An Entropy-based Model for a Fast Computation of SSIM

Vittoria Bruni[1,2] and Domenico Vitulano[2]
[1]*Dept. of SBAI, University of Rome La Sapienza, Via A. Scarpa 16, Rome, Italy*
[2]*Istituto per le Applicazioni del Calcolo, CNR, Via dei Taurini 19, Rome, Italy*

Keywords: Information Theory, SSIM, Image Quality Assessment, Typical Set.

Abstract: The paper presents a model for assessing image quality from a subset of pixels. It is based on the fact that human beings do not explore the whole image information for quantifying its degree of distortion. Hence, the vision process can be seen in agreement with the Asymptotic Equipartition Property. The latter assures the existence of a subset of sequences of image blocks able to describe the whole image source with a prefixed and small error. Specifically, the well known Structural SIMilarity index (SSIM) has been considered. Its entropy has been used for defining a method for the selection of those image pixels that enable SSIM estimation with enough precision. Experimental results show that the proposed selection method is able to reduce the number of operations required by SSIM of about 200 times, with an estimation error less than 8%.

1 INTRODUCTION

A wide literature has definitely proved that embedding and translating HVS concepts in image processing based applications promote the optimization of several applications in terms of efficiency, precision, automaticity and, sometimes, computing time (Bruni et al., 2012; Bruni et al., 2013a; Hontsch and Karam, 2002; Hou and Yau, 2010; Jourlin and Pinoli, 1998; Lee and Lee, 2006; Panetta et al., 2008; Wang and Li, 2011). In this context, the definition of measures for image quality assessment that correlate more with human visual system plays a fundamental role (Bruni et al., 2013b; Ferzli and Karam, 2009; Moorthy and Bovik, 2009; Sheikh et al., 2005; Wang et al., 2004; Wang and E.P.Simoncelli, 2005; Wang and Li, 2011). Despite its recognized lack of correlation with human perception, the classical mean squared error (MSE) is still used in many applications, especially in optimization problems, due to its simplicity, low computational effort and nice mathematical properties. The Structural SIMilarity index (SSIM) (Wang et al., 2004) revealed to be a robust competitor of MSE thanks to its discrete correlation with HVS, its definition through very simple operations and, as recenty proved, its interesting mathematical properties that can promote its use, for example, in regularization methods. Unfortunately, the computational cost required by SSIM is higher than the one required by MSE. SSIM is a pixelwise measure but it involves block-based operations for each pixel[1].

The aim of this paper is to speed up the computation of SSIM by computing it on a reduced number of blocks. This strategy mainly relies on the fact that humans are able to assign a score to the image just looking at few specific points, known as fixation points (Monte et al., 2005; Frazor and Geisler, 2006). This way of selecting information is closely related to some concepts in Information Theory and, in particular, to the Asymptotic Equipartition Property (Cover and Thomas, 1991). This principle states that for a given source, there exists a subset of sequences able to represent the whole source — i.e. with entropy close to the source entropy. Accordingly, in the context of vision, there exists more than one sequence of fixation points of a given length that is able to code the whole image information content. Hence, by defining the visual distortion typical set as in (Bruni and Vitulano, 2014), we want to develop a method for extracting at least one sequence belonging to this set from which

[1]For the $i-$th pixel SSIM, is defined as follows:

$$SSIM(b_i,d_i) = \frac{2\mu_{b_i}\mu_{d_i}+C_1}{\mu_{b_i}^2+\mu_{d_i}^2+C_1}\,\frac{\sigma_{b_i d_i}+C_2}{\sigma_{b_i}^2+\sigma_{d_i}^2+C_2},$$

where b_i and d_i are blocks centered at i respectively in the original and distorted image, μ_* and σ_* respectively are the mean and the standard deviation of $*$, $\sigma_{b_i d_i}$ is the correlation between b_i and d_i, while C_1 and C_2 are numerical stabilizing constants.

Bruni, V. and Vitulano, D.
An Entropy-based Model for a Fast Computation of SSIM.
DOI: 10.5220/0005730002260233
In *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2016) - Volume 4: VISAPP*, pages 226-233
ISBN: 978-989-758-175-5

assessing the quality of the whole image. With respect to the literature concerning the selection of the best pooling weights for an image quality assessment measure (Moorthy and Bovik, 2009; Park et al., 2011; Wang and Li, 2011)), the proposed method can also be seen as a binary pooling, preserving some blocks while discarding the others.

The proposed method consists of the following main steps:

1. Luminance based image segmentation: the image is split into a finite number of distinct regions having different characteristics;
2. Finite random walk on a connected and weighted graph whose nodes are the regions given by the segmentation. This step provides a sequence of points belonging to the typical set of length K. K is automatically determined for each image using the Minimum Description Length principle (MDL) (Grunwald, 2004).

It will be shown that the mean value of the SSIM evaluated on blocks centered at these points gives a faithful estimation of SSIM of the whole image with a considerable computational saving. Experimental results on test images from TID2013 database (Ponomarenko et al., 2015) show that it is possible to reach a speed up for SSIM, evaluated for different distortion levels, over 200:1 with a relative estimation error lower than 8%.

The outline of the paper is the following. The next Section gives some preliminary results on the visual distortion typical set. Section 3 presents a method for determining a sequence of points belonging to this set; details about the algorithm and its computational cost will also be given. Section 4 presents some experimental results obtained on TID2013 database while the last section draws the conclusions.

2 SOME PRELIMINARY RESULTS

In (Bruni and Vitulano, 2014), the *visual distortion typical set* A_M^ε has been defined as a subset of all sequences composed of samples of the original image I (and the corresponding ones in the degraded image I_d) such that they give an approximated value $\hat{M}$ of the expected value of the measure M (i.e. $\bar{M}$) within an error ε, i.e.: $|\hat{M}-\bar{M}|<\varepsilon$, where M is the reference quality measure. In our case, M is the pointwise SSIM, $\bar{M}$ is the mean of M computed using all image pixels, while $\hat{M}$ is the mean of M computed on a reduced number of image pixels. More formally, A_M^ε is the set of sequences of fixed size whose entropy is close to the entropy of the source. The existence of A_M^ε is guaranteed by the Asymptotic Equipartition Property (AEP) (Cover and Thomas, 1991), that states that for i.i.d. r.v.s X_i it holds: $\frac{1}{n}\log\frac{1}{p(X_1,X_2,..,X_n)} \to H(X) \quad n\to\infty$.

AEP is the entropic version of the weak law of large numbers. However, the entropy based version is more mathematically tractable as entropy increases as the number of samples grows (Cover and Thomas, 1991), while it is not so for the mean value. Based on these concepts, in (Bruni and Vitulano, 2014) the authors gave some guidelines for an optimized extraction of the visual distortion typical set from the couple of images (I,I_d). Specifically, it has been formally proved that:

1. Not all information in I and I_d is really important; it is sufficient to select just a part of it for assessing image quality. In addition, an entropy based criterion can be applied for selecting the significant information. Specifically, it has been proved the following result:

   Proposition 1. Let $X\sim Q$ with a positive and numerical alphabet χ and $\{X_1\}\sim p_1,\{X_1,X_2\}\sim p_2,\ldots,\{X_1,X_2,\ldots,X_n\}\sim p_n$. Let μ_n be the mean of p_n, μ be the mean of Q and D_{KL} the Kullbach-Leibler divergence. Then

   (a) the sequence $\{\mu_n\}$ is not monotonic for increasing n;

   (b) $|\mu_n-\mu|^2\leq 2M_nD_{KL}(p_n||Q) \quad \forall n$, with $M_n=\max_{x\in\chi}x$.

   This Proposition along with the known results on the monotonicity of the entropy per element of a stationary stochastic process (Cover and Thomas, 1991), support the use of the entropy as fundamental measure to use in the selection of sequences belonging to the visual distortion typical set.

2. In the construction of the sequence of interest, it is more convenient to select non overlapping local regions (for instance, blocks) as samples of I (and I_d) rather than to randomly select isolated pixels.
3. It is more convenient to extract significant information from M rather than from the couple of images I and I_d.

What was missing in (Bruni and Vitulano, 2014) is a constructive method for determining the typical set: only its existence along with some criteria and guidelines for its best search have been provided. That is why, in the sequel we will give an answer to the following question

4. How to find a sequence belonging to A_M^ε using a fast procedure.

It is worth outlining that there is a wide literature concerning *fixation points* (Frazor and Geisler, 2006; Monte et al., 2005; Raj et al., 2005), i.e. those points that allow to sinthesize and understand scene information in the preattentive phase. Several approaches for the determination of a subset of scene information mainly rely on the construction of *saliency maps* (Benabdelkader and Boulemden, 2005; Bruni et al., 2011; Wang et al., 2010). However, to the best of authors' knowledge, there are not complete theoretical formalisms that lead to a specific subset that can be extracted in a limited time (Raj et al., 2005), as the proposed approach does. In addition, unlike existing methods that provide empirical and computationally demanding strategies that lead to a specific solution (i.e. a specified walk in the scene under exam), the proposed approach proves the existence of more than one walk given I, I_d, M and ε, in agreement with the concept of *typical set* in Information Theory (Cover and Thomas, 1991).

3 THE PROPOSED MODEL

Fixation points vary from observer to observer since they depend on personal cognitive experience, the scope of the observation and image content. However, if we restrict to the class of natural images, some rules of visual system, that guide the saccadic movements in the preattentive phase, can be modelled in an easier way. The characteristics of natural scenes guided the adaptation of the visual system over time; hence, their sources are the ones with which the visual system is more familiar. In the first milliseconds of scene inspection, fixation points are not conditioned by the observer, but mainly by image features; that is why only global distortions (affecting all image pixels) will be considered in the remaining part of the paper. In fact, a local distortion would strongly orient the path of fixations, that cannot be easy predictable without additional information on the distortion kind.

The proposed method consists of the following main steps:

1. Luminance based segmentation of the image I. The output is a partition of the image in 2^L regions $R_i, \quad i = 1, \dots, 2^L$ having different characteristics. To this aim the Successive Mean Quantization Transform (SMQT) (Nilsson et al., 2005) applied to the approximation band of the wavelet expansion of the image has been employed.
2. Finite random walk on a connected and weighted graph whose nodes are the regions $R_i, \quad i = 1, \dots, 2^L$. This step provides a sequence belonging to the typical set of length K. K is automatically determined for each image using the Minimum Description Length principle.

Figure 1: From left to right top to bottom: Original image; Image affected by additive gaussian noise; pointwise SSIM map; Image affected by gaussian blur; pointwise SSIM map.

3.1 Luminance based Segmentation

The first step aims at discriminating image regions in agreement with the visibility of distortion. In fact, global distortions are not perceived in the same way in the whole image. For example, as shown in Fig. 1 random noise is more visible in flat regions while it is masked in textured regions. On the contrary, blurring is more visible in textured regions than in flat regions. Hence, the proposed method segments the image according to this visibility criterion. Specifically, the luminance value at a given fixed resolution has been selected as the visibility criterion. Luminance is one of the two measures that regulate the adaptation process in the preattentive phase. The second one is the contrast that has not been considered here for simplicity. The resolution aims at simulating early vision process, that essentially is a low pass filter whose cutoff frequency depends on the viewing distance. In order to speed up the segmentation process, the approximation band (low-pass component) at level J ($\mathcal{A}^J$) of the dyadic wavelet expansion of the image I has been computed (Mallat, 1998), since its dimension is $\frac{1}{2^{J+1}}$ of the original image size. For segmenting $\mathcal{A}^J$, the Successive Mean Quantization Transform (SMQT) has been adopted due to its simplicity and reduced computational effort. SMQT builds a binary tree using the following rule: given a set of data $\mathcal{A}^J$ and a real parameter L (number of levels), split $\mathcal{A}^J$ into two subsets,

$$\mathcal{A}_0^J = \left\{ x \in \mathcal{A}^J | \mathcal{A}^J(x) \leq \overline{\mathcal{A}^J} \right\}$$

and

$$\mathcal{A}_1^J = \left\{ x \in \mathcal{A}^J | \mathcal{A}^J(x) > \overline{\mathcal{A}^J} \right\},$$

where $\overline{\mathcal{A}^J}$ is the mean value of $\mathcal{A}^J$. $\mathcal{A}_0^J$ and $\mathcal{A}_1^J$ are the first level of the SMQT. The same procedure is recursively applied to $\mathcal{A}_0^J$ and $\mathcal{A}_1^J$ until the L^{th} level, that is composed of 2^L subsets (regions) that will be denoted with $R_1, R_2, \ldots, R_{2^L}$.

3.2 Random Walk on a Connected Graph

The fixation path is determined by suitably extracting points from these regions. To this aim, the observation process has been modeled as a Markov chain, i.e. random walk on a connected weighted graph whose nodes are the 2^L regions $R_1, R_2, \ldots, R_{2^L}$, with weights $W_{ij} \geq 0$ on the edge joining node i to node j. The graph is undirected, i.e $W_{ij} = W_{ji}$, and $W_{ij} = 0$ if there is not an edge joining the node i to the node j.

Hence, given a point randomly extracted from the region R_i, the successive point in the walk is a random point in the region R_j chosen among the nodes connected to R_i with a probability

$$P_{ij} = \frac{W_{ij}}{\sum_{i \sim k} W_{ik}} \tag{1}$$

that is proportional to the weight W_{ij}. By denoting with n_i the number of pixels in the region R_i, the weights are defined as follows

$$W_{ij} = \begin{cases} n_i & i = j \\ \frac{Z_{ij}+Z_{ji}}{2} & i \neq j \end{cases} \tag{2}$$

where $Z_{ij} = n_j \frac{\sum_{i \sim k, k \neq i} n_k}{\sum_{k=1}^{2^L} n_k}$. W_{ij} takes into account the representativeness of the region R_j in the image and also as neighbouring region of R_i. Even though a more refined definition of the weights could be used, this choice is simple but enough significant for our preliminary study.

The initial point of the walk is extracted on the basis of the stationary distribution of the process, as described in (Cover and Thomas, 1991). On the contrary, the last point of the walk is determined on the basis of the minimum descritpion length principle (Grunwald, 2004), as shown in the sequel.

3.3 MDL for Blocks Number

This principle allows the selection of a good model for approximating the data with the least complexity. It is based on the concept that good compression means good approximation, in agreement with the definition of Kolmogorov complexity. Specifically, the simpler version of MDL, namely crude-MDL, selects a model from a set of candidates $\mathcal{M}^{(1)}, \mathcal{M}^{(2)}, \ldots$ by minimizing the following cost

$$L(\mathcal{M}^{(k)}) + L(X|\mathcal{M}^{(k)}) \tag{3}$$

where $L(\mathcal{M}^{(k)})$ is the cost (in terms of bits) required for coding the model $\mathcal{M}^{(k)}$, while $L(X|\mathcal{M}^{(k)})$ is the number of bits required for coding the data X given the model. In general, the better the model the higher its cost but the smaller the approximation error. That is why the selection of the best model is a trade off between complexity and good approximation. In our case the model $\mathcal{M}^{(k)}$ is the fixation path containing the SSIM value of k points whose average gave an approximation of SSIM of the whole image. The data X are corresponding blocks in I and I_d centered at the selected pixels that are involved in SSIM computation. The cost is measured as entropy per element. More precisely, by indicating with M_1, M_2, M_k the value of SSIM computed in the first k points selected during the random walk on the graph described above, and with $(b_1, b_2, \cdots, b_k)$ the blocks used for the evaluation of SSIM, we have $L(X|\mathcal{M}^{(k)}) = \frac{H(M_1, M_2, M_k)}{k}$ and $L(X|\mathcal{M}^{(k)}) = \frac{H(b_1, b_2, \ldots, b_k) + 2log_2(k)+1}{2w^2}$ where H is the entropy, w^2 is the dimension of a block and $2log_2(k)+1$ is the cost for coding the integer k. By coding the blocks independently, $H(b_1, b_2, \cdots, b_k) = kH(b_i), \quad i = 1, 2, \ldots, k$ and by considering a compression ratio 8 : 1, eq. (4) can be rewritten as

$$K = argmin_k \frac{H(M_1, M_2, M_k)}{k} + \frac{k + 2log_2(k) + 1}{2w^2} \tag{4}$$

where K gives the length of the optimal path, i.e. the length of a sequence in the visual distortion typical set.

3.4 Algorithm

1. Compute the wavelet approximation band $\mathcal{A}_J$ at $J-th$ level of the image I
2. Apply L levels of the SMQT transform to $\mathcal{A}_J$ and extract the regions $R_1, R_2, \ldots, R_{2^L}$
3. Compute the cardinality $n_1, n_2, \ldots, n_{2^L}$ of the segmented regions and evaluate the weights of the graph as in eq. (2)
4. Extract a point from a region R_1 according to the stationary distribution of the graph as defined eq. (2)
5. Compute M_1, i.e. SSIM on a block of dimension $w \times w$ centered at the selected point and set $k = 2$
6. Extract a point in the region R_j selected according to the probability $P_{i,j}$ defined in eq. (1)

7. Compute M_k, i.e. SSIM on a block of dimension $w \times w$ centered at the selected point
8. Evaluate the argument of eq. (4) and assign its value to the variable $\mathcal{L}_k$
9. If $\mathcal{L}_k > \mathcal{L}_{k-1}$, set $K = k-1$ and $\hat{M} = \frac{\sum_{k=1}^{K} M_k}{K}$ and stop; otherwise set $k = k+1$ and go to step 6.

$\hat{M}$ is the approximation for SSIM given by the model, while K is the number of blocks used for getting it.

3.5 Model's Complexity

By denoting with $\mathbf{C}_{log}$ the cost for the calculation of the logarithm of a number, with N the image size and with $|\chi|$ the cardinality of the alphabet of SSIM, it is possible to prove that the proposed algorithm requires

$$\left(\frac{7}{3}\left(1-\frac{1}{2^{2J}}\right)+2L+1\right)N-2^L-L+1+2^{2L})+$$

$$+\left(8w^2+30+\frac{3}{2}\mathbf{C}_{log}+log_2|\chi|\right)K+(4+\mathbf{C}_{log})\frac{K^2}{2}$$

operations, i.e. multiplications, algebraic sums, divisions and comparisons, while the computation of SSIM using all image pixels requires

$$(8w^2+18)N$$

operations. Hence, by comparing the number of operations given above, it is possible to determine the maximum value for K, which gives a gain in the computation of SSIM. This value depends on the parameters of the proposed method and the image size.

4 EXPERIMENTAL RESULTS

The proposed method has been tested on several images affected by different distortion kinds and levels. In this paper we will give some results obtained from natural images extracted from TID2013 database (Ponomarenko et al., 2015) affected by global distortions like additive and multiplicative gaussian noise, high freuency noise, gaussian blurring, jpeg and jpeg 2000 compression, mean shift and contrast change. For each distortion, four levels have been considered. In all tests the following parameters have been used. The level J of the wavelet transform has been set equal to 3 and a Daubechies with 2 vanishing moments has been adopted; the levels L of the SMQT have been fixed to 3 in order to have 8 regions; the dimension of the blocks for SSIM computation of SSIM has been fixed to 17×17, since it corresponds to a visual angle equal to 0.56 degrees (Monte et al., 2005) — however smaller dimensions provide similar results; the cardinality of the alphabet for SSIM has been set equal to 200, that corresponds to a quantization step equal to 0.01. Table 1 provides the results achieved on 512×384 Ocean image (image I16 in TID2013). It is worth outlining that each run of the proposed algorithm provides a different sequence in the visual distortion typical set of the image. That is why the average value of SSIM estimations obtained by 30 runs of the algorithm has been given in Table 1. The same table includes the standard deviation of the estimation as well as the average number of blocks used for computing it and the corresponding standard deviation. As it can be observed, the estimation error increases as the distortion level increases but it does not overexceed 8% and the standard deviation is quite small. For some distortion kinds, like gaussian noise and gaussian blur this percentage is less than 5% and for distortions like mean shift and contrast change it is does not overexceed 1.2%. The average number of blocks is less than 50, it means that the number of operations required for the computation of SSIM of Ocean image is reduced of about 200 times. It is worth outilining that similar results have been obtained for the other images in the database; for some of them the average number of blocks is smaller than 50, while the average estimation error is still less than 8%. It is also worth stressing that the proposed procedure does not involve an exhaustive search of points of interest, as required by the contrast-based procedure in (Raj et al., 2005).

Figure 2 shows the segmentation used for Ocean image. As it can be observed, the segmentation is not far from the one given by the SSIM map except for the edges. It is due to the fact that the criterion used for the segmentation is based just on the luminance and then a region based segmentation has been employed. Nonethless, the optimal point selected by the MDL principle on the entropy curve corresponds to a good value of SSIM, providing acceptable estimation errors. The same figure shows the blocks belonging to the selected fixation path. As it can be observed, more blocks are selected in regions where blurring is more visible.

5 CONCLUSIONS AND FUTURE RESEARCH

This paper has presented a method for the estimation of the Strucutral SIMilarity index from a reduced number of suitably selected image pixels. It models the observation process in the preattentive phase as a random walk on a graph whose nodes are im-

Table 1: Ocean image; I16 in TID2013 database. SSIM ($\bar{M}$), estimated SSIM ($\hat{M}$) using the proposed method, mean value of the estimation error (%) over 30 runs (ε), standard deviation of the estimation error (σ_ε), mean value of the number of blocks used ($\bar{K}$), standard deviation of the number of blocks (σ_K).

distortion	level	$\bar{M}$	$\hat{M}$	ε	σ_ε	$\bar{K}$	σ_K
high frequency noise	2	0.8661	0.8672	1.1222	0.0113	48.00	2.92
	3	0.7034	0.7086	2.6788	0.0237	48.87	4.54
	4	0.4829	0.4842	4.9297	0.0275	48.70	3.51
	5	0.2722	0.2627	8.9767	0.0284	47.80	4.43
Gaussian noise	2	0.8673	0.8728	1.4447	0.0132	45.67	3.87
	3	0.7781	0.7821	2.1462	0.0217	48.50	4.55
	4	0.6614	0.6607	3.4739	0.0308	48.87	3.50
	5	0.5276	0.5272	4.3112	0.0260	49.30	3.37
Gaussian blur	2	0.9513	0.9542	0.6528	0.0069	37.37	6.15
	3	0.8805	0.8828	1.4811	0.0159	44.47	5.10
	4	0.7925	0.8045	2.9532	0.0260	46.00	6.59
	5	0.7012	0.7128	4.1474	0.0346	48.40	4.67
JPEG compression	2	0.9451	0.9448	0.5260	0.0061	41.87	5.17
	3	0.8891	0.8895	0.8761	0.0097	47.33	3.94
	4	0.7578	0.7518	2.1654	0.0198	48.33	3.56
	5	0.6320	0.6257	4.1073	0.0311	49.37	4.10
JPEG2K compression	2	0.8516	0.8553	1.9288	0.0183	46.73	3.59
	3	0.6942	0.6939	4.0191	0.0326	49.43	4.35
	4	0.5394	0.5529	6.4984	0.0395	47.73	3.55
	5	0.4799	0.4827	7.6406	0.0426	49.53	2.83
Mean shift	2	0.9951	0.9950	0.1259	0.0015	21.00	5.52
	3	0.9778	0.9779	0.2198	0.0028	34.37	6.13
	4	0.9620	0.9644	0.5506	0.0059	31.27	7.36
	5	0.8929	0.8930	1.0167	0.0111	45.53	4.57
Contrast change	2	0.9829	0.9832	0.3674	0.0042	29.23	5.70
	3	0.9713	0.9711	0.1596	0.0019	33.53	5.51
	4	0.9349	0.9392	0.8867	0.0086	40.20	5.70
	5	0.8726	0.8749	0.7490	0.0081	44.60	3.28
Multiplicative gaussian noise	2	0.8594	0.8697	2.3503	0.0220	45.80	5.25
	3	0.7730	0.7851	3.3095	0.0286	47.40	4.33
	4	0.6615	0.6627	4.0346	0.0327	48.60	3.75
	5	0.5376	0.5346	5.0814	0.0322	50.77	3.18

age regions having distinct visual characteristics and whose edges are weighted accounting for the representativeness of the region in the whole image and also in the neighborhood of proximal regions. The length of the sequence is automatically determined for each image by means of the minimum description length that selects the number of blocks able to guarantee a good tradeoff between good estimation error and reduced computational complexity. The proposed method makes some assumptions on the class of analysed images (natural images) and distortion kind (global distortion); in addition, it uses some simple criteria and fixed parameters in the segmentation step. Nonetheless, even though in its simpler form, the results are satisfying and promising. Very few blocks provide SSIM estimation with errors less than 8%; this worst case is reached in very particular cases.

Future research will be devoted to the use of more refined criteria in the segmentation process and to make adaptive and automatic the choice of the parameters involved in the segmentation step (resolution of the wavelet transform, number of regions of image partition). Furthermore, some dependency on region content will be introduced in the definition of the edge weights of the graph that is used for defining the fixation path. In fact, such an approach may also allow: *i)* to improve the design of existing QA measures, *ii)* to design novel and possibly more precise QA measure, *iii)* to build novel HVS based regularization functions, *iv)* to add some novel elements to Visual Information Theory with possible effects on the definition of new visive image coding schemes.

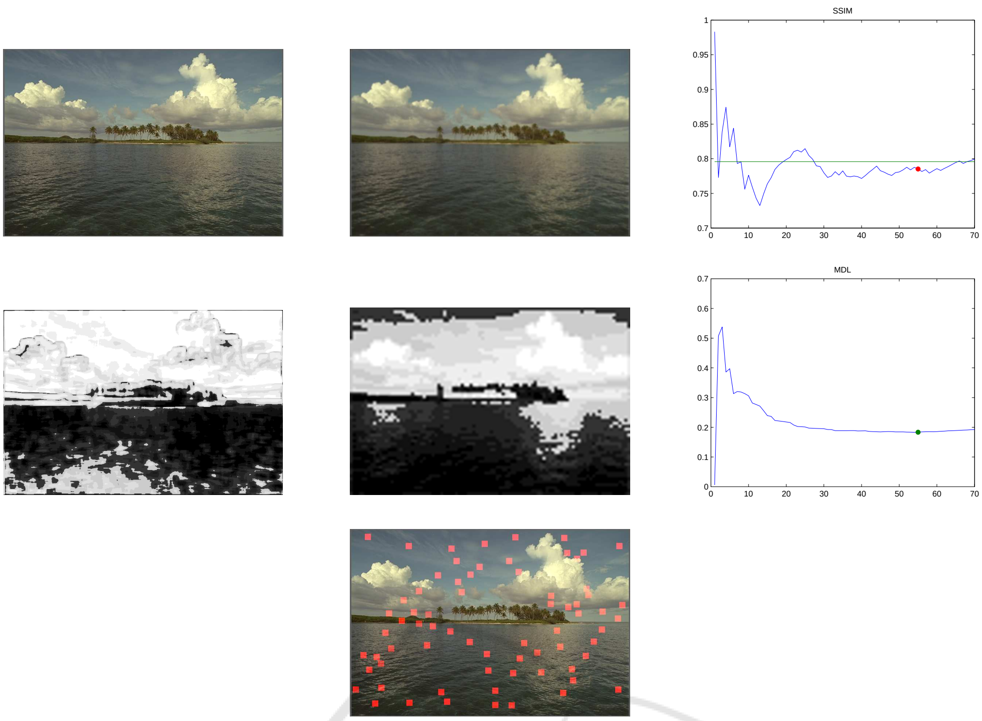

Figure 2: *First row* Original Ocean image (*left*); Blurred image (*middle*); SSIM value estimated for an increasing number of blocks (the optimal point has been marked) (*right*). *Second row* SSIM map (*left*); Segmentation provided by the SMQT (*middle*); entropy per sample used in the MDL based procedure — the optimal point has been marked (*right*). *Last row* selected image blocks.

ACKNOWLEDGEMENTS

The Authors would like to thank Simone Guarracino for the development of part of the Matlab code of the proposed method.

REFERENCES

Benabdelkader, S. and Boulemden, M. (2005). Recursive algorithm based on fuzzy 2-partition entropy for 2-level image thresholding. In *Pattern Recognition*. Elsevier.

Bruni, V., Crawford, A., Kokaram, A., and Vitulano, D. (2013a). Semi-transparent blotches removal from sepia images exploiting visibility laws. In *Signal Image and Video Processing, 7(1), 11-26*.

Bruni, V., Rossi, E., and Vitulano, D. (2012). On the equivalence between jensen shannon divergence and michelson contrast. In *IEEE Trans. on IInformation Theory, Vol. 58, No. 7*. IEEE.

Bruni, V., Rossi, E., and Vitulano, D. (2013b). Jensen-shannon divergence for visual quality assessment. In *Signal Image and Video Processing, Vol. 7, No. 3*. Springer.

Bruni, V. and Vitulano, D. (2014). A fast computation method for iqa metrics based on their typical set. In *Proc. of ICPRAM 2014*.

Bruni, V., Vitulano, D., and Ramponi, G. (2011). Image quality assessment through a subset of the image data. In *Proc. of ISPA 2011*. IEEE.

Cover, T. M. and Thomas, J. A. (1991). *Elements of Information Theory*. John Wiley sons.

Ferzli, R. and Karam, L. J. (2009). A no-reference objective image sharpness metric based on the notion of just noticeable blur (jnb). In *IEEE Trans. Image Processing, Vol. 18, No. 4*. IEEE.

Frazor, R. and Geisler, W. (2006). Local luminance and contrast in natural in natural images, 46. In *Vision Research*.

Grunwald, P. D. (2004). A tutorial introduction to the minimum description length principle. In *Advances in Minimum Description Length: Theory and Applications*. Myung Grunwald, Pitt.

Hontsch, I. and Karam, L. (2002). Adaptive image coding with perceptual distortion control. In *IEEE Trans. on Image Processing*. IEEE.

Hou, Z. and Yau, W. (2010). Visible entropy: A measure for image visibility. In *Proc. of ICPR*.

Jourlin, M. and Pinoli, J. C. (1998). A model for logarithmic image processing. In *J. Microsc., Vol. 149*.

Lee, H. and Lee, S. (2006). Visual entropy gain for wavelet image coding. In *IEEE Sig. Proc. Letters*. IEEE.

Mallat, S. (1998). *A wavelet tour of signal processing*. Academic Press.

Monte, V., Frazor, R., Bonin, V., Geisler, W., and Corandin, M. (2005). Independence of luminance and contrast in natural scenes and in the early visual system 8(12). In *Nature Neuroscience*.

Moorthy, A. and Bovik, A. (2009). Visual importance pooling for image quality assessment. In *IEEE Journal on Special Topics in Sig. Proc., 3(2)*.

Nilsson, M., Dahl, M., and Claesson, I. (2005). The successive mean quantization transform. In *Proc. of ICASSP05*.

Panetta, K. A., Wharton, E. J., and Agaian, S. S. (2008). Human visual system-based image enhancement and logarithmic contrast measure. In *IEEE Transaction on Systems, Man, and Cybernetics-Part B, Vol. 38, No. 1*. IEEE.

Park, J., Sshadrinathan, K., Lee, S., and Bovik, A. C. (2011). Spatio-temporal quality pooling accounting for transients severe impairments and egomotion. In *Proc. of ICIP 2011*. IEEE.

Ponomarenko, N., Jin, L., Ieremeiev, O., Lukin, V., Egiazarian, K., Astola, J., Vozel, B., Chehdi, K., Carli, M., Battisti, F., and Kuo, C. J. (2015). Image database tid2013. In *Image Communication, Vol. 30*. Elsevier Science Inc.

Raj, R., Geisler, W., Frazor, R., and Bovik, A. (2005). Contrast statistics for foveated visual systems: fixation selection by minimizing contrast entropy. In *J Opt Soc Am A, Vol. 20, No. 10*. Opt Image Sci Vis.

Sheikh, H. R., Bovik, A. C., and Veciana, G. D. (2005). An information fidelity criterion for image quality assessment using natural scene statistics. In *IEEE Trans. on Image Proc., Vol. 14, No. 12*. IEEE.

Wang, W., Wang, Y., Huang, Q., and Gao, W. (2010). Measuring visual saliency by site entropy rate. In *Proc. of GVPR 2010*. IEEE.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. In *IEEE Trans. on Image Proc., Vol. 13, No. 4*. IEEE.

Wang, Z. and E.P.Simoncelli (2005). Reduced-reference image quality assessment using a wavelet-domain natural image statistic model. In *Proc. of SPIE Human Vision and Electronic Imaging X, vol. 5666*. SPIE.

Wang, Z. and Li, Q. (2011). Information content weighting for perceptual image quality assessment. In *IEEE Trans. on Image Proc., Vol. 20, No. 5*. IEEE.