Information Content in Nonlinear Local Normalization Processing of Digital Images

Nesim Halyo, Zia-ur Rahman, Stephen K. Park
College of William & Mary, Department of Computer Science, Williamsburg, Virginia 23187.

ABSTRACT
In this paper, we investigate the rate at which information is transferred from the scene to the digital processed image in imaging systems using nonlinear Local Normalization. While a formula for the Rate of Information has been used in many studies of linear systems with Gaussian (Normal) signals, a formula applicable to nonlinear/non-Gaussian systems has not been available until now. We discuss a new formula for the Rate of Information developed by the authors for systems with nonlinear Local Normalization processing. The Local Normalization algorithm, which is similar to the Retinex algorithm, is formulated and discussed. As the name implies, this algorithm forms a local average signal and normalizes the detail (high spatial frequency) scene signal by the local average. The scene detail information transferred to the locally normalized image is seen to be at least as high as that transferred to the acquired digital image. Thus, no information loss occurs in using Local Normalization. A case study of the new Rate of Information formula shows that designing the image-gathering system cut-off frequency near or slightly below the Nyquist frequency maximizes the scene detail information transferred to the locally normalized image for all the conditions considered.

1. INTRODUCTION
Images, whether obtained by electro-optical means or by purely optical instruments, provide information about the scene they depict. Typically, the image-gathering system converts the radiance coming from the scene into a sampled digital signal which is then processed for enhancement, compression and restoration. The amount of information about the original scene which is contained in the processed image depends on the image-gathering system and the digital processing algorithms used in obtaining the image. Thus, the rate at which information is transferred from the scene to the processed digital image is a measure of the fidelity and the efficiency of the image-gathering system and the digital processing algorithms. The rate of information can then be used as a design criterion in the development of the end-to-end imaging system.

In the last few decades, the information rate has been has been applied only to a special case of systems, namely linear systems with Gaussian (Normal) inputs. The main reason for concentrating on this class of linear/Gaussian systems has been the availability of a well-known formula with which the rate of information can be computed. A formula for nonlinear systems with non-Gaussian inputs has not been available. In this paper, we will investigate the rate of information transferred from the scene to the digital image for systems containing nonlinear processing such as Local Normalization processing and for signals whose distributions are not necessarily Gaussian. We will use a formula that has recently been developed by the authors to obtain the rate of information for systems using nonlinear Local Normalization (LN) processing.

Local Normalization processing is a center/surround operation based mainly on two observations of human perception. The first observation is that our perception of gray levels depends more on local characteristics rather than the absolute magnitude of the image signals. For example, we perceive a pixel with the same absolute gray level as darker if it is surrounded by light pixels, while we perceive the same pixel as light when it is surrounded by dark pixels (Figure 1). The second observation used in Local

Dr. Halyo is also the president of Information & Control Systems. Direct correspondence to ZR: zrahman@cs.wm.edu
Figure 1. The pixel in the middle of the bright area has the same absolute gray level value as the pixel in the middle of the dark area.

Normalization is that the scene information is mostly contained in the image detail or the high spatial-frequency portion of the digital image; the assumption being that the slowly-varying, low spatial-frequency portion of the image is mostly due to variations in the illumination of the scene. To incorporate these observations into the processing of the image, Local Normalization separates the image into a local average or a low-frequency signal, and a surface detail or high-frequency signal. The locally normalized signal is then obtained by normalizing (i.e., dividing) the detail signal by the local average. It is of interest to note that Local Normalization and Retinex processing\cite{9,10} are equivalent up to first-order terms.

From the point of view of information theory, it is of interest know how the rate of information changes when we use nonlinear processing such as Local Normalization. In particular, is information gained or lost by using Local Normalization? Does the rate of information increase or decrease when we use this type of nonlinear processing? What are the effects of assuming non-Gaussian signals on the rate of information? We will discuss these questions the following sections.

2. THE IMAGE-GATHERING SYSTEM

Consider the end-to-end imaging system shown in Figure 2. The scene radiance field, $L(x)$, is captured by an image-gathering system with the point-spread function (PSF), $\tau(x)$, producing the analog signal

$$S_g(x) = \int_{R^2} \tau(x - x') L(x') dx' + N_e(x) = L(x) \ast \tau(x) + N_e(x), \quad x \in R^2$$

(1)

where $N_e(x)$ is the random noise signal introduced by the device electronics, $R^2$ is the two-dimensional (real) image plane and "$\ast$" denotes the two-dimensional convolution operator. For convenience, we will use the notation $x = (x_1, x_2)$ where $x_1, x_2$ denote the axes of a Cartesian coordinate system describing the image plane.

The signal obtained by the image-gathering system is then uniformly sampled and then digitized resulting in quantization errors which we model as an additive and independent random process.

$$s(kX) = q \lfloor S_g(kX) \rfloor = S_g(kX) + N_q(x), \quad k_1, k_2 = 0, \pm 1, \ldots$$

(2)
where \( q \) represents the quantization operation and \( N_q(kX) \) is the quantization error introduced by limiting the representation of the signal to a given finite number of bits. For notational convenience, we define the product \( kX \) by the following equations:

\[
k = (k_1, k_2), \quad X = (X_1, X_2), \quad \text{and} \quad kX = (k_1X_1, k_2X_2).
\]

(3)

where \( X_1 \) and \( X_2 \) represent the intersample distances in the \( x_1 \) and \( x_2 \) directions respectively. Assuming that the scene radiance field, \( L(x) \), the electronic noise, \( N_e(x) \), the quantization noise, \( N_q(x) \), are mutually independent, 2nd order stationary processes continuous in quadratic mean, the Power Spectral Density (PSD) of the digital signal \( s(kX) \) is given by

\[
\Phi_s(\omega_1, \omega_2) = |\hat{\tau}(\omega_1, \omega_2)|^2 \Phi_L(\omega_1, \omega_2) + \Phi_{N_e}(\omega_1, \omega_2) + \Phi_{N_q}(\omega_1, \omega_2) + \Phi_{N_0}(\omega_1, \omega_2)
\]

(4)

We will use the following notation. The superscripts "\(" and "\)" will denote the continuous Fourier transform and the discrete Fourier transform, respectively. Thus, \( \Phi_L(\omega) \) denotes the continuous Fourier transform of the auto-correlation function \( \Phi_L(x) \), representing the Power Spectral Density (PSD) or the Wiener spectrum of the scene radiance field \( L(x) \); and \( \hat{\Phi}_s(\omega) \) denotes the discrete Fourier transform of the auto-correlation function, \( \Phi_s(kX) \), representing the PSD of the discrete quantized signal, \( s(kX) \). By \( \Phi_L(x) \), we will denote the auto-correlation function of the process, \( L(x) \), with the mean subtracted. The spatial frequency variable used in the Fourier transforms is denoted as \( \omega = (\omega_1, \omega_2) \). Thus the frequency passband \( \mathcal{W}_0 \), is given by

\[
\mathcal{W}_0 = \{(\omega_1, \omega_2) \in \mathbb{R}^2 : |\omega_1| \leq 1/2X_1, |\omega_2| \leq 1/2X_2 \}.
\]

(5)

The last 3 terms on the RHS of Equation 4 correspond to the PSD’s of the sampled electronic noise, the discrete quantization error and the discrete aliasing noise. The sampled electronic noise and the aliasing noise can be expressed in the form,

\[
\Phi_{N_e}(\omega_1, \omega_2) = \sum_{k=-\infty}^{\infty} \Phi_{N_e}(\omega - k/X), \quad \omega \in \mathcal{W}_0.
\]

(6)

where the summation over the two-dimensional integer \( k = (k_1, k_2) \) represents double sums, \( k = 0 \) denotes \((k_1, k_2) = (0,0)\), and \( k/X = (k_1/X_1, k_2/X_2) \).

\[
\Phi_{N_e}(\omega_1, \omega_2) = \sum_{k=-\infty}^{\infty} \Phi_{N_e}(\omega - k/X), \quad \omega \in \mathcal{W}_0.
\]

(7)

Thus the PSD of the digital quantized signal, \( s(kX) \), is given by

\[
\Phi_s(\omega) = \sum_{k=-\infty}^{\infty} |\hat{\tau}(\omega - k/X)|^2 \Phi_L(\omega - k/X) + \Phi_{N_e}(\omega - k/X) + \Phi_{N_q}(\omega), \quad \omega \in \mathcal{W}_0.
\]

(8)

Figure 2. Image Gathering and processing system

![Diagram of Image Gathering and Processing System](image-url)
3. LOCAL NORMALIZATION

By observation of Equation 4, we note that the digital image, \( s(kX) \), contains a portion which depends on the scene radiance, namely the first term on the RHS of Equation 4, and a portion which arises from the various noise sources, including the aliasing error, namely the last three terms in Equation 4. In previous studies using information theory\(^{1-6}\) the signal containing the desired information has been defined as this first term in Equation 4. In this paper, we take a different approach.

The Local Normalization algorithm is mainly based on two observations of the human perception of images which have been discussed in connection with Retinex theory by Land and McCann.\(^{11}\) The first observation is that our perception of image features depends more on the average local characteristics near a pixel rather than the absolute signal magnitude at that pixel. For example, we perceive a pixel with the same absolute gray level as being darker if it is surrounded by light pixels, while we perceive the same pixel as light when it is surrounded by dark pixels.\(^{8,11}\) This is depicted in Figure 1.

The color and intensity we perceive tends to depend more on the spectral reflectance of the surface rather than the spectrum and amplitude of the radiance falling on our retina. Our “eye-brain” or the “retina-cortex” processes the incoming signal and distinguishes between the surface properties and the light illuminating the scene and being reflected toward our eyes. The precise algorithm by which we process images is beyond the scope of this paper. Here, we assume that the local average intensity of a monochromatic image depends largely on the characteristics of the source illuminating the scene and its geometric relation to the scene. Thus, if an area of the image darker than other areas, we assume that this area is in a shadow due to geometric reasons and is not being illuminated as much as other areas. The second assumption we make is that the effects due to scene illumination are relatively slowly varying in comparison to features due to the scene surface reflectance.

To incorporate these assumptions and observations of human perception into the Local Normalization algorithm, we form a slowly-varying signal by low-pass filtering the digital image, and a high-frequency signal which is assumed to contain the effects of the surface reflectance properties. We refer to the low-frequency signal as the “local average” which is mostly due to the effects of the scene illumination. The high-frequency signal is referred to as the detail signal which contains the desired information. To cancel the amplitude effects of the scene illumination, we normalize the signals by dividing them by the local average.

In this formulation, the signal containing the desired information is the detail signal normalized by the local average rather than the total signal used in the previous studies mentioned above. Thus, the scene information which is to be transferred to the processed image is the high-pass portion of the acquired signal normalized by the local average.

To obtain the local average signal, we use a spatial low-pass filter. The local average may simply be a weighted average of a given neighborhood around each image pixel or it may be obtained by a more elaborate low-pass filtering algorithm including a filter with a Gaussian Point-Spread Function (PSF) or an ideal low-pass PSF.

\[
s_{LP}(kX) = \sum_{j=-\infty}^{+\infty} \tau_{LP}(jX)s((k-j)X), \quad j \equiv (j_1, j_2), \quad -\infty \leq k \leq +\infty. \quad (9)
\]

where \( s_{LP}(kX) \) is the local average signal and \( \tau_{LP}(kX) \) is the PSF of the spatial low-pass filter. To produce an average of the digital image signal, we will assume that filter PSF is normalized to unity as follows.

\[
\sum_{j=-\infty}^{\infty} \tau_{LP}(jX) = 1. \quad (10)
\]
Note that if the low-pass PSF vanishes outside a selected neighborhood of the pixel and is constant within that neighborhood, we have a true local average of the pixels in the neighborhood. However, any PSF of the low-pass type can be selected.

Applying the discrete Fourier transform (DFT) to Equation 9, we obtain

\[ \tilde{s}_{LP}(\omega) = \tilde{\tau}_{LP}(\omega) \tilde{s}(\omega), \quad \omega \in W_0 \]  

where \( \tilde{\tau}_{LP}(\omega) \) is the DFT of the PSF \( \tau_{LP}(kX) \); i.e.,

\[ \tilde{\tau}_{LP}(\omega) = \sum_{k=-\infty}^{\infty} \tau_{LP}(kX) \exp[-i2\pi \omega \cdot kX], \quad \omega \in W_0, \]  

The high-pass portion of the digital image, \( s(kX) \), can be obtained by

\[ \tilde{s}_{HP}(\omega) = \tilde{s}(\omega) - \tilde{s}_{LP}(\omega) = \tilde{s}(\omega) [1 - \tilde{\tau}_{LP}(\omega)] \quad \omega \in W_0 \]  

with the corresponding spatial representation (Figure 3)

\[ s_{HP}(kX) = s(kX) - s_{LP}(kX). \]  

In order to define the signal containing the desired information which we want to eventually transmit to the processed image, and the noise terms which corrupt the signal, we express the digital image signal in the spatial frequency domain in the form corresponding to Equation 4 where the signal PSD’s are shown as:

\[ \tilde{s}(\omega) = \hat{L}(\omega) \hat{\tau}(\omega) + \tilde{N}_e(\omega) + \tilde{N}_a(\omega) + \tilde{N}_q(\omega), \quad \omega \in W_0 \]  

Substituting 16 into Equation 14, we obtain

\[ \tilde{s}_{HP}(\omega) = [1 - \tilde{\tau}_{LP}(\omega)] \hat{L}(\omega) \hat{\tau}(\omega) + [1 - \tilde{\tau}_{LP}(\omega)] [\tilde{N}_a(\omega) + \tilde{N}_e(\omega) + \tilde{N}_q(\omega)], \quad \omega \in W_0. \]  

We can see from Equation 17, that the high spatial frequency portion of the digital image is the sum of two components, the first depending on the scene radiance field, the second depending on the noise and aliasing error terms. We can write the high spatial frequency portion as

\[ \tilde{s}_{HP}(\omega_1, \omega_2) = \tilde{s}_d(\omega_1, \omega_2) + \tilde{N}_d(\omega_1, \omega_2) \quad (\omega_1, \omega_2) \in W_0, \]  

![Figure 3. Local Normalization processing](image-url)
where \( \tilde{s}_d(\omega_1, \omega_2) \) is the detail portion of the signal before normalization and \( \tilde{N}_d(\omega_1, \omega_2) \) is the high-frequency portion of the noise and aliasing error. Thus, the high spatial frequency portion of the digital image can be written in the spatial domain as

\[
s_{HP}(kX) = s_d(kX) + N_d(kX).
\]

(19)

Although this high-pass signal describes the surface reflectance characteristics more accurately, it still depends highly on the scene illumination spectral characteristics and amplitude or intensity. To cancel, or at least reduce, this dependence, we normalize the high spatial frequency portion by the local average which also contains the amplitude of the illumination source. Thus, we obtain the locally normalized signals.

\[
S_{LN}(kX) = \frac{s_{HP}(kX)}{s_{LP}(kX)} = \frac{s_d(kX)}{s_{LP}(kX)} + \frac{N_d(kX)}{s_{LP}(kX)}
\]

(20)

In the Local Normalization formulation, we consider the normalized portion of the high-pass signal as the signal containing the desired information corrupted by an additive noise term, which we rewrite as

\[
S_{LN}(kX) = S(kX) + \mathcal{N}(kX)
\]

(21)

Finally, we can express the measured raw signal, or the sampled and quantized digital image signal, \( s(kX) \), in terms of the locally normalized signals, defined above.

\[
s(kX) = s_{LP}(kX) \left[ 1 + S_{LN}(kX) \right]
\]

(22)

\[
s(kX) = s_{LP}(kX) \left[ 1 + S(kX) + \mathcal{N}(kX) \right]
\]

(23)

From an information theoretic point of view, the problem can be expressed as one of determining the image-gathering system parameters and Local Normalization processing parameters which maximize the rate of information transferred from the signal \( S(kX) \) to the locally normalized signal, \( S_{LN}(kX) \).

4. INFORMATION RATE IN LOCALLY NORMALIZED IMAGES

One of the original motivations in the development of the Local Normalization algorithm has been to find whether information is lost when we use Retinex or Local Normalization to process the digital image. Of course, an equally important objective is to find a new formula for the rate of information for nonlinear systems with non-Gaussian signals which is computable and which provides some insight into the problem. While a general formula which is applicable to arbitrary nonlinearities and arbitrary random processes would be highly desirable, this generality also makes a solution elusive. However, a formula for the information rate for systems using the nonlinear Local Normalization algorithm with a class of signals has been developed by the authors. These results will be discussed in this section.

Consider a second order stationary random process, continuous in quadratic mean, band-limited to \( W_0 \) defined in Equation 5. Let \( \{P(x), x \in \mathbb{R}^2\} \) and \( \{Z(x), x \in \mathbb{R}^2\} \) be such processes. The rate of information from the signal \( P(x) \) to the measured (or processed) signal \( Z(x) \) is defined as

\[
R(P, Z) = \mathcal{E}(P) - \mathcal{E}(P|Z)
\]

(24)

where \( \mathcal{E}(P) \) is the entropy (or information content) of the signal \( P(x) \) and \( \mathcal{E}(P|Z) \) is the conditional entropy of \( P(x) \) given the measured or processed signal \( Z(x) \). The entropy of a signal \( P(x) \) is

\[
\mathcal{E}(P) = \lim_{(n_1, n_2) \to \infty} \frac{1}{(2n_1 + 1)(2n_2 + 1)} \mathcal{H}(P_{(n_1, n_2)}) = \lim_{n \to \infty} \frac{1}{(2n + 1)} \mathcal{H}(P_n)
\]

(25)

\[
\mathcal{H}(P_n) = -E \{ \log f_{P_n}(p_n) \} = - \int_{R^{2(n+1)}} f_{P_n}(p_n) \log \left[ f_{P_n}(p_n) \right] \, dp_n
\]

(26)
where \( \mathcal{P}_n \) is a collection of \([2n + 1]\) uniform samples of \( \mathcal{P}(x) \), \( \mathcal{H}(\mathcal{P}_n) \) is the joint entropy of a finite number of random variables as defined by Shannon\(^{12} \) (p.91), \( E \) is the statistical expectation operator and \( f_{\mathcal{P}_n}(p_n) \) is the joint probability density function of the random variables in \( \mathcal{P}_n \). The conditional entropy is defined similarly using the conditional probability density function of \( \mathcal{P} \) given \( \mathcal{Z} \).

Note that the entropy and the rate of information depend only on the uniform samples of the processes since these samples determine the continuous band-limited process. Thus, we can extend the definitions to sequences. The entropy of a process as defined above can be interpreted as the average information content or the average uncertainty in a pixel of the process. Similarly, the conditional entropy of \( \mathcal{P} \) given \( \mathcal{Z} \) can be interpreted as the information or uncertainty still remaining in \( \mathcal{P} \) when we know the signal \( \mathcal{Z} \), i.e., given \( \mathcal{Z} \).

Accordingly, the rate of information from \( \mathcal{P} \) to \( \mathcal{Z} \), \( \mathcal{R}(\mathcal{P}, \mathcal{Z}) \), can be interpreted as the information in \( \mathcal{P} \) minus the portion of this information still remaining in \( \mathcal{P} \) after we have measured \( \mathcal{Z} \). In other words, this is the information about the signal \( \mathcal{P} \) that was actually brought by measuring \( \mathcal{Z} \). Thus, the rate of information from \( \mathcal{P} \) to \( \mathcal{Z} \) is the portion of the information in the signal \( \mathcal{P} \) that has been transferred to the measured or processed signal \( \mathcal{Z} \). In general, although the desired information is contained in \( \mathcal{Z} \), it may still be necessary to do further processing (or decoding) to restore the image. However, if the information were not present in \( \mathcal{Z} \), no amount of processing or decoding could restore the image. Accordingly, maximizing the rate of information by appropriate selection of the image-gathering system parameters and the digital processing algorithm parameters provides a design criterion of end-to-end performance.

In the Local Normalization problem, the signal containing desired scene information is \( \mathcal{S}(kX) \) given in Equation 21. The measurement is the sampled digital image, \( s(kX) \), shown in Equations 2 and 22. The processed or locally normalized signal, \( \mathcal{S}_{LN}(kX) \) is shown in Equation 21. We are interested to know whether the locally normalized signal or the digital image contains more scene information. The following result\(^7\) (Theorem 2) provides an answer.

**Theorem 4.1.** Let Equations 21 and 22 hold with all the sequences being second order stationary, then

\[
\mathcal{R}(\mathcal{S}, \mathcal{S}_{LN}) \geq \mathcal{R}(\mathcal{S}, s) \tag{27}
\]

This is a striking result and provides a general answer. Whether the signals are Gaussian or not, the nonlinear Local Normalization signal contains at least as much information about the scene detail signal as the measured digital image. Alternately, we lose no information about the scene detail by using the locally normalized signal. Given our definition of the scene detail as the desired signal in Equation 25, the rate of information for the nonlinear Local Normalization signal will be at least as high as for the digital image signal whether the scene or noise statistics have a Gaussian distribution or not, whatever the Wiener spectrum of the scene or the noises, whatever the parameters of the image-gathering system.

While the result above is very general and informative, it does not provide a quantitative way of estimating the rate of information for a given set of conditions. For the case of linear systems with linear digital processing algorithms and signals with Gaussian statistics, a formula for the rate of information has been available for several decades\(^5,12\) and has been used effectively to analyze imaging systems.\(^1\)–\(^6\) However, until Halyo\(^7\) et al (Theorem 3) such a formula has not been available for nonlinear systems.

To obtain a computable formula for the rate of information for systems using nonlinear Local Normalization processing, we will consider the class of sequences which admit a representation of the form

\[
\mathcal{Z}(kX) = \sum_{j=-\infty}^{\infty} \tau_{\mathcal{Z}}((k - j)X)w_{\mathcal{Z}}(jX) \tag{28}
\]
where \( \{w_z(jX), j_1, j_2 = 0, \pm 1, \pm 2, \ldots \} \) is a sequence of statistically independent random variables having the same probability density function. However, this independent noise process need not have Gaussian statistics; it may have an arbitrary PDF. Note that this a large class of random processes. Almost any sequence of random variables can be represented in the form of Equation 28 where \( \{w_z(jX)\} \) is a white noise process using spectral factorization. In general, this white noise process has uncorrelated samples, but not necessarily independent samples as is assumed here. Thus the restriction comes in with the assumption of independence which is stronger than only lack of correlation. However, when one considers all the processes that can be obtained using an arbitrary PSF and an arbitrary independent sequence in Equation 28, the class is seen to be quite large. We will say that a sequence satisfying Equation 28 is normal.

**Theorem 4.2.** Let the radiance signal, \( \{L(x), x \in \mathbb{R}^2\} \), the noise process, \( \{N_e(x), x \in \mathbb{R}^2\} \) and the quantization noise \( \{N_q(kX), k_1, k_2 = 0, \pm 1, \ldots\} \) be statistically independent and strictly stationary second order random processes that satisfy Equations 1 and 2. Let \( \{s_{LP}(kX), k_1, k_2 = 0, \pm 1, \ldots\} \) be independent of \( \{s_d(kX), k_1, k_2 = 0, \pm 1, \ldots\} \) and \( \{N_d(kX), k_1, k_2 = 0, \pm 1, \ldots\} \), given by Equations 9, 18, and 19. Furthermore, let \( \{S_{LN}(kX), k_1, k_2 = 0, \pm 1, \ldots\} \) and \( \{N(kX), k_1, k_2 = 0, \pm 1, \ldots\} \), given by Equations 20 and 21 have independent bases \( w_{LN}(kX) \) and \( w_N(kX) \), respectively. Then the local normalization rate of information satisfies:

\[
\mathcal{R}(S, S_{LN}) \geq \mathcal{E}(S_{LN}) - \mathcal{E}(N) = [\mathcal{H}(w_{LN}(0)) - \mathcal{H}(w_N(0))] + \frac{1}{|W_0|} \int_{W_0} \log_2 \left[ 1 + \frac{\tilde{\Phi}_{s_2}^*(\omega) \ast \tilde{\Phi}_{s_{LP}}^* (\omega) + s_{LP}^{-1} \tilde{\Phi}_{s_2}^* (\omega) + \tilde{s}_{LP}^{-1} \tilde{\Phi}_{s_{LP}}^* (\omega)}{\tilde{\Phi}_{N_2}^*(\omega) \ast \tilde{\Phi}_{s_{LP}}^* (\omega) + s_{LP}^{-1} \tilde{\Phi}_{N_2}^* (\omega)} \right] d\omega
\]

with equality if, and only if, \( \{S(kX), k_1, k_2 = 0, \pm 1, \ldots\} \) and \( \{N(kX), k_1, k_2 = 0, \pm 1, \ldots\} \) are independent, where

\[
\tilde{\Phi}_{s_d}^*(\omega) = |1 - \tilde{\tau}_{LP}^2(\omega)|^2 \tilde{\tau}(\omega)^2 \tilde{\Phi}_L^* (\omega), \quad \omega \in W_0
\]

\[
\tilde{\Phi}_{N_d}^*(\omega) = |1 - \tilde{\tau}_{LP}^2(\omega)|^2 \left[ \tilde{\Phi}_{N_d}^* (\omega) + \tilde{\Phi}_{N_e}^* (\omega) + \tilde{\Phi}_{N_q}^* (\omega) \right], \quad \omega \in W_0
\]

\[
\tilde{\Phi}_{s_{LP}}^* (\omega) = \sum_{k=-\infty}^{\infty} \left[ \int_0^\infty \int_0^\infty \frac{1}{y y_k} f_{LP,k}(y, y_k) dy dy_k - s_{LP}^{-1} \right] \exp[-i2\pi\omega \cdot kX], \quad \omega \in W_0
\]

\[
\tilde{s}_{LP}^{-1} = \mathcal{E} \left\{ \frac{1}{S_{LP}(kX)} \right\}, \quad \tilde{s}_d = \mathcal{E}\{s_d(kX)\}, \quad \tilde{s}_d = \mathcal{E}\{N_d(kX)\} = 0
\]

where \( f_{LP,k}(y, y_k) \) is the joint PDF of \( s_{LP}(jX) \) and \( s_{LP}(j+k)X \) for any \( j \), and \( * \) denotes the convolution.

\[
\tilde{\Phi}_1^*(\omega) \ast \tilde{\Phi}_2^*(\omega) = \frac{1}{|W_0|} \int_{W_0} \tilde{\Phi}_1^*(\omega') \tilde{\Phi}_2^*(\omega - \omega') d\omega', \quad \omega \in W_0.
\]

While the new formula for the rate of information in Equation 29 has some similarity to the one for the Linear/Gaussian (LG) case, it has several new terms which account for the various assumptions which are no longer made. First note that in the LG case, the mean values of the signals do not enter into the rate of information. Since the entropy of a signal does not depend on its mean and since linear operations on Gaussian signals result in still Gaussian signals, it is reasonable that the rate of information in the LG case will not have mean terms. However, since Local Normalization introduces nonlinearities, the mean value of the signals does have an impact on the rate of information. Thus, note how the mean of the
inverse of the local average and the mean of the scene detail signal prior to normalization enter into the new formula. The mean of the noise is not shown as it vanishes; however, in situations where bias errors are an issue, the mean of the noise would have to be included in the formula. Note that the mean of the inverted local average is both in the numerator and denominator and produces an impact proportional to the signal-to-noise ratio. However, the mean of the scene detail prior to normalization simply adds to the numerator, thus increasing the rate of information. The mean of the noise would produce a similar term in the denominator.

The normalization by the local average further modifies the signals’ spectral character. This is seen in Equation 29 in the form of the convolution of the signal PSD in the numerator and the convolution of the noise PSD in the denominator. Since the local average is a slowly-varying signal with low spatial frequency content, the impact of the convolution is to smooth out the signal and introduce some low-frequencies where none may have been present before.

In particular, note the case when the local average is a constant (i.e., it is a global average), the PSD in Equation 32 approaches a delta function, resulting in an identity operation which eliminates the convolution. Therefore, when the signals have zero mean, and the normalization is simply a division by a constant (i.e., a linear operation), the new LN formula reduces to the old formula for the LG case, as would be expected.

Finally, the effect of non-Gaussian distributions is further seen in the first term of Equation 29 the difference between the entropies of the bases for the LN signal and the noise signal. Since, the basis is normalized to unity variance, when both signals are Gaussian, the entropy (which depends only on their variance) is identical for both basis terms and cancels out. Thus, the formula for the LG case does not have this term which accounts for non-Gaussian distributions in the signals.

While the new formula for the nonlinear case has extra terms as would be expected, it is still sufficiently simple to be easily computable. The new terms also add to our understanding of the impact of various cases as discussed above.

5. COMPUTATIONAL RESULTS

We can use the new formula for the information rate given in Equations 29–34 to evaluate the impact of the nonlinear digital processing introduced by the Local Normalization algorithm. As the variables in the problem are numerous, a full evaluation of all the effects introduced by nonlinear processing require further investigation. Here we will consider a specific case which can be obtained with relative computational ease.

To investigate the impact of the local averaging, we will consider the class of ideal filters which pass the low spatial frequencies from zero to the $\alpha$. We will also consider how the cut off frequency in the image-gathering PSF impacts the amount of information transferred to the processed image. We will make parameter selections and assumptions to make the computational burden as low as possible. The following variable selections describe to the case studied here.

\[
X_1 = X_2 = 1 \\
\hat{\Phi}_L(\omega_1, \omega_2) = \frac{2\pi\sigma_L^2 \mu^2}{[1 + 4\pi^2 \mu^2 (\omega_1^2 + \omega_2^2)]^{3/2}} \quad \sigma_L = 1, \quad \mu \in R \\
\hat{\tau}(\omega_1, \omega_2) = \exp[-(\omega_1^2 + \omega_2^2)/\rho_c^2], \quad \rho_c \in R \\
\tilde{\tau}_{LP}(\omega_1, \omega_2) = \begin{cases} 
1, & |\omega_1| \leq \alpha_1, \quad |\omega_2| \leq \alpha_2 \\
0, & \text{otherwise} 
\end{cases} \quad \alpha_1, \alpha_2 \in R
\]
For the initial case we will use a white noise process for \( \hat{N}_e(\omega) \). Then,

\[
\hat{\Phi}_{N_e}(\omega) = \begin{cases} 
\sigma^2_{N_e} & \omega \in \mathcal{W}_0, \\
0 & \text{else} 
\end{cases}, \quad \sigma^2_{N_e} \in R 
\]

(39)

\[
\hat{\Phi}_{N_q}(\omega) = \begin{cases} 
\frac{2^{-2(n+1)}}{3} & \omega \in \mathcal{W}_0 \\
0 & \text{else} 
\end{cases}, \quad n = 4, 6, 8, \ldots 
\]

(40)

In addition,

\[
\mathcal{H}(w_{LN}(0)) = \int_{-\infty}^{\infty} f_{w_{LN}}(s) \left[ \log[f_{w_{LN}}(s)] \right] ds 
\]

(41)

\[
f_{w_{LN}}(s) = \frac{1}{\sqrt{2\pi}} \exp \left[ -s^2/2 \right] 
\]

(42)

\[
f_{w_n}(s) = \begin{cases} 
1/\sqrt{12}, & |s| \leq \sqrt{3} \\
0, & \text{else} 
\end{cases} 
\]

(43)

\[
\Phi_{s_{LP}}^{-1}(k_1, k_2) = \sigma^2_{s_{LP}} \exp \left[ -\alpha_1 k_1 - \alpha_2 k_2 \right], \quad k_1, k_2 = 0, \pm 1, \pm 2, \ldots 
\]

(44)

\[
\tilde{\Phi}_{s_{LP}}^{-1}(\omega_1, \omega_2) = \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} \Phi_{s_{LP}}^{-1}(k_1, k_2) \exp \left[ -i2\pi(\omega_1 k_1 + \omega_2 k_2) \right], \quad (\omega_1, \omega_2) \in \mathcal{W}_0 
\]

(45)

\[
= \sigma^2_{s_{LP}} \left[ \frac{\exp[2\alpha_1 X_1] - 1}{\exp[2\alpha_1 X_1] - 2 \exp[\alpha_1 X_1] \cos(2\pi \omega_1 X_1) + 1} \right] 
\]

(46)

\[
\sigma^2_{s_{LP}} = \alpha_0^2 \exp\left[ -b_0^2(\alpha_1^2 + \alpha_2^2) \right], 
\]

(47)

To normalize the problem sampling geometry, we will choose the intersample distance to be unity in both directions. Similarly, for normalization purposes, we assume the scene radiance field to have a variance of unity. The average zero-crossing parameter \( \mu \) is also selected as one for this paper, although it would be of interest to see its impact on the rate of information.

In Figure 4, the Power Spectral Density (PSD), or Wiener Spectrum of the detail signal is seen for various values of the image-gathering cut-off frequency, \( \rho \), and various values of the low-pass cut-off frequency, \( \alpha \). These spectra are computed using Equations 30 and 35–38. As can be noted from the plots, as the image-gathering cut-off frequency is decreased, the detail signal PSD falls off more sharply and nearer the zero frequency, reducing aliasing errors and other noises but increasing the blurring. On the other hand, as the low-pass filter cut-off frequency, \( \alpha \), increases, the PSD of the signal vanishes between zero and \( \alpha \), reducing its low-frequency content. Note that as the low-pass cut-off reaches the image-gathering cut-off frequency, the detail signal which has been decreasing monotonically, becomes almost negligible.

Figure 5 shows the PSD of the inverse low-pass filter output or the local average signal for various values of the low-pass cut-off frequency. We use Equations 44–47 for this purpose. As \( \alpha \) approaches zero, the PSD approaches a Dirac delta function corresponding to a constant value of the auto-correlation function. As \( \alpha \) moves away from zero, the PSD widens. Figure 6 shows the PSD of the noise for various parameters.

Figure 7 shows the rate of information, \( \mathcal{R}(\mathcal{S}, \mathcal{S}_{LN}) \), from the normalized scene detail signal, \( \mathcal{S} \), to the locally normalized digital image, \( \mathcal{S}_{LN} \), as a function of the image-gathering system cut-off frequency, \( \rho \), for various low-pass filter cut-off frequencies, \( \alpha \), and for various signal-to-noise ratios (SNRs). In all cases, when the image-gathering system blur is small, the rate of information is also low because the image-gathering system is blurring the signal too much even though it is reducing, almost eliminating,
Figure 4. The PSD $\tilde{\Phi}_{sd}(\omega)$ of the detail signal as a function of the parameter $\alpha$ which defines the extent of $\tilde{\tau}_{LP}$, for several cutoff-frequency values $\rho_c$. As $\alpha$ increases, more and more of $\tilde{\Phi}_{sd}$ gets truncated to 0.

Figure 5. The PSD of the inverse low-pass filter, $\tilde{\Phi}_{s_{LP}}^{-1}(\omega)$ for various values of $\alpha$, with $a_0 = 1.0$, and $b_0 = 2.0$. 

$\rho_c = 0.80$

$\rho_c = 0.40$

$\rho_c = 0.30$
the aliasing error. As \( \rho_c \) increases, the scene detail signal is less blurred and less attenuated while some aliasing is allowed, more information is transferred from the scene to the LN image, \( S_{LN} \). When the image-gathering system cut-off frequency, \( \rho_c \), reaches the 0.4 to 0.5 area, the rate of information reaches a maximum level depending on the SNR and the value of \( \alpha \). Since we have normalized the sampling lattice to an intersample distance of unity, or alternately we have normalized the frequency passband to \([-0.5, 0.5] \times [-0.5, 0.5] \), we note that the maximum information is being transferred when the image-gathering system cut-off frequency, \( \rho_c \), is near the Nyquist frequency of the system. When \( \rho_c \) goes beyond the Nyquist frequency, the aliasing error increases significantly without a correspondingly large reduction in blurring or attenuation. Accordingly, less information is transferred from the scene to the locally normalized image and the rate of information starts to decline. It is important to note that these trends are present in all the plots, for every value of the low-pass filter cut-off frequency, \( \alpha \); i.e., the maximum information rate is reached near the Nyquist frequency whatever the local averaging. This robustness is a desirable quality for design considerations.

Further study of the new formula for the information rate applied to nonlinear Local Normalization problems is needed to determine further trends and design criteria for the image-gathering and digital processing algorithm parameters.

**Figure 6.** The PSD \( \hat{\Phi}_{Nd} \) of the noise signal as a function of the parameter \( \alpha \) which defines the extent of \( \hat{\tau}_{LP} \). As \( \alpha \) increases, more and more of \( \Phi_{Nd} \) gets truncated to 0.
The rate of information $R$ in LN processing as a function of $p$ for several values of the low pass filter parameter $\alpha$. $s_{LP}^2 = 1.00; \sigma_a^2 = 0.00; \sigma_s^{2L} = a_0^2 \exp[-b_0^2(\alpha_1^2 + \alpha_2^2)]; a_0 = 1.00; b_0 = 2.00$

Acknowledgements

The authors would like to acknowledge several stimulating discussions of nonlinear Retinex and Local Normalization information theory concepts with Friedrich O. Huck, Daniel J. Jobson, C. L. Fales, Richard D. Davis, and Glenn A. Woodell of NASA, Langley Research Center. This research was supported by NASA Langley Research Center Contract #70511d.

REFERENCES


