Spatial Domain Synthetic Scene Statistics

Debarati Kundu and Brian L. Evans
Embedded Signal Processing Laboratory
The University of Texas at Austin, Austin, TX
Email: debarati@utexas.edu, bevans@ece.utexas.edu

***Abstract*—Natural Scene Statistics (NSS) has been applied to natural images obtained through optical cameras for automated visual quality assessment. Since NSS does not need a reference image for comparison, NSS has been used to assess user quality-of-experience, such as for streaming wireless image and video content acquired by cameras. In this paper, we take an important first step in using NSS to automate visual quality assessment of synthetic images found in video games and animated movies. In particular, we analyze NSS for synthetic images in the spatial domain using mean-subtracted-contrast-normalized (MSCN) pixels and their gradients. The primary contributions of this paper are (1) creation of a publicly available ESPL Synthetic Image database, containing 221 color images, mostly in high definition resolution of 1920 $\times$ 1080, and (2) analysis of the statistical distributions of the MSCN coefficients (and their gradients) for synthetic images, obtained from the image intensities. We find that similar to the case for natural images, the distributions of the MSCN pixels for synthetic images can be modeled closely by Generalized Gaussian and Symmetric α-Stable distributions, with slightly different shape and scale parameters.**

***Index Terms*—Natural Scene Statistics, Mean subtract contrast normalization, log contrast, Generalized Gaussian Distribution, Symmetric α-Stable**

I. Introduction

Recent years have seen a huge growth in the acquisition, transmission, and storage of videos. In addition to videos captured with optical cameras, video traffic also consists of synthetic scenes, such as animated movies, cartoons and video games. The burgeoning popularity of multiplayer video games (esp. on handheld platforms) is causing an exponential increase in synthetic video traffic. In all these cases, the ultimate goal is to provide the viewers with a satisfactory quality-of-experience. For video frames and other still images, many objective quality metrics for quality-of-experience have been proposed.

Full-reference image quality assessment metrics quantify the distortions present in an image, in comparison to a reference "pristine" image. However, this approach is rendered unusable in applications where the ground-truth reference image is not available. For these cases, blind or no-reference image quality assessment metrics are better suited, where, the only information available is the distorted image. The no-reference algorithms are primarily based on studying the overall statistical properties possessed by the pristine images, which for natural images tend to appear irrespective of the image content, and are based on the assumption that distortions tend to deviate the Natural Scene Statistics (NSS). [1], [2], [3] are some of the best performing no-reference image quality metrics for natural images.

However, these metrics for evaluating the quality of natural images have not been studied in the context of images generated using computer graphics. Compared to natural images, no-reference quality evaluation of computer graphics images is even more important, because of the non-availability of the ground truth. It has been mentioned in literature that both computer graphics and image distortions tend to deviate the statistics from NSS [3]; however, the extent to which computer graphics images deviate from their natural counterparts has not been explicitly quantified. Also, the deviation in "naturalness" caused by the distortions can be very different from the way computer graphics deviates properties of natural scenes. In fact, with the improvement of rendering technology, rendered images are becoming more and more photo-realistic, which has made us believe that with some adjustments, the NSS models can potentially be applied in the domain of computer graphics with some modifications. A body of work aims at detecting synthetic images from natural ones, primarily applied in forgery detection [4], but dedicated learning of the scene statistics for detecting distortions in synthetic images is relatively new. [5] proposes a machine learning based no-reference metric for quantifying rendering distortions, but in this case, the features were chosen rather heuristically, instead of analyzing the properties of the synthetic images under test.

II. Database of Synthetic Images

Getting access to a substantial number of images is crucial for the development of any learning based quality evaluation scheme, and the performance of the metrics tend to become better with the availability of more training data. For the purpose of this study, 221 synthetic images have been chosen from video games and animated movies, which reflect sufficient diversity in the image content. We collected these images in the ESPL Synthetic Image Database [6]. These high quality color images from the Internet are mostly 1920$\times$1080 pixels in size. Some video games which were considered were multiplayer role playing games (such as War of Warcraft), first person shooter games (such as Counter

Strike), motorcycle and car racing games, and games with more realistic content (such as FIFA). Some of the animated movies, from which the images were collected, are, The Lion King, the Tinkerbell series, Avatar, The Beauty and the Beast, Monster series, Ratatouille, the Cars series etc. [1]

Care has been taken to provide as varied a range of images as possible, by incorporating both natural and non-photorealistic renderings of human figures, man-made objects, fantasy figures like fairies, and monsters, close up shots, wide angle shots, images showing both high and low degrees of color saturation, background textures with no foreground object etc. Fig. 1 shows a subset of the images considered.

We generated several categories of distorted images from the pristine images for the ESPL Synthetic Image Database, such as images containing high frequency noise, ringing, aliasing and banding artifacts, Gaussian blurring, JPEG blocking artifacts, and artifacts due to overtly saturated colors. However, as the premise of this paper is analyzing the statistics of undistorted synthetic images, detailed description of the methods of generating the artifacts have been omitted.

III. Spatial Domain Natural Scene Statistics

Previous work has quantified NSS both in the spatial domain and transform domains, such as using wavelets or DCT. For this preliminary work, only spatial domain features of natural scenes are considered. One of the earliest attempts at quantifying the statistical properties of the distribution of the natural images in the spatial domain can be found in [7]. The author verified the scale invariant properties of natural images. Since the natural images belong to a very special category of 2D signals, they possess unique properties which tend to appear irrespective of image content.

A. Mean Subtracted Contrast Normalized Image Patches

As in [3], the pixels of the image are preprocessed by mean subtraction and divisive normalization. Let $M \times N$ be the dimension of the image I, $I(i,j)$ be the pixel value in the (i,j)-th spatial location, $i \in \{1,2,..,M\}$, $j \in \{1,2,...,N\}$. The mean subtracted contrast normalized (MSCN) image values are generated by

$$\hat{I}(i,j) = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + 1} \tag{1}$$

where the local mean $\mu(i,j)$ and standard deviation $\sigma(i,j)$ are defined as:

$$\mu(i,j) = \sum_{k=-K}^{k=K} \sum_{l=-L}^{l=L} w_{k,l} I(i+k, j+l) \tag{2}$$

$$\sigma(i,j) = \sqrt{\sum_{k=-K}^{k=K} \sum_{l=-L}^{l=L} w_{k,l} [I(i+k, j+l) - \mu(i,j)]^2} \tag{3}$$

$w = \{w_{k,l} | k = -K,..,K, l = -L,...,L\}$ is a symmetric local convolution window centered at the (i,j)-th pixel. K and L determine the size of local patch considered in the calculation of the mean and standard deviation. In [3], the authors have considered 7×7 image patches, and a circularly symmetric 2D Gaussian kernel $w_{k,l}$; however, experiments show that the distribution of the MSCN patches are not very sensitive to the size of the window, or the convolution kernel.

In their original work [7], the authors propose that, as the visual systems adapt to the mean background value, it should be removed by considering the logarithmic intensity fluctuations above the mean value. Let $I(x)$ be the image intensity at spatial location x. The 'log-contrast' $\phi(x)$ is defined as

$$\phi(x) = \ln \left[\frac{I(x)}{I_0} \right] \tag{4}$$

where, I_0 is defined in such a way that $\sum_x \phi(x) = 0$. This causes the histogram to have zero mean. The log-contrast values are normalized with respect to their local standard deviations. By this operation, patches of small image contrast are expanded, while regions of high contrast are toned down. The resulting field is given by

$$\psi(x) = \frac{\phi(x) - \bar{\phi}(x)}{\sigma(x)} \tag{5}$$

The variance normalized image $\psi(x)$ tend to be more uniform than the original image, and almost looks like a noise pattern, except at the object boundaries. Also, their histograms seem to show a Gaussian-like distribution. Compared to $\psi(x)$, the standard deviation image $\sigma(x)$ looks more like the original image, highlighting the object boundaries, and attenuating the textures. $\psi(x)$, and $\sigma(x)$ for sample images have been shown in Section IV.

The procedure can be iterated over multiple scales. The standard deviation image is considered as the original image, and its log-contrast is defined as

$$\xi(x) = \ln \left[\frac{\sigma^2(x)}{\sigma_0^2} \right] \tag{6}$$

where, σ_0^2 is analogous to I_0. The statistics of ξ are found to be similar to those of ψ obtained by (5). The above can be further re-applied for a multiscale characterization, this time considering $\xi(x)$ as the original image, and finding its log contrast. It yields two images: the variance normalized image $\zeta(x)$, and the standard deviation image $\Sigma(x)$. Experimental results show that $\psi(x)$ and $\zeta(x)$ show similar Gaussian-like intensity distributions, and $\sigma(x)$ and $\Sigma(x)$ show image like distributions. The multiscale generation of the MSCN coefficients has been outlined in Fig. 2.

[1] All images are copyright of their rightful owners, and the authors do not claim ownership. No copyright infringement is intended. The database is to be used strictly for non-profit educational purposes.

(a) (b) (c)

Fig. 1: Sample Synthetic Images in the ESPL database [6]

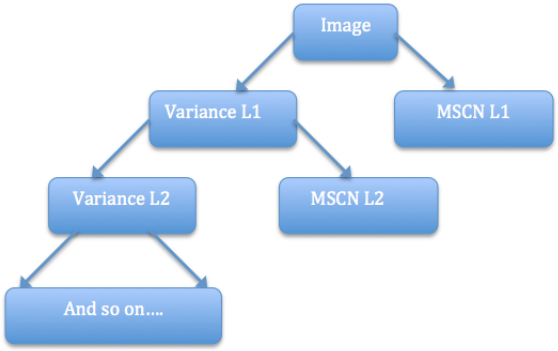


Fig. 2: Multiscale generation of the MSCN coefficients by iterating over the variance images.

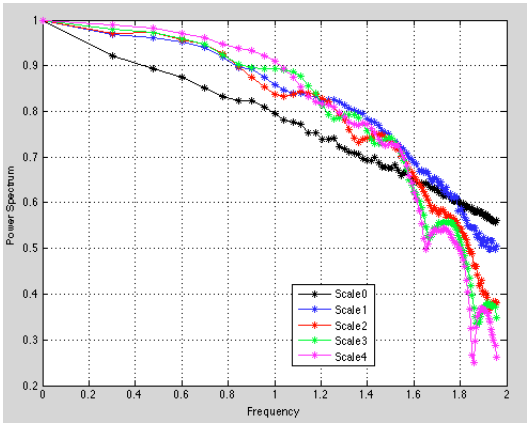


Fig. 3: Spectrum of the original image (scale 0), and variance images obtained by iterating the MSCN process over multiple scales. The variance images become more correlated as we iterate the MSCN operation multiple times. Structural correlation between the original image and scale 1 variance image is 0.508, whereas, that between the variance image at level 3 and level 4 is 0.99.

IV. Experimental Results

This section shows some of the experimental results obtained by modeling the distribution of the transformed pixel intensities, obtained by (1) and (5). The analysis has been carried out on the luminance component of the color images. For the calculation of the mean and standard deviation of the pixels, a window of size 7×7 was considered at each pixel location (although, the results were found to be almost independent of the size of the selected local window).

As mentioned in [8], we begin our analysis by looking at the spatial structural correlation between image pixel densities and the neighboring pixels, computed as [9]:

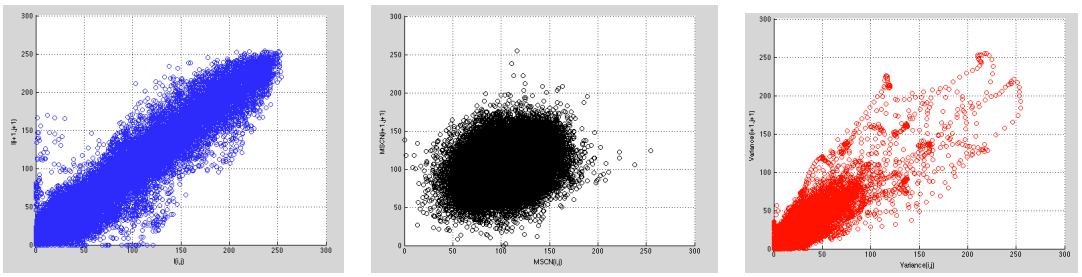
$$\rho = \frac{2\sigma_{xy} + C_1}{\sigma_x^2 + \sigma_y^2 + C_1} \tag{7}$$

where σ_{xy} is the cross-covariance between the two local image patches, σ_x^2 and σ_y^2 are their respective variances and C_1 is a constant to prevent numerical instabilities arising if the denominator becomes close to zero. The value of C_1 has been specified in [9]. The spatial correlation is observed for the original intensity images, the modified MSCN coefficients, as well the variance images. Fig. 4 illustrates the scatter diagram between the pixels, and one of their diagonal neighbors. The values of the structural correlation (ρ) have also been mentioned. We also plot the structural correlation between the original image, the generated MSCN image, and the variance image in Fig. 5.

For natural images, the power spectrum is found to roughly follow a $\frac{1}{f^\gamma}$ relation, where γ varies over a small range for natural images. Fig. 3 shows the power spectrum of a synthetic image on a log-log scale. The spectrum of the variance images has been found to align with the power spectrum of the original image.

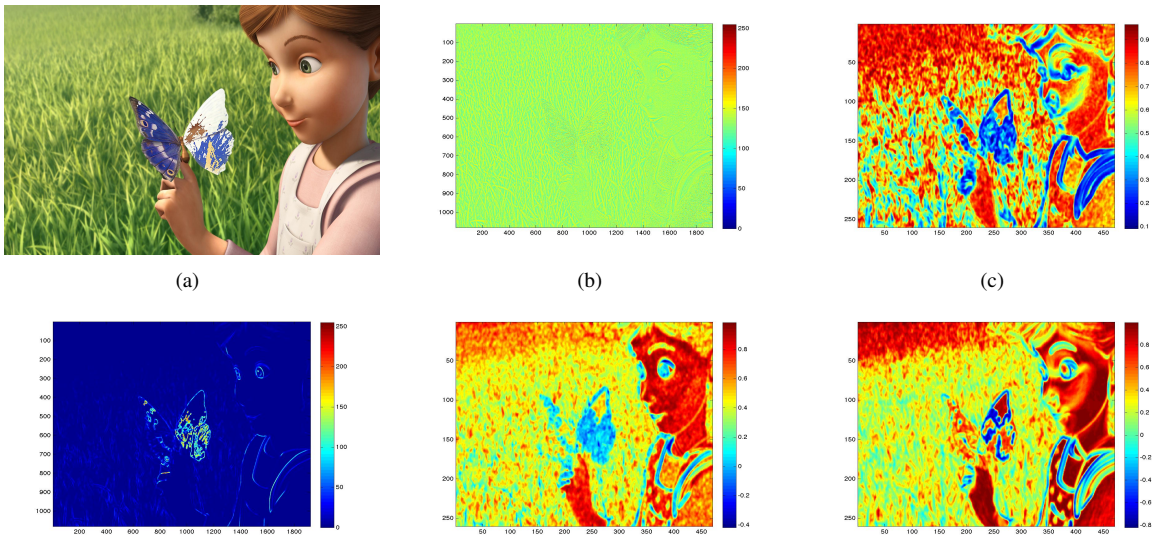
Next, in order to get an idea of the type of distribution which would be most suitable for modeling the MSCN coefficients obtained from (1), skewness and excess kurtosis values are studied. Generalized Gaussian density [10], and Symmetric α-Stable distribution [11] have been considered. Fig. 6 shows the scatter plot of skewness and excess kurtosis values computed from the empirical histograms of the MSCN coefficients obtained from synthetic images in the ESPL database, and natural images obtained from the Berkeley segmentation dataset [12].

The empirical histogram skewness values are mostly clustered around the zero value, with some showing small amounts of positive shifts. This shows that a



(a) (b) (c)

Fig. 4: Scatter Diagram between pixel values, and one of their diagonal neighbors, shown for one of the images from the ESPL database (a) Original Synthetic image, $\rho = 0.967$, (b) Modified MSCN coefficients, $\rho = 0.296$ (c) Variance image, $\rho = 0.921$. The original image shows a high degree of correlation between the neighboring pixels. But the MSCN coefficients are much less correlated with their neighbors. However, the correlation of the variance image shows more image-like properties, such as, a high degree of correlation with neighboring pixels. This behavior is observed to be repeated if a new set of MSCN coefficients is generated by treating the variance image as the original image, and iterating the process over multiple scales.



(a) (b) (c)

(d) (e) (f)

Fig. 5: Correlations between original image, MSCN coefficients, and the variance image (a) Original image (b) MSCN image (c) Normalized windowed correlation between (a) and (b), average $\rho = 0.596$ (d) Variance image (e) Normalized windowed correlation between (b) and (d), average $\rho = 0.435$ (f) Normalized windowed correlation between (a) and (d), average $\rho = 0.504$. The correlation between the MSCN and variance images is somewhat lower than that between the original images and the variance images.

symmetric non-skewed distribution should be able to model the variation in most of the images. However, when compared to the natural images, some of the synthetic images tend to show a higher degree of excess kurtosis. This is common if the images show large textureless regions, and abrupt change of contrast, e.g., those occurring across sharp boundaries. This is also found to be common feature of cartoon images. In this case, most of the MSCN coefficients tend to be zero, and hence, a sharp spike is observed near the origin. For modeling these type of images, the Symmetric α-Stable distribution with small values of α is found to be a better model compared to the GGD models.

The next step is to estimate the GGD mean μ, scale α, and shape β parameters from the sample histograms. This is done by the method of maximum likelihood estimation [10]. In order to understand how much these parameters differ for natural and synthetic images in the ESPL database, we plot the histogram of the scale and shape parameters. Fig. 7 shows the histogram of the GGD shape parameter β. A substantial overlap in the distribution of β is found among natural and synthetic images, which shows that the value of β is not discriminative enough to classify computer generated imagery and natural images. In fact, a natural scene and a highly non realistic synthetic scene may show the same distribution of the MSCN coefficients. For natural images, β tends to cluster around 2, which corresponds

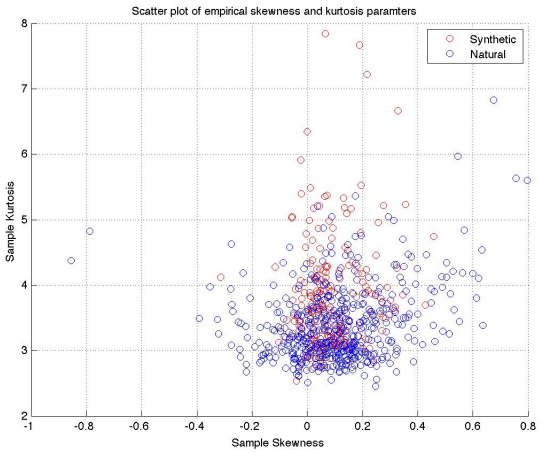


Fig. 6: Scatter plot of skewness(X-axis) and kurtosis(Y-axis) of 221 synthetic and 500 natural images. Note that the while most the synthetic images show zero or very small skewness values, some of them might exhibit high excess kurtosis, indicating heavily peaked distribution of the MSCN coefficients.

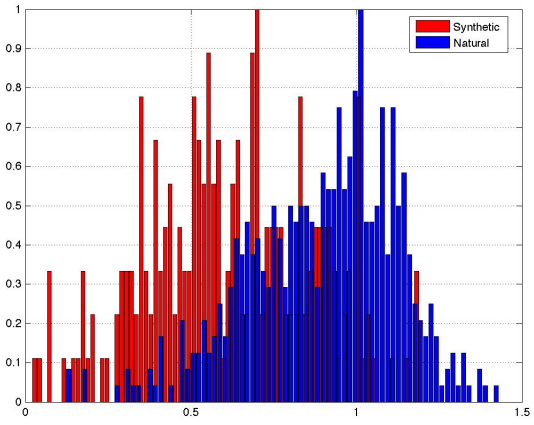


Fig. 8: Normalized histogram of the scale parameter α obtained over 221 synthetic and 500 natural images. For natural images, since the histogram is shifted to the right, this means that on an average, the variance of the distribution of the MSCN coefficients is more, compared to synthetic images.

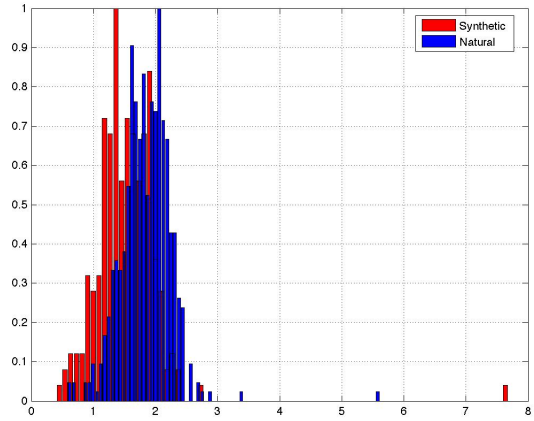


Fig. 7: Normalized histogram of the shape parameter β obtained over 221 synthetic and 500 natural images. Note how β for natural images tend to cluster around 2, indicating a Gaussian-like distribution of the MSCN coefficients. The synthetic images show more variability in the value of β.

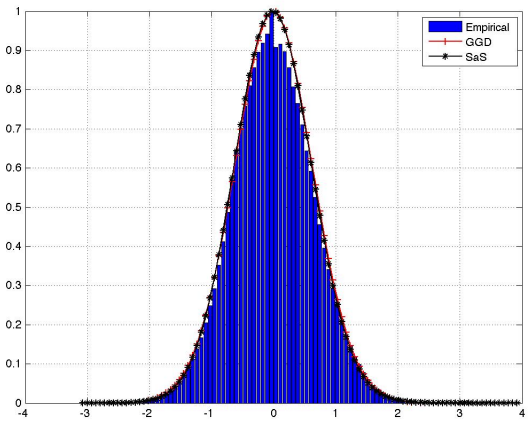


Fig. 9: Empirical distribution of the MSCN coefficients obtained from (1), and the fitted GGD and SαS distributions, after parameter estimation from the sample histogram. The distribution shows a Gaussian signature resulting is $\alpha = 0.8871$, $\beta = 1.9440$ for the GGD fit, and $\alpha = 1.9801$ for the SαS fit.

to the shape parameter of a Gaussian distribution. For synthetic images, the peak of the distributions occurs for $\beta < 2$, which means that more leptokurtic GGDs are needed to model the MSCN coefficients.

Fig. 8 shows the histogram of the GGD scale parameter α. For a fixed β, the variance of a GGD is proportional to α^2. For natural images, the distribution of α is found to be have a mean higher than the corresponding distribution for synthetic images.

Fig. 9 shows the empirical distribution of the MSCN coefficients of an image from the database. The GGD and SαS models are overlaid on top to show the model match. If the X and Y components of a 2D vector are independent Gaussian distributions having a zero mean, the vector amplitude is distributed according to the Rayleigh distribution. So, we also chose to observe how accurately the Rayleigh distribution can model the amplitude of the gradient of the MSCN coefficients, if the MSCN coefficients themselves are distributed close to a Gaussian distribution. Fig. 10 shows both the empirical distribution of the gradient of the MSCN coefficients (computed by the Sobel operator), and the best-fitting Rayleigh distribution overlaid on top of it. We also chose to study how other asymmetric distributions, like the Weibull and Nakagami, can account for the distribution of the MSCN gradients.

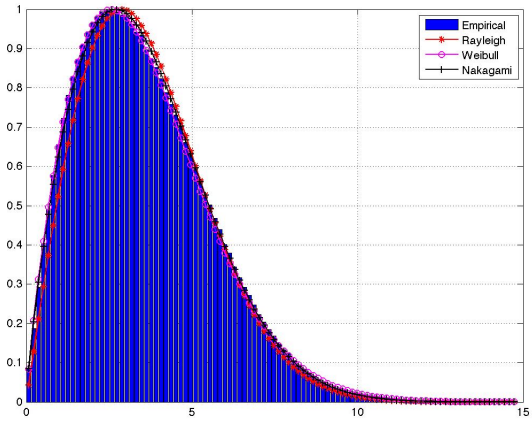


Fig. 10: Empirical distribution of the gradient of the MSCN coefficients obtained from (1), and the fitted Rayleigh, Weibull, and Nakagami distributions, after parameter estimation from the sample histogram, shown for more Gaussian-like distribution of the MSCN coefficients.

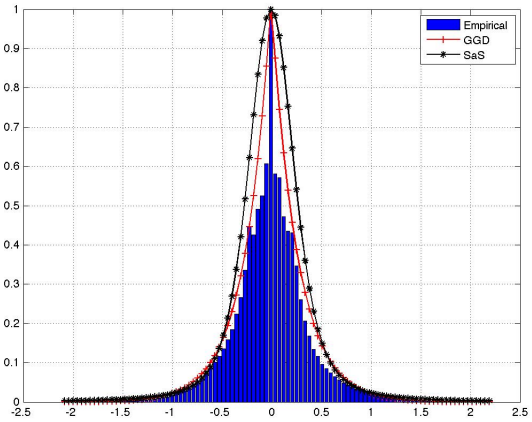


Fig. 11: Empirical distribution of the MSCN coefficients obtained from (1), and the fitted GGD and SαS distributions, after parameter estimation from the sample histogram. The distribution shows a Laplacian signature resulting is $\alpha = 0.2667$, $\beta = 1.0194$ for the GGD fit, and $\alpha = 1.4795$ for the SαS fit.

[7] considered natural scenes, which were primarily photographed in a natural environment, such as a forest, and contained thick foliage, streams, and rocks. For natural or synthetic scenes containing man-made objects, the distribution of the MSCN coefficients are found to deviate from the Gaussian distribution, and this is manifested in the value of the shape parameter β of the GGD model, which best fits the data. One example of a Laplacian type distribution of the MSCN coefficients has been shown in Fig. 11. The shape parameter is close to 1, which indicates a more leptokurtic distribution.

However, as the distribution of the magnitude of the MSCN gradients tend to deviate from a Gaussian model, the corresponding distribution of the gradients of MSCN coefficients also start to deviate from the Rayleigh distribution. Fig. 12 shows the mismatch in the empirical distribution of the gradient magnitude, and the attempts of fitting a Rayleigh, Weibull, or Nakagami distribution.

For modeling the distribution of the 'log-contrast' MSCN coefficients obtained by (5), we consider the generic α-stable and the Skewed Gaussian distribution as shown in Fig. 13, since the empirical distributions were found to show some negative skew. However, since the Skewed Gaussian distribution can only model moderately skewed distributions, with sample skewness lying between $[-1, 1]$, it gives a bad fit for heavily skewed distributions.

In order to quantify the extent to which the probability models fit the empirical distributions, we used the mean-squared error, and the J-divergence. For two probability distributions, the J-divergence between them is defined as the sum of the two possible Kullback-Leibler distances (provided they exist). We also performed χ^2

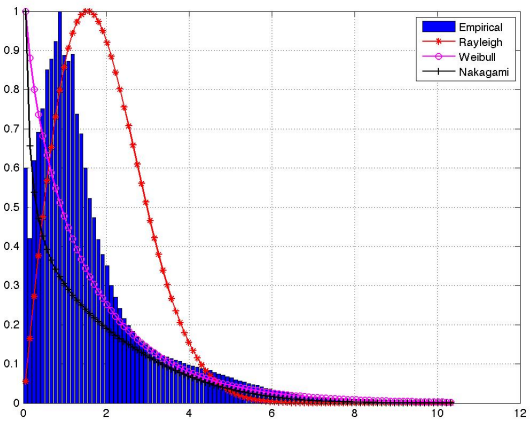


Fig. 12: Empirical distribution of the gradient of the MSCN coefficients obtained from (1), and the fitted Rayleigh, Weibull, and Nakagami distributions, after parameter estimation from the sample histogram, shown for more Laplacian like distribution of the MSCN coefficients.

tests at 1% confidence interval. The null hypothesis was assumed to be the distribution which we were trying to fit to the empirical spatial domain data, and for all the cases, the null hypothesis got accepted. The chi-square values for all the cases were found to be smaller than the upper cut off of the χ^2 distribution, $\chi^2_{(0.01)} = 6.635$ with degree of freedom = 1, which indicates that the values were generated from the fitted distributions, instead of by chance. Tables I, II, and III below show the values of the mean square error, J-divergence, and Pearson's χ^2 values for the distributions, calculated over the synthetic image database. Table I shows that the GGD gives a slightly better fit to the empirical distribution of the MSCN pixels

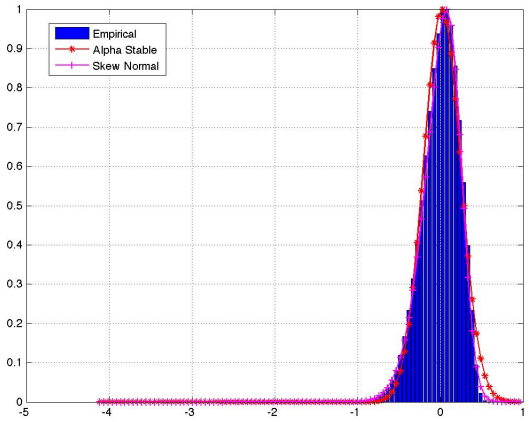


Fig. 13: Empirical distribution of the log contrast of the MSCN coefficients obtained from (6), and the fitted α-stable, and Skewed Gaussian distribution, after parameter estimation.

compared to the SαS distribution, both in terms of the mean square error and the J-divergence. For the gradient of the MSCN pixels, the Rayleigh distribution is found to fit the empirical distribution best, followed closely by the Nakagami distribution, as indicated by Table II. Table III shows that for the log-contrast MSCN pixels, the α-Stable distribution yields a better fit compared to the Skewed Gaussian distribution.

TABLE I: Mean square error, J-Divergence, and Pearson's χ^2 values for the distributions fitted to the histogram of the MSCN coefficients of an image, obtained from (1) for all the considered parametric families, averaged over the entire database

	MSE	J	χ^2
GGD	**0.00257**	**0.0772**	0.00252
SαS	0.00264	0.0948	0.00174

TABLE II: Mean square error, J-Divergence, and Pearson's χ^2 values for the distributions fitted to the histogram of gradients of the MSCN coefficients of an image, obtained from (1) for all the considered parametric families, averaged over the entire database

	MSE	J	χ^2
Rayleigh	**0.00891**	**4.730**	0.769
Weibull	0.0251	5.00432	0.663
Nakagami	0.00916	5.304	0.892

V. Conclusion

This paper aims to analyze how well the spatial domain natural scene statistics, which have been used in natural image and video quality assessment, apply to synthetic scenes. The results show that in the spatial domain, for pristine images, synthetic scene statistics can also be modeled in a fashion similar to natural scene statistics. The difference in the perception of natural and synthetic concerns higher level cognition factors of the human brain. The next step would be to study how distortions of synthetic images affect the spatial domain statistics. Future avenues of work also include analysis of color images, and the joint spatial distribution of the color channels using a suitable color space. They might prove to be useful in designing no-reference metrics aimed towards color specific distortions, such as those arising from tonemapping of high dynamic range images.

TABLE III: Mean square error, J-Divergence, and Pearson's χ^2 values for the distributions fitted to the histogram of log-contrast MSCN coefficients obtained from (6) for all the considered parametric families, averaged over the entire database

	MSE	J	χ^2
α-Stable	**0.00949**	**0.245**	0.0327
Skewed Gaussian	0.0435	1.360	0.0397

References

[1] A. Moorthy and A. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. on Image Processing*, vol. 20, no. 12, pp. 3350–3364, Dec 2011.

[2] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain." *IEEE Trans. on Image Processing*, vol. 21, no. 8, pp. 3339–3352, 2012.

[3] A. Mittal, A. Moorthy, and A. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. on Image Processing*, vol. 21, no. 12, pp. 4695–4708, Dec 2012.

[4] T.-T. Ng, S.-F. Chang, J. Hsu, L. Xie, and M.-P. Tsui, "Physics-motivated features for distinguishing photographic images and computer graphics," in *Proc. ACM Int. Conf. on Multimedia*, 2005, pp. 239–248.

[5] R. Herzog, M. Cadk, T. O. Aydin, K. I. Kim, K. Myszkowski, and H.-P. Seidel, "NoRM: No-Reference image quality metric for realistic image synthesis." *Comput. Graph. Forum*, vol. 31, no. 2, pp. 545–554, 2012.

[6] D. Kundu and B. L. Evans, "ESPL synthetic image database," April 2014, http://signal.ece.utexas.edu/~bevans/synthetic/index.html.

[7] D. L. Ruderman and W. Bialek, "Statistics of natural images: Scaling in the woods," in *Neural Information Processing Systems Conference and Workshops*, 1993, pp. 551–558.

[8] E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," *Annual Review of Neuroscience*, vol. 24, pp. 1193–1216, 2001.

[9] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[10] M. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance," *IEEE Trans. on Image Processing*, vol. 11, no. 2, pp. 146–158, Feb 2002.

[11] J. McCulloch, "Simple consistent estimators of stable distribution parameters," *Communications in Statistics - Simulation and Computation*, vol. 15, no. 4, pp. 1109–36, 1986.

[12] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. Int'l Conf. Comp. Vision*, vol. 2, July 2001, pp. 416–423.