https://doi.org/10.2352/EI.2023.35.15.COLOR-188

Spatiochromatic and temporal natural image statistics modelling: applications from display analysis to neural networks

Scott Daly, Timo Kunkel, Guan-Ming Su, Anustup Choudhury; Dolby Labs, Inc.; San Francisco and Sunnyvale, USA

Abstract

Using synthetic test patterns that are simple in geometry or mathematical description is common when assessing algorithms, models, and display systems. A more complex approach is the use of 'real-world' content that is captured for example by a camera. For the application of display system assessments, an additional test pattern approach was recently proposed [1]. This approach uses test sequences following spatio-chromatic noise properties inspired by natural image statistics. Such sequences can offer complex image and video properties that are statistically more common with content captured by a camera while maintaining properties that can be more easily quantified as is typical with synthetic test pattern. In this paper we investigate additional applications for spatio-chromatic noise sequences beyond the ones used with display assessment. In particular, the applications of image data compression and neural fields are discussed in detail.

Introduction

Developing algorithms, models and systems are a common task in many areas of research and engineering. One crucial component of this task is the evaluation of an algorithm or model's performance. In the field of electronic imaging, this includes image and video compression, decompression, signal modification such as tone-mapping, as well as rendering and display. In addition, methods to visually characterize these systems are important for example to carry out psychophysical assessments.

Over the past decades, various approaches on how to evaluate the various elements of an imaging pipeline have been developed. Many of them require test signals that can be processed, sampled, optically measured, and otherwise analyzed.

With display characterization and psychophysics, there has been a long tradition of assessing performance via synthetic test patterns, which are generally simple in geometry or mathematical description. The use of such test signals is for example common when designing new display systems or characterizing them later [2]. Another common test material consists of 'real world' content captured with a camera, either as still image or as video sequence. Such 'practical imagery', that is, images or videos that will be encoded, decoded, and viewed on a display by non-expert consumers includes natural, civilization and entertainment imagery typically viewed for entertainment and educational purposes. Depending on the application it can also include x-ray, MRI, and histology for medical imaging or with satellite and radar imagery used in the geosciences. Such practical imagery is commonly used when assessing the performance of image and video encoders but has also found use in other areas such as display system assessment [3,4].

To obtain objectively quantifiable results, various types of image statistics are typically computed, which are then analyzed to deduct an algorithm or model's performance. While image statistics computed and measured from synthetic evaluation content can be straight forward and give detailed information such as peak luminance or average picture level (APL) [2], the scope of these results is often low level and typically far from the typical use case of the elements in an imaging pipeline. Practical imagery is challenged by the opposite problem: Even through the content is typically closer to the actual use cases, there is a lack of consistently quantifiable parameters. There has been some effort to quantify such imagery e.g., by content color volume, noise level [4], and no-reference metrics such as sharpness [5] or by carefully selecting practical test images [6,7]. Another approach was to replace test targets with a standardized test video that had luminance statistics matching 48 hours of broadcast content [3]. More recently, large data sets of images have been used to train various neural network models for tasks such as super resolution, bit rate compression, and dynamic range mapping [7]. However, images are not ergodic, and possibly not even wide-sense stationary (ergodicity describes a system that cannot be reduced into smaller components and requires the variance and other statistical properties to be constant, while wide-sense stationarity only requires the 1st and 2nd moments to be constant).

Consulting image statistics that are based on averages or manufacturing deviation criteria (e.g., 3σ) bears a risk as these averages might not cover or reflect the properties of corner cases. For example, a single movie may already contain several corner cases over its duration. Therefore, one key desire is to not only design and test for average content but also include imagery that challenges the algorithms, models, and technology. With that, algorithm development and testing for image and video processing has often been ad hoc, with a mixture of geometric test targets and hand selected test images, sometimes aiming to be corner cases, sometimes not.

An option to bridge this gap of robustness quantification and stress testing practicality that lies between simple synthetic test targets and content captured with cameras has been proposed by Kunkel and Daly [1] for the context of display performance assessment. They proposed the use of spatio-temporal noise sequences based on $1/f^{\alpha}$ distributions following natural image statistics.

In this work, we propose to extend the use of such imagery to compactly probe the wide variety of image possibilities for algorithmic development. While we don't suggest replacing actual practical imagery, we believe such noise fields can augment image algorithm analysis.

To address the problem of non-ergodicity, we allow the basic power term α in the natural image statistic model to vary over a large range in a video (see [1]), such that it includes the extremes of white noise and low frequency gradients, not only achromatically, but also by using color image statistics models that include decorrelated colors to generate the RGB video. We will present results for traditional adaptive data compression (with chromatic subsampling), as well as a more contemporary neural network approach using Neural Fields [8] as applied to upscaling and denoising.

Properties of Natural Image Statistics

Stochastic noise and its quantifiable statistics, such as the noise power spectra, can provide a rigorous framework such as white noise analysis with its history in engineering and vision science [9]. Natural image statistics are well known to having spatial frequency power spectra that has a $1/f^{\alpha}$ behavior. The $1/f$ family of statistics exhibit a simple relationship between spectral density S(f) and frequency f [10] where the value of α affects the frequency distribution of the noise. The α values related to natural mage statistics are typically stated as being between 2 and 4. This indicates an invariance to scale. Further work has theorized how the visual system is tuned for such statistics in the visual cortex (V1) [11]. Color image statistics also show an invariance to scale. The luminance histogram is typically understood to be log normal with respect to luminance, although for HDR images, a subcomponent with skew toward much higher luminance levels is observed. Color statistics were initially described at the simplest level via the gray world hypothesis [12], but more details are available, even at the hyperspectral [13]. The power function for HDR was found to increase from the lower values of 2 to more typical values of 4 and 5 [14].

For temporal statistics, data tends to be measured primarily for media with a focus on the motion statistics via optical flow. Another example can be given by scene cut statistics also following a 1/f α distribution [15]. Statistics for purely natural and human made environments (e.g., buildings and the resulting perspective geometry) have been studied, having different orientation statistics [16].

For color aspects of statistics, there is work looking at histogram statistics, such as in cone space [17], hyperspectral [18], or the non-natural TV broadcast space [19]. For color spatial frequencies we only found one study using RGB analysis and one looking at opponent color. The findings of the RGB study [20] are shown in Figure 1a. We can see the slopes of the amplitude frequency spectra are similar and close (α = -1) across red, green, and blue channels. There is a slightly different slope for the blue channel, which may relate the chromatic aberration in the lens used to capture the scenes, or the blue-yellow vector of daylight illumination.

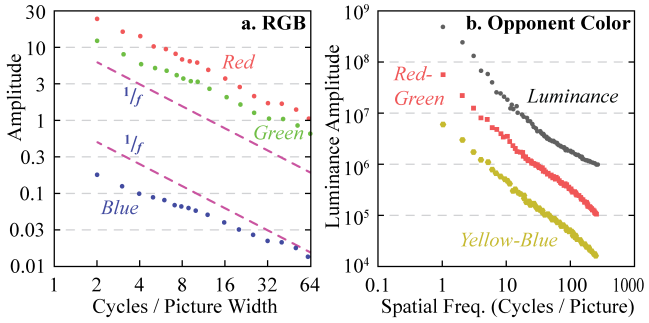


Figure 1: a. Amplitude spectra averaged across 19 images (dots) for the red, green, and blue channel (after Burton [20]). b. Spatio-chromatic statistics in the achromatic and opponent color domain (after Parraga [21]).

Another approach is to look at the spatio-chromatic frequencies across opponent colors, as motivated by the concept of physiology becoming tuned to the real-world spatio-chromatic statistics or motivated by common video signal formats that can take advantage of this information. The study of frequency statistics in the opponent color domain [21] is shown in Figure 1b., where achromatic properties are assessed from the luminance channel, and the red-green and yellow-blue opponency vectors are calculated from LMS cone responses. The behavior in terms of slopes was similar across all three channels, similarly, having an approach of α = -1, but with some slope reduction at the highest frequencies. In one image they found the red-green channel to have a steeper slope, so this is a field where more data is certainly needed. A similarity of slopes is expected as the opponent color channel responses are linear matrix transforms of RGB but could be different if such matrices occur after nonlinearities. For our work, we will assume them having the same slope, and that they translate to the nonlinear opponent color space of CIELAB.

Preparing the Test Imagery

To replicate properties of natural image statistics, we computed noise sequences as described by Kunkel and Daly [1], including achromatic and chromatic noise fields.

Figure 2 illustrates the high-level concept of generating individual 2D noise images from a random value seed φ (a.). Based on the selected α value (b.) the specific spatial frequency distribution can be computed (c.), which can be converted to a 2D image (e.) using an inverse Fourier transform (d.).

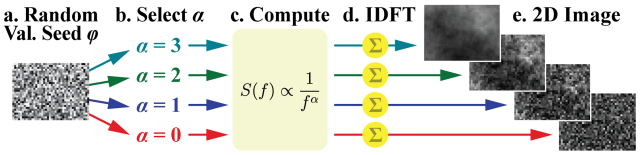


Figure 2: High Level Concept to generate 2D noise images as described by Kunkel & Daly [1].

The temporal component of the noise sequence was generated with a new seed value φ for each frame. The final noise images were then computed using α values ranging from 0 (white noise) to 5 (very low pass consisting of smooth gradients) followed by blending the frames together using a Gaussian temporal blend kernel to form a smooth continuous motion (see [1] for more details). For the chromatic noise fields, we computed three independent noise sequences and assigned them to the CIELAB opponent channels L*, a* and b* as shown in Figure 3.

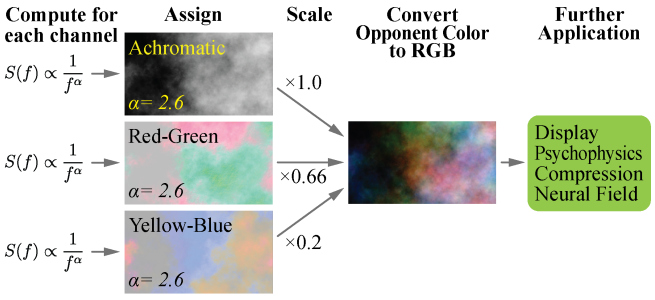


Figure 3: Generating opponent color noise sequences that after channel scaling are converted to RGB and provided to further applications.

Based on referenced studies of chromatic natural image statistics, the ratio of achromatic to red-green and blue-yellow opponent channels were scaled in the ration of 1:0.66:0.2 followed

by a conversion to sRGB. In total, a sequence of 360 images was computed (15 sec. at 24 fps) with each image having a size of 512x512px. This test image sequence can then be used by further applications.

Key applications that benefit from this approach include Display assessment, psychophysics, compression analysis and Neural Networks. The application in the context of display assessment has been described in previous work [1,3]. In the context of psychophysics and psychovisual research, the noise images and sequences can offer several beneficial properties such as scale invariance and non-descript texture. With that, such images can serve as stimulus background, resting target and luminance anchor. Figure 4 shows an example setup using a $1/f^{\alpha}$ target as background for a visual stimulus.

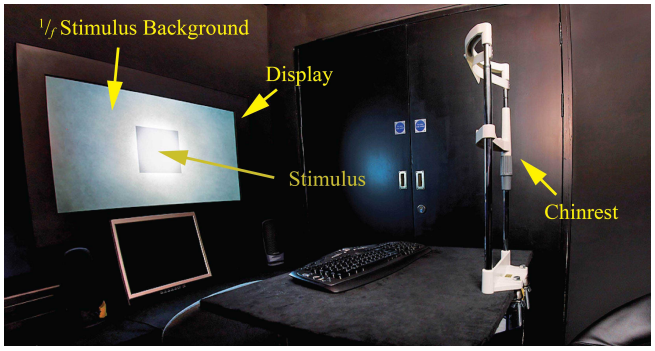


Figure 4: Example use of spatial noise targets for psychovisual research

Further, image data compression and neural fields are beneficial applications for the noise sequences. In the following, these latter two areas are discussed in detail.

Application to Image Data Compression

Two key phenomena in image data compression are the diametrically opposed aspects of entropy and perceptual masking. For a fixed quantizer codec, increasing entropy increases the bitrate (i.e., bit total per image). However, increasing entropy also increases perceptual masking, which allows the signal to mask the distortions. This allows the quantization to be increased, which in general lowers the bitrate. It is currently unknown which of these two trends dominates. Of course, such an answer would be specific to the actual image, which may consist of low and high entropy regions. Remember, that we are treating these noise fields as surrogates of signals, so please do not get confused with the typical signal to noise ratio (SNR) terminology.

Further, the spatial frequency spectra of the content will play a role in the perceptual masking because spectra with high alpha values result in low amplitudes for the higher spatial frequencies. Such lower amplitudes mean the signals are closer to threshold.

Figure 5 examines perceptual masking further where the threshold for seeing the compression distortion is plotted on the vertical axis, and the signal causing the masking on the horizontal axis. We can see two key regions, a low noise asymptote and a high noise asymptote (using conventional psychophysical terminology; the noise being referred to on the asymptotes is the signal to be compressed, which is the noise field). The low noise asymptote indicates that there is no masking, and the threshold is unaffected. The high-noise asymptote (slope of 1 in a log-log plot) means the masking can be modeled as a simple SNR ratio. We can also see a transition region between the two asymptotes where the masking gradually reduces as the signal contrast reduces.

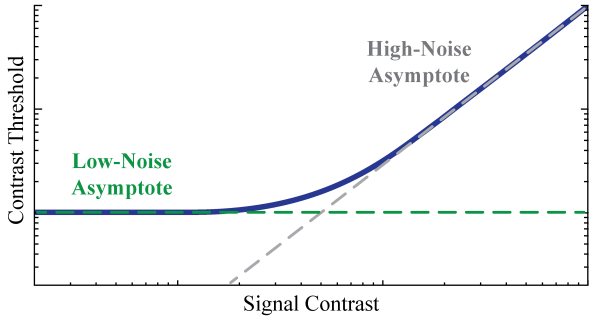


Figure 5: Masking basics

For the high spatial frequencies with higher values of α, the signal contrast is reduced compared to image with lower α. This means the masking is likely reduced for such frequencies as compared to the case with a lower α values.

Masking Study

Studying masking behavior using these noise fields can be instructive for quantifying such complex effects. We carried out two studies to investigate this further. The first study kept the quantization parameter constant in a JPEG still image compression scheme, while the second allowed the quantization parameter to increase as needed to keep a constant bitrate. These two cases are referred to as constant quantization or constant bit total (analogous to bit-rate if the frames occurred at a temporal rate).

We found for achromatic noise fields, that in the constant quantization case the distortions were increasingly more visible as the alpha value increased, and less visible as the value of α decreased. In this case we conclude that masking is a stronger effect in improving image quality than entropy is in degrading quality. Of course, for the constant quantization case, the resulting bit totals steadily increased with decreasing α (& increasing entropy). In the case for constant bit totals, and thus variable quantization, we found the opposite trend. Quality was reduced as the α value decreased. The masking was not able to overcome the increasing distortions caused by the increasing entropy caused by the flatter spectra of lower α values. Even though the noise fields with lower high spatial frequency amplitudes (i.e., the higher α values) had some signals transitioning to the lower noise asymptote, and thus causing less masking, this reduced masking did not enable the distortions to become more visible. This is due to a significantly reduced entropy level enabling the entropy coder to easily encode. With that, the reduced entropy issue dominated over the reduced masking. This was generally expected, and therefore the results are not shown.

However, for chromatic noise, the results were not as simple. Figure 6 shows the results for the chromatic noise for a series of α values using constant quantization, including a magnification of an image region for improved visibility of potential artifacts. The bit totals for each image are also indicated in the figure. The distortions were more visible for the images with high α values, and in the form of horizontal and vertical distortions due to partial blocking visibility, as in the achromatic case. However, there were also visible distortions for the case with the lowest α value of 0 (a.) but manifested in a different form. In this case, very high spatial color features that were present in the reference image were lost due to the

compression. We understand this to be due to the lower amplitude red-green and blue-yellow channel's signal being quantized to zero.

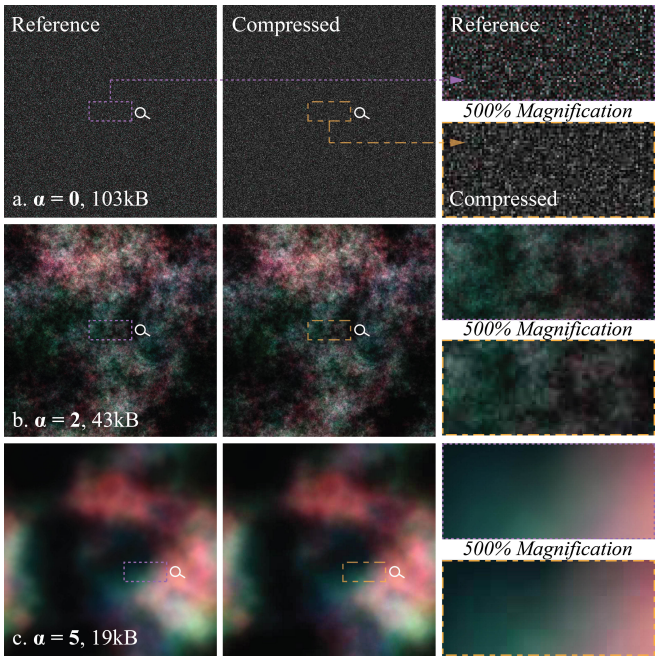


Figure 6: Results for constant quantization of spatio-chromatic noise fields of increasing alpha (decreasing entropy). The bit totals are for the full 512x512px images. The full-size images are intended to be viewed at 3280 pixels (~6H) so that the Nyquist is approximately 30 cycles/deg.

With that, there appears to be a sweet spot of α values between 0 and 5 that gave the best results in terms of visibility for fixed quantization compression, which does not occur for the achromatic versions.

For the study using an adaptive quantizer set so that the bit totals were approximately constant, the results favored the higher α values shown in Figure 7 using the same reference images as in Figure 6.

One can see in the image with the lowest α value that the high spatial frequency chromatic features ('color dots') are entirely missing, and the image looks completely achromatic, as was the case for the constant quantization criteria. In addition, structural patterns are visible within the noise (viewing with a dark surround may be required). With α = 2, a slight loss of detail is visible in comparison to the reference, while the highest α value of 5 consisting entirely of low frequency chromatic gradients shows no visible distortion and is in fact visually lossless at even closer viewing distances than 6H (6 picture heights viewing distance).

For both studies, constant quantization, and constant bit total, we computed results for integer values of α ranging from 0 to 5, but due to space constraints only show the endpoints and the value of alpha most often cited for natural image statistics (α = 2).

This preliminary study into the application to image data compression shows the value of spatiochromatic noise fields in understanding the most basic trends. For examples we can observe interactions of basic image features (the α parameter) with the perceptual issues of masking and CSF (i.e., visibility as a function of viewing distance). Instead of relying of the arbitrary regions of differing α values in optically captured real world imagery, we can isolate and draw conclusions specific to potential localized image statistics via the α parameter.

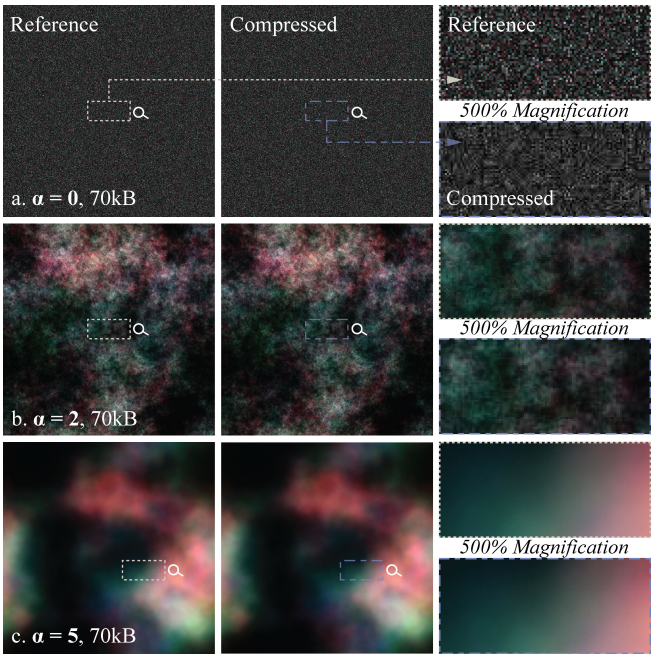


Figure 7: Results for constant bit totals by using adaptive quantization of spatiochromatic noise fields of increasing alpha (decreasing entropy). The alpha values listed are for the achromatic channel. The opponent color red-green and blue-yellow channels have the same alpha values but reduced amplitudes in the ratio of 1:0.66:0.2. The resulting bit totals for the full 512x512 images are all approximately 70kb. The reference images and viewing conditions are the same as in Figure 6.

Application to Neural Fields

Another application can be found in neural fields. A neural field [8] is a type of coordinate based neural network that parameterizes physical properties of scenes or objects across space and time. It has applications in visual computing problems such as 3D image reconstruction, image synthesis, etc. In neural field frameworks, field quantities are produced by sampling coordinates and feeding them to a neural network. For example, consider the 3D scene representation 'Neural Radiance Field' (NeRF). A typical NeRF takes an input from the spatial location (x, y, z) and the viewing direction (θ, ϕ) and outputs the volume density and view dependent emitted radiance of that coordinate. The output volume density and RGB color are followed by volume rendering to construct a 2D projected image. The 'Local Implicit Image Function' [22] is one of the recent applications using neural fields. An image is represented as set of latent code to predict the RGB value of a given pixel from the (x, y) coordinate and the 2D deep feature.

Positional Encoding

Neural fields are known to suffer from a loss of high frequency information during rendering. To mitigate that, one common approach is to apply positional encoding. For that, the inputs to the neural fields, which are typically pixel co-ordinates, are mapped to a higher dimensional space [23]. For neural scene representation, the performance of a neural network is significantly improved by

mapping the position coordinates p from R to R^{2L} where L is the number of frequencies.

A typical mapping γ [24] acting on a coordinate p can be represented as:

$$\gamma(p) = [\sin(2^{l_0}\pi p) \ \cos(2^{l_0}\pi p) \ \sin(2^{l_1}\pi p) \ \cos(2^{l_1}\pi p) \ \ldots \ \sin(2^{l_{L-1}}\pi p) \ \cos(2^{l_{L-1}}\pi p)] \quad (1)$$

where $\{l_0, l_1, \ldots, l_{L-1}\}$ are integers. In the typical setting, $l_k = k$.

Multiple Layer Perceptron Network (MLP)

The core component of the neural field is the multi-layer perceptron (MLP). The basic building block of an MLP layer consists of one linear sub-layer followed by a 'Rectified Linear Unit' (ReLU) activation sub-layer. Each linear layer has different number of neurons. A deep MLP network can be constructed by stacking multiple building blocks sequentially. The details of each MLP layer are illustrated in Figure 8.

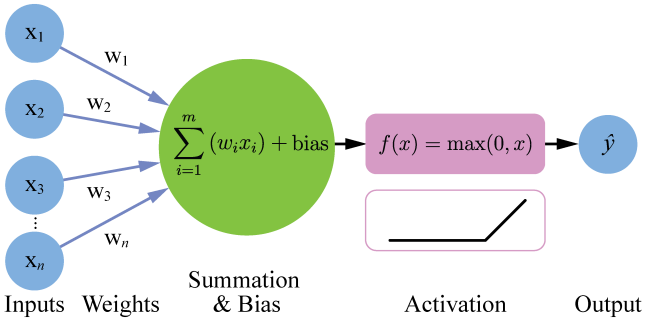


Figure 8: Basic building block using weighted linear sub-layer and activation layer

The MLP layer takes several inputs (blue circle) and multiplies each input with an associated *weight* and adds a *bias* (green circle). The parameters at k^{th} layer weights ($\boldsymbol{W}_k$) and *bias* ($\boldsymbol{b}_k$), need to be trained to optimize the loss function. The output will be adjusted by an activation function (rectangle), such as ReLU. The equation for ReLU activation function is given by:

$$ReLU(x) = \max(0, x) \quad (2)$$

Sometimes there is a requirement that the final network output needs to be in the range [0,1]. A sigmoid layer can then be added as the final layer of the MLP which normalizes every real number to between 0 and 1. The equation for the sigmoid is given by:

$$S(x) = \frac{1}{1+e^{-x}} \quad (3)$$

Experimental Setup

In the following, we describe how the neural field can be applied to model a spatio-chromatic noise image sequence. We are using the same three-channel color sequences as in the previous section. This means that there are three color values for every pixel and an additional dimension for time t (since the source is an image sequence), making the field effectively 4 dimensional. The neural field takes the co-ordinate pixel position as input and predicts the color values per pixel for every frame in the sequence. With the 2D coordinate (x, y) and time (dt), the neural field problem can be represented as follows:

$$(\hat{R}, \hat{G}, \hat{B}) = MLP_{\Phi}(\gamma(x), \gamma(y), dt) \quad (4)$$

where γ is the positional encoding function applied on every pixel of the image and $\hat{R}, \hat{G}, \hat{B}$ are the red, green and blue values of every pixel.

The optimal MLP parameters is formulated as

$$\Phi^* = arg\min_{\Phi} D(\{(\hat{R}, \hat{G}, \hat{B})\}, \{R^{gt}, G^{gt}, B^{gt}\}) \quad (5)$$

where R^{gt}, G^{gt}, B^{gt} are the red, green, and blue values of the reference image of the sequence.

The loss function is the overall normalized root mean squared error from all three color channels of all pixels for the entire image sequence.

Network architecture and implementation details

For our experiments, we use a MLP with 5 hidden layers. The structure of the MLP is shown in Table 1.

Table 1: Network architecture of MLP

Layer	Input Channels	Output Channels	Input
Layer 1	41	256	$\gamma(I)$
Layer 2	256	128	Layer 1
Layer 3	128	64	Layer 2
Layer 4	64	32	Layer 3
Layer 5	32	16	Layer 4
Layer 6	16	3	Layer 5

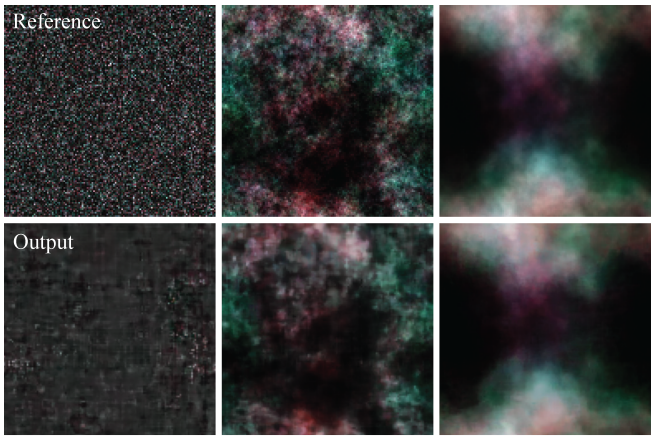
The input to the first layer is denoted as $\gamma(I)$. Here, I is the input image and γ is the positional encoding applied to the input pixels. We use Equation 1 as the mapping function for γ using a value of $L = 10$ (resulting in 40 dimensions) and an additional value of time ($t = 1$) resulting in 1 additional dimension. The final layer has a sigmoid function and outputs the color value (Red, Green and Blue) for every pixel of every frame of the sequence.

We implement the framework in PyTorch [25] and use the Adam optimizer [26] with a learning rate of 0.001. The other Adam hyper-parameters are left at default values The optimization for a single sequence takes 3500 epochs to converge on a single NVIDIA A100 GPU.

Results

For efficient/faster training to demonstrate the application, instead of creating a neural field using the entire sequence consisting of 360 frames, we create a neural field to model 11 frames from the spatio-chromatic temporal noise sequence. Starting from frame 0, we pick every 20th frame up to frame 200. The MLP takes the coordinates and the time interval as an input and predicts the output frame. Visual results of the reconstruction for alpha values of 0.8, 2.6 and 4.3 based on frames 10, 80 and 120, respectively, are shown in Figure 9.

The image quality differences are the strongest for Figure 9a. showing the output of a low α value input as neural fields are not the best in modelling high special frequencies. The higher α values of Figure 9b. and c. show a loss of detail suggesting localized blurring although the general appearance resembles the reference. Larger neural fields can potentially further improve the performance especially for the higher α values. For the lower α values, we might need a better positional encoding mapping.



a. $\alpha = 0.8$ PSNR = 18.97 b. $\alpha = 2.6$ PSNR = 26.85 c. $\alpha = 4.3$ PSNR = 36.34

Figure 9: Visual results of modeling spatio-chromatic noise using neural fields along with the corresponding PSNR values. The top row corresponds to the frames from the reference image sequence and the bottom row corresponds to the reconstruction using the neural field.

The PSNR results for the sequence frames are shown in Figure 10.

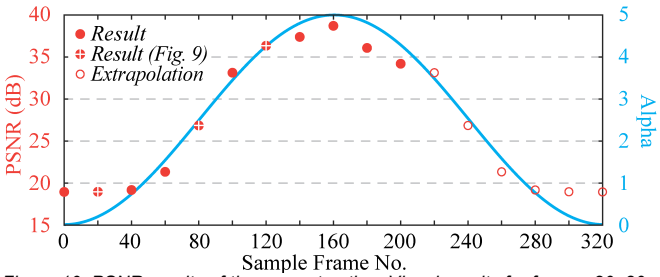


Figure 10: PSNR results of the reconstruction. Visual results for frames 20, 80 and 120 are shown in Figure 9. The results after frame 200 are extrapolated and were not computed by the neural field. As reference, the alpha values used to compute the reference image sequence are shown as well.

As can be seen in Figure 9 and Figure 10, the objective measure, PSNR is high (> 30dB) for the frames with higher α values. However, for the frames with lower α values, the PSNR is not very high. There seems to be a strong correlation between the α values and the reconstruction using the neural field. This shows that although the neural fields can be used to model the spatio-chromatic temporal noise sequence, there still seems to be some scope of improvement, especially for the lower α values.

Summary

We have shown that spatiochromatic noise fields are useful beyond display assessment tasks, both as static images and as image sequences. In particular, the changes of spatial frequency distribution facilitated by adjusting the alpha value provide interesting opportunities to assess models and systems that are part of an imaging pipeline. This work represents initial investigations of using these sequences in combination with image compression and neural field applications. Nevertheless, further studies are needed to extend our understanding of how spatiochromatic noise sequences can be of use.

References

[1] T. Kunkel & S. Daly, “57-1: Spatiotemporal Noise Targets Inspired by Natural Imagery Statistics”. SID Symposium Digest of Technical Papers, 51:842-845, 2020.

[2] IEC 62087-1. Audio, video, and related equipment – Determination of power consumption. 2015

[3] T. Kunkel, & F. Friedrich, “Utilizing advanced spatio-temporal backgrounds with dynamic test signals for high dynamic range display metrology”. J Soc Inf Display. 2022; 30(5): 423– 432. https://doi.org/10.1002/jsid.1125

[4] M. Smith & M. Zink, “Film Grain and Film Scanner Noise in HDR Video”. SMPTE Annual Meeting. 2019

[5] N. Nill & B. Bouzas, “Objective Image Quality Derived from Digital Image Power Spectra”. SPIE Optical En., 1992, 31(4).

[6] A. Lewandowska Tomaszewska, “Scene Reduction for Subjective Image Quality Assessment”. JEI, 2016; 25(1)

[7] Farnand SP, Fairchild MP. Designing pictorial stimuli for perceptual experiments. Appl. Opt. 2014; 53, C72-C82

[8] Y Xie, T Takikawa, S Saito, O Litany, S Yan, N Khan, F Tombari, J Tompkin, V Sitzmann, S Sridhar, "Neural Fields in Visual Computing and Beyond", Eurographics / CGF State-of-the-Art Report, 2022

[9] Sakai HM, Naka K, Korenberg MJ. White-noise analysis in visual neuroscience. Vis Neurosci. 1988;1(3):287-96.

[10] Voss RF. Fractals in Nature: From Characterisation to Simulation. In Peitgen HO, Saupe D (eds) The science of fractal images. Springer, 1988; pp 21-90.

[11] Field DJ. Relations between the statistics of natural images and the response properties of cortical cells. J. Opt. Soc. Am. A, 1987; 4:2379-2394

[12] R. M. Evans, “Fluorescence and gray content of surface colors”. JOSA V 49 #11, 1049-1059, 1959

[13] A. Chakrabarti & T. Zickler, “Statistics of real-world hyperspectral images”. CVPR 2011.

[14] R.O. Dror, A.S. Willsky & E.H. Adelson, “Statistical character-ization of real-world illumination”. J. of Vision, 4(9):11, 2004.

[15] J. E. Cutting, J.E. DeLong & C.E. Nothelfer CE, “Attention and the Evolution of Hollywood Film”. Psych. Sci., 21(3):432-9, 2010.

[16] A. Torralba & A. Oliva, “Statistics of natural image categories”, Network: Computational Neural Systems, 14 391-412, 2003.

[17] M. Webster & J. Mollon, “Adaptation and the color statisictis of natural images”. Vis Res. V37, 3282-3298, 1987.

[18] A.Chakrabarti & T. Zickler, “Statistics of real-world hyperspectral images”. IEEE CVPR, 2011.

[19] F. Moss, F. Zhang, R. Baddeley & D. Bull, “What’s on TV: Large scale quantitative characterization of modern broadcast content”. ICIP, 2016.

[20] G.J. Burton & I.R. Moorhead, “Color and spatial structure in natural scenes”. Applied optics. V26 #1, 1987.

[21] C. A. Parraga, T. Troscianko & D.J. Tolhurst, “Spatiochromatic properties of natural images and human vision”. Current Biology, V12 , 483-487, 2002.

[22] Y. Chen, S. Liu & X. Wang, "Learning Continuous Image Representation with Local Implicit Image Function", CVPR 2021

[23] N. Rahaman, A. Baratin, D. Arpit, F. Draxler, M. Lin, F.A. Hamprecht, Y. Bengio & A.C. Courville, “On the spectral bias of neural networks”. ICML, 2018.

[24] B. Mildenhall, P.P. Srinivasan, M. Tancik, J.T. Barron, R. Ramamoorthi & R. Ng. “Nerf: Representing scenes as neural radiance fields for view synthesis”. In European conference on computer vision, pp. 405-421. Springer, Cham, 2020.

[25] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai & S. Chintala, “Pytorch: An imperative style, high-performance deep learning library,” in Advances in Neural Information Processing Systems 32, pp. 8024–8035. 2019.

[26] D.P. Kingma & J. Ba, “Adam: A method for stochastic optimization”, International Conference on Learning Representations, ICLR, 2015.

Author Biography

Scott Daly is an applied vision scientist at Dolby Laboratories, with specialties in spatial, temporal, and chromatic vision. He has significant experience in applications toward display engineering, image processing, and video compression with over 100 technical papers. Current focal areas include high dynamic range, motion appearance, physiological assessment and preserving artistic intent. He has a B.S. in electrical engineering from NCSU and an M.S. in bioengineering from the University of Utah. Past accomplishments led to the Otto Schade award from SID in 2011, a SMPTE Motion Imaging Journal best paper award in 2021, a SID Distinguished Paper award in 2013 and a team technical Emmy in 1990. He is a member of IEEE, SID, and SMPTE. He recently completed the 100-patent dash in just under 30 years

Timo Kunkel is Director of Image Technology & Standards in the CTO office of Dolby Laboratories. Over the past 18 years, he has been investigating the technical and perceptual aspects of HDR and wide color gamut imaging with focus on advanced display approaches. He has published and taught about HDR concepts and technologies throughout the industry for many years. He is also a member and technical expert with the CIE, ICC, SID ICDM and IEC and currently serves as vice president on the board of the IS&T. He holds a Ph.D. in Computer Science from the University of Bristol, UK, and a M.S. from the University of Freiburg, Germany.

Anustup Choudhury received the Ph.D. and M.S. degrees in Computer Science from University of Southern California (USC), Los Angeles in 2012 and 2007 respectively. He is currently a Research Scientist at Dolby Laboratories. His research interests include image/video analysis and processing, pattern recognition, computer vision, computational photography, and machine learning. His work has been published in the proceedings of reputed conferences and journals such as ICIP, ICPR, CVPR, ICCV, CSVT and JOSA A. His papers have received special recognition as the best journal paper in SMPTE Motion Imaging Journal 2021, as one of the best papers in IEEE BTAS 2007 and as a top 10% paper at IEEE ICIP 2015. He is an inventor of over a dozen patents and is a member of IEEE.

Guan-Ming Su is the Director of Image Tech Dev at Dolby Labs, Sunnyvale, CA, USA. He served as an associate editor in Asia Pacific Signal and Information Processing Association (APSIPA) Transactions on Signal and Information Processing and IEEE MultiMedia Magazine, and Director of review board and R-Letter in IEEE Multimedia Communications Technical Committee. He also served in multiple IEEE international conferences such as TPC Co-chair in MIPR 2019, TPC Co-Chair in ICME 2021, and General Co-Chair in MIPR 2022. He also served as VP Industrial Relations and Development in APSIPA 2018 – 2019. He serves as Vice Chair for Conference in IEEE Technical Committee on Multimedia Computing (TCMC) since 2021. He obtained his Ph.D. degree from University of Maryland, College Park.