ANALYSIS OF IMAGE INFORMATIVENESS MEASURES

Màrius Vila*  Anton Bardera*  Miquel Feixas*  Philippe Bekaert†  Mateu Sbert*

*Graphics and Imaging Laboratory, University of Girona, Spain
†iMinds - Multimedia Technologies Department, Diepenbeek, Belgium

ABSTRACT

Shannon entropy has been commonly used to quantify the image informativeness. The main drawback of this measure is that it does not take into account the spatial distribution of pixels. In this paper, we analyze four information-theoretic measures that overcome this limitation. Three of them (entropy rate, excess entropy, and erasure entropy) consider the image as a stationary stochastic process, while the fourth (partitional information) is based on an information channel between image regions and histogram bins. Experimental results, applied to natural and synthetic images, show the performance of these measures to characterize several informativeness aspects of an image. We also analyze their behavior under some image effects such as blurring, contrast change, and noise.

Index Terms—Information theory, Shannon entropy, entropy rate, erasure entropy, excess entropy, mutual information

1. INTRODUCTION

What is image information? How can we quantify the information content of an image? In this paper, we analyze the information content of an image from the perspective of information theory. Information theory, created by Shannon in 1948, deals with the information transmission, storage, and processing. It has been applied to many different fields such as physics, economics, and neurology. The most basic information measure is the Shannon entropy, which quantifies the information content or uncertainty of a random variable.

Given a color space (e.g., CIELab), the information associated with the lightness of the whole image can be computed from the entropy of the lightness histogram. The same procedure can be applied to any color component of a color space. The main drawback of histogram entropy, called image entropy, is the fact that it does not take into account the spatial distribution of pixels. This means that after a simple swapping of pixels, the image entropy would give the same result.

In other words, Shannon entropy is not an adequate measure to characterize, for instance, the structure of an image.

Previous works have used image information-theoretic measures to quantify visual quality by comparing a distorted image against a reference image [1, 2]. In this paper, we present a set of information measures that capture several aspects of image information, from a local perspective, taking into account the vicinity of pixels, to an evolutionary perspective, based on the difficulty of extracting or discovering the image information. Thus, we focus our attention not on how the image information varies when distortion is applied but on the quantification of the information of a single image. This fact enables us to evaluate the image quality and also to provide a set of image features that could be used to deal with several image processing problems such as image classification and optimization of image acquisition parameters.

The main novelties of this paper are the proposal of a new informativeness measure, called partitional information, which is based on the difficulty of gaining information, and the application of the erasure entropy in the context of image processing.

2. IMAGE INFORMATION MEASURES

In this paper, an image is considered as a random variable $B$ taking intensity bin values $b$ from a finite set $B$. Each value $b \in B$ represents a bin of intensity values that can be either composed by a single intensity value or by a set of similar intensity values depending on the used quantization. The probability distribution of the random variable $B$ is given by $p(b) = Pr[B = b]$. The Shannon entropy $H(B)$ of the random variable $B$ is defined by

$$H(B) = -\sum_{b \in B} p(b) \log p(b).$$

(1)

The term $-\log p(b)$ represents the information content associated with the intensity bin $b$. Thus, the entropy gives us the average amount of information of the intensity values in the image. From Equation 1, we can see that the entropy depends only on the probabilities of the intensity values, but not on their spatial distribution. Therefore, as we have mentioned in Section 1, two perceptually different images can give the same image entropy value.
In order to consider the spatial structure in the image information computation, two different approaches are presented. First, an image is modeled as a stationary stochastic process to quantify the image information from the vicinity of pixels and, second, an information channel between image regions and histogram bins is used to study the difficulty of extracting the information of an image.

The first approach models an image as a stationary stochastic process \( \{ B_i \} \), which is an indexed sequence of random variables characterized by the joint probability distribution \( p(b_1, b_2, \ldots, b_n) = \text{Pr}\{(B_1, B_2, \ldots, B_n) = (b_1, b_2, \ldots, b_n)\} \) with \( (b_1, b_2, \ldots, b_n) \in B^n \) for \( n \geq 1 \) [3, 4]. In our case, the sequence of states will be determined by consecutive positions in the image, considering, thus, the spatial information. Following the notation used in the work of Feldman and Crutchfield [5], a block of \( L \) consecutive random variables is denoted by \( B^L = B_1 \ldots B_L \). The probability that the particular \( L \)-block \( b^L \) occurs is denoted by the joint probability \( p(b^L) = p(b_1, b_2, \ldots, b_L) \). The Shannon entropy of length-\( L \) sequences or \( L \)-block entropy is defined by

\[
H(B^L) = - \sum_{b^L \in B^L} p(b^L) \log p(b^L),
\]

(2)

where the sum runs over all possible \( L \)-blocks. The entropy rate is defined by

\[
h = \lim_{L \to \infty} \frac{H(B^L)}{L} = \lim_{L \to \infty} h(L),
\]

(3)

where

\[
h(L) = H(B^L) - H(B^{L-1}) = H(B_L|B_1, \ldots, B_{L-1})
\]

(4)

is the entropy of a symbol conditioned on a block of \( L - 1 \) adjacent symbols. The entropy rate of a sequence quantifies the average amount of information per symbol \( h \) and the optimal achievement for any possible compression algorithm [3]. The entropy rate is always equal or lower than the Shannon entropy and is only equal when there is no correlation between consecutive symbols. Entropy rate has been introduced to image processing by Rigau et al. [6].

A complementary measure to the entropy rate is the excess entropy, which is a measure of the structure of a system. The excess entropy is defined by

\[
E \equiv \sum_{L=1}^{\infty} (h(L) - h)
\]

(5)

and captures how \( h(L) \) converges to its asymptotic value \( h \). If Equation (4) is inserted into Equation (5), the sum telescopes and an alternate expression for the excess entropy [5] is obtained:

\[
E = \lim_{L \to \infty} [H(B^L) - h \cdot L].
\]

(6)

It is important to note that, when we take into account only a few number of symbols in the entropy computation, the system appears more random that it actually is. This excess randomness measures how much additional information must be gained about the configurations in order to reveal the actual uncertainty \( h \) [7]. Excess entropy has been introduced to image processing by Bardera et al. [8].

As we have previously mentioned, the minimum compression length converges to the entropy rate, where the entropy of a symbol is conditioned by the previous (past) ones. However, what if we condition on both the past and the future? This idea is carried out by the erasure entropy. The erasure entropy measures the information content of each symbol knowing its context. For any stationary process, the erasure entropy is given by

\[
H^- = \lim_{L \to \infty} H(B_0|B_{-1}^L, B_1^L),
\]

(7)

where \( B_{-1}^L \) symbolizes the previous samples (past) and \( B_1^L \) the posterior samples (future).

The second approach introduces spatial information into the information measure by considering an information channel \( R \to B \) between the random variables \( R \) (input) and \( B \) (output), which represent, respectively, the set of regions \( \mathcal{R} \) of an image and the set of intensity bins \( B \). This channel is defined by a conditional probability matrix \( p(B|R) \) which expresses how the pixels corresponding to each region of the image are distributed into the histogram bins. Thus, each row \( r \) of the \( p(B|R) \) matrix corresponds to the normalized histogram of the region \( r \). The input distribution \( p(R) \), which represents the probability of selecting each image region, is defined by \( p(r) = \frac{n(r)}{N} \) (i.e., the relative area of region \( r \)), where \( N \) and \( n(r) \) are, respectively, the number of pixels of the image and the region \( r \). As it was proposed in [9, 10], the information bottleneck method can be applied to this information channel. Following a top-down partition procedure, a greedy algorithm based on a binary space partitioning (BSP) can be used to find a partition that maximizes the mutual information of the channel. The BSP partitioning algorithm can be represented by an evolving binary tree where each leaf corresponds to a terminal region of the image. At each partitioning step, the tree gains information from the original image such that each internal node \( \tilde{r} \) contains the information \( \delta I_{\tilde{r}} \), gained by splitting \( \tilde{r} \) in \( r_1 \) and \( r_2 \), which is given by

\[
\delta I_{\tilde{r}} = I(R, B) - I(\tilde{R}, B) = p(\tilde{r}) JS(\pi_1, \pi_2; p(B|r_1), p(B|r_2)),
\]

(8)

where \( \pi_1 = \frac{n(r_1)}{n(\tilde{r})} \) and \( \pi_2 = \frac{n(r_2)}{n(\tilde{r})} \). \( \tilde{R} \) represents the variable \( R \) before partitioning, and \( JS(\pi_1, \pi_2; p(B|r_1), p(B|r_2)) \) is the Jensen-Shannon divergence between two regions that can be interpreted as a measure of dissimilarity between them with respect to the intensity values. At a given moment, \( I(R, B) \) can be obtained adding up the information available at the
The entropy of the image \( H \) regions reach the number of pixels, the locally. If this procedure is performed until the number of re-

is important to stress that the best partition can be decided 

\[
\sum_{k=1}^{T} p(k) \delta I_{k},
\]

where \( T \) is the number of internal nodes of the tree weighted by \( p(k) \), where \( p(k) = \frac{n(k)}{N} \) is the relative area of the region associated with node \( k \) and \( n(k) \) is the number of pixels of this region. Thus, the mutual information of the channel is given by \( I(R, B) = \sum_{k=1}^{T} p(k) \delta I_{k} \). It is important to stress that the best partition can be decided locally. If this procedure is performed until the number of regions reach the number of pixels, the \( I(R, B) \) curve leads to the entropy of the image \( H(B) \). The integral of this curve can be seen as a measure of difficulty of describing the spatial distribution of the intensities. Thus, a new information measure, which we call partitional information, can be defined as 

\[
PI = \frac{\sum_{k=1}^{N} I(R, B)}{N \cdot H(B)}.
\]

This measure takes values in \([0, 1]\), leading to high values when the image is simple to describe and high values when the image is complex.

### 3. RESULTS

In this section, we analyze the information content of a group of synthetic and real images using the image information measures presented in Section 2. All images used in our experiments have been converted to grayscale in order to obtain a single intensity value between 0 and 256 for each pixel. Thus, probabilities of \( H \) have been computed using 256 intensity bins. To convert \( RGB \) values to grayscale values, we have used the CIE 1931 transformation 

\[
Y = 0.2126R + 0.7152G + 0.0722B,
\]

where \( R, G \) and \( B \) are, respectively, the values of red, green, and blue channels, and \( Y \) is the luminance obtained. To compute \( h, E \) and \( H^{-} \), the intensity values of an image are captured at evenly spaced positions over uniformly distributed random lines, also called global lines (see [8] for more details). To compute \( h \) and \( E \) we have used the Equations 3 and 6, respectively, taking the neighbor intensity values in \( L \)-blocks of size 3. \( H^{-} \) has been computed using Equation 7 with \( L \)-blocks of size 1.

First, we use four synthetic images with pixel resolution of \( 256 \times 256 \) to illustrate the behavior of the measures (see Fig. 1). All images have been created to have the same entropy value. The two first images 1(a-b) represent the same scene with the colors interchanged. In this case, \( h, E, H^{-} \) and \( PI \) values are equal since these measures are not dependent on the colors themselves, but only on their probability and spatial distribution. In the third image 1(c), some shapes are moved with respect to the original image 1(b). The last image 1(d) has been generated by randomly swapping 200000 pixels of image 1(b). Thus, both images 1(c-d) alter the structure with respect to the original image 1(b) but keep the same probability for each color. Observe that \( h \) and \( H^{-} \) increase their value with the increase of uncertainty and variability in the image. This fact complicates the prediction of the value of a pixel from the spatial distribution and, therefore, the value of these measures increase. On the contrary, \( E \) and \( PI \) decrease with the decrease of the spatial structure. Since Fig. 1(d) has no spatial structure, the knowledge of the spatial distribution does not improve the capability of predicting a pixel value and, therefore, \( h \) and \( H^{-} \) are very close to the \( H \) value.

<table>
<thead>
<tr>
<th>a)</th>
<th>b)</th>
<th>c)</th>
<th>d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H = 0.9427 )</td>
<td>( H = 0.9427 )</td>
<td>( H = 0.9427 )</td>
<td>( H = 0.9427 )</td>
</tr>
<tr>
<td>( h = 0.0932 )</td>
<td>( h = 0.0932 )</td>
<td>( h = 0.1830 )</td>
<td>( h = 0.9215 )</td>
</tr>
<tr>
<td>( E = 0.8551 )</td>
<td>( E = 0.8551 )</td>
<td>( E = 0.7657 )</td>
<td>( E = 0.0003 )</td>
</tr>
<tr>
<td>( H^{-} = 0.0253 )</td>
<td>( H^{-} = 0.0253 )</td>
<td>( H^{-} = 0.0591 )</td>
<td>( H^{-} = 0.9212 )</td>
</tr>
<tr>
<td>( PI = 0.9580 )</td>
<td>( PI = 0.9580 )</td>
<td>( PI = 0.9337 )</td>
<td>( PI = 0.7669 )</td>
</tr>
</tbody>
</table>

Fig. 1. Synthetic images and their entropy \( H \), entropy rate \( h \), excess entropy \( E \), erasure entropy \( H^{-} \) and partitional information \( PI \) values (a-d).

<table>
<thead>
<tr>
<th>h</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.9870</td>
</tr>
<tr>
<td>4.0776</td>
</tr>
<tr>
<td>4.3085</td>
</tr>
<tr>
<td>4.3853</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.8514</td>
</tr>
<tr>
<td>4.3185</td>
</tr>
<tr>
<td>4.3383</td>
</tr>
<tr>
<td>4.4422</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( H^{-} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5367</td>
</tr>
<tr>
<td>3.7335</td>
</tr>
<tr>
<td>3.7517</td>
</tr>
<tr>
<td>4.0256</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PI</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8971</td>
</tr>
<tr>
<td>0.8634</td>
</tr>
<tr>
<td>0.8580</td>
</tr>
<tr>
<td>0.8497</td>
</tr>
</tbody>
</table>

Fig. 2. Values of the measures of eight natural images. Each row corresponds to a measure and is sorted from the lowest to the highest value, except the partitional information that is sorted from the highest to the lowest value.
Fig. 3. (a-f) Mean values of each measure for five levels of distortion with respect to the original image using Gaussian white noise, Gaussian blurring, global contrast decrements, additive Gaussian pink noise, JPEG and JPEG2000 compression, respectively. (g) Mean values of each measure for three different image resolutions.

Second, we use a group of real images belonging to the Categorical Image Quality (CSIQ) Database [11] developed at Oklahoma State University. It consists of 30 original images with pixel resolution of $512 \times 512$ corresponding to five different subjects: animals, landscapes, people, plants, and urban environments. Each image is distorted using six types of distortions at five different levels, obtaining a total of 930 images. The distortion types used in CSIQ are JPEG and JPEG2000 compression, global contrast decrements, additive Gaussian pink noise, additive Gaussian white noise, and Gaussian blurring. Furthermore, with the aim of analyzing the behavior of all measures at different image resolutions, we have reduced the original images to pixel resolutions of $256 \times 256$, $128 \times 128$, and $64 \times 64$. Fig. 2 shows the absolute values of the measures of four images belonging to the CSIQ Database. Each row corresponds to a measure and is sorted from the lowest to the highest value, except $PI$ which is sorted from the highest to the lowest value.

Figs. 3(a-f) show several plots resulting from the application of additive Gaussian white noise, Gaussian blurring, global contrast decrements, additive Gaussian pink noise, JPEG and JPEG2000 compression distortion, respectively. Each plot shows the mean values of each measure at five levels of distortion with respect to the value in the original image. These six types of image effects can be divided into two main groups. On the one hand, at each new level of distortion, the additive Gaussian pink and white noise increase the variability of the image while the spatial structure decreases. On the other hand, the rest of image effects have a completely opposite behavior. In the first group, $H, h, H^{-}$ increase with the distortion level, while $E$ and $PI$ decrease. In the second group, the behavior is complementary. The only exception corresponds to the $E$ value with the global contrast reduction (see Fig. 3(c)). In this case, the decrease of $E$ value is due to the fact that the reduction of contrast keeps the spatial structure but also produces a global loss of information. Observe that these results are consistent with the ones obtained by using the synthetic images.

Finally, Fig. 3(g) shows how the image resolution affects the value of the different measures with respect to the value in the original image. We can observe that $H, PI$ and $E$ are nearly invariant to image resolution, while the other measures increase when the resolution is reduced. This reduction in the image resolution can be seen as a distortion that increases the variability of the image. Entropy rate and erasure entropy have obtained high values for images with high variability and low spatial structure, while excess entropy and partitional information show a complementary behavior. We can observe that excess entropy and partitional information are less sensitive to image resolution than the rest of measures.

4. CONCLUSIONS

In this paper, we have presented two different approaches to quantify the information content of an image taking into account the spatial distribution of pixels. In the first approach, entropy rate, excess entropy, and erasure entropy have been used to quantify the image information from the vicinity of pixels and, in the second approach, an information channel between image regions and histogram bins has been applied to study the difficulty of extracting the information of an image. The measures have been applied to several synthetic and natural images, analyzing, in the latter case, the behavior of the measures under several types of image effects. In our future work, these measures will be used to automatically adjust different camera effects such as focus, contrast or sharpness. We will also study the correlation between our measures and the mean opinion score (MOS) contained in the used database.
5. REFERENCES


