# An Information-Theoretic Framework for Image Complexity

J. Rigau, M. Feixas, and M. Sbert

Institut d'Informàtica i Aplicacions, Universitat de Girona, Spain

**Abstract**

*In this paper, we introduce a new information-theoretic approach to study the complexity of an image. The new framework we present here is based on considering the information channel that goes from the histogram to the regions of the partitioned image, maximizing the mutual information. Image complexity has been related to the entropy of the image intensity histogram. This disregards the spatial distribution of pixels, as well as the fact that a complexity measure must take into account at what level one wants to describe an object. We define the complexity by using two measures which take into account the level at which the image is considered. One is the number of partitioning regions needed to extract a given ratio of information from the image. The other is the compositional complexity given by the Jensen-Shannon divergence of the partitioned image.*

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computing Methodologies]: Computer GraphicsPicture/Image Generation; I.4.0 [Computing Methodologies]: Image Processing and Computer VisionImage Processing Software; I.4.6 [Computing Methodologies]: Image Processing and Computer VisionSegmentation

## 1. Introduction

In this paper, we introduce a new framework based on information theory and image segmentation to study the complexity of an image. Different authors have established a relationship between aesthetics and complexity. In 1928, G.D. Birkhoff introduced the concept of the *aesthetic measure*, defined as the ratio between *order* and *complexity* [Bir33, Bir50]. The complexity is roughly the number of elements that the image consists of and the order is a measure for the number of regularities found in the image [SB93]. Using information theory, M. Bense transformed Birkhoff's measure into an informational measure: redundance divided by statistical information. To compute the complexity, he introduced the assumption that an input pattern can always be described as a two dimensional grid of discrete symbols from a pre-defined repertoire. On the other hand, he observed that order corresponds to the possibility of perceiving large structures [Ben65, SB93]. A. Moles held that an aesthetic measure is closely related to image complexity, and based his measure of image complexity on information theory [Mol71, MC98]. P. Machado and A. Cardoso established that an aesthetic visual measure depends on two factors: processing complexity and image complexity [MC98]. They consider that images that are simultaneously visually complex and easy to process are the images that have a higher

aesthetic value. From the above discussed works, it appears that complexity is at the core of aesthetics. With the guideline that understanding complexity can shed light on aesthetics, we will explore image complexity from an information-theoretic perspective.

Image complexity has also been related to entropy of the image intensity histogram. However, this measure does not take into account the spatial distribution of pixels, neither the fact that a complexity measure must take into account at what level one wants to describe an object. For instance, a random sequence requires a long description if all details are to be described, but a very short one if a rough picture is required [Li97].

In image processing, an image is segmented by grouping the image's pixels into units that are homogeneous in respect to one or more characteristics, or features. Segmentation of nontrivial images is one of the most difficult tasks in image processing. Image segmentation algorithms are generally based on one of two basic properties of intensity values: discontinuity and similarity. In the first category, the approach is to partition the image based on abrupt changes in intensity, such as edges in an image. The principal approaches in the second category are based in partitioning an image into regions that are similar according to a set of predefined criteria. Thresholding, region growing, and region

| (*a*) Baboon | (*b*) Earth rise | (*c*) Flowers | (*d*) Lena | (*e*) NY | (*f*) Peppers |

**Figure 1:** *In your opinion, how complex are these images?*

splitting and merging are examples of methods in this category [BB82, GW02].

This paper is organized as follows. In Section 2, we present an algorithm which splits an image in relatively homogeneous regions using a binary space partition (BSP) or a quad-tree. In Section 3, complexity is defined by using two measures which take into account the level at which the image is considered. Finally, in Section 4, we present our conclusions and future research.

## 2. Previous Work

In this section, the most fundamental definitions and inequalities of information theory [CT91] are reviewed. In addition, the meaning of complexity and its diverse interpretations are presented.

### 2.1. Information Theory

#### 2.1.1. Entropy and Mutual Information

The *Shannon entropy* $H(X)$ of a discrete random variable $X$ with values in the set $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$ is defined as

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i, \qquad (1)$$

where $n = |\mathcal{X}|$, $p_i = Pr[X = x_i]$ for $i \in \{1, \ldots, n\}$. As $-\log p_i$ represents the *information* associated with the result $x_i$, the entropy gives us the *average information* or *uncertainty* of a random variable. The logarithms are taken in base 2 and entropy is expressed in bits. We use the convention that $0\log 0 = 0$. We can use interchangeably the notation $H(X)$ or $H(p)$ for the entropy, where $p = \{p_1, p_2, \ldots, p_n\}$ is the corresponding probability distribution.

If we consider another random variable $Y$ with marginal probability distribution $q$, corresponding to values in the set $\mathcal{Y} = \{y_1, y_2, \ldots, y_m\}$, the *conditional entropy* is defined as

$$H(X|Y) = -\sum_{j=1}^{m} \sum_{i=1}^{n} p_{ij} \log p_{i|j}, \qquad (2)$$

where $m = |\mathcal{Y}|$ and $p_{i|j} = Pr[X = a_i|Y = b_j]$ is the conditional probability. $H(X|Y)$ corresponds to the uncertainty in the channel input from the receiver's point of view, and vice versa for $H(Y|X)$. Note that in general $H(X|Y) \neq H(Y|X)$ and $H(X) \geq H(X|Y) \geq 0$.

The *mutual information* (MI) between two random variables $X$ and $Y$ is defined as

$$
\begin{aligned}
I(X,Y) &= H(X) - H(X|Y) = H(Y) - H(Y|X) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} \log \frac{p_{ij}}{p_i q_j}, \qquad (3)
\end{aligned}
$$

$p_{ij} = Pr[X = x_i, Y = y_j]$ is the joint probability. Mutual information represents the amount of information that one random variable, the output of the channel, contains about a second random variable, the input of the channel, and vice versa. $I(X,Y)$ is a measure of the *shared information* or *dependence* between $X$ and $Y$.

#### 2.1.2. Basic Inequalities

The following inequalities are fundamental to develop the most basic ideas in this paper.

**2.1.2.1. Jensen's Inequality** If $f$ is a convex function on the interval $[a, b]$, then

$$\sum_{i=1}^{n} \lambda_i f(x_i) - f\left(\sum_{i=1}^{n} \lambda_i x_i\right) \geq 0 \;, \qquad (4)$$

where $0 \leq \lambda \leq 1$, $\sum_{i=1}^{n} \lambda_i = 1$, and $x_i \in [a, b]$. If $f$ is a concave function, the inequality is reversed. Hence, if $f$ is substituted by the Shannon entropy, which is a concave function, we obtain the Jensen-Shannon divergence [BR82]:

$$J(\{\Pi_i\} : \{\pi_i\}) = H(\sum_{i=1}^{n} \pi_i \Pi_i) - \sum_{i=1}^{n} \pi_i H(\Pi_i) \geq 0, \qquad (5)$$

where $\Pi_1, \Pi_2, \ldots, \Pi_n$ are a set of probability distributions and $\pi_1, \pi_2, \ldots, \pi_n$ are the priori probabilities or weights, fulfilling $\sum_{i=1}^{n} \pi_i = 1$. The Jensen-Shannon divergence coincides with $I(X,Y)$ when $\{\pi_i\}$ is the marginal probability distribution $\{p_i\}$ of $X$ and $\{\Pi_i\}$ are the rows $\{P_i\}$ of the conditional probability matrix of the channel, i.e., $P_i = (p_{1|i}, p_{2|i}, \ldots, p_{m|i})$.

**2.1.2.2. Data processing inequality** If $X \rightarrow Y \rightarrow Z$ is a Markov chain, i.e., $p(x, y, z) = p(x)p(y|x)p(z|y)$, then

$$I(X,Y) \geq I(X,Z). \qquad (6)$$

This inequality demonstrates that no processing of $Y$, deterministic or random, can increase the information that $Y$ contains about $X$ [CT91].

**2.1.2.3. Fano's inequality** Suppose we have two correlated random variables $X$ and $Y$ and we wish to measure the probability of error in guessing $X$ from the knowledge of $Y$. Fano's inequality gives us a tight lower bound on this error probability in terms of the conditional entropy $H(X|Y)$. As $H(X|Y)$ is zero if and only if $X$ is a function of $Y$, we can estimate $X$ from $Y$ with zero probability of error if and only if $H(X|Y) = 0$. Intuitively, we expect to be able to estimate $X$ with a low probability of error if and only if $H(X|Y)$ is small [CT91].

If $X$ and $Y$ have the joint distribution $p(x,y) = p(x)p(y|x)$, from $Y$ we calculate a function $g(Y) = \tilde{X}$ which is an estimate of $X$. The probability of error is defined by

$$P_e = Pr[\tilde{X} \neq X] \tag{7}$$

and the Fano's inequality is given by $H(X|Y) \leq H(P_e) + P_e \log(n-1)$ or equivalently by

$$I(X,Y) \geq H(X) - H(P_e) - P_e \log(n-1), \tag{8}$$

where $H(P_e)$ is the binary entropy from $\{P_e, 1 - P_e\}$. Thus, Fano's inequality bounds the probability that $\tilde{X} \neq X$.

## 2.2. Complexity

In the last two decades, the study of *complexity* has become a very active research area in many different fields (automata, information theory, computer science, physics, biology, neuro-science, etc.) [BP97]. But, what is complexity? Webster's dictionary (1986) defines a complex object to be 'an arrangement of parts, so intricate as to be hard to understand or deal with.' According to W. Li's, the meaning of this quantity should be very close to certain measures of *difficulty* concerning the object or the system in question: the difficulty in constructing an object, the difficulty in describing a system, the difficulty in reaching a goal, the difficulty in performing a task, and so on [Li91]. There are many definitions of complexity [Gra86, Li91, BP97] corresponding to the different ways of quantifying these difficulties.

A list of complexity measures provided by Seth Lloyd [Llo02] is grouped under three questions: how hard is it to describe, how hard is it to create, and what is its degree of organization? In the first group, entropy is widely applicable for indicating randomness. It also measures uncertainty, ignorance, surprise, or information. In the second group, the computational complexity quantifies the amount of computational resources (usually time or space) needed to solve a problem [HU79]. Finally, in the third group, mutual information expresses the concept of complexity that quantifies the degree of structure or correlation of a system [Li91, FC98] or the amount of information shared between the parts of a system as a result of this organizational structure.

To our knowledge, the only framework existing until now dealing with image complexity is defined in [PS90], which deals with comparing the performance of ATR applications.

In this context, image complexity is defined as a measure of the inherent difficulty of finding a true target in a given image. Such a metric should predict the performance of a large class of ATRs on diverse imagery, without advanced knowledge of the targets. A split and merge segmentation algorithm is first applied that partitions an image into compact regions of uniform gray-level, no larger than the expected target size. Recursive thresholding determines the splits. After the segmentation procedure is applied, the target similarity of each region is estimated. The distribution of this similarity is taken as a basis for complexity measurement. For instance, if there are many regions with target similarity near the maximum the image is relatively complex. Three complexity measures are then given. The first is the number of regions whose target-similarity exceeds a given threshold, the second measures the distance from the body of the distribution to the most significant outlier, and the third is the weighted average of the distance to all outliers.
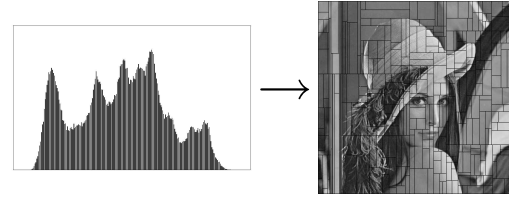


**Figure 2:** *Input and output distributions for the partitioning of channel.*

## 3. Method

Given an image with $N$ pixels and an intensity histogram with $n_i$ pixels in bin $i$, we define a discrete information channel where input $X$ represents the bins of the histogram, with probability distribution $\{p_i\} = \{\frac{n_i}{N}\}$, output $Y$ the pixel-to-pixel image partition, with uniform distribution $\{q_j\} = \{\frac{1}{N}\}$, and conditional probability $\{p_{j|i}\}$ of the channel is the transition probability from bin $i$ of the histogram to pixel $j$ of the image [RFS04]. This *information channel* can be represented by

$$X \longrightarrow Y \tag{9}$$
$$\{p_i\} \xrightarrow{\{p_{j|i}\}} \{q_j\}$$

As we have seen in Section 2.1.1, mutual information $I(X,Y)$ is a measure of the dependence or shared information between $X$ and $Y$. It can be seen that in the previous channel (9), given a pixel, there is no uncertainty about the corresponding bin of the histogram, or equivalently $I(X,Y) = H(X)$. From the data processing inequality (6), any clustering over $X$ or $Y$ will reduce $I(X,Y)$.

An information channel can be defined for each color component of an image. All the algorithms presented in this paper can be applied to each channel component of a color system.

*(a.i)* $P_e = 0.8$    *(b.i)* $P_e = 0.6$    *(c.i)* $P_e = 0.4$    *(d.i)* $P_e = 0.2$

*(a.ii)* $MIR = 0.2$    *(b.ii)* $MIR = 0.4$    *(c.ii)* $MIR = 0.6$    *(d.ii)* $MIR = 0.8$

**Figure 3:** Lena *image with luminance* $Y_{709}$ *for differents values of* $P_e$ *(*i*) and MIR (*ii*). The (RMSE, PSNR) values for (*i*) are (45.47, 14.66), (30.02, 18.27), (14.03, 24.88), and (8.21, 29.54), respectively. For (*ii*) we have (32.98, 17.45), (16.23, 23.61), (9.71, 28.07), and (6.25, 31.90), respectively.*



*(a)* R    *(b)* G    *(c)* B    *(d)* RGB

**Figure 4:** Lena *image in RGB system with* $P_e = 0.4$. *We obtain (*a*) 1835 (*$r = 0.70$*), (*b*) 3692 (*$r = 1.41$*), and (*c*) 4179 (*$r = 1.57$*) regions. The merging image (*d*) has RMSE=13.20 and PSNR=25.32.*



*(a)* BSP *(*$r = 0.010$, $MIR = 42.34$*)    (*b*) Quad-tree (*$r = 0.010$, $MIR = 39.10$*)    (*c*) Quad-tree (*$r = 0.014$, $MIR = 42.34$*)
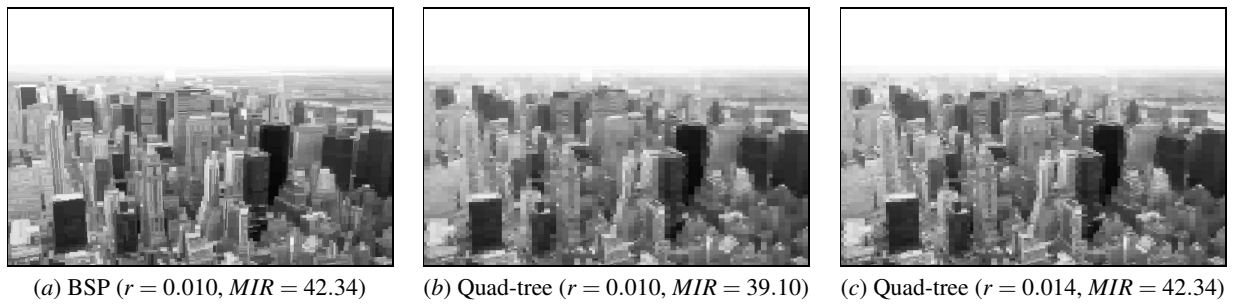
**Figure 5:** *BSP vs quad-tree splitting for* NY *image (Fig.* 1*.e) where (*a*) and (*b*) have 5002 regions and (*c*) 6859. The values of (RMSE, PSNR) for each image are (27.43, 19.37), (30.72, 18.38), and (29.59, 18.71), respectively.*

## 3.1. Image Partitioning

In this section, we present a greedy algorithm which partitions an image in quasi-homogeneous regions. The optimal partitioning algorithm is NP-complete. To do this partition, a natural approach could consider the above channel (9) as the starting point for the image partitioning, designing a pixel clustering algorithm which minimizes the loss of MI. This process can be described by a Markov chain, $X \rightarrow Y \rightarrow \widehat{Y}$, where $\widehat{Y} = f(Y)$ represents a clustering of $Y$.

However, due to the computational cost of this algorithm, a completely opposite strategy has been adopted: a top-down splitting algorithm takes the full image as the unique initial partition and progressively subdivides it with vertical or horizontal lines (BSP) chosen according to the maximum MI gain for each partitioning step. Note that other types of lines could be used, obtaining a varied polygonal subdivision. Our splitting process is represented over the channel (see Fig. 2)

$$X \longrightarrow \widehat{Y}. \tag{10}$$

The channel varies at each partition step because the number of regions is increased and, consequently, the marginal probabilities of $\widehat{Y}$ and the conditional probabilities of $\widehat{Y}$ over $X$ also change. This process can be interpreted in the following way: the choice of the partition which maximizes the MI increases the chances of guessing the intensity of a pixel chosen randomly from the knowledge of the region it pertains to.

The algorithm proposed here generates a partitioning tree for a given probability of error $P_e$ by maximizing the mutual information gain at each partitioning step. This algorithm is based on Fano's inequality and was introduced by Sethi and Sarvarayudu [SS82] in the context of pattern recognition. Similar algorithms with different split criteria have been used in learning [KLV98] and DNA segmentation [BOR99].

Given the error probability $P_e$ allowed in partitioning, Fano's inequality (8) provides us with a lower bound for the gain of mutual information. Taking the equality, we obtain the minimum value of MI needed in the partitioning algorithm for a given probability of error:

$$I_{min}(X,Y) = H(X) - H(P_e) - P_e \log(B-1), \tag{11}$$

where $B$ is the number of bins of the histogram. Note that $I_{min}(X,Y)$ is calculated from the initial channel (9).

The partitioning process can then be seen as follows. At each partitioning step, the tree acquires information from the original image. The total $I(X,\widehat{Y})$ captured by the tree can be obtained adding up the mutual information available at the non-terminal nodes of the tree weighted by the relative area of the region, i.e., the relative number of pixels, corresponding to each node. The mutual information $I_i$ of an interior node $i$ is *only* the information gained with its corresponding splitting. Thus, the total mutual information acquired in the process is given by

$$I(X,\widehat{Y}) = \sum_{i=1}^{T} \frac{n_i}{N} I_i, \tag{12}$$

where $T$ is the number of non-terminal nodes and $n_i$ is the number of pixels corresponding to node $i$. It is important to stress that this process of extracting information enables us to decide locally which is the best partition. Partitioning stops when $I(X,\widehat{Y}) \geq I_{min}(X,Y)$. Alternatively, a predefined ratio of mutual information (*MIR*) can be given as a stopping criterion. Note that $I(X,\widehat{Y})$ is the MI of the channel obtained at the end of the process.

This process can also be visualized from equation

$$H(X) = I(X,\widehat{Y}) + H(X|\widehat{Y}), \tag{13}$$

where the acquisition of information increases $I(X,\widehat{Y})$ and decreases $H(X|\widehat{Y})$, producing a reduction of uncertainty due to the equalization of the regions. Observe that the maximum mutual information that can be achieved is $H(X)$.

## 3.2. Results

Throughout this paper, the color channels used are $Y_{709}$, $R$, $G$, and $B$, although any other color space could be used with our algorithms. Also, the regions in all the partitioned images are shown with their average intensity. The default partition tree is BSP and the main test image is *Lena* in Fig. 1.*d*.

The performance of our partitioning approach is shown in Figures 3-6. A set of partitions over the test image, shown in Fig. 3, illustrates the behavior of the BSP partitioning algorithm. The first row (*i*) has been obtained using the error probability $P_e$ stopping criterion and the second (*ii*) the *MIR* criterion. The behavior of both root mean square error (RMSE) and peak signal-to-noise ratio (PSNR) values is as expected, decreasing the RMSE (respectively increasing the PSNR) with decreasing $P_e$ (respectively increasing *MIR*).

The partition of the test image in the RGB space for error probability $P_e = 0.4$ is shown in Fig. 4. Each channel is independently partitioned and the merging of Fig. 4.*a-c* is shown in Fig. 4.*d*. The ratio $r = \frac{R}{N}$, where $R$ is the number of regions obtained and $N$ is the number of pixels, is shown in this figure.

The quality of the two splitting variants, BSP and quad-tree, is analyzed in Fig. 5. Observe that, for a given error probability, the quad-tree solution has more partitions than the BSP one. In general, the quad-tree needs more regions than the BSP to extract the same quantity of information. In addition, for the same *MIR*, the quality of the BSP option is better than for the quad-tree one. Observe in Fig. 5 that both RMSE and PSNR values are ranked accordingly with the visual quality.

The ratio $r$ obtained from the processing of the six images in Fig. 1 is presented in Fig. 6. Observe that, for instance, the

*Baboon* image (Fig. 1.*a*) requires 7.45 times more regions than the *Earth rise* image (Fig. 1.*b*) for the same *MIR* = 0.9.
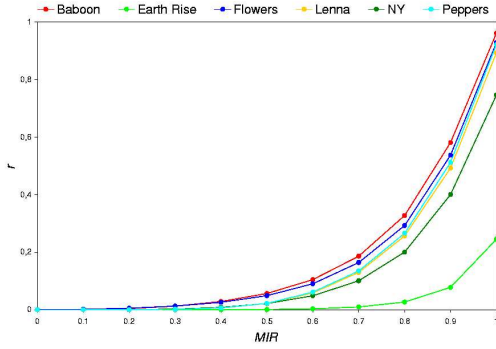


**Figure 6:** *Ratio of the number of regions r with respect to MIR for the images of Fig. 1 with luminance $Y_{709}$.*

## 4. Image Complexity

### 4.1. Complexity Measures

According to Li [Li97], a measure of complexity of an object is a measure of complexity of a task performed on that object. As we have seen in Sec. 2.2, the concept of complexity is closely related to the difficulty of understanding an object, which, at the same time, is related to the accuracy of the description of it [BP97]. On the other hand, the measure of complexity must take into account at what level one wants to describe the object. Thus, we can describe every detail of an object or only its non-random regularities [Li97]. According to this, an important group of complexity measures tries to capture the organizational structure or the degree of regularity versus randomness. In this section, we are going to present two complexity measures rooted in these criteria and based on image partitioning.

To introduce our complexity framework, we will reinterpret the previous partitioning approach from the point of view of the maximization of the Jensen-Shannon divergence. This perspective, although equivalent to the maximization of mutual information, is more appropriate to deal with image complexity and has been introduced in the study of the DNA complexity [RBO98].

First, we define a complexity measure, the Jensen-Shannon divergence, which expresses the *image compositional complexity* (ICC) of an image. This measure can be interpreted as the spatial heterogeneity of an image from a given partition. From (5), the Jensen-Shannon divergence applied to an image is given by

$$JS(X,\widehat{Y}) = H(X) - \sum_{i=1}^{R} \frac{n_i}{N} H(X_i)$$
$$= H(X) - H(X|\widehat{Y}) = I(X,\widehat{Y}) \quad (14)$$

where $R$ is the number of regions of the image, $X_i$ is the random variable associated with region $i$, representing the intensity histogram of this region, $n_i$ is the number of pixels of region $i$, and $N$ is the total number of pixels of the image. Observe that for the information channel (10), the Jensen-Shannon divergence coincides with the MI. The compositional complexity (14) fullfils the following properties:

- It increases with a finer partition.
- It is null for a single partition.
- For a random image and a coarse resolution it would be close to 0.
- For a random image and the finest resolution it would be maximum and equal to $H(X)$.

Thus, given an image partition, we can express the heterogeneity of an image using the JS-divergence applied to the probability distribution of each region.

We can also ask which partition maximizes the compositional complexity, for a given number of regions. As we have seen in Sec. 3, this partition should extract the maximum information of the image and create the maximum heterogeneity between the generated parts. Finding this partition is an NP-complete problem. We have approached the solution to this problem in Sec. 3 using a greedy algorithm.

Our second measure is the number of needed regions in the partitioned image to extract a given ratio of information. It is related to the complexity in describing an image, and depends on the accuracy level given by $P_e$ or *MIR*. The justification for our measure is that the number of regions is the number of leaves of the tree created in the partitioning process. The coding of this tree (or equivalently the description of the image) will be clearly dependent on this number. This is further justified by taking into account that our algorithm tries to create homogeneous regions with the minimum splitting. In this case, the error probability of the channel is interpreted as the compression error and thus the number of regions is also related to the difficulty of compression.

### 4.2. Results

We use a uniform partition to test the compositional complexity on the images in Fig. 1. The results obtained are shown in Fig. 7 for the number of partitions running from $2 \times 2$ to the number of pixels in the respective images. We observe that the relative ordering of the complexities depends on the resolution level (number of partitions). For instance, the earth rise image appears to be the most complex at resolution $4 \times 4$ while the wild flowers appears as the least one. However, this behavior is reversed at high resolution.

In Figure 6 we can analyze the behavior of the second proposed complexity measure. While the lines in the graph in Fig 7 cross themselves, the ones in Figure 6 keep a regular ordering. Observe their exponential growing with *MIR* that is due to the increasing cost of the MI extraction. It is important to note that for *MIR* = 0.5 we obtain a good quality
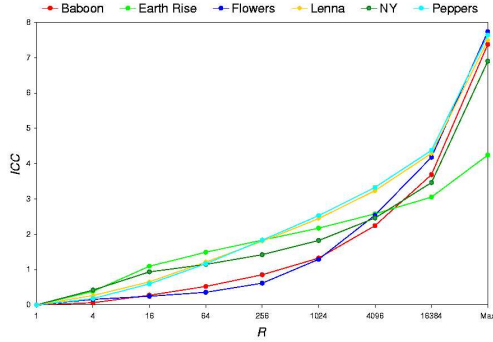
**Figure 7:** *Compositional complexity ICC over the number of regions R of the partitioned images of Fig. 1 with luminance $Y_{709}$. The number of partitions goes from $2 \times 2$ to the number of pixels N in the respective images.*

with a few number of regions. With respect to the number of regions, the most complex image appears to be the *Baboon* and the least one is the *Earth rise*.

It can also be shown (Figure 8) that while blurring an image will cause a loss of complexity, increasing the contrast causes the opposite effect. For instance, for a $MIR = 1$ and the luminance channel $Y_{709}$, the contrasted Lena image of Figure 8.b ($r = 91.7$) needs more regions than the original Lena image ($r = 89.4$) and the blurred image of Figure 8.a ($r = 48.3$) needs less regions.



(a)             (b)

**Figure 8:** Lena *image: (a) Out of focus and (b) more contrasted than its original.*

## 5. Conclusions and Future Research

We have introduced in this paper a new framework to study the complexity of an image, based on information theory. The framework is based on the segmentation of an image. We defined a generic information channel that takes an image and its histogram as its input and outputs a partitioned image. The channel evolves with the output, which at the beginning is formed by the image as a single region, the root of the partitioning tree. The mutual information between the

input and output variables drives the partitioning so that the next splitting is chosen to maximize the gain in mutual information. This process stops when the accumulated gain in mutual information ensures, by the data processing inequality, that we have reached a given error probability. At the end of the process we have a segmented image that provides us with two complexity measures. The first represents the compositional complexity, and is given by the Jensen-Shannon divergence of the partitioned image. The second is the number of regions in which the image was partitioned for a given information gain, and gives us the difficulty in describing the image.

In our future work, the relationship between image complexity and aesthetic measures will be further investigated in line with the work started by Birkhoff. Short and long correlations in an image will be studied, and also their relation with image compression.

## References

[BB82] BALLARD D. H., BROWN C. M.: *Computer Vision*. Prentice Hall, Englewood Cliffs (NJ), USA, 1982. 2

[Ben65] BENSE M.: *Aesthetica. Einführung in die neue Aesthetik*. Agis-Verlag, Baden-Baden, Germany, 1965. 1

[Bir33] BIRKHOFF G. D.: *Aesthetic Measure*. Harvard University Press, Cambridge (MA), USA, 1933. 1

[Bir50] BIRKHOFF G. D.: *Collected Mathematical Papers*. American Mathematical Society, New York (NY), USA, 1950. 1

[BOR99] BERNAOLA P., OLIVER J. L., ROMÁN R.: Decomposition of DNA sequence complexity. *Physical Review Letters 83*, 16 (October 1999), 3336–3339. 5

[BP97] BADII R., POLITI A.: *Complexity. Hierarchical Structures and Scaling in Physics*. Cambridge University Press, 1997. 3, 6

[BR82] BURBEA J., RAO C. R.: On the convexity of some divergence measures based on entropy functions. *IEEE Transactions on Information Theory 28*, 3 (May 1982), 489–495. 2

[CT91] COVER T. M., THOMAS J. A.: *Elements of Information Theory*. Wiley Series in Telecommunications, 1991. 2, 3

[FC98] FELDMAN D. P., CRUTCHFIELD J. P.: *Discovering Noncritical Organization: Statistical Mechanical, Information Theoreticand Computational Views of Patterns in One-Dimensional Spin Systems*. Working Paper 98–04–026, Santa Fe Institute, Santa Fe (NM), USA, April 1998. 3

[Gra86] GRASSBERGER P.: Toward a quantitative theory of self-generated complexity. *International Journal of Theoretical Physics 25*, 9 (1986), 907–938. 3

[GW02]  GONZALEZ R. C., WOODS R. E.: *Digital Image Processing*. Prentice Hall, Upper Saddle River (NJ), USA, 2002. 2

[HU79]  HOPCROFT J. E., ULLMANN J. D.: *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley, Reading (MA), USA, 1979. 3

[KLV98]  KULKARNI S. R., LUGOSI G., VENKATESH S. S.: Learning pattern classification – a survey. *IEEE Transactions on Information Theory 44*, 6 (1998), 2178–2206. 5

[Li91]  LI W.: On the relationship between complexity and entropy for markov chains and regular languages. *Complex Systems 5*, 4 (1991), 381–399. 3

[Li97]  LI W.: The complexity of DNA. *Complexity 3*, 2 (1997), 33–37. 1, 6

[Llo02]  LLOYD S.: Measures of complexity: a non-exhaustive list, 2002. 3

[MC98]  MACHADO P., CARDOSO A.: Computing aesthetics. In *Proceedings of XIVth Brazilian Symposium on Artificial Intelligence (SBIA'98)* (Porto Alegre, Brazil, November 1998), LNAI, Springer-Verlag, pp. 219–229. 1

[Mol71]  MOLES A.: *Art et ordinateur*. Casterman, Tournay, Belgium, 1971. 1

[PS90]  PETERS II R. A., STRICKLAND R. N.: Image complexity metrics for automatic target recognizers, October 1990. invited paper. 3

[RBO98]  ROMÁN R., BERNAOLA P., OLIVER J. L.: Sequence compositional complexity of DNA through an entropic segmentation method. *Physical Review Letters 80*, 6 (February 1998), 1344–1347. 6

[RFS04]  RIGAU J., FEIXAS M., SBERT M.: An information theoretic framework for image segmentation. In *Proceedings of IEEE International Conference on Image Processing (ICIP 2004)* (October 2004). Held in Singapore, Republic of Singapore. 3

[SB93]  SCHA R., BOD R.: Computationele esthetica. *Informatie en Informatiebeleid 11*, 1 (1993), 54–63. 1

[SS82]  SETHI I. K., SARVARAYUDU G.: Hierarchical classifier design using mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence 4*, 4 (July 1982), 441–445. 5