

# Subjective Quality Assessment of Ground-based Camera Images

Lucie L ev eque<sup>1</sup>, Soumyabrata Dev<sup>2,3,4</sup>, Murhaf Hossari<sup>2</sup>, Yee Hui Lee<sup>5</sup>, and Stefan Winkler<sup>6</sup>

<sup>1</sup>Xi'an Jiaotong-Liverpool University, Suzhou, China

<sup>2</sup>ADAPT SFI Research Centre, Dublin, Ireland

<sup>3</sup>School of Computer Science, University College Dublin, Ireland

<sup>4</sup>UCD Earth Institute, University College Dublin, Ireland

<sup>5</sup>Nanyang Technological University Singapore, 639798, Singapore

<sup>6</sup>Department of Computer Science, National University of Singapore, 117417, Singapore

**Abstract**— Image quality assessment is critical to control and maintain the perceived quality of visual content. Both subjective and objective evaluations can be utilised, however, subjective image quality assessment is currently considered the most reliable approach. Databases containing distorted images and mean opinion scores are needed in the field of atmospheric research with a view to improve the current state-of-the-art methodologies. In this paper, we focus on using ground-based sky camera images to understand the atmospheric events. We present a new image quality assessment dataset containing original and distorted nighttime images of sky/cloud from SWINSEG database. Subjective quality assessment was carried out in controlled conditions, as recommended by the ITU. Statistical analyses of the subjective scores showed the impact of noise type and distortion level on the perceived quality.

## 1. INTRODUCTION

With the tremendous advancement in the photogrammetric sensors, there has been a paradigm shift in the manner we analyse the atmosphere. Traditionally, satellite images were the only sources of visual information of the earth's atmosphere [1]. However, these satellite images have low temporal and low spatial resolutions. Therefore, nowadays, researchers use ground-based imaging systems to understand the various events in the atmosphere. These imaging systems are popularly called Whole Sky Imagers (WSIs) [2]. The WSIs capture the images of the atmosphere at regular time intervals, and provide rich information about the clouds and other related atmospheric events.

Such ground-based sky cameras are generally equipped with a wide-angle lens, in order to capture the entire horizon. In addition to the inherent distortion due to the lens, several other noises get inadvertently incorporated in the captured sky/cloud images. This creates a plethora of problems for the remote sensing analysts, while processing the captured images during the post-analysis stage. During the curation and creation of sky/cloud image datasets, the remote sensing analysts choose the images carefully that are considered noise-free. However, no consideration is applied in understanding the impact of noise in the subjective quality of the considered sky/cloud images in the dataset. Such subjective assessment of a noisy sky/cloud image are generally ignored in the literature. Generally, the images in the popular sky/cloud image datasets are chosen carefully, such that they are noise-free. Dandini et al. in [3] used ground-based sky cameras to compute the scattering phase function (SPF) in cirrus clouds. In [4], Pawar et al. proposed a methodology to detect clear sky images in ground-based sky cameras. Also, Dev et al. in [5] provided an overview of the various machine learning techniques in ground-based sky imaging. However, none of these works provide an analysis on the impact of noise in the subjective quality of the ground-based camera images. The visual distortions in the captured sky/cloud image can impact several post-analysis of sky/cloud images viz. cloud coverage computation [6], cloud-type recognition [7] and cloud-base height computation [8]. None of the existing works in the literature provide an understanding on the subjective image quality evaluation of the ground-based observation. And since the quality of the sky images have a direct impact on the information extracted, and also on decisions made based on the analysis of these images, we wanted to introduce a subjective assessment in this field. In this paper, we attempt to bridge this gap, by providing a subjective assessment of the quality of ground-based sky camera images.

The use of image quality assessment is necessary to evaluate the quality of visual content [9]. Two different approaches can be considered for image and video quality assessment, i.e., objective and subjective evaluations [10]. Objective assessment is based on mathematical algorithms which

provide global or local quality measures. This method is reproducible and does not need human observation, as the main goal of an image or video quality assessment metric is to automatically predict the quality of a content. On the contrary, subjective image quality assessment requires observers for a visual study of perceived quality [11, 12]. As human observers are the ultimate receivers of visual content, subjective quality assessment is considered the most reliable approach. Furthermore, subjective data is needed to develop more accurate objective measures. The International Telecommunication Union (ITU) established standardised methods for subjective quality evaluation of images and videos [13]. In this paper, we present a new image quality database composed of 70 stimuli, which consists of different types of noise and levels of degradation. Thirteen observers rated the quality of all images in controlled laboratory conditions.

The main contributions of this paper are as follows:

- We provide the first-of-its-kind subjective image quality assessment of sky/cloud images captured by sky cameras;
- We propose and release a new image quality dataset comprising 70 images with different noise levels; and
- In the spirit of reproducible research, the source code used in generating the stimuli is available online<sup>1</sup>.

## 2. VISUAL QUALITY ASSESSMENT

### 2.1. Stimuli

Figure 1 illustrates the stimuli used for our experiment. It consists of ten nighttime images of sky scene. These ten images were randomly selected from SWINSEG (Singapore Whole sky Nighttime Imaging SEGmentation database) [14] dataset. The images were captured using a ground-based whole sky imager, called Wide Angle High Resolution Sky Imaging System (WAHRISIS) [15]. This ground-based camera captures the images of the atmosphere at regular time intervals, with an interval of five minutes. It captures the images of the sky horizon throughout the 24 hours of the day. In this paper, we focus our attention on only nighttime sky/cloud images. Indeed, nighttime images are noisier in nature because of the high exposure time of the camera, and the low lighting conditions in the surroundings.



Figure 1: Illustration of the original stimuli used in our experiment.

The ten original (i.e., reference) stimuli were distorted using three different types of image noise, i.e., Gaussian noise, salt-and-pepper noise, and speckle noise, at two different levels of distortion. We term the two different levels of noise as *low* noise and *high* noise, depending on the amount of added noise. We create noisy versions of the original stimuli by adding random noise to image with varying magnitude. In the case of images added with Gaussian noise, we add noise with varying degree of standard deviation ( $\psi$ ) in the Gaussian distribution. We set  $\psi = 5$  and  $\psi = 12$  for low- and high-noisy images respectively. We also create noisy images that are noised with salt and

<sup>1</sup><https://github.com/Soumyabrata/cloud-quality>.

pepper noise. In such noisy images, a random proportion of the pixels in a sky/cloud image are assigned to 0 or 255. The salt and pepper noisy images are generated only for grayscale images. We assign varying degree of probabilities to create the noisy version. We set this value of probability  $p$ , as  $p = 0.01$  and  $p = 0.03$  for low- and high-noisy images respectively. Finally, we also generate images noised with speckle noise. We define speckled noisy image as  $\hat{\mathbf{X}} = \mathbf{X} + \Phi \cdot \mathbf{X}$ , where  $\hat{\mathbf{X}}$  is the noisy image,  $\mathbf{X}$  is the reference image, and  $\Phi$  is a matrix of random Gaussian-distributed values with the same dimensions of the reference image. We also set varying values of  $\sigma$  to generate the matrix  $\Phi$ . We set  $\psi = 0.25$  and  $\psi = 0.5$  for the cases of low- and high-noisy images respectively.

Figure 2 represents all the distortion configurations for a given sample stimulus. The salt and pepper noised images are defined only for grayscale images, whereas the Gaussian-noised and speckled noisy images are defined for RGB images. We can visually observe and discern the difference in the amount of added noise for both low- and high-noisy images.

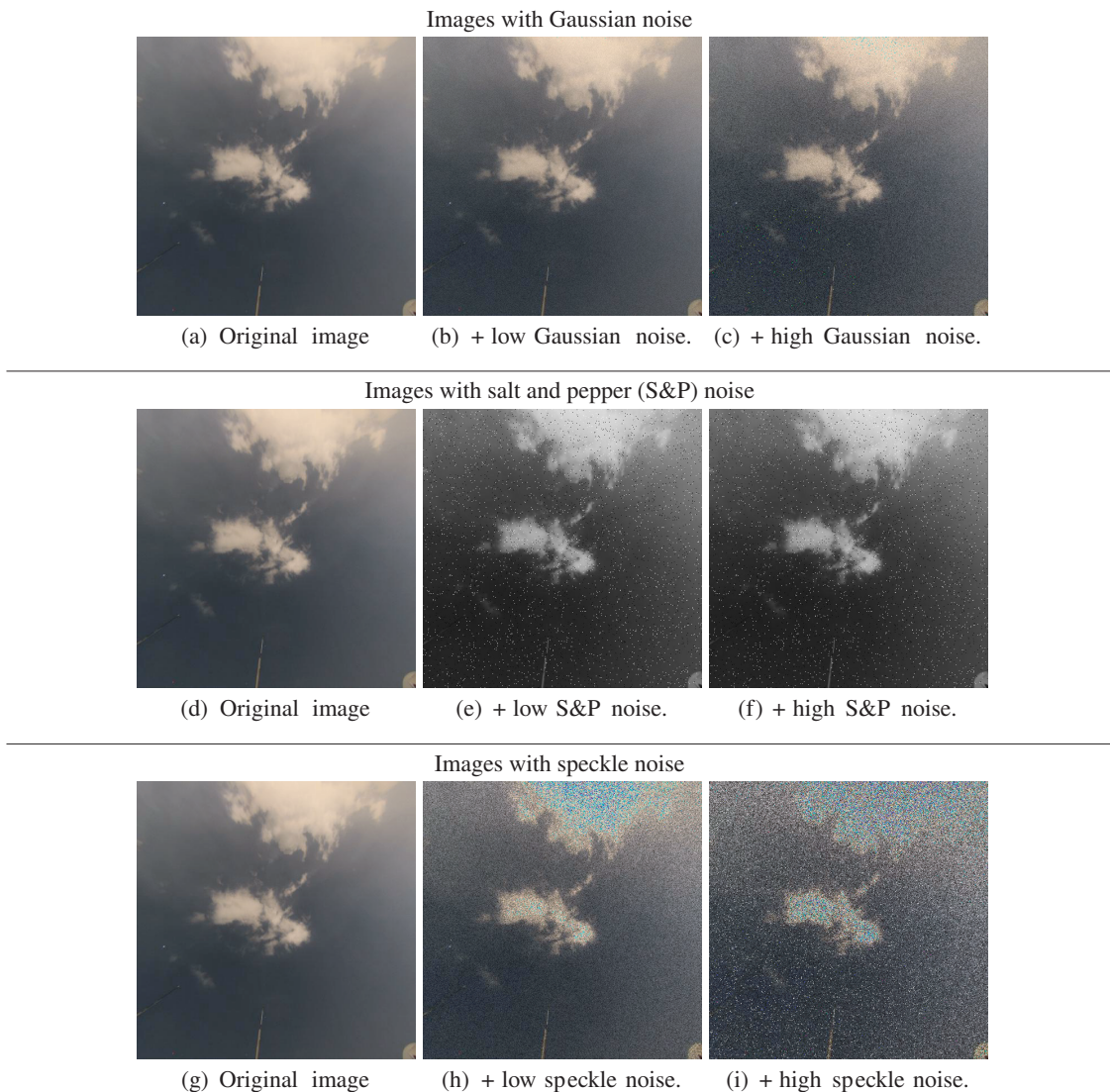


Figure 2: Illustration of various distortion configurations for a given nighttime sky/cloud image.

## 2.2. Experimental Procedure

We conducted a visual perception experiment using a multi-stimulus method [16], where human subjects were asked to score the overall quality of the 70 stimuli by inserting a slider mark on a continuous scale (i.e., from 0 to 100). The quality scale was divided into five semantic portions for scoring image quality: Bad (0–20), Poor (20–40), Fair (40–60), Good (60–80), and Excellent (80–100). Figure 3 illustrates the scoring interface used for the experiment. Figure 3 also represents an example of the test organisation for each source sequence where an explicit reference and six compressed versions placed in a different random order are included. Subjects were allowed to view

and grade any stimulus in any order per scene; and each stimulus can be viewed and assessed as many times as the subject wishes (note the final score remained recorded).

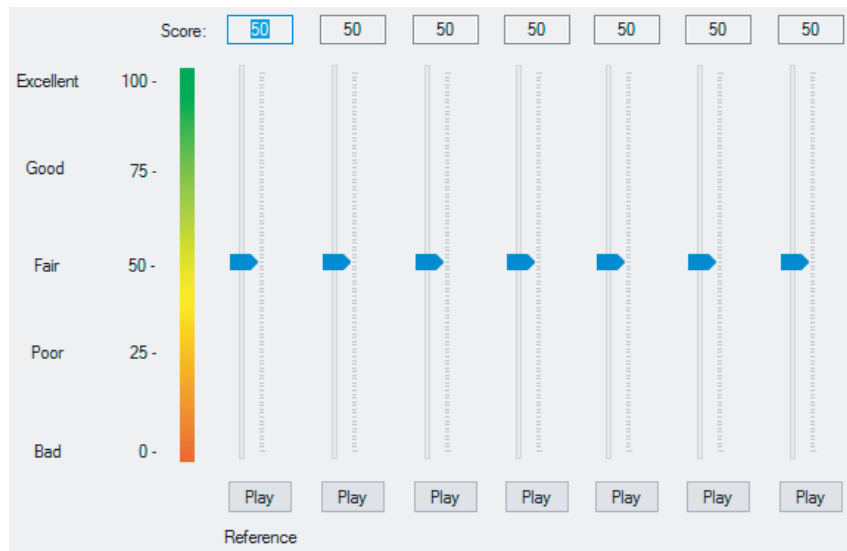


Figure 3: Illustration of the interface used in our experiment.

The experiment was conducted in a standard office environment at Xi’an Jiaotong-Liverpool University, Suzhou, China. The venue represented a controlled viewing environment to ensure consistent experimental conditions, i.e., low surface reflectance and approximately constant ambient light. The test stimuli were displayed on a 19-inch LCD monitor screen, with a native resolution of  $1920 \times 1080$  pixels. No image adjustment, e.g., zoom or window level, was allowed during the experiment.

Thirteen participants, including six men and seven women, participated in the experiment. Before the start of the actual experiment, each participant was provided with instructions on the procedure of the experiment (e.g., explaining the type of assessment and the scoring interface). A training session was conducted in order to familiarise the participants with the visual distortions involved and with how to use the range of the scoring scale. After training, all test stimuli were shown to each participant.

### 3. EXPERIMENTAL RESULTS

#### 3.1. Data Pre-processing

To process the raw data, an outlier detection and subject exclusion procedure was applied to the scores [17]. An individual score would be considered an outlier if it was outside an interval of two standard deviations around the mean score for that image. A subject would be rejected if more than 20% of their scores were outliers. As a result of the above procedure, one participant was rejected from the database.

#### 3.2. Impact of Distortion on Perceived Quality

Figure 4 illustrates the mean opinion score (MOS) averaged over all participants for each image in our experiment. It can be seen clearly that both the distortion type and distortion level seem to

Table 1: Results of the ANOVA to evaluate the effects of ‘Participant’, ‘Image’, ‘Distortion type’, and ‘Distortion level’ on the perceived quality.

Factor	df	F	p-value
Participant	11	1.74	0.125
Image	9	1.96	0.110
Distortion type	2	9.89	0.029
Distortion level	1	4.39	0.174
Distortion type * level	2	23.38	0.000

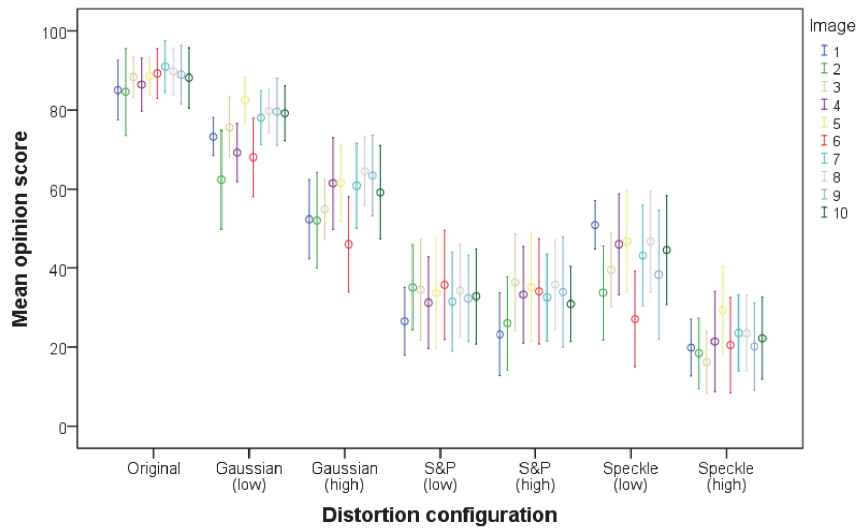


Figure 4: Illustration of the mean opinion score (MOS) averaged over all subjects for each distorted image. Error bars indicate a 95% confidence interval.

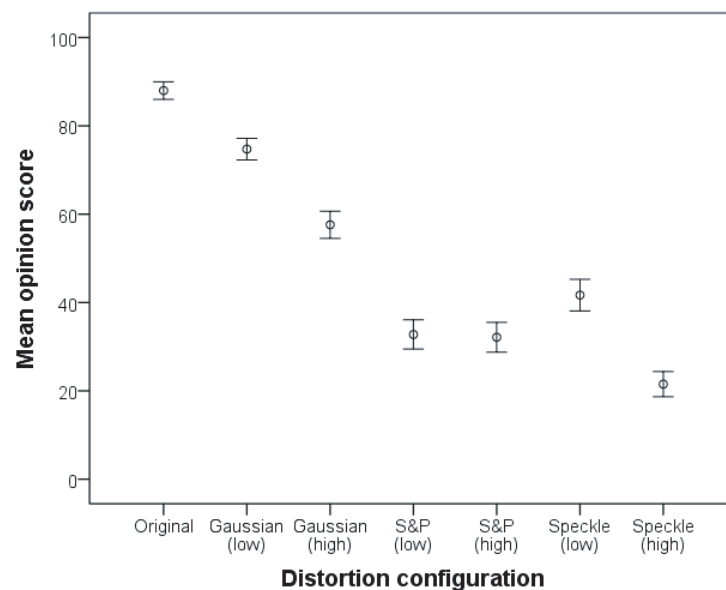


Figure 5: Illustration of the mean opinion score (MOS) averaged over all subjects and all images. Error bars indicate a 95% confidence interval.

affect the perceived quality.

The observed tendencies are further statistically analysed using the Analysis of Variance (ANOVA). The results are summarised in Table 1 where the F-statistic and its associated degrees of freedom and significance are included, along with the two-way interactions. The perceived quality is selected as the dependent variable, the image, distortion type, and distortion level as fixed independent variables, and the participant as random independent variable. The results show that there is no statistically significant difference between the participants in scoring the image quality (i.e.,  $p > 0.05$ ). Due to the consistency in scoring, there is little need to calibrate the scores using z-scores as conventionally required in image quality [18].

Similarly, the results show no statistically significant difference between the images (i.e.,  $p > 0.05$ ). Figure 5 represents the mean opinion score averaged over all participants and over all images. The results of the ANOVA show a significant impact of the distortion configuration (i.e., distortion type and level) on the perceived quality (i.e.,  $p < 0.005$ ). The post-hoc test reveals the following order in quality (note that commonly underlined entries are not significantly different from each

other):

Original ( $\langle \text{MOS} \rangle = 88.0$ ) > Gaussian low ( $\langle \text{MOS} \rangle = 74.7$ ) > Gaussian high ( $\langle \text{MOS} \rangle = 57.6$ ) > Speckle low ( $\langle \text{MOS} \rangle = 41.7$ ) > S&P low ( $\langle \text{MOS} \rangle = 32.8$ ) > S&P high ( $\langle \text{MOS} \rangle = 32.1$ ) > Speckle high ( $\langle \text{MOS} \rangle = 21.5$ ).

It is interesting to notice that there is a statistically significant difference between the low and high distortion levels for both Gaussian noise and speckle noise, whereas there is no difference in scoring for the salt-and-pepper noise.

#### 4. CONCLUSIONS AND DISCUSSION

In this paper, we presented a new subjective image quality assessment database with 70 night-time/cloud images. Statistical analyses showed the impact of the distortion type, i.e., Gaussian noise, salt-and-pepper noise, and speckle noise on the human perceived quality, as well as the impact of the distortion level as far as Gaussian noise and speckle noise are concerned.

This work can be further used as a better understanding of the impact of visual distortion on ground-based camera images and, subsequently, improve the current sky imaging detection methodologies. We will also evaluate the impact of noise in the computation of cloud coverage and cloud-type recognition frameworks.

#### ACKNOWLEDGMENT

The ADAPT Centre for Digital Content Technology is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund.

#### REFERENCES

1. Arking, A. and J. D. Childs, "Retrieval of cloud cover parameters from multispectral satellite images," *Journal of Climate and Applied Meteorology*, Vol. 24, No. 4, 322–333, 1985.
2. Shields, J. E., M. E. Karr, R. W. Johnson, and A. R. Burden, "Day/night whole sky imagers for 24-h cloud and sky assessment: History and overview," *Applied Optics*, Vol. 52, No. 8, 1605–1616, 2013.
3. Dandini, P., Z. Ulanowski, D. Campbell, and R. Kaye, "Halo ratio from ground-based all-sky imaging," *Atmospheric Measurement Techniques*, Vol. 12, No. 2, 1295–1309, 2019.
4. Pawar, P., C. Cortés, K. Murray, and J. Kleissl, "Detecting clear sky images," *Solar Energy*, Vol. 183, 50–56, 2019.
5. Dev, S., B. Wen, Y. H. Lee, and S. Winkler, "Ground-based image analysis: A tutorial on machine-learning techniques and applications," *IEEE Geoscience and Remote Sensing Magazine*, Vol. 4, No. 2, 79–93, 2016.
6. Dev, S., Y. H. Lee, and S. Winkler, "Color-based segmentation of sky/cloud images from ground-based cameras," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 10, No. 1, 231–242, Jan. 2017.
7. Dev, S., Y. H. Lee, and S. Winkler, "Categorization of cloud image patches using an improved texton-based approach," *Proc. IEEE International Conference on Image Processing (ICIP)*, 422–426, Sep. 2015.
8. Savoy, F. M., S. Dev, Y. H. Lee, and S. Winkler, "Stereoscopic cloud base reconstruction using high-resolution whole sky imagers," *2017 IEEE International Conference on Image Processing (ICIP)*, 141–145, Sep. 2017.
9. Lévêque, L., W. Zhang, C. Cavaro-Ménard, P. Le Callet, and H. Liu, "Study of video quality assessment for telesurgery," *IEEE Access*, Vol. 5, 9990–9999, 2017.
10. Mohammadi, P., A. Ebrahimi-Moghadam, and S. Shirani, "Subjective and objective quality assessment of image: A survey," *CoRR*, Vol. abs/1406.7799, 2014.
11. Lévêque, L., H. Liu, S. Baraković, J. Baraković Husić, M. Martini, M. Outtas, L. Zhang, A. Kumcu, L. Platasa, R. Rodrigues, A. Pinheiro, and A. Skodras, "On the subjective assessment of the perceived quality of medical images and videos," *Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, May 2018.
12. Aflaki, P., M. M. Hannuksela, and M. Gabbouj, "Subjective quality assessment of asymmetric stereoscopic 3D video," *Signal, Image and Video Processing*, Vol. 9, No. 2, 331–345, Feb. 2015.
13. International Telecommunication Union (ITU), "Subjective video quality assessment methods for multimedia applications," Recommendation ITU-T P.910, Apr. 2008.

14. Dev, S., F. M. Savoy, Y. H. Lee, and S. Winkler, "Nighttime sky/cloud image segmentation," *Proc. IEEE International Conference on Image Processing (ICIP)*, 345–349, IEEE, 2017.
15. Dev, S., F. M. Savoy, Y. H. Lee, and S. Winkler, "WAHRIS: A low-cost high-resolution whole sky imager with near-infrared capabilities," *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXV*, Vol. 9071, 90711L, International Society for Optics and Photonics, 2014.
16. Kozamernik, F., V. Steinmann, P. Sunna, and E. Wyckens, "SAMVIQ — A new EBU methodology for video quality evaluations in multimedia," *SMPTE Motion Imaging Journal*, Vol. 114, 152–160, Apr. 2005.
17. Sheikh, H. R., M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *Trans. Img. Proc.*, Vol. 15, No. 11, 3440–3451, Nov. 2006.
18. Hemminger, B. M., R. Eugene Johnston, J. P. Rolland, and K. E. Muller, "Introduction to perceptual linearization of video display systems for medical image presentation," *Journal of Digital Imaging*, Vol. 8, No. 1, 21–34, Feb. 1995.