

Article

Mutual Information Loss in Pyramidal Image Processing [†]

Jerry Gibson ^{*,‡}  and Hoontaek Oh [‡]

Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106-9560, USA; hoontaek@ucsb.edu

* Correspondence: gibson@ece.ucsb.edu

† This paper is an extended paper presented at the Multimedia Information Processing and Research conference in San Jose, CA, USA, in April 2019.

‡ These authors contributed equally to this work.

Received: 19 May 2020; Accepted: 4 June 2020; Published: 15 June 2020



Abstract: Gaussian and Laplacian pyramids have long been important for image analysis and compression. More recently, multiresolution pyramids have become an important component of machine learning and deep learning for image analysis and image recognition. Constructing Gaussian and Laplacian pyramids consists of a series of filtering, decimation, and differencing operations, and the quality indicator is usually mean squared reconstruction error in comparison to the original image. We present a new characterization of the information loss in a Gaussian pyramid in terms of the change in mutual information. More specifically, we show that one half the log ratio of entropy powers between two stages in a Gaussian pyramid is equal to the difference in mutual information between these two stages. We show that this relationship holds for a wide variety of probability distributions and present several examples of analyzing Gaussian and Laplacian pyramids for different images.

Keywords: Gaussian pyramids; Laplacian pyramids; pyramidal image processing; multiresolution image processing; log ratio of entropy powers

1. Introduction

Gaussian and Laplacian pyramids allow images to be viewed, stored, analyzed, and compressed at different levels of resolution, thus minimizing browsing time, required storage capacity, and required transmission capacity, in addition to facilitating comparisons of images across levels of detail [1–3]. Different resolution pyramids play a major role in image and speech recognition in machine learning and deep learning [4,5].

The formation of Gaussian and Laplacian pyramids requires the signal processing steps of filtering/averaging, decimation and interpolation. Interestingly, the filters and the interpolation steps are usually left unspecified, and in applications such as image storage and compression, the performance of the various steps in the Gaussian pyramid is primarily based on mean squared error. The cascade of signal processing steps performed in obtaining a Gaussian pyramid forms a Markov chain, and as such, satisfies the Data Processing Inequality from information theory. Based on this fact, we utilize the quantity entropy rate power from information theory and introduce a new performance indicator, the log ratio of entropy powers, to characterize the loss in mutual information at each signal processing stage in forming the Gaussian pyramid.

In particular, we show that the log of the ratio of entropy powers between two signal processing steps is equivalent to the difference between the differential entropy of the two stages and equal to twice the difference in the mutual information between the two stages. However, in order to calculate

the entropy power, we need an expression for the differential entropy, but the accurate computation of the differential entropy can be quite difficult and requires considerable care [6].

We show that for i.i.d. Gaussian and Laplacian distributions, and note that for logistic, uniform, and triangular distributions, the mean squared estimation error can replace the entropy power in the log ratio of entropy power expression [7]. As a result, the log ratio of entropy power and the difference in differential entropies and mutual informations can be much more easily calculated.

Note that the mean squared error and the entropy power are only equal when the distribution is Gaussian, and that the entropy power is the smallest variance possible for a random variable with the same differential entropy. However, if the distributions in the log ratio of the entropy powers are the same, then the constant multiplier of the mean squared error in the expressions for entropy power will cancel out and only the mean squared errors are needed in the ratio. We also note that for the Cauchy distribution, which has an undefined variance, the log ratio of entropy powers equals the log ratio of the squared Cauchy distribution parameters.

The concept of entropy power and entropy rate power are reviewed in Section 2. Section 3 sets up the basic cascade signal processing problem being analyzed, the well known inequalities for mutual information in the signal processing chain are stated, and the inequalities that follow from the results in Section 2 are given. The log ratio of entropy powers is developed in Section 4, explicitly stating its relationship to the changes in mutual information as a signal progresses through the signal processing chain, and wherein a discussion of calculating entropy power and log ratio of entropy powers is presented. Section 5 provides results using the log ratio of entropy powers to characterize the mutual information loss in bits/pixel for Gaussian and Laplacian image pyramids. In Section 6 we compare log ratio of entropy powers to a direct calculation of the difference in mutual information. Section 7 contains a discussion of the results and conclusions.

This paper is an expansion and extension of the conference paper presented in 2019 [8] with expanded details of the examples and more extensive analyses of the results.

2. Entropy Power/Entropy Rate Power

Given a random variable X with probability density function $p(x)$, we can write the differential entropy

$$h(X) = - \int_{-\infty}^{\infty} p(x) \log p(x) dx \quad (1)$$

where the variance $\text{var}(X) = \sigma^2$. Since the Gaussian distribution has the maximum differential entropy of any distribution with mean zero and variance σ^2 [9],

$$h(X) \leq \frac{1}{2} \log 2\pi e \sigma^2 \quad (2)$$

from which we obtain

$$Q = \frac{1}{(2\pi e)} \exp 2h(X) \leq \sigma^2 \quad (3)$$

where Q was defined by Shannon to be the *entropy power* associated with the differential entropy of the original random variable [10]. In addition to defining entropy power, this equation shows that the entropy power is the *minimum variance* that can be associated with the not-necessarily-Gaussian differential entropy $h(X)$.

If we let \mathbf{X} be a stationary continuous-valued random process with samples $X^n = [X_i, i = 1, 2, \dots, n]$, then the differential entropy rate of the process \mathbf{X} is [9,11]

$$\bar{h} = \lim_{N \rightarrow \infty} \frac{1}{N} h(X_1, \dots, X_N) \quad (4)$$

which is the long term average differential entropy in bits/symbol for the sequence being studied. An alternative definition of differential entropy rate is

$$\bar{h} = \lim_{N \rightarrow \infty} h(X_N | X_{N-1}, \dots, X_1) \tag{5}$$

These two definitions are equal for stationary processes and we drop the overbar notation and use $h = \bar{h}$. Using the entropy rate in the definition of entropy power yields the nomenclature entropy rate power.

3. Cascaded Signal Processing

Figure 1 shows a cascade of N signal processing operations with the Estimator blocks at the output of each stage as studied by Messerschmitt [12]. He used the conditional mean at each stage and the corresponding conditional mean squared errors to obtain a representation of the distortion contributed by each stage. We analyze the cascade connection in terms of information theoretic quantities, such as mutual information, differential entropy, and entropy rate power. Similar to Messerschmitt, we consider systems that have no hidden connections between stages other than those explicitly shown. Therefore, we conclude directly from the Data Processing Inequality [9] that

$$I(X; Y_1) \geq \dots \geq I(X; Y_{N-1}) \geq I(X; Y_N) \geq I(X; \hat{X}) \tag{6}$$

Since $I(X; Y_n) = h(X) - h(X|Y_n)$, it follows from Equation (6) that

$$h(X|Y_1) \leq \dots \leq h(X|Y_{N-1}) \leq h(X|Y_N) \leq h(X) \tag{7}$$

For the optimal estimators at each stage, the basic Data Processing Inequality also yields $I(X; Y_n) \geq I(X; \hat{X}_n)$ and thus $h(X|Y_n) \leq h(X|\hat{X}_n)$.

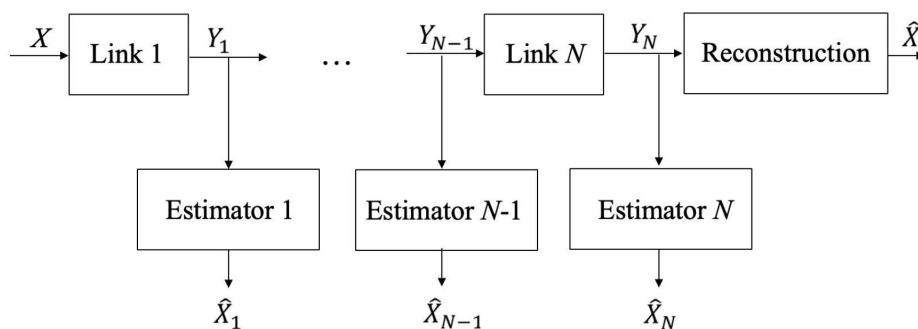


Figure 1. N-Link System Block Diagram (Adapted from [12]).

These are the fundamental results that additional processing cannot increase the mutual information.

Now we notice that the series of inequalities in Equation (7) along with the entropy power expression in Equation (3) gives us the series of inequalities in terms of entropy power at each stage in the cascaded signal processing operations

$$Q_{X|Y_1} \leq \dots \leq Q_{X|Y_{N-1}} \leq Q_{X|Y_N} \leq Q_X \tag{8}$$

We can also write that

$$Q_{X|Y_n} \leq Q_{X|\hat{X}_n} \tag{9}$$

In the context of Equation (8), the notation $Q_{X|Y_n}$ denotes the minimum variance when reconstructing an approximation to X given the sequence at the output of stage n in the chain.

4. Log Ratio of Entropy Powers

We can use the definition of the entropy power in Equation (3) to express the logarithm of the ratio of two entropy powers in terms of their respective differential entropies as

$$\log \frac{Q_X}{Q_Y} = 2[h(X) - h(Y)] \quad (10)$$

We can write a conditional version of Equation (3) as

$$Q_{X|Y_n} = \frac{1}{(2\pi e)} \exp 2h(X|Y_n) \leq \text{Var}(X|Y_n) \quad (11)$$

and from which we can express Equation (10) in terms of the entropy powers at successive stages in the signal processing chain as

$$\frac{1}{2} \log \frac{Q_{X|Y_n}}{Q_{X|Y_{n-1}}} = h(X|Y_n) - h(X|Y_{n-1}) \quad (12)$$

If we add and subtract $h(X)$ to the right hand side of Equation (12), we then obtain an expression in terms of the difference in mutual information between the two stages as

$$\frac{1}{2} \log \frac{Q_{X|Y_n}}{Q_{X|Y_{n-1}}} = I(X; Y_{n-1}) - I(X; Y_n) \quad (13)$$

From the series of inequalities on the entropy power in Equation (8), we know that both expressions in Equations (12) and (13) are greater than or equal to zero.

These results are from [13] and extend the Data Processing Inequality by providing a new characterization of the information loss between stages in terms of the entropy powers of the two stages. Since differential entropies are difficult to calculate, it would be particularly useful if we could obtain expressions for the entropy power at two stages and then use Equations (12) and (13) to find the difference in differential entropy and mutual information between these stages.

More explicitly, we are interested in studying the change in the differential entropy brought on by different signal processing operations by investigating the log ratio of entropy powers. However, in order to calculate the entropy power, we need an expression for the differential entropy! Thus, why do we need the entropy power?

First, entropy power may be easy to calculate in some instances, as we show later. Second, the accurate computation of the differential entropy can be quite difficult and requires considerable care [6,14]. Generally, the approach is to estimate the probability density function (pdf) and then use the resulting estimate of the pdf in Equation (1) and numerically evaluate the integral.

Depending on the method used to estimate the probability density, the operation requires selecting bin widths, a window, or a suitable kernel [6], all of which must be done iteratively to determine when the estimate is sufficiently accurate. The mutual information is another quantity of interest, as we shall see, and the estimate of mutual information also requires multiple steps and approximations [14–16]. These statements are particularly true when the signals are not i.i.d. and have unknown correlation.

In the following we highlight several cases where Equation (10) holds with equality when the entropy powers are replaced by the corresponding variances. The Gaussian and Laplacian distributions often appear in studies of speech processing and other signal processing applications[17–19], so we show that substituting the variances for entropy powers in the log ratio of entropy powers for these distributions satisfies Equations (10), (12), and (13) exactly.

Interestingly, using mean squared errors or variances in Equation (10) is accurate for many other distributions as well. It is straightforward to show that Equation (10) holds with equality when the entropy powers are replaced by mean squared error for the logistic, uniform, and triangular

distributions as well. Further, the entropy powers can be replaced by the ratio of the squared parameters for the Cauchy distribution.

Therefore, the satisfaction of Equation (10) with equality occurs not just in one or two special cases. The key points are first that the entropy power is the smallest variance that can be associated with a given differential entropy, so the entropy power is some fraction of the mean squared error for a given differential entropy. Second Equation (10) utilizes the ratio of two entropy powers, and thus, if the distributions corresponding to the entropy powers in the ratio are the same, the scaling constant (fraction) multiplying the two variances cancels out. So, we are not saying that the mean squared errors equal the entropy powers in any case but for Gaussian distributions.

It is the new quantity, the log ratio of entropy powers that enables the use of the mean squared error to calculate the loss in mutual information at each stage [13].

4.1. Gaussian Distributions

For two i.i.d. Gaussian distributions with zero mean and variances σ_X^2 and σ_Y^2 , we have directly that $Q_X = \sigma_X^2$ and $Q_Y = \sigma_Y^2$, so

$$\frac{1}{2} \log \frac{Q_X}{Q_Y} = \frac{1}{2} \log \frac{\sigma_X^2}{\sigma_Y^2} = [h(X) - h(Y)] \quad (14)$$

which satisfies Equation (10) exactly. Of course, since the Gaussian distribution is the basis for the definition of entropy power, this result is not surprising.

4.2. Laplacian Distributions

For two i.i.d. Laplacian distributions with parameters λ_X and λ_Y [20], their corresponding entropy powers $Q_X = 2e\lambda_X^2/\pi$ and $Q_Y = 2e\lambda_Y^2/\pi$, respectively, so we form

$$\frac{1}{2} \log \frac{Q_X}{Q_Y} = \frac{1}{2} \log \frac{\lambda_X^2}{\lambda_Y^2} = [h(X) - h(Y)]. \quad (15)$$

since $h(X) = \ln(2e\lambda_X)$ so the Laplacian distribution also satisfies Equation (10) exactly [7]. We thus conclude that we can substitute the variance, or for zero mean Laplacian distributions, the mean squared value for the entropy power in Equation (10) and the result is the difference in differential entropies [7].

5. Pyramidal Image Processing

Pyramidal decimation to several different levels is fundamental to multiresolution image processing and compression and is a common first step in many machine learning algorithms for image recognition [21]. To illustrate the utility of the new log ratio of entropy powers expression, we consider an experiment conducted on several different images. We examine the results of passing four different images, Campus, News1, News2, and City, shown in Figure 2, through the given pyramidal signal processing chain. The sequence of image processing operations is illustrated in Figure 3 for the Campus image.



Figure 2. News1, News2, Campus, and City Images.

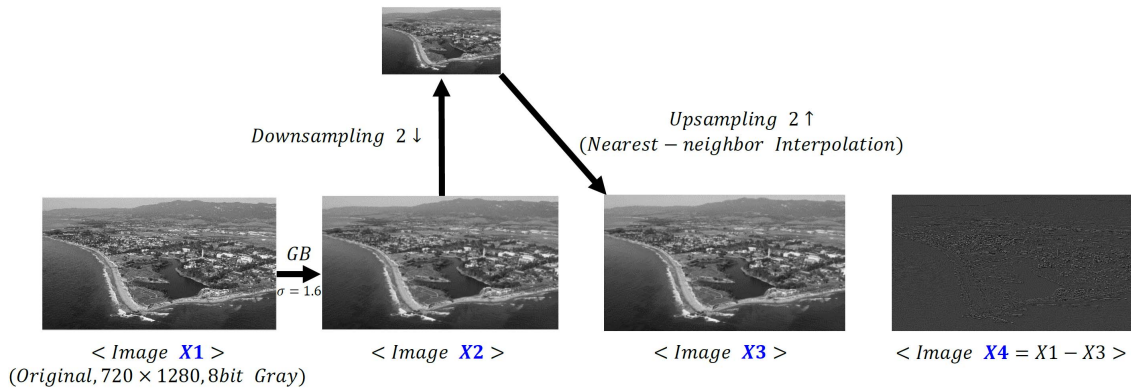


Figure 3. Image Processing Pyramid Level 1.

We start with an original 720 by 1280 gray level image (called Image X1) to which we apply Gaussian blurring to produce Image X2. We then downsample the Gaussian blurred image X2 by a factor of 2, as indicated in Figure 4, to generate a 360 by 640 image that is upsampled using nearest neighbor interpolation, as illustrated in Figure 5, to generate image X3. We compare both sides of Equation (13), but where $Q_{X|X2}$ is replaced with $MSE(X1, X2)$ and $Q_{X|X3}$ is replaced with $MSE(X1, X3)$. We emphasize that this substitution is known from Section 4 to always be possible for i.i.d. Gaussian, Laplacian, Logistic, triangular, and uniform distributions, as long as both quantities in the ratio have the same distribution.

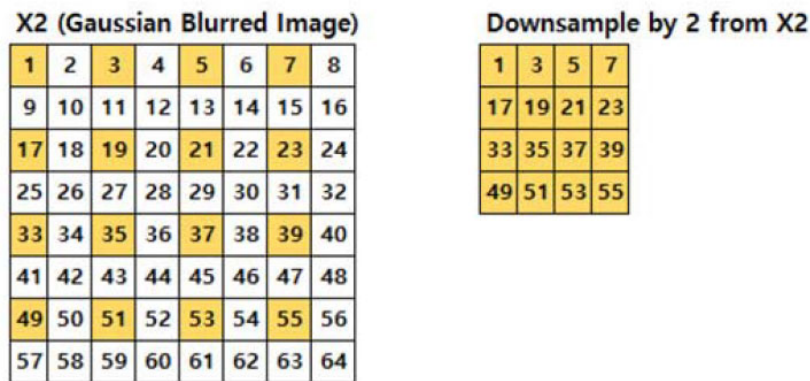


Figure 4. Downsampling Operation.

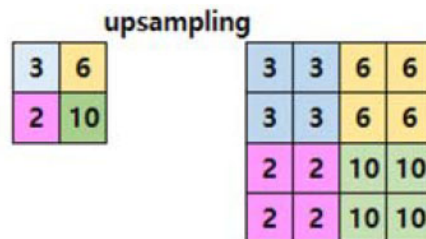


Figure 5. Upsampling Operation.

Table 1 shows the mutual information loss with respect to the original image X1 in going from image X2 to image X3 in the Gaussian pyramid for the City Image. Based on $\frac{1}{2} \log \frac{MSE(X1, X3)}{MSE(X1, X2)}$, we see that the simple operations of Gaussian blurring, downsampling, and then upsampling have produced a loss of about 0.2 bits/pixel for the 720 x 1280 resolution that increases to about 0.3 bits/pixel for the 360 x 640 and 180 x 320 images. As shown in Tables 2–4, for the 720 x 1280 resolution, the loss in the first layer of the Gaussian Pyramid is 0.3366 bits/pixel for News1, 0.3507 bits/pixel for News2, and

0.1661 bits/pixel for Campus, respectively. The losses in mutual information for the Gaussian Pyramid for the lower resolutions are also shown in the tables.

Table 1. Mutual Information Loss for the City Image: $P1 = I(X1; A) - I(X1; B)$ and $P2 = \frac{1}{2} \log \frac{MSE(X1,B)}{MSE(X1,A)}$; $A = X2$ and $B = X3$ for Gaussian Pyramid and $A = X3$ and $B = X4$ for Laplacian Pyramid.

Resolution	Gaussian Pyramid		Laplacian Pyramid	
Pixels	P1	P2	P1	P2
720 × 1280	0.2006	0.1989	1.2002	2.8224
360 × 640	0.3083	0.2881	1.2782	3.1357
180 × 320	0.2884	0.2703	1.0628	3.1638

Table 2. Mutual Information Loss for the News1 Image: $P1 = I(X1; A) - I(X1; B)$ and $P2 = \frac{1}{2} \log \frac{MSE(X1,B)}{MSE(X1,A)}$; $A = X2$ and $B = X3$ for Gaussian Pyramid and $A = X3$ and $B = X4$ for Laplacian Pyramid.

Resolution	Gaussian Pyramid		Laplacian Pyramid	
Pixels	P1	P2	P1	P2
720 × 1280	0.2077	0.3366	3.4145	3.3883
360 × 640	0.2957	0.4237	3.2072	3.1685
180 × 320	0.2553	0.3760	2.5955	2.786

Table 3. Mutual Information Loss for the News2 Image: $P1 = I(X1; A) - I(X1; B)$ and $P2 = \frac{1}{2} \log \frac{MSE(X1,B)}{MSE(X1,A)}$; $A = X2$ and $B = X3$ for Gaussian Pyramid and $A = X3$ and $B = X4$ for Laplacian Pyramid.

Resolution	Gaussian Pyramid		Laplacian Pyramid	
Pixels	P1	P2	P1	P2
720 × 1280	0.2667	0.3507	4.0786	5.1496
360 × 640	0.5012	0.3764	3.8745	4.9123
180 × 320	0.5048	0.3398	3.3294	4.7132

Table 4. Mutual Information Loss for the Campus Image: $P1 = I(X1; A) - I(X1; B)$ and $P2 = \frac{1}{2} \log \frac{MSE(X1,B)}{MSE(X1,A)}$; $A = X2$ and $B = X3$ for Gaussian Pyramid and $A = X3$ and $B = X4$ for Laplacian Pyramid.

Resolution	Gaussian Pyramid		Laplacian Pyramid	
Pixels	P1	P2	P1	P2
720 × 1280	0.1462	0.1661	1.9348	2.8614
360 × 640	0.2111	0.2471	2.4519	3.3809
180 × 320	0.2378	0.2836	2.5494	3.5067

We thus have a characterization of the loss in mutual information in a Gaussian pyramid decomposition of images that is easily calculated and is valid for a wide range of distributions. Therefore, other filtering, decimation, and interpolation methods thus can be easily compared in terms of the loss in mutual information produced in forming the Gaussian pyramid.

Table 1 shows results for the difference images in the Laplacian pyramid for the City image. We see that the mutual information loss is substantially higher than for the Gaussian pyramids for all images. This intuitively makes sense, but we note that the differencing operation means that the Markov chain property and the data processing inequality are no longer satisfied. We discuss the Laplacian pyramid results in Section 6.1.

6. Comparison to Direct Calculation

Tables 1–4 also show the difference in mutual information $I(X1; X2) - I(X1; X3)$, where the mutual informations are determined by estimating the probability density functions needed using

histograms of the image pixels and then using these results in the standard expression for the mutual information as if the pixels were i.i.d. Since the images certainly have spatial correlations, this is clearly just an approximation.

In any event, the quantity $I(X1; X2) - I(X1; X3)$ is 0.2006, 0.2077, 0.2667, and 0.1462 bits/pixel for the City, News1, News2, and Campus images (720 × 1280 resolution), respectively. Comparing to the loss in mutual information obtained from the expression $\frac{1}{2} \log \frac{MSE(X1,X3)}{MSE(X1,X2)}$, we see that the two quantities agree well for the City image and are close for the Campus image, but are significantly different for News1 and News2. It is thus of interest to investigate these differences further.

Specifically, we compare $\frac{1}{2} \log \frac{MSE(X1,X3)}{MSE(X1,X2)}$ with $I(X1; X2) - I(X1; X3)$ to determine when Equation (13) is satisfied with equality, viz, when

$$\frac{1}{2} \log \frac{MSE(X1, X3)}{MSE(X1, X2)} = I(X1; X2) - I(X1; X3) \tag{16}$$

In Table 5, we tabulate a normalized comparison between $I(X1; X2) - I(X1; X3)$ and $\frac{1}{2} \log \frac{MSE(X1,X3)}{MSE(X1,X2)}$, when normalized by the latter. From Table 5, we see that for the 720 × 1280 City image, Equation (16) is satisfied with near equality, less than 1% error, and for the Campus image, the one half log ratio of entropy powers is within 12% of the difference in the mutual information between the two stages. For the 720 × 1280 News1 this discrepancy is 38.3% and for News2 the difference is 24%. These trends are fairly similar across all three resolutions.

Table 5. Approximation Error : Per Cent Error = $\frac{|P2-P1|}{P2} \times 100$ where $P1 = I(X1; A) - I(X1; B)$ and $P2 = \frac{1}{2} \log \frac{MSE(X1,B)}{MSE(X1,A)}$; $A = X2$ and $B = X3$ for Gaussian Pyramid and $A = X3$ and $B = X4$ for Laplacian Pyramid.

	News1		News2		Campus		City	
Resolution Pixels	Gaussian Pyramid	Laplacian Pyramid	Gaussian Pyramid	Laplacian Pyramid	Gaussian Pyramid	Laplacian Pyramid	Gaussian Pyramid	Laplacian Pyramid
720 × 1280	38.30	2.72	23.98	25.45	11.98	37.30	0.90	59.85
360 × 640	30.22	4.44	33.14	27.44	14.59	32.67	7.00	61.76
180 × 320	32.10	16.66	48.55	39.41	16.13	34.76	6.71	70.64

Since we know that the one-half the log ratio of entropy powers equals one half the log ratio of mean squared errors when the distributions of the two quantities being compared are the same for i.i.d. Gaussian, Laplacian, Logistic, triangular, and uniform distributions, we are led to investigate the histograms of the four images to analyze these discrepancies. For the Gaussian and Laplacian distributions, we conducted Kolmogorov-Smirnov (K-S) goodness-of-fit tests for the four original images at all three resolutions. The results are shown in Figures 6–9.

The K-S statistic that we use is the distance of the cumulative distribution function from the hypothesized distribution given by

$$D = \sup_x |F_n(x) - F(x)| \tag{17}$$

where $F_n(x)$ is the distribution being tested, after correction for mean and variance, and $F(x)$ is the distribution being hypothesized. For this measure, a smaller value indicates a better match to the hypothesis.

For the News1 image results in Figure 6, the K-S statistic values are nearly the same for the Gaussian Pyramid for all three resolutions, but for the Laplacian Pyramid difference image the K-S statistic for a Gaussian distribution increases and becomes larger than the K-S value for a Laplacian distribution, suggesting that the difference image is more Laplacian than Gaussian. For the News2 image results in Figure 7, the behavior is similar although the K-S statistic for both the Gaussian and Laplacian distribution increase substantially for the difference image, and that the values of the K-S

statistics in Figure 7 for the Gaussian pyramid are smaller than for News1 but the K-S values are very nearly the same as News1 for the difference image.

The characteristics of the K-S test values start to change when it comes to the Campus and City images as shown in Figures 8 and 9. The magnitude of the K-S test statistic drops for the Gaussian distribution for the Campus image and is significantly lower for the Gaussian pyramid resolutions than the Laplacian K-S statistic. Indicating that the Campus image is more Gaussian than Laplacian. For the difference image, the K-S statistic for the Gaussian distribution increases and is larger than the K-S statistic for the Laplacian distribution. For the City image results in Figure 9, the K-S statistics for both the Gaussian and Laplacian distributions become smaller and the K-S statistic for the Gaussian pyramid images indicates that these images have a Gaussian distribution. For the difference image, again the K-S statistics for the Gaussian and Laplacian distributions change places with the K-S statistic suggesting that the difference image is more Laplacian than Gaussian.

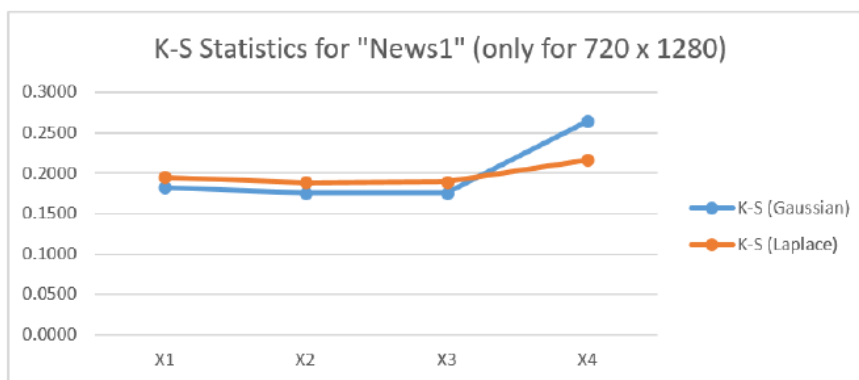


Figure 6. Kolmogorov-Smirnov Test Statistics for News1.

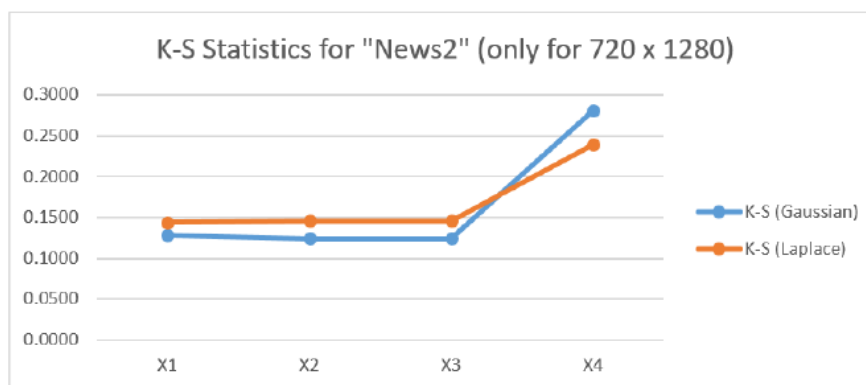


Figure 7. Kolmogorov-Smirnov Test Statistics for News2.

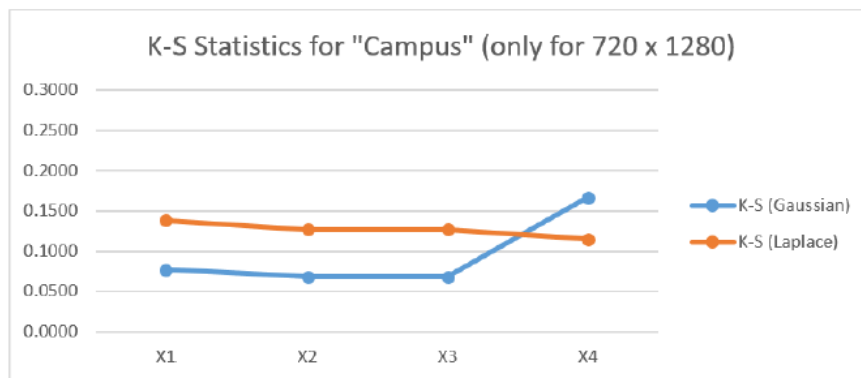


Figure 8. Kolmogorov-Smirnov Test Statistics for Campus.

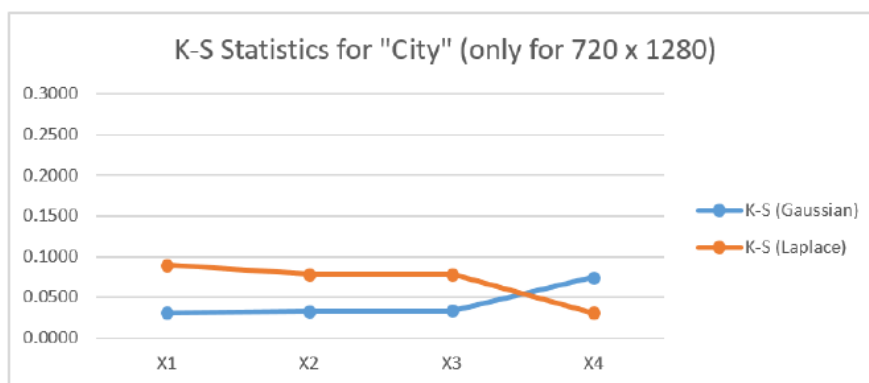


Figure 9. Kolmogorov-Smirnov Test Statistics for City.

Before applying these Kolmogorov-Smirnov goodness-of-fit test results to the analysis of Table 5, it is perhaps instructive to show the histograms and cumulative distribution functions of the four original 720 by 1080 resolution images in relation to the Gaussian density and Gaussian cumulative distribution as shown in Figures 10–13, for the four images. Visually it appears evident that News1 and News2 are not Gaussianly-distributed (nor Laplacian for that matter) and that the Campus image, while having multiple peaks in its histogram, has a more concentrated density than News1 and News2. The City image, however, appears remarkably Gaussian, at least visually.

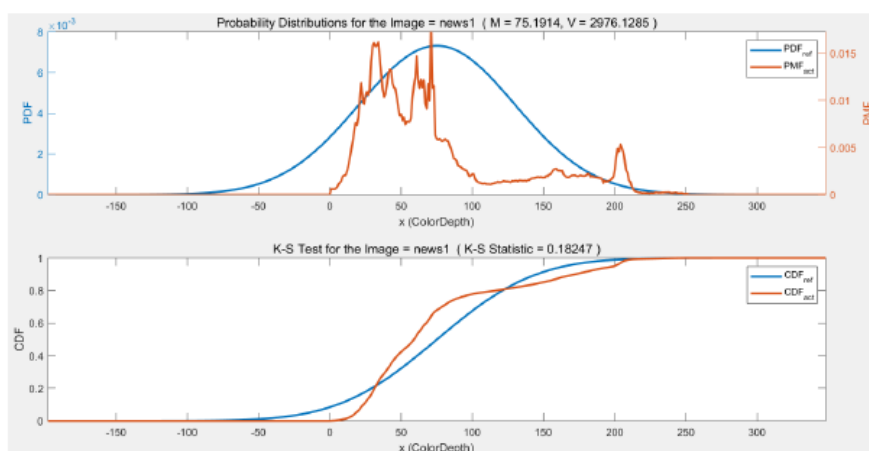


Figure 10. Histogram and Cumulative Distribution for News1.

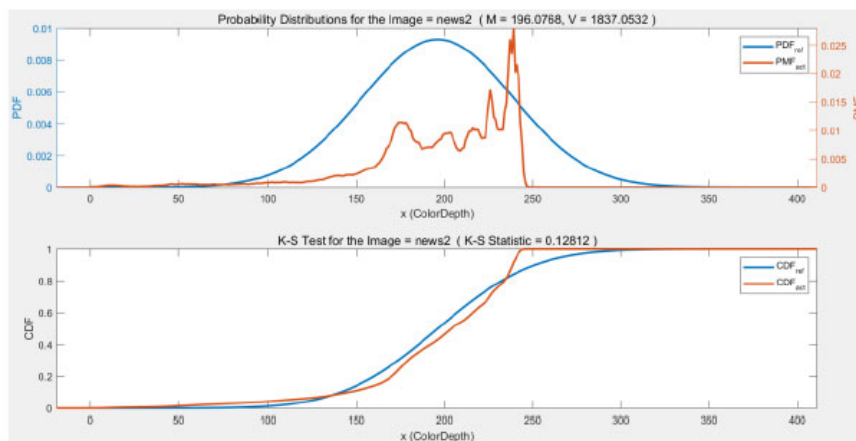


Figure 11. Histogram and Cumulative Distribution for News2.

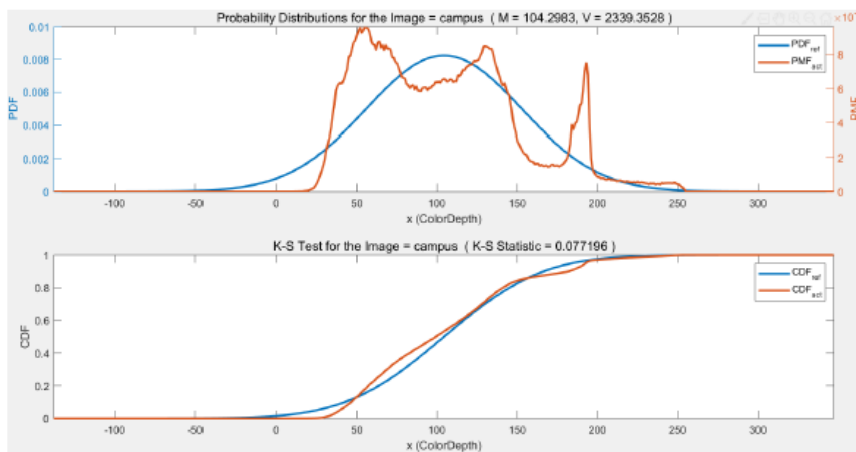


Figure 12. Histogram and Cumulative Distribution for Campus.

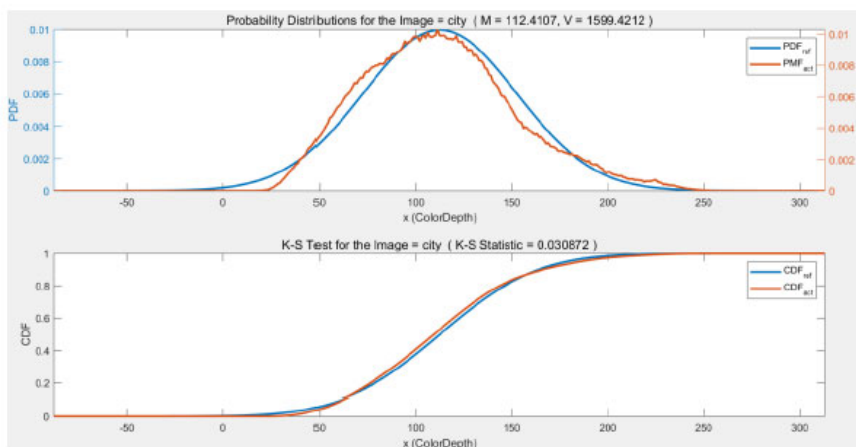


Figure 13. Histogram and Cumulative Distribution for City.

Returning to an analysis of the results shown in Table 5, we see that for the Gaussian pyramid, the accuracy of the approximation of the change in mutual information of the processing in going from Image X2 to X3 is best for the City image, which is as it should be since we know that the expression in Equation (16) should hold exactly for a Gaussian distributed sequence. The accuracy of the approximation to Equation (16) is progressively poorer for the Campus, News2, and News1

images in that order. More explicitly, for the Campus image, we see that it tests to be close to Gaussian, and so again the close approximation shown in the table is to be expected. Neither the News1 or News2 images test to be reliably Gaussian or Laplacian and so at least for these two distributions, the relatively poor agreement between the two quantities in the table is as to be expected.

We note that the approximation for the Laplacian pyramid is better in the reverse order to the Gaussian pyramid analyses just presented. These results are investigated further in the next section.

6.1. Laplacian Pyramid Analysis

Table 1 shows results for the difference images in the Laplacian pyramid for the City image. While the results are striking, as noted earlier, the differencing operation means that the Markov chain property and the data processing inequality are no longer satisfied. While the data processing inequality does not hold, Equations (10), (14), and (15) can hold if the distributions in the ratio are the same.

The results in Table 1 contrast comparing X1 to X3 in relation to a comparison between X1 and X4 using the quantity $\frac{1}{2} \log \frac{MSE(X1,X4)}{MSE(X1,X3)}$. We see from the table for the City image that the Laplacian difference image formed by subtracting X3 from X1 has a loss of mutual information of 2.8 bits/pixel for the largest resolution and a loss of slightly greater than 3 bits/pixel for the smaller images. The same result for the News1 image in Table 2 shows a loss of about 3.4 bits/pixel for the 720 × 1280 resolution. While one would expect the loss in going from X3 to the difference image X4 to entail a significant loss of information, here we have been able to quantify the loss in mutual information using an easily calculated quantity.

As shown in Tables 2 and 4 for the other images studied, News2 and Campus, $\frac{1}{2} \log \frac{MSE(X1,X4)}{MSE(X1,X3)} = I(X1; X3) - I(X1; X4)$ reveals an information loss of about 5.1 bits/pixel and 2.9 bits/pixel, respectively, in the Laplacian Pyramid for the 720 × 1280 resolution.

Intriguingly, the approximation accuracy for the Laplacian pyramid shown in Table 5 is much better for the image News1 than for the other images, and the accuracy of the approximation for the Laplacian pyramid difference image for the City image is the poorest. To analyze this behavior, we consider the Gaussian and Laplacian K-S test results for the four original 720 by 1080 resolution images X1 in Figure 14 and compare to the 720 by 1080 resolution Laplacian pyramid difference images X4 in Figure 15.

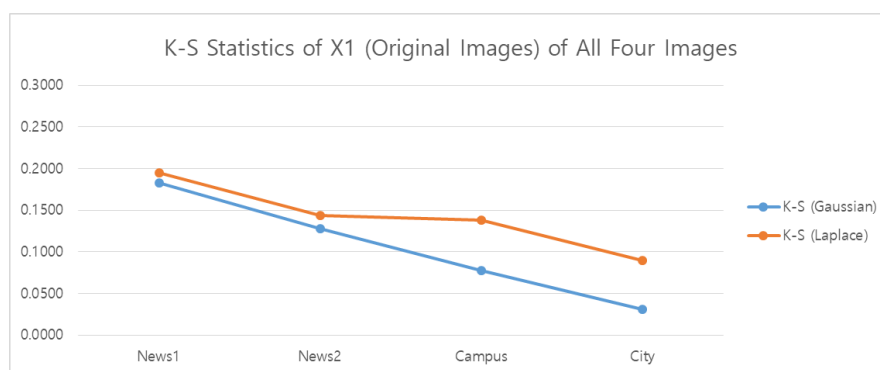


Figure 14. Kolmogorov-Smirnov Test Statistics for X1.

The entropy powers in Equation (13) can be replaced with the minimum mean squared estimation errors only if the distributions of X1 and X4 are the same. An inspection of Figures 14 and 15 implies that this is not true. That is, we see from the K-S test results in Figure 15 that while all of the difference images test to be more Laplacian than Gaussian, the difference image for the City image has the smallest K-S statistic of all the images for the Laplacian distribution hypothesis. At the same time, from Figure 14, all of the four test images are closer to the Gaussian hypothesis distribution than the Laplacian, and the City image is the closest to a Gaussian distribution of them all. Hence, the poor

approximations shown in Table 5 provided by the log ratio expression for the City image (and also for the News2 and Campus images) can be explained by the differing distributions of X_1 and X_4 .

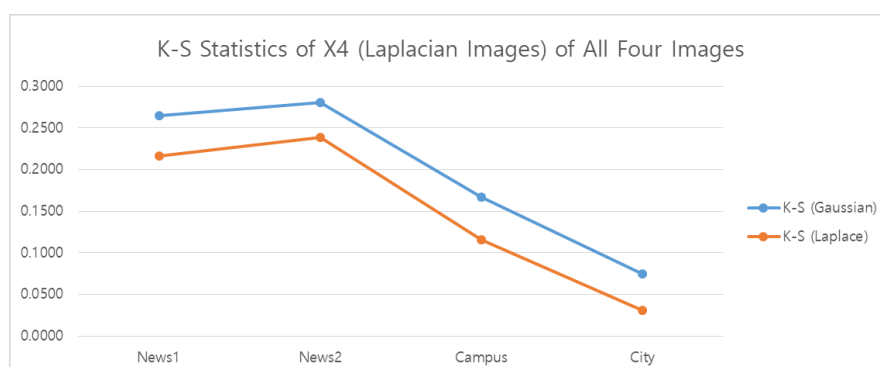


Figure 15. Kolmogorov-Smirnov Test Statistics for X_4 .

This analysis does not explain the very positive result shown for the Laplacian pyramid of News1 in Table 5. We conjecture that this unexpected result, namely that the approximation in Table 5 for the Laplacian pyramid is much better for the image News1 than for the other images, is the consequence of inherent correlation between the Laplacian pyramid image X_4 for News1 and the original News1 image. We elaborate on this idea as follows. The comparisons in the approximation table, namely Table 5, are with respect to the difference in mutual information between X_1 and X_3 and the mutual information between X_1 and X_4 . Generally, since X_4 is a difference image, it is expected that the mutual information between X_1 and X_4 should be less than the mutual information between X_1 and X_3 . For the City image the differencing does not enhance the similarities between X_1 and X_4 , whereas for the News1 image the difference enhances the strong edges present. See Figure 16. From these figures of the difference images, it is evident that all are correlated to some degree with the originals, but the correlation between X_1 and X_4 is considerably larger for the News1 image than for the City image and the other two images.

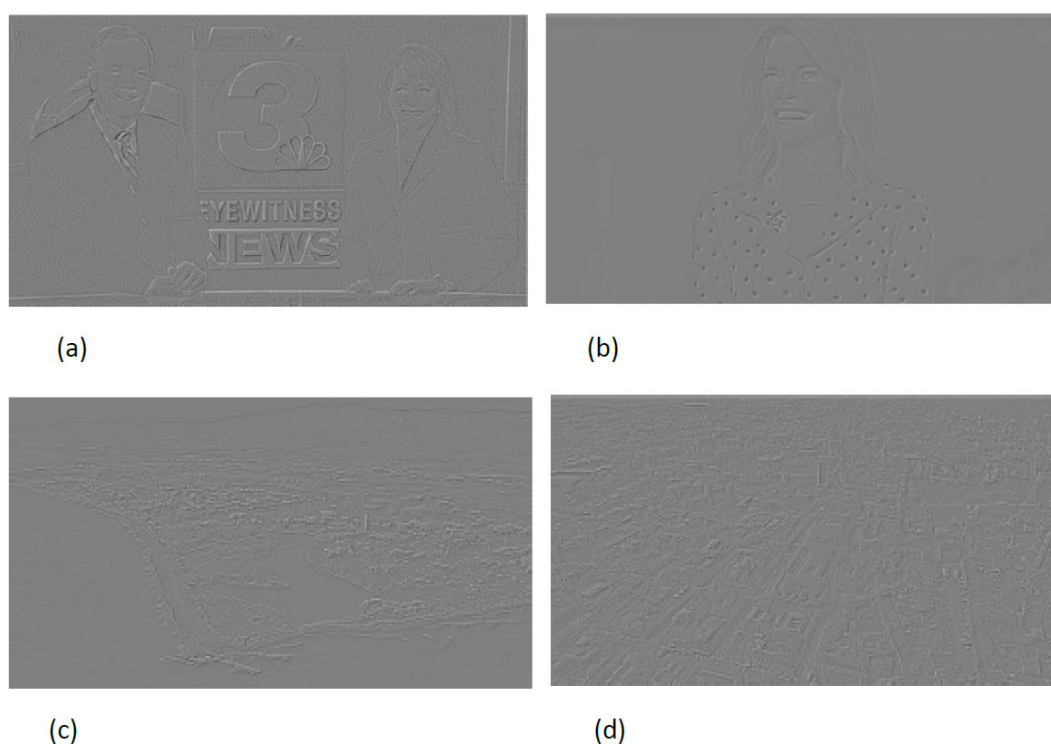


Figure 16. Level1 Difference Images for Laplacian Pyramids (a) News1, (b) News2, (c) Campus, (d) City.

To further elaborate on the effect of the correlation, if $X1$, $X3$, and $X4$ are Gaussian (which for an i.i.d. assumption, they are not), a well known result yields that

$$\frac{1}{2} \log \frac{MSE(X1, X4)}{MSE(X1, X3)} = \frac{1}{2} \log \frac{(1 - \rho^2(X1, X4))}{(1 - \rho^2(X1, X3))} \quad (18)$$

which is in fact $I(X1; X3) - I(X1; X4)$. Thus, the fact that $X3$ is more correlated with $X1$ than $X4$ for the City image, makes the denominator of Equation (18) smaller and the numerator larger, yielding the results shown in Table 5. Whereas for the News1 image, the differencing enhances the correlation between $X1$ and $X4$ and thus reduces the numerator in Equation (18).

A more detailed analysis of the mutual information loss in Laplacian pyramids is left for further study.

7. Discussion and Conclusions

The formation of Gaussian and Laplacian pyramids requires the signal processing steps of filtering/averaging, decimation and interpolation. The cascade of signal processing steps performed in obtaining a Gaussian pyramid forms a Markov chain, and therefore satisfies the Data Processing Inequality from information theory. This fact allows the introduction of a new performance indicator, the log ratio of entropy powers, to characterize the loss in mutual information at each signal processing stage in the Gaussian pyramid. Further, for i.i.d. Gaussian, Laplacian, logistic, uniform, and triangular distributions, the mean squared estimation error can replace the entropy power in the log ratio of entropy power expression as long as the two distributions in the ratio are the same. Therefore, the difference in differential entropies and mutual informations can be much more easily calculated using the log ratio of entropy powers in the chain.

The utility of these theoretical results is tested by comparing the images produced after the filtering stage and the images after decimation and upsampling with the original image for four different images. It is observed that under the correct assumptions, the log ratio of entropy powers captures the loss in mutual information in progressing through the filtering, decimation, and upsampling sequence. It is also seen that the images in the Laplacian pyramid do not form a Markov chain and therefore do not exhibit the expected behavior in the log ratio of entropy powers and mutual information loss.

There are two caveats to be stated here. First, the mutual informations for $I(X1; X2) - I(X1; X3)$ are calculated under an i.i.d. assumption as noted earlier. Further, the Kolmogorov-Smirnov goodness-of-fit tests are also for i.i.d. distributions. However, it is our conjecture that one half the log ratio of mean squared errors is perhaps a more accurate estimate of the difference in mutual informations when the underlying theoretical assumptions are satisfied than the calculated values of the difference in mutual informations because of the i.i.d. assumptions in obtaining the mutual informations directly from the first order histograms.

Author Contributions: Conceptualization, J.G.; data curation, H.O.; formal analysis, J.G.; software, H.O.; writing — original draft, J.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

K-S	Kolmogorov-Smirnov
MSE	Mean squared error
Q	Entropy power

References

1. Rabbani, M.; Jones, P.W. *Digital Image Compression Techniques*; SPIE Press: Bellingham, WA, USA, 1991; Volume 7.
2. Akansu, A.N.; Haddad, P.A.; Haddad, R.A.; Haddad, P.R. *Multiresolution Signal Decomposition: Transforms, Subbands, and Wavelets*; Academic Press: San Diego, CA, USA, 2001.
3. Vetterli, M.; Kovacevic, J. *Wavelets and Subband Coding*; Prentice-Hall: Englewood Cliffs, NJ, USA, 1995.
4. Yang, J.; Yu, K.; Gong, Y.; Huang, T. Linear spatial pyramid matching using sparse coding for image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 1794–1801.
5. Ogunfunmi, T.; Ramachandran, R.P.; Togneri, R.; Zhao, Y.; Xia, X. A primer on deep learning architectures and applications in speech processing. *Circuits Syst. Signal Process.* **2019**, *38*, 3406–3432. [[CrossRef](#)]
6. Silverman, B.W. *Density Estimation for Statistics and Data Analysis*; Routledge: Boca Raton, FL, USA, 2018.
7. Gibson, J.D. Entropy Power, Autoregressive Models, and Mutual Information. *Entropy* **2018**, *20*, 750. [[CrossRef](#)]
8. Gibson, J.D.; Oh, H. Characterizing Mutual Information Loss in Pyramidal Image Processing Structures. In Proceedings of the IEEE International Conference on Multimedia Information Processing and Retrieval, San Jose, CA, USA, 28–30 March 2019; pp. 1–6.
9. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; Wiley-Interscience: Hoboken, NJ, USA, 2006.
10. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [[CrossRef](#)]
11. El Gamal, A.; Kim, Y.H. *Network Information Theory*; Cambridge University Press: Cambridge, UK, 2011.
12. Messerschmitt, D.G. Accumulation of Distortion in tandem Communication links. *IEEE Trans. Inf. Theory* **1979**, *IT-25*, 692–698. [[CrossRef](#)]
13. Gibson, J.D. Log Ratio of Entropy Powers. In Proceedings of the UCSD Information Theory and Application Workshop, San Diego, CA, USA, 11–16 February 2018.
14. Xiong, W.; Faes, L.; Ivanov, P.C. Entropy measures, entropy estimators, and their performance in quantifying complex dynamics: Effects of artifacts, nonstationarity, and long-range correlations. *Phys. Rev. E* **2017**, *95*, 062114. [[CrossRef](#)]
15. Darbellay, G.A.; Vajda, I. Estimation of the information by an adaptive partitioning of the observation space. *IEEE Trans. Inf. Theory* **1999**, *45*, 1315–1321. [[CrossRef](#)]
16. Hudson, J.E. Signal Processing Using Mutual Information. *IEEE Signal Process. Mag.* **2006**, *23*, 50–54. [[CrossRef](#)]
17. Rabiner, L.R.; Schafer, R.W. *Digital Processing of Speech Signals*; Pearson: New York, NY, USA, 2011.
18. Chu, W.C. *Speech Coding Algorithms*; Wiley: Hoboken, NJ, USA, 2003.
19. Gray, R.M. *Linear Predictive Coding and the Internet Protocol*; NOW: Hanover, MA, USA, 2010.
20. Shynk, J.J. *Probability, Random Variables, and Random Processes: Theory and Signal Processing Applications*; John Wiley & Sons: Englewood Cliffs, NJ, USA, 2012.
21. Grauman, K.; Darrell, T. The Pyramid Match Kernel: Efficient Learning with Sets of features. *J. Mach. Learn. Res.* **2007**, *8*, 725–760.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).