BULLETIN OF THE POLISH ACADEMY OF SCIENCES
TECHNICAL SCIENCES, Vol. 59, No. 1, 2011
DOI: 10.2478/v10175-011-0003-4

Measurement of noise in the Monte Carlo point sampling method

K. GUZEK* and P. NAPIERALSKI

Institute of Information Technology, Technical University of Lodz, 215 Wólczańska, 90-924 Łódź, Poland

Abstract. This paper gives a concise overview of the techniques we have used to find out the degree of measuring the quality of rendered images and a level of noise in particular. First part of the paper presents designed and conducted psychophysical experiment involving human subjective judgment. Then, two of the existing numerical image comparison methods are considered in the context of assessing the level of noise produced by global illuminations algorithms. The results of the participants' subjective responses are correlated with the data obtained from objective mathematical metrics. The main goal of research was to determine the objective and perceptual measure of quality for images with fixed sampling strategy. The results will help to establish the measures of identifying the human perception in the assessment of images generated with global illumination algorithms.

Key words: photorealism, image evaluation methods, image quality metrics, visual perception, global illumination.

1. Introduction

Images rendered using global illumination algorithms often appear more photorealistic than images rendered using only direct illumination algorithms. However, such images are much slower to generate and consequently more expensive. Ironically, what we perceive as being „realistic" in an image has more to do with our expectations of it, rather than a faithful account of the real world. We get used to certain conventions, and any deviation from them brings the risk of being regarded as unnatural. For instance, despite the fact that the photographic grain does not exist in the physical world, we can consciously claim that the concrete image is realistic exactly because it contains so called noise. Therefore, investigation of image features determining the level of perceived visual realism is crucial for achieving photorealistic rendering. Separating specific image features in complex scenes is a difficult task, but beyond any doubt, the level of noise is the key feature, that affects perception of any scene.

Despite the fact, that several computational methods for assessing the quality of computer generated images and the level of noise in particular have been proposed, precise scales and threshold are still not defined. Most of the existing objective quality metrics are algorithms designed to characterize the quality of video compression and predict viewer MOS (Mean Opinion Score). Two of them will be presented, and the main intention of this work is to review them for suitability in global illuminations algorithms.

Another approach to noise evaluation is to use standard psychophysical test. The most typical subjective methods compare images based on perceptual appearance. The idea for this investigation is to design and conduct such experiment involving human subjective judgment, and compare obtained results with the numerical data. If the subjects' responses are correlated with the mathematically calculated values, then such numerical methods for noise measurement could be successfully used to speed up existing rendering techniques directly targeting human perception limitations.

2. Variance in the estimates in the global illumination algorithms

A synthesis of photorealistic images becomes possible in the latter half of the 90s. Algorithms for simulating the physics of light and light transport can be divided into two major techniques: point sampling and finite elements. The physically based simulation of all light scattering in the synthetic model is called Global Illumination.

Methods based on finite elements compute the equilibrium of the light exchange between surfaces of geometry model. This is done by discretization of the model into small patches that can form a basis for the final light distribution. The lighting distribution is found by solving a set of linear equations for the light exchange between all patches. This approach is impractical in complex models due to the division of the geometry into the large number of patches.

Actually the most popular methods in the computer graphics is based on point sampling. Point sampling methods have been extended with Monte Carlo Methods (MC). The basic algorithms for this group are: Path Tracing, Light Tracing, Bidirectional Path Tracing and Metropolis Light Tracing. Many of rendering engines use follow techniques to reproduce behavior of light. These algorithms permit simulation of all types of light scattering. In MC the rays are distributed stochastically to simulate all paths from the light source. Stochastic sampling gives possibility to compute effects such as soft shadows, motion blur, and depth of field. MC point sampling methods is a straightforward extension to ray tracing that makes it possible to compute lighting effects that requires evaluation of integration problems such as area lights and indirect light reflected by a diffuse surface. In this methods the unknown lighting distribution function is sampled by tracing rays sto-

*e-mail: krzysztof.guzek@p.lodz.pl

chastically along all possible light paths. By averaging a large number of sample rays for a pixel we get an estimate of the integral over all light paths through that pixel. Mathematically, it is a continuous Markov chain random walk technique for solving the rendering equation. When the rendering equation is solved by stochastic methods we get a variance seen as the noise in rendering images (Fig. 1).

Fig. 1. Reference photograph (top), rendered color (middle) and grayscale image (bottom)

Eliminating this noise requires a large number of sample-rays. The number of sample-rays depends on how much noise is acceptable in an image, and geometry complex. The rays are generated randomly with the same distribution as the emitted power of the light. The rays are also sent from points distributed on the surface of the light. The noise level in a Monte Carlo solution decreases with the square root of the number of samples taken. To reduce the noise by a‘ factor of 2, 4 times more samples are needed [1]. The goal is to find a compromise between a reasonable rendering time and quality of rendered images. The best way is to find an automatic quantity to measure the quality of images connected with noise.

3. Perceptual test for noise evaluation

This part of the paper describes the proposed experimental method for subjective evaluation of the variance in the estimates. The perceptual method, as well as numerical approach presented in the next paragraph, can be performed by rating the set of images with different scale of manipulation of certain feature [2, 3], in this case the level of noise. The exact purpose of conducting this test is to see whether there is a correlation between the subjects' responses and the objective mathematical metrics. While performing this test we should also assess the perceived distance between visually indistinguishable level of noise and acceptable quality of the rendered images [4].

3.1. Creating test images. For the purpose of overall experiment we have created the 3D representation of the conference room, containing tables and chairs, with different types of materials but without glass surfaces (Fig. 1).

The geometry of the scene must be sufficiently complicated in order to produce a broad scope of noise. The simplicity of a typical cornell box itself causes that the image rendered using light-tracing algorithm with even less than 10 rays per pixel contains an imperceptible variance in the estimates. All images used in the experiment present the same scene renderd with the ligh-tracing algorithm with the resolution of 800x600 considering parallel processor scheduling algorithms [5].

3.2. Test design. The experiment was undertaken under constant and controlled illumination conditions with the 15.4” monitors set to 1280×800. The distance between the eyes and the screen was approximately 0.4 meter. The subjects were presented with two series of controlled images grouped into pairs. The first series contains only grayscale images and the second set only RGB color images. Each pair of images was preceded with the black screen with the assigned number. Subsequent images in each series varied according to increasing number of sample rays used in rendering process, and thereby decreasing level of noise. All other image factors were constant [2]. The image presentation was automated with 2 seconds of time interval between the images. The subjects ran their test in one sitting with short breaks between the grayscale and the color series.

Before the start of the test the participants were given minimum information about the context of the experiment, in order to avoid responses biased toward what they were told. For instance, to prevent situation, when responses are more a reflection of subjects expectation of average score, rather than their actual perception, subjects were not informed about a number of pairs (150) in the series.

Each of the participants was tested separately and was asked to answer the following questions concerning both the grayscale and the color series:

1. Please indicate the number of the first pair for which you cannot see the difference in the level of noise between the pictures.
2. Please indicate the number of the first pair for which the level of noise is negligible.

According to the first question, it was necessary to insert a blank black screen displayed for approximately 0.5 second between the pair of the images. Because of the stochastic nature of the light-tracing algorithm, subjects could still observe the relocation of the grain and interpret this shift as an expected difference, even if the actual difference in noise intensity between the images was imperceptible. The applied interruption eliminated potential degeneration of subjects' responses.

The second question considered subject's toleration of noise in computer-generated image. Due to the lack of a fixed reference image there might be a considerable distribution of responses. Another difficulty is the lack of a clear definition of what is meant by "negligible" in term of visual realism [6]. In this case prior experience and predilection of the subjects may have played an important role. We want to investigate the acceptable range of noise depending on the image characteristics. The analysis and discussion on the obtained data is presented in the final paragraph.

4. Numerical metrics for noise measurement

There are many numerical approaches to compare synthetic images [7, 8]. They are mostly designed for digital images created by digital camera. They focus on measuring the distance between images. Most of the methods compute distance between the two images is computed by finding the MSE (mean squared error) [9]

$$d(X,Y) = \frac{\sum\limits_{i=0,j=0}^{m,n} (x_{ij} - y_{ij})^2}{mn}, \tag{1}$$

where X, Y are current images and x_{ji}, y_{ji} color values of pixels.

This method is based on raw error measures and works best when the distortion is due to an additive noise contamination. Therefore it seems reasonable to use such metrics to examine the noise associated with the global illumination algorithms. Of course it does not necessarily correspond to all aspects of the observer's visual perception of the errors degree of compressed image deformation [10]. We will verify the usefulness of metrics for global illumination algorithms.

4.1. Peak-to-peak signal-to-noise ratio. The metric which is used often in practice is PSNR(Peak-to-peak Signal-to-Noise Ratio) [11], this metric is equivalent to Mean Square Error, extended by logarithmic scale. It has the same disadvantages as the MSE metric:

$$\mathrm{PSNR} = 10 \cdot \log_{10} \frac{\mathrm{MaxErr}^2 \cdot w \cdot h}{\sum\limits_{i=0,j=0}^{w,h} (x_{ij} - y_{ij})^2}, \tag{2}$$

where MaxErr – maximum possible absolute value of color components difference, w – image width, h – image height.

The main application of PSNR is comparing the compressed images. The value of PSNR function is measured in db, the bigger PSNR – the lesser is the difference between images. Some analytical methods of determining extreme dynamic errors have been proposed in [12]. Our main goal is to investigate whether there needs to be a consistency between the levels of noise of the different elements within a images. It seems reasonable to use PSNR to compare the noise measurement between images. We use this method to find differences in noise level (Fig. 2).

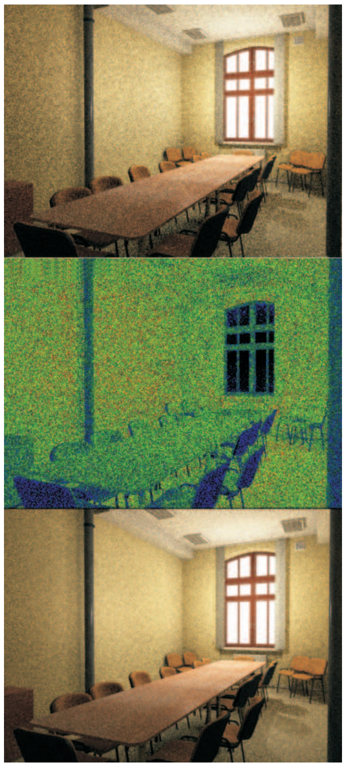

Fig. 2. Results of PSNR – difference between images with different ray-samples (red pixels indicate the biggest difference, dark blue pixels indicate the least difference)

A PSNR method for measuring the perceptual equivalence between different ray-samples images of the same scene was tested for 150 images with successively increased number of ray samples. Results of noise in decibels between all current images are presented below in Fig. 3.

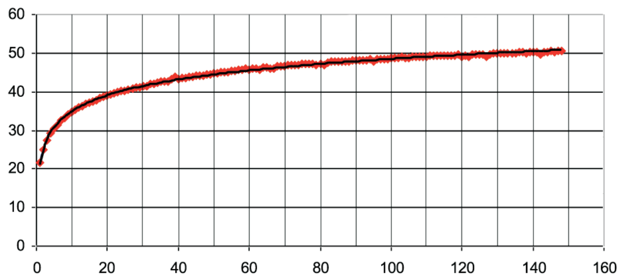

Fig. 3. PSNR – difference between images with different ray-samples. Log(psnr) = 5.894Ln(x), R^2 = 0.9975

In the figure above we can notice a degradation of the noise. This degradation will depend on the number of ray-samples. It seems reasonable to find demarcation of quality between the compared images. The first limit is made where a number of samples indicate slight changes between the qualities of compared images. The second limit will show the smallest increase in the quality of the test images. The least-squares fitting process give Equation of Trend estimation for samples [13]:

$$y = 5.894Ln(x) + 21.384. \tag{3}$$

Coefficient of determination for analyzed data:

$$r^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = \frac{\hat{a}covxy}{vary} = 0.9975. \tag{4}$$

The collected data will be reviewed in terms of the tests carried out on a random group of people in the final chapter.

4.2. DCT-based video quality evaluation. For multimedia applications, there has been an increase in the use of quality measures based on human perception. Method VQM uses DCT to correspond to human perception [14].

Before the described Methods Pixel-based Root Mean Square Error is the dominant metric in practice. However, it doesn't take into account the spatial-temporal property of a human's visual perception that is the reason why it fails under many circumstances. DCT-based video quality metric (VQM) is based on Watson's proposal, which exploits the property of visual perception [11]. This method frequently used in compression technology (quantization matrix, spatial scalability, temporal scalability) affects video distortion. The same group of images has been tested for VQM metrics, as it is shown in Figs. 4 and 5.

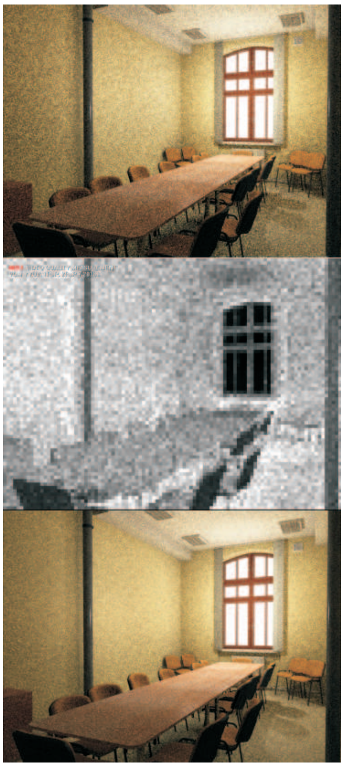

Fig. 4. Results of VQM – difference between images with different ray-samples (the brightest areas indicate the biggest difference, darkest areas indicate the least difference)

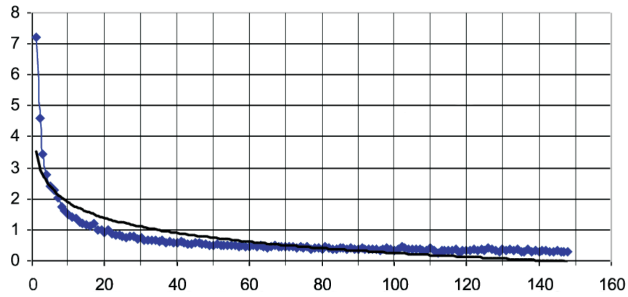

Fig. 5. VQM – difference between images with different ray-samples. Log (vqm) = - 0.7085 Ln(x)+3.5129, R^2 = 0.7162

The method checks the better large lighting areas. In the case of noise it seems more reasonable to use metrics like PSNR. The least-squares fitting process gives Equation of Trend estimation for samples [13]:

$$y = -0.7085 Ln(x) + 3.5129. \tag{5}$$

5. Results and conclusions

The performance of the objective quality measurement algorithm is improved by comparing to the results of the subjective test and the results of a PSNR measurement. Experimental data was obtained from 19 subjects aged from 18 to 40 including authors. Subjects had either normal or corrected-to-normal vision. After performing the test, the preferences of statistical viewer were obtained. The coefficient of consistency for individual test subjects was measured using the pair of images. The answers of respondents give series of four values. The received data for the first pair for which the level of noise is negligible will be marked as: Qc in color, Qbw in grayscale. Analogically, DNc and DNbw for the number of the first pair, for which the difference in the level of noise between the pictures cannot be seen. The results for the average of the values obtained in four series are as follows:

$$\begin{aligned} AM(DN_c) &= 8.8125, \\ AM(DN_{bw}) &= 8.75, \\ AM(Q_c) &= 68.875, \\ AM(Q_{bw}) &= 64.25. \end{aligned} \tag{6}$$

The results for the median of the values obtained in four series are as follows:

$$\begin{aligned} \mu_{1/2}(DN_c) &= 5.5, \\ \mu_{1/2}(DN_{bw}) &= 6.5, \\ \mu_{1/2}(Q_c) &= 68.5, \\ \mu_{1/2}(Q_{bw}) &= 61. \end{aligned} \tag{7}$$

The relative increase of the PSNR function can be found by:

$$d_{t/t-1} = \frac{y_t - y_{t-1}}{y_{t-1}}. \tag{8}$$

The most significant differences were observed between the pairs from 1 to 9 ($d_{9/8} = 0.1627$). After this value a rise between subsequent images is decidedly gentle Fig. 6. The value indicated by the respondents fluctuates around 8 rays per sample. Without loss of generality, we can assume that the PSNR of noise between 8 and 9 rays per sample (33 [db]) can represent the critical point of distinguishing noise for human perception ($AM(DNc) = 8$). Hence the results of the perceptual test are coherent to the curve of PSNR function. Therefore the data metrics such as PSNR can be used to analyze noise in the images generated with global illuminations methods based on the point sampling algorithms.

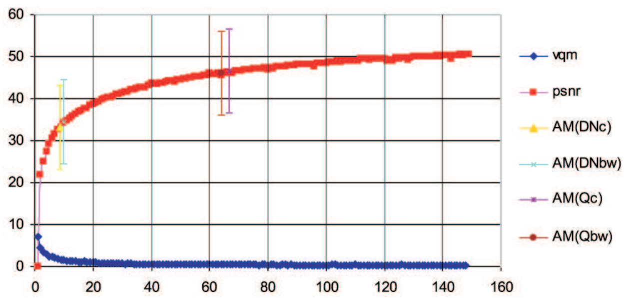

Fig. 6. VQM and PSNR – difference between images with different ray-samples marked with points obtained in test

In response to the second question, received values fluctuate around the 70 rays per sample. Thus, we can theoretically calculate PSNR for 1 to 70 ray-sample, it gives 22 [db] or for 69 to 70 it gives 46 [db]. But it is difficult to determine the amount of this value only for one test scene. In future work, it will be important to identify other checkable necessary conditions for different tests scenes. More measurements and user tests, similar to the one proposed, should be performed. The data collected from such experiments can be used to further validate or refine the outcomes.

In our experiments we found that results of numerical metrics used for noise measurement did correlate with subjects' responses. The received measure can be helpful to speed up rendering algorithms, based on modified metrics that can directly target observers' perception.

REFERENCES

[1] D. Banks, "A basic guide to global illumination", *SIGGRAPH 98* 1, CD-ROM (1998).

[2] G. Levine and S. Parkinson, *Experimental Methods in Psychology*, Lawrence Erlbaum Associates, New Jersey, 1994.

[3] K. Chiu and P. Shirley, "Rendering, complexity, and perception", 5^{th} *Eurographics Workshop on Rendering* 1, 19–33 (1994).

[4] A. McNamara, A. Chalmers, T. Troscianko, and E. Reinhard, "Fidelity of graphics reconstructions: a psychophysical investigation", *Proc. 9th Eurographics Rendering Workshop* 1, 237–246 (1998).

[5] A. Janiak, W.A. Janiak, and R. Januszkiewicz, "Algorithms for parallel processor scheduling with distinct due windows and unit-time jobs", *Bull. Pol. Ac.: Tech.* 57 (3), 209–215 (2009).

[6] P.M. Rademacher. "Measuring the perceived visual realism of images", *Ph.D. Thesis*, Chapel Hill University, Chapel Hill, 2002.

[7] H.R. Sheikh, M.F. Sabir, and A.C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms", *IEEE Trans. Image Processing* 15 (11), 3440–3451 (2006).

[8] R. de Freitas Zampolo and R. Seara, "A comparison of image quality metric performances under practical conditions", *ICIP 2005* 3, 1192–1195 (2005).

[9] H. Rushmeier, G. Ward, C. Piatko, P. Sanders, and B. Rust, "Comparing real and synthetic images: some ideas about metrics", *6th Eurographics Workshop on Rendering* 1, CD-ROM (1995).

[10] İ. Avcıbaş, "Statistical evaluation of image quality measures", *J. Electronic Imaging* 11, 206 (2002).

[11] A.B. Watson, "Toward a perceptual video quality metric", *Human Vision, Visual Processing, and Digital Display* VIII 3299, 139–147 (1998).

[12] H. Górecki and M. Zaczyk, "Extremal dynamic errors in linear dynamic systems", *Bull. Pol. Ac.: Tech.* 58 (1), 99–105 (2010).

[13] M. Bianchi, M. Boyle, and D. Hollingsworth, "A comparison of methods for trend estimation", *Applied Economics Letters* 6 (2), 103–109 (1997).

[14] F. Xiao, "DCT-based video quality evaluation", *Final Project for EE392J Winter* 1, CD-ROM (2000).