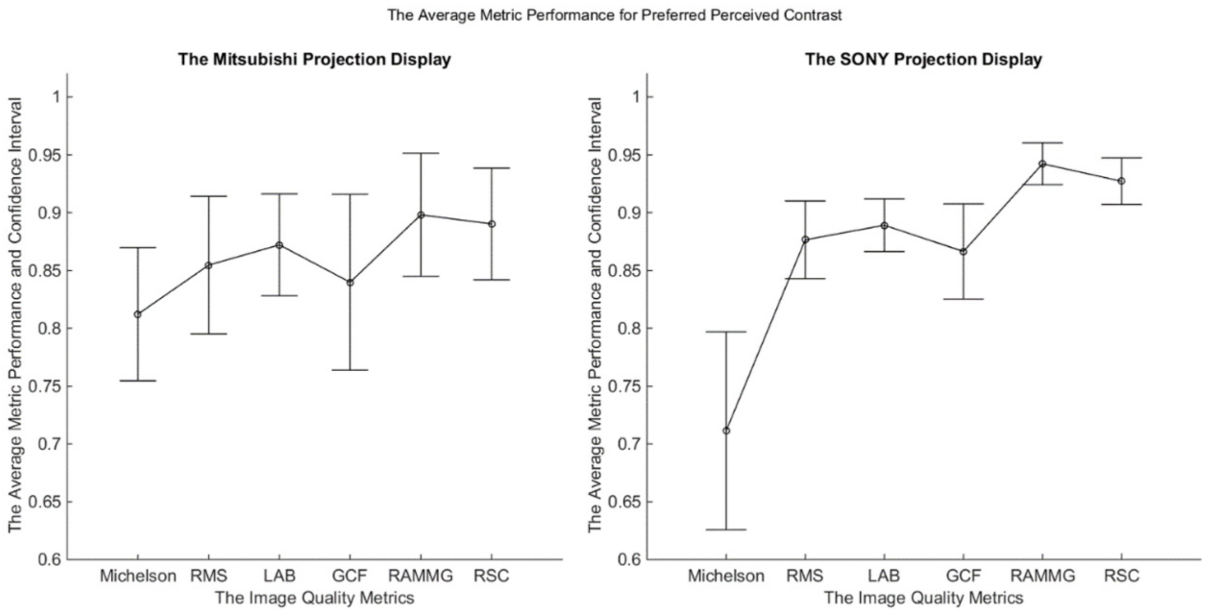


Figure 10. The average measurement performance for ranked perceived contrast with respect to the distribution of their Pearson correlation coefficients. The circles indicate the mean of the correlation coefficients which are calculated on a per image basis. The bars stand for the 95% confidence interval.

Table II. The Pearson correlations between the average correlation coefficients without the Michelson contrast measure.

Contrast/Projector	Ranked, Mit.	Ranked, SONY	Preferred, Mit.	Preferred, SONY
Ranked, Mit.	1	0.9608	0.9409	0.8424
Ranked, SONY	0.9608	1	0.9354	0.8829
Preferred, Mit.	0.9409	0.9354	1	0.9696
Preferred, SONY	0.8424	0.8829	0.9696	1

Table I increase by a certain amount, and their values are all above 0.9 in Table II. However, the correlation coefficients of both ranked and preferred contrast between different projection displays show no significant improvement. This observation suggests that human preference on the perceived contrast has a closely linear relationship with the ranked perceived contrast. In this circumstance, we conclude that the most preferred perceived contrast corresponds to the highest ranked perceived contrast; even for test images 1, 2, 4, and 5 in Fig. 6(a) the highest preferred perceived contrast corresponds to the second highest ranked perceived contrast. For related research in the future it is unnecessary to explicitly distinguish them and do the experiments twice.

CONCLUSION AND FUTURE WORKS

In this paper, we have proposed an objective approach to measure the relative contrast of projection displays based on pictures taken with a calibrated digital camera in a controlled environment. To the best knowledge we have, this is the first research regarding evaluating the perceived contrast on projection displays based on the images captured with a calibrated camera. This objective approach can be easily extended to measure other image quality attributes such as sharpness and non-uniformity for all types of information displays. The metric performance evaluation is based on two separate projection displays, so the validity and reproducibility of the research have been enhanced. The research feasibility is supported by the high correlations between subjective and objective experimental results, as well as the correlations between the two projection displays. We classified the contrast measures into local and global categories. For each category, we selected the representative contrast measures and evaluated their performance with respect to the Pearson correlations between subjective and objective assessment results. The experimental results based on two separate projection displays suggest that the Michelson contrast measure has very low performance over all test images, as expected. Other global contrast measures (RMS and LAB) also perform relatively better than the Michelson contrast measure, but they have less correlation with the perceptual ratings compared to the local contrast measures. The local contrast measure GCF has similar performance to the RMS and LAB measures, but it performs worse than other local contrast measures (RAMMG and RSC). The contrast measures RAMMG and RSC perform the best overall, and they have very close performance on contrast measurements for almost all test images. With respect to the 95% confidence interval of the average measurement performance over all test images, RAMMG has slightly improved correlations with the preferred contrast. It is interesting to see that many contrast measures do not perform well on the test images 1 and 6. These two images either have large area of low luminance component or dominant color component, and they do not have obvious texture area. We recommend local contrast measures incorporating low-level human visual system models since they have better overall

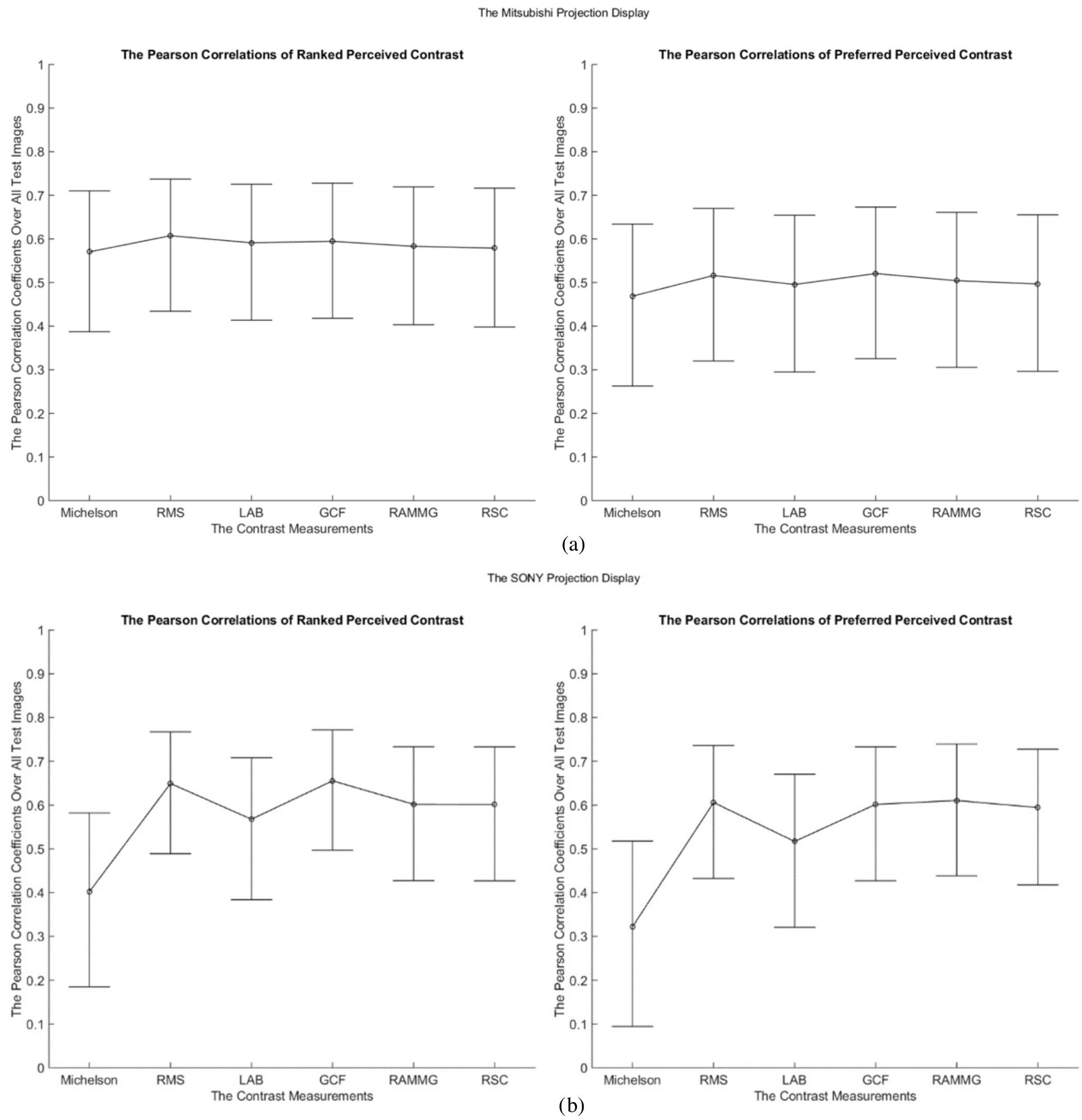


Figure 11. The average measurement performance for ranked perceived contrast with respect to the distribution of their Pearson correlation coefficients over all test images. The circles indicate the mean of correlation coefficients which are calculated on a per image basis. The bars stand for the 95% confidence interval calculated based on Fisher Z transformation.[47] (a) The Mitsubishi projection display. (b) The SONY projection display.

performance over global contrast measures in terms of both contrast prediction accuracy and stability regardless of the image content. Since the average correlations and stability of local contrast measures are good for many test images, we do not need to propose a new contrast measure, but rather to improve the models of human visual system to predict the image contrast better in future research.

REFERENCES

[1] O. Bimber, A. Emmerling, and T. Klemmer, "Embedded entertainment with smart projectors," IEEE Comput. **38**, 48–55 ACM Press, New York, USA (2005).

[2] W. Zou and H. Xu, "Colorimetric color reproduction framework for screen relaxation of projection display," Displays **32**, 313–319 (2011).

[3] O. Bimber and E. Andreas, "Multifocal projection: a multiprojector technique for increasing focal depth," IEEE Trans. Vis. Comput. Graphics **12**, 658–667 (2006).

[4] B. Sajadi, M. Lazarov, A. Majumder, and M. Gopi, "Color seamlessness in multi-projector displays using constrained gamut morphing," IEEE Trans. Vis. Comput. Graphics **15**, 1317–1325 (2009).

[5] M. Pedersen, N. Bonnier, J. Y. Hardeberg, and F. Albregtsen, "Attributes of image quality for color prints," J. Electron. Imaging **19**, 011016 (2010).

[6] J. Gille, L. Arend, and J. Larimer, "Display characterization by eye: contrast ratio and discrimination throughout the grayscale," Proc. SPIE **5292**, 218–233 (2004).

[7] D. H. Brainard, D. G. Pelli, and T. Robson, "Display characterization," *Imaging Science and Technology* (Wiley, 2002), pp. 172–188.

[8] M. Fairchild and D. R. Wyble, *Colorimetric Characterization of The Apple Studio Display (Flat panel LCD)* (New York, 1998).

9. A. M. Bakke, J.-B. Thomas, and J. Gerhardt, "Common assumptions in color characterization of projectors," *Gjøvik Color Imaging Symposium* (Gjøvik, 2009), pp. 45–53.
10. A. Majumder, "Contrast enhancement of multi-displays using human contrast sensitivity," *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition 2* (Providence, RI, 2005), pp. 377–382.
11. J.-B. Thomas, *Colorimetric Characterization of Displays and Multi-display Systems* (Université de Bourgogne, 2009).
12. A. Majumder, "Modeling color properties of tiled displays," Comput. Graph. Forum **24**, 149–163 (2005).
13. J. Y. Hardeberg, I. Farup, and G. Stjernvang, "Color quality analysis of a system for digital distribution and projection of cinema commercials," SMPTE Motion Imaging **114**, 146–151 (2005).
14. E. Peli, "Contrast in complex images," J. Opt. Soc. Am. A **7**, 2032–2040 (1990).
15. M. Pedersen, N. Bonnier, J. Y. Hardeberg, and F. Albregtsen, "Attributes of a new image quality model for color prints," *Proc. IS&T/SID Seventeenth Color and Imaging Conf.* (IS&T, Springfield, VA, 2009), pp. 204–209.
16. G. Simone, M. Pedersen, and J. Y. Hardeberg, "Measuring perceptual contrast in digital images," J. Vis. Commun. Image Represent. **23**, 491–506 (2012).
17. A. Majumder and R. Stevens, "Color nonuniformity in projection-based displays: analysis and solutions," IEEE Trans. Vis. Comput. Graphics **10**, 177–188 (2003).
18. M. Brown, A. Majumder, and R. Yang, "Camera-based calibration techniques for seamless multiprojector displays," IEEE Trans. Vis. Comput. Graphics **11**, 193–206 (2005).
19. J.-B. Thomas, "Webcam based display calibration," *Proc. IS&T/SID Twentieth Color and Imaging Conf.* (IS&T, Springfield, VA, 2012), pp. 82–87.
20. G. Simone, M. Pedersen, J. Y. Hardeberg, and A. Rizzi, "Measuring perceptual contrast in a multi-level framework," Proc. SPIE **7240**, 72400Q (2009).
21. M. A. Abraham, *Studies in Optics* (Dover, 1927).
22. P. Whittle, "The psychophysics of contrast brightness," *Lightness, Brightness, and Transparency* (Lawrence Erlbaum Associates, New Jersey, USA, 1994), pp. 35–110.
23. P. E. King-Smith and J. J. Kulikowski, "Pattern and flicker detection analysed by subthreshold summation," J. Physiol. 519–548 (1975).
24. D. A. Burkhardt, J. Gottesman, D. Kersten, and G. E. Legge, "Symmetry and constancy in the perception of negative and positive luminance contrast," J. Opt. Soc. Am. A **1**, 309–316 (1984).
25. M. Pavel, G. Sperling, T. Riedl, and A. Vanderbeek, "Limits of visual communication: the effect of signal-to-noise ratio on the intelligibility of American sign language," J. Opt. Soc. Am. **4**, 2355–2365 (1987).
26. M. Pedersen, A. Rizzi, J. Y. Hardeberg, and G. Simone, "Evaluation of contrast measures in relation to observers perceived contrast," *Proc. IS&T CGIV2008: Fourth European Conf. on Colour in Graphics, Imaging and Vision* (IS&T, Springfield, VA, 2008), pp. 253–256.
27. E. B. Goldstein, *Sensation and Perception*, 8th ed. (Cengage Learning, Belmont, CA, USA, 2009).
28. R. L. DeValois and K. K. DeValois, *Spatial Vision* (Oxford University Press, 1988).
29. M. D. Fairchild, *Color Appearance Models*, 2nd ed. (John Wiley & Sons, Ltd, 2005).
30. "TCO certified displays 6.0," pp. 1–125, TCO Development AB, Stockholm, Sweden (2012).
31. "SPWG notebook panel specification 3.8," pp. 1–59, Standard Panel Working Group, Newark, USA (2007).
32. "Information display measurements standard 1.03," Campbell, CA, USA (2012).
33. "ISO9241 Ergonomics of human-system interaction - Part 307: Analysis and Compliance Test Methods for Electronic Visual Displays," pp. 1–217, International Standard Organization (2008).
34. G. Boccignone, M. Ferrearo, and T. Caelli, "Encoding visual information using anisotropic transformations," IEEE Trans. Pattern Anal. Mach. Intell. **23**, 1–16 (2001).
35. A. J. Ahumada Jr., "Simplified vision models for image quality assessment," *SID Int. Symp. Dig. Tech. Pap. 27* (1996), pp. 397–402.
36. K. Matkovic, L. Neumann, A. Neumann, T. Psik, and W. Purgatholer, "Global contrast factor – a new approach to image contrast," *Eurographics Workshop on Computational Aesthetics in Graphics, Visualization and Imaging – Computational Aesthetics*, edited by L. Neumann, M. S. Casasayas, B. Gooch and W. Purgathofer (Girona, Spain, 2005), pp. 159–167.
37. Y. Tadmor and D. J. Tolhurst, "Calculating the contrasts that retinal ganglion cells and LGN neurones encounter in natural scenes," Vision Res. **40**, 3145–3157 (2000).
38. A. Rizzi, T. Algeri, G. Medeghini, and D. Marini, "A proposal for contrast measure in digital images," *Proc. IS&T CGIV2004: Second European Conf. on Colour in Graphics, Imaging, and Vision* (IS&T, Springfield, VA, 2004), pp. 187–192.
39. X. Liu, M. Pedersen, and J. Y. Hardeberg, "CID:IQ – a new image quaity database," *Int'l Conf. on Image and Signal Processing 8509*, edited by A. Elmoataz, O. Lezoray, F. Nouboud and D. Mammass (Springer International Publishing, Cherbourg, Normandy, France, 2014), pp. 193–202.
40. K. Van Ngo, J. Storvik Jr., C. Andre Dokkeberg, I. Farup, and M. Pedersen, "QuickEval: a web application for psychometric scaling experiments," Proc. SPIE **9396**, 93960Q (2015).
41. I. Sprow, Z. Baranczuk, T. Stamm, and P. Zolliker, "Web-based psychometric evaluation of image quality," Proc. SPIE **7242**, 72420A (2009).
42. C. Keimel, J. Habigt, C. Horch, and K. Diepold, "Qualitycrowd – a framework for crowd-based quality evaluation," *Picture Coding Symposium* (IEEE, Krakow, Poland, 2012), pp. 245–238.
43. "ISO 12233:2014 photography – electronic still picture imaging – resolution and spatial frequency responses" (2014).
44. P. Zhao, M. Pedersen, J.-B. Thomas, and J. Y. Hardeberg, "Perceptual spatial uniformity assessment of projection displays with a calibrated camera," *Proc. IS&T Twenty-second Color and Imaging Conf.* (IS&T, Springfield, VA, 2015), pp. 159–164.
45. P. G. Engeldrum, *Psychometric Scaling: A Toolkit for Imaging Systems Development* (Imcotek Pr, 2000).
46. J. Morovic, *To Develop a Universal Gamut Mapping Algorithm* (University of Derby, 1998).
47. The Video Quality Experts Group, "Final Report from The Video Quality Experts Group on The Evaluation of Objective Models of Video Quality Assessment—Full Reference Television Phase I" (2000).