Weighted Shannon Entropy's Additivity for Image Segmentation

C. D. Gallão, H. Erdmann, and P. S. Rodrigues
Signal Processing Group
Centro Universitário da FEI
São Bernardo do Campo, São Paulo, Brazil
celsogallao@hotmail.com, horste@fei.edu.br, psergio@fei.edu.br

***Abstract*—Thresholding based image segmentation using classical Shannon entropy does not find an optimal separation on images whose luminance probability distribution can be approximated by bimodal gaussian functions. In this work we demonstrate that an appropriate pondering of Shannon's entropy additivity property allows us to achieve a segmentation with as good as the one observed when the optimal threshold is analytically calculated over these Gaussians. Furthermore, we show that the quality of the results remains the same even under different probability distribution characteristics.**

I. Introduction

Image segmentation is a task with applications in several areas related to digital image processing. It can be done by estimating the number of thresholds used to partition an image into regions of interest. The most simple thresholding technique is to split the image in two regions, binarizing the search space.

In 1948, Shannon [1], deduced the Equation (1), joining statistical mechanics concepts created by Boltzmann and Gibbs [2], which later became the Boltzmann-Gibbs-Shannon entropy, or BGS entropy given by

$$S = -\sum p_i \log p_i. \tag{1}$$

In 1981, T. Pun [3] proposed an image thresholding method based on Shannon's entropy additive property [1]. Formally, this property can be described as follows: The probability distribution of the n gray levels is given by $P = \{p_1; p_2; ...; p_n\}$. The method proposed in [3] splits P in two normalized distributions, one being the foreground ($P_P = \frac{p_1}{p_P}, \frac{p_2}{p_P}, ..., \frac{p_t}{p_P}$), and the other being the background ($P_Q = \frac{p_{t+1}}{p_Q}, \frac{p_{t+2}}{p_Q}, ..., \frac{p_n}{p_Q}$), where $p_P = \sum_{i=1}^{t} p_i$, $p_Q = \sum_{i=t+1}^{n} p_i$, and t is a value that divide P into P_P and P_Q distributions. After applying the Equation (1), the optimal threshold calculated through the BGS entropy's additive property is given by:

$$S(P,Q) = S(P) + S(Q), \tag{2}$$

where $S(P)$ and $S(Q)$ are the entropy values obtained by Equation (1).

So, the t value which best balances the information quantity between the two regions is the same that maximizes the Equation (2), given by:

$$T_S = max[\ S(P,Q)\ ]. \tag{3}$$

This technique became frequently used in several practical applications because it's simple and easy to implement and has been largely studied until the present date. Several works using algorithms inspired by this idea where applied to various areas and domains with promising results. These works indicates the efficiency of entropy thresholding for image segmentation and are described below.

Kapur et al. [4] maximized the upper threshold of the maximum entropy to obtain the optimal threshold, and Abutaleb [5] improved the method using bidimensional entropies. Furthermore, Li and Lee [6] and Pal [7] used the direct Kullback-Leibler divergence to define the optimal threshold. And some years before, Sahoo et al. [8] used the *Reiny*-entropy seeking the same objective. More details about these approaches can be found in [9], which presents a review of entropy-based methods for image segmentation.

Despite its use in several physics problems for over a century, it is known that this old formalism is only an approximation when the subjects of analysis are non-extensive physical systems. In example, such phenomena can be observed in turbulence, galaxies expansions and social networks interactions [10].

The BGS entropy describes extensive systems well. But the same cannot be said about non-extensive systems. To describe these systems, Constantino Tsallis proposed in 1988 the so-called *q*-entropy [11], a new kind of entropy that is considered as a generalization of Shannon entropy through the inclusion of a real parameter q, called "non-extensive parameter". Considering the restrictions of Shannon entropy, Albuquerque et al. [12] proposed an image segmentation method based on Tsallis non-extensive entropy [13]. The work of Albuquerque [12] showed promising results and a vast literature demonstrating the performance of this method for image segmentation.

By observing Equations (1)-(3), one can remark that the automatically obtained threshold calculated with BGS entropy is inflexible since it results in only one threshold value for separating the regions. Thus, a measure of efficiency to evaluate this method is hard to find because the quality of the

result is subjective and dependent on cognitive factors that are outside the scope of this paper. An early study of the amount of thresholds required to best separate the objects of the image can be found at [14]

Thus, we propose in this work an alternative to the BGS entropy in order to overcome its inflexibility. We added weights on the additivity by including a new parameter $0 \leq \alpha \leq 1$. This allows us to change the balance between the background and the foreground of the image. This will be formally presented in the next section. In this paper, we may address it as weighted Shannon entropy (WSE).

To demonstrate the effectiveness of weighting on the additivity of BGS entropy for image thresholding, we performed experiments with artificial gaussian distributions so we could control their parameters and study the new method's behavior under a controlled environment.

II. Weighted Shannon Entropy (S_w)

In this section, we present the formalism of the proposed methodology to redefine the image thresholding with BGS entropy as described below.

Setting the P_P distribution weighted by an α value and the P_Q distribution weighted by $(1-\alpha)$, where $0 \leq \alpha \leq 1$ and $n = [1, ..., 254]$, results in:

$$S_w(P) = -\sum_{i=1}^{n} \alpha p_i \log \alpha p_i, \tag{4}$$

$$S_w(Q) = -\sum_{j=1}^{255-n} (1-\alpha) q_j \log(1-\alpha) q_j, \tag{5}$$

Thus, we redefine Shannon's additivity property as an Weighted Shannon Pseudo-Additivity:

$$S_w(P,Q) = 2\ [\ \alpha S_w(P)\ +\ (1-\alpha) S_w(Q)\ ]. \tag{6}$$

The Equation (6) can be reduced back to Equation (2) when $\alpha = 0.5$. Henceforth , we will call the additivity property given by Equation (6) as Weighted-Pseudo-Additivity (WPA).

Similarly to the work of T. Pun [3] the threshold chosen for the image segmentation will be the one with highest WPA value given by:

$$T_{SW} = max[\ S_w(P,Q)\ ]. \tag{7}$$

First, we consider the two subsystems with equal weights ($\alpha = 0.5$) for the WSE calculation on the P distribution using Equation (4). Thus we have:

$$-\sum_{i=1}^{n} p_i \log p_i = -\sum_{i=1}^{n} 2\alpha p_i \log 2\alpha p_i$$

Developing the right side of the equation, we can conclude that for distribution P, the WSE $S_w(P)$ is equal to the BGS entropy when $\alpha = 0.5$:

$$-\sum_{i=1}^{n} p_i \log p_i = -2\sum_{i=1}^{n} (\alpha p_i \log \alpha p_i) - 2\alpha \tag{8}$$

This can be represented by $S(P) = 2S_w(P) - 2\alpha$, and results in the variation of the WSE distribution P, given by:

$$S(P) = 2[\ S_w(P) - \alpha\ ] \tag{9}$$

Considering $\alpha = 0.5$ in the calculation of WSE for the Q distribution through the Equation (5), results in:

$$-\sum_{j=1}^{255-n} q_j \log q_j = -\sum_{j=1}^{255-n} 2(1-\alpha) q_j \log 2(1-\alpha) q_j$$

Developing the right side of the equation, we can conclude that for distribution Q, the WSE $S_w(Q)$ is equal to the BGS entropy when $\alpha = 0.5$:

$$-\sum_{j=1}^{255-n} q_j \log q_j = -2\sum_{j=1}^{255-n} [\ (1-\alpha) q_j \log(1-\alpha) q_j\ ] - 2(1-\alpha) \tag{10}$$

This can be represented by $S(Q) = 2S_w(Q) - 2(1-\alpha)$ and results in the variation of the WSE of the Q distribuction given by:

$$S(Q) = 2[\ S_w(Q) - (1-\alpha)\ ] \tag{11}$$

Finally, the WPA Equation (6) analytically suggests that one can be consider the two subsystems with different weights, even if they are dependent on one another. This makes it possible to extract information with low *a priori* probability just by balancing between the subsystems changing the value of α. The same cannot be done with the BGS entropy's addictive property because of the method's assumption that both subsystems have equal weights.

III. Optimal Threshold Calculation (T_{opt})

To evaluate our proposed method, we needed to establish an efficient comparison between the thresholds obtained with the WSE and BGS entropy. First, we calculate the optimal threshold (T_{opt}) and define it as the reference threshold that optimally separates the regions of interest of the image.

In order to guarantee complete control over our tests, and also to address the subjectiveness of image segmentation, we generated syntectic data which simulates a histogram of normalized luminance similar to the ones obtained from digital images. These generated histograms are composed of two gaussian functions, that together form a unique bimodal gaussian function.

Formally, we define the generated functions as $G_1$and G_2, discretized in the range $i = [0, 1, .., 255]$. We also define c_1 and c_2 as the corresponding amplitudes of G_1 and G_2 with initial values $c_1 = c_2 = 1$. We also define σ_1 and σ_2 as their standard deviations, with initial values of $\sigma_1 = \sigma_2 = 30$. Finally, we define $\mu_1 = \frac{k}{3}\ 255$ and $\mu_2 = \frac{3-k}{3}\ 255$ as the means of G_1 and G_2 where $k = [0, .., 1.5]$. So, the normalized mean of G_1 and G_2 is given by:

$$G = \frac{G_1(i) + G_2(i)}{\sum (G_1 + G_2)}. \tag{12}$$

Thus, we consider that the gaussian G represents the system to be analyzed. Since G is completely synthetic, we are able to change any of the three parameters c, σ and μ for each gaussian G_1 and G_2 consisting of 6 variations of control parameters. We also consider the analytical optimal threshold (T_{opt}) to be the intersection between the subsystems $G1$ and $G2$.

As stated before, the optimum analytical threshold is given at the intersection of G_1 and G_2. When both gaussians have the same amplitude ($c_1 = c_2 = c$) and standard deviation ($\sigma_1 = \sigma_2 = \sigma$) but have different means (μ_1 and μ_2), the equality of $G_1 = G_2$ is given by:

$$c\frac{\exp^{-\frac{(i-\mu_1)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} = c\frac{\exp^{-\frac{(i-\mu_2)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}, \tag{13}$$

From which we obtain the optimum analytical threshold:

$$i = T_{opt}^{means} = \frac{(\mu_1 + \mu_2)}{2} \tag{14}$$

When the amplitudes of G_1 and G_2 are different but with the same standard deviations ($\sigma_1 = \sigma_2 = \sigma$), the equality of $G_1 = G_2$ is given by:

$$c_1\frac{\exp^{-\frac{(i-\mu_1)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} = c_2\frac{\exp^{-\frac{(i-\mu_2)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}, \tag{15}$$

Which results in the following expression for the optimum analytical threshold:

$$i = T_{opt}^{amplitude} = \frac{ln(\frac{c_1}{c_2})2\sigma^2 \ - \ \mu_1^2 \ + \ \mu_2^2}{2(\mu_2 \ - \ \mu_1)} \tag{16}$$

Finally, when the two subsystems have different standard deviations, but equal amplitudes ($c_1 = c_2 = c$), the equality of $G_1 = G_2$ is given by:

$$c\frac{\exp^{-\frac{(i-\mu_1)^2}{2\sigma_1^2}}}{\sqrt{2\pi}\sigma_1} = c\frac{\exp^{-\frac{(i-\mu_2)^2}{2\sigma_2^2}}}{\sqrt{2\pi}\sigma_2} \tag{17}$$

The Equation (17) has two solutions, called i_1 and i_2. After calculating the roots, the choice of the best threshold obtained with them is done as in Algorithm 1, where T_1 and T_2 are the maximum values of the gaussian G_1 and G_2.

IV. The Proposed Methodology

According to the previous section, the experiments with synthetic gaussian functions allows us to control the parameters σ_1, σ_2, μ_1, μ_2, c_1 and c_2, which govern the probability distributions. Thus allowing us to study the behavior of the proposed method against the effects of the variation of such parameters.

Algorithm 1 Algorithm to select the threshold T_{opt}^{stdev}, between gaussian with different standard deviations.

[H]

Input: i_1 and i_2, σ_1, σ_2, T_1, T_2.
Output: $T_{opt}^{stdev} \in \{i_1, i_2\}$.
if $\sigma_1 < \sigma_2$ **then**
  if $i_1 \geq T_1$ **then**
    $T_{opt}^{stdev} \leftarrow i_1$
  else
    $T_{opt}^{stdev} \leftarrow i_2$
  end if
else
  if $i_1 \leq T_2$ **then**
    $T_{opt}^{stdev} \leftarrow i_1$
  else
    $T_{opt}^{stdev} \leftarrow i_2$
  end if
end if
return T_{opt}^{stdev}

For each discrete value of t, where $1 \leq t \leq 255$, we calculate:

- The optimal analytical threshold, described by T_{opt}, for each parameter variation of the gaussian distributions described in section III.
- The additive property of the BGS entropy, described by $S(P, Q)$, for each parameter variation of the gaussian distributions according to Equation (3), to find the BGS threshold (T_S).
- The WPA, described by $S_w(P, Q)$, for each parameter variation of the gaussian distributions, starting with $\alpha = 0.5$ to find the highest WPA, like weigthed Shannon's threshold (T_{SW}) according to Equation (7). Next, the value of α is varied in the search space of $\alpha = [0, ..., 1]$ to find which value makes T_{SW} closest to T_{opt} and also closes to $\alpha = 0.5$.

After the calculation of T_{opt}, $S(P, Q)$ and $S_w(P, Q)$, we compare the thresholds T_S and T_{SW} with T_{opt}. In this paper, we considered as best results the ones on which T_S and T_{SW} are the closest to the analytical optimal threshold (T_{opt}).

To calculate the amount of iteration to find the value of α which makes $T_{SW} = T_{opt}$, we defined the search space as $SS_\alpha = (\alpha_{max} - \alpha_{min})10^\lambda$, where α_{max} is the maximum possible value for α, α_{min} is the minimal and λ is the number of decimal places. As $\alpha = [0, 1]$, then the amount of iteration performed to find the best α is $SS_\alpha = 10^\lambda$.

V. Experiments

In our experiments, we use the methodology previously presented in Section IV. This section shows the results of amplitude, standard deviation and mean variations of the subsystems G_1 and G_2.

A. Amplitude Variation

Altogether, 60 experiments were performed varying the c_2 amplitude of G_2. For each variation, the value of T_{opt} has been calculated through Equation (16). For the WPA calculations, we used $\lambda = 3$, which gives us $SS_\alpha = 10^3$ tests for each amplitude variation. The result from one of these experiments is shown in Fig. 1. We notice that T_S deviates from the intersection between G_1 and G_2 towards the subsystem with the highest amplitude, whereas T_{SW} remains at the intersection, indicating that the WSE can find the optimal analytical threshold. In such case, the calculated α was 0.521. The results shown in Fig. 1 were: $T_{opt} = 134,8392$ discretized to $T_{opt} = 135$, $T_S = 128$, and $T_{SW} = 135$ with $\alpha = 0.521$.

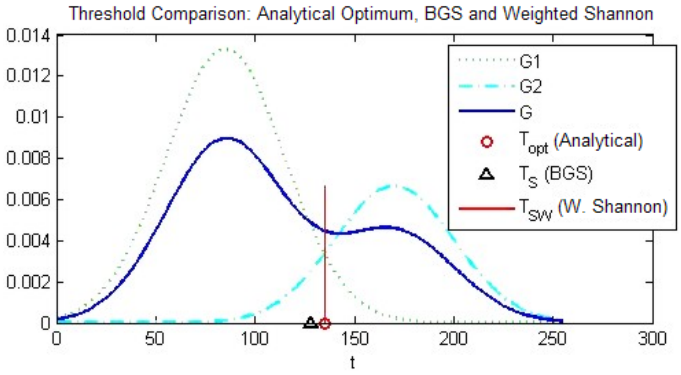

Fig. 1: Example of analytical optimal thresholds of BGS and WSE obtained when varying only the amplitude of the gaussian.

The graphs in Fig. 2 show the results from a set with 20 of the 60 experiments conducted for amplitude variation of G_2 with $c_2 = [0.1, 0.2, ..., 2]$. One can note in Fig. 2b that the classical BGS entropy was not able to find the analytical optimum threshold in any case ($T_S \neq T_{opt}$), even though the difference between the two thresholds tend to diminish when $c_2 = 1$. However, it is possible to observe that there was always an α value that allowed the WSE to find $T_{SW} = T_{opt}$ in all tests.

The same behavior was observed with the other 40 experiments: in all cases there was always an α value which guarantee that $T_{keptSW} = T_{opt}$. The classical BGS entropy also did not find T_{opt} in any case. These results demonstrate the benefits of the WSE's flexibility over the BGS entropy.

B. Standard Deviation Variation

A total of 60 experiments varying only the standard deviation σ_2 of G_2 were performed. The value of T_{opt} was calculated for each variation using Algorithm 1. As in the previous experiment, for the WPA calculations, we used $\lambda = 3$, which gives us $SS_\alpha = 10^3$ tests for each standard deviation. The result of one of these tests is shown in Fig. 3. Observing this graph, we note that T_S is displaced from the intersection between G_1 and G_2 towards the subsystem with the highest standard deviation. Meanwhile, parameter $\alpha = 0.552$ enabled T_{SW} to find the intersection of the subsystems. The results

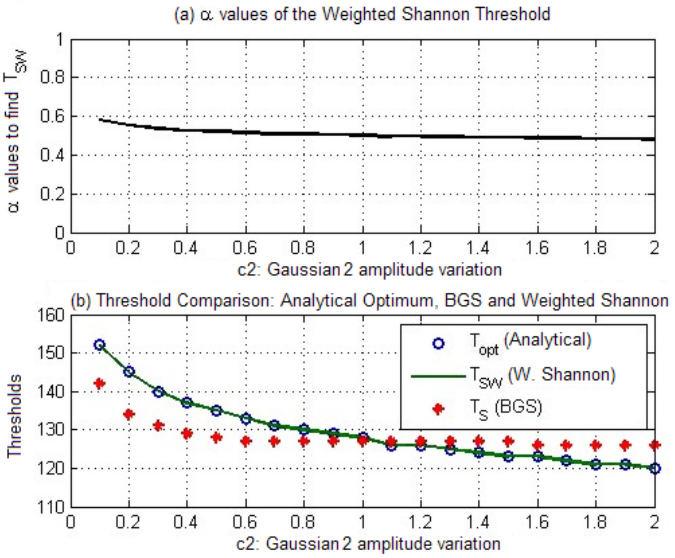

Fig. 2: Experiments with amplitude variation of G_2. (a) Values of α found on each variation such that $T_{SW} = T_{opt}$. (b) Values of T_{opt}, T_{SW} e T_S, found on each variation.

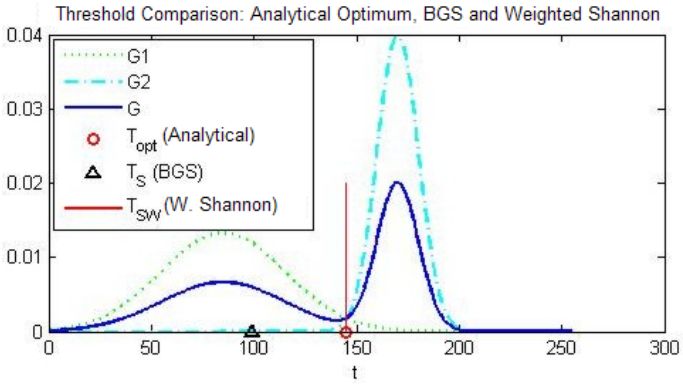

Fig. 3: Example of analytical optimal thresholds of BGS and WSE obtained when varying only the standard deviation of the gaussian.

shown in Fig. 3 were: $T_{opt} = 134.8392$ discretized to $T_{opt} = 135$, $T_S = 128$, and $T_{SW} = 135$ with $\alpha = 0.552$.

The graphs in Fig. 4 show the results from a set with 20 of the 60 experiments conducted for standard deviation of G_2 $\sigma_2 = [10, 20, ..., 200]$. From Fig. 4b we can observe that the classical BGS entropy still did not find the analytical optimum threshold in any case ($T_S \neq T_{opt}$), even though the two thresholds tend to approach when $\sigma_2 = 30$ and when $160 \leq \sigma_2 \leq 180$. Once again, it is possible to observe that there was always an α value which allowed the WSE to find $T_{SW} = T_{opt}$ in all tests and can be verified by the overlap of curves for T_{opt} e T_{SW}. The remaining 40 experiments with standard deviation variation presented the same behavior as the first 20 discussed.

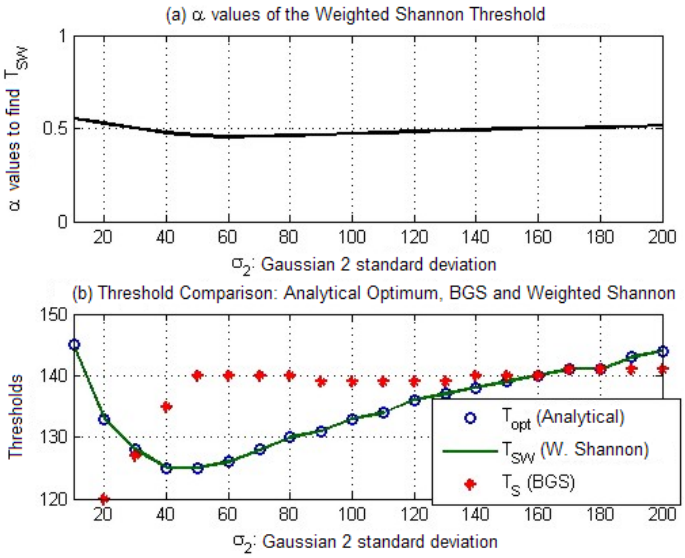

Fig. 4: Experiments with standard deviation variation of G_2. (a) Values of α found on each variation such that $T_{SW} = T_{opt}$. (b) Values of T_{opt}, T_{SW} e T_S, found on each variation.

C. Variation of the Means

A total of 31 tests were performed varying only $k = [0, 0.05, ..., 1.5]$, which is used to calculate the means of G_1 and G_2, μ_1 and μ_2. The value of T_{opt} was calculated for each variation through Equation (14). Unlike the previous experiment, for the WPA calculations, we used $\lambda = 6$ and $\alpha = [0.49, ..., 0.51]$, which gives us $SS_\alpha = 0.02^6 = 20.000$ tests for each k variation. The result of one of these experiments is shown in Fig. 5, where one can observe that T_S was once again different from $T_{opt} = 135$ while T_{SW} had the same value as the analytical optimum threshold. The results shown in Fig. 5 were: $T_{opt} = 127$, $T_S = 154$, e $T_{SW} = 127$ with $\alpha = 0.49807$.

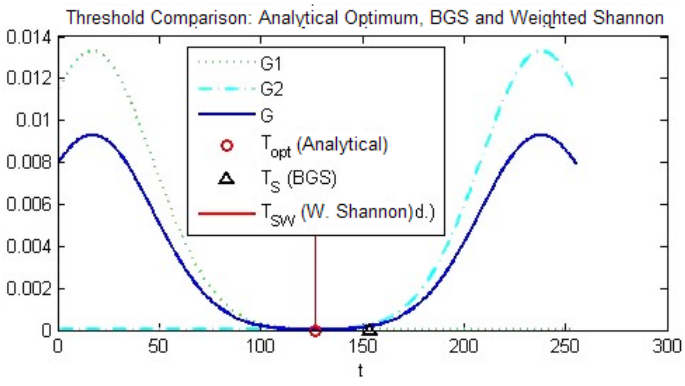

Fig. 5: Example of analytical optimal thresholds of BGS and WSE obtained when varying only the mean of the gaussians.

Looking at the graphics from Fig. 6, it is possible to observe the results of the 31 tests conducted through the variation of value $k = [0, 0.05, ..., 1.5]$ that generated the means $\mu_1 = \frac{0.2}{3}255$ and $\mu_2 = \frac{2,8}{3}255$. The other parameters of the gaussian functions where kept with the default values. This time, the thresholds obtained with the BGS entropy (T_S) converged to the analytical optimal threshold (T_{opt}) when $k > 0.5$. As in the previous experiments, the adjustments to the parameter α were effective and maintained $T_{SW} = T_{opt}$ in all tests.

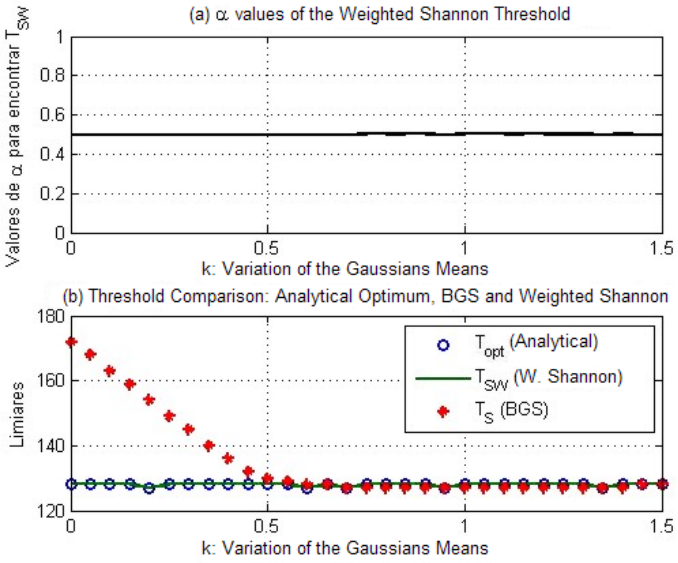

Fig. 6: Experiments varying the means μ_1 and μ_2. (a) Values of α found on each variation such that $T_{SW} = T_{opt}$. (b) Values of T_{opt}, T_{SW} and T_S, found on each variation.

Finally, Fig.7 shows examples of the WSE-based thresholding applied to digital images, with variations of the parameter $\alpha = [0.1, 0.2, ..., 0.9]$. As previously stated, for $\alpha = 0.5$, the result is the same as the one obtained by thresholding with BGS entropy. By observing Fig.7 in detail, one can notice that when the value of α is lower than 0.5, the region of interest of the image, represented by the white color, is given a greater weight than the background which is represented by the black color. But when $\alpha > 0.5$, then the background receives a greater prominence. Although the result of a segmentation is subjective, it can be said that, visually, the segmentation obtained with $\alpha = 0.4$ is the best for both images.

VI. Conclusions

In this paper presented an extension of the Shannon entropy addictive property. Our proposal allows us to improve its result for image segmentation. To improve control of the experiments, we use syntectical gaussians to simulate image histograms. These experiments where executed varying the amplitude, standard deviation and mean of the gaussian functions.

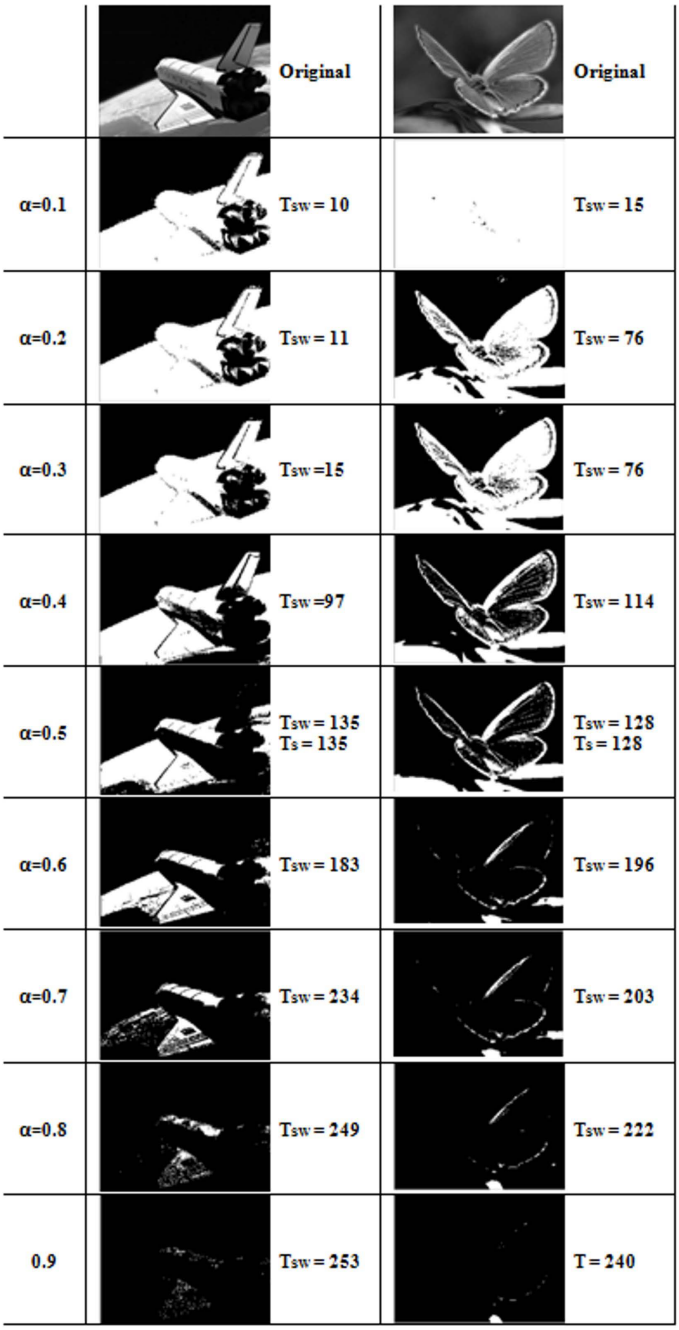

Fig. 7: Examples of the weighted Shannon entropy based thresholding for several values of α.

The results indicate that for the weighted Shannon entropy, there is always an α value which makes the calculated threshold T_{SW} equal to the optimal threshold T_{opt}. This suggests a correlation between the α value and some distribution parameter dependent value.

The variations on the gaussian function's parameters makes the BGS entropy calculated threshold (T_S) different from the optimal threshold (T_{opt}) in most cases. This occurs because in the BGS entropy there are no parameters that can be adjusted in order to make $T_S = T_{opt}$.

The experiments indicate that the values of α showed little variation and tend to stabilize in the range of $\alpha = [0.4, ..., 0.6]$. Also, the results show the effectiveness of the proposed method, which allows us to adjust the α parameter that ponders the both subsystems. This opens the possibility to analyze information represented by small probabilities that normally would be ignored by the BGS entropy. This means that by using the WSE method, one can assign different weights to enhance the foreground or the background of the image.

REFERENCES

[1] C. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, 1948.

[2] E. P. Borges, "Irreversibilidade, desordem e incerteza: Três visões da generalização do conceito de entropia," *Revista Brasileira de Ensino de Física*, vol. 21, no. 4, p. 453, Dezembro 1999.

[3] T. Pun, "Entropic thresholding: A new approach," *Computer Graphics and Image Processing*, vol. 16, no. 3, pp. 210–239, July 1981.

[4] J. N. Kapur, P. K. Sahoo, and A. K. C. Wong, "A new method for gray-level picture thresholding using the entropy of the histogram," *Computer Graphics Image Process*, vol. 29, pp. 273–285, 1985.

[5] A. S. Abutaleb, "A new method for gray-level picture thresholding using the entropy of the histogram," *Comput. Graphics Image Process*, vol. 47, pp. 22–32, 1989.

[6] C. H. Li and C. K. Lee, "Minimum cross entropy thresholding," *Pattern Recognition*, vol. 26, pp. 617–625, 1993.

[7] N. R. Pal, "On minimum cross entropy thresholding," *Pattern Recognition*, vol. 26, pp. 575–580, 1996.

[8] P. Sahoo, S. Soltani, A. Wong, and Y. Chen, "A survay of thresholding techniques," *Computer Vision Graphics Image Processing*, vol. 41, no. 1, pp. 233–260, 1988.

[9] C.-I. Chang, Y. Du, J. Wang, S.-M. Guo, and P. Thouin, "Survey and comparative analysis of entropy and relative entropy thresholding techniques," *The Institution of Engineering and Technology*, 2006.

[10] C. Tsallis, *Nonextensive Statistical Mechanics and its Applications*, ser. Lecture Notes in Physics, S. Abe and Y. O. (Eds.), Eds. Berlin, Germany: Springer, 2001, vol. 560. [Online]. Available: http://tsallis.cat.cbpf.br/biblio.htm

[11] C. TSALLIS, "Possible generalization of boltzmann-gibbs statistics," *Journal of Statistical Physics*, vol. 52, pp. 479–487, 1988. [Online]. Available: http://dx.doi.org/10.1007/BF01016429

[12] M. P. Albuquerque, M. P. Albuquerque, I. Esquef, and A. Mello, "Image thresholding using tsallis entropy," *Pattern Recognition Letters, Elsevier B.V.*, 2004.

[13] C. Tsallis, "Nonextensive statistics: Theoretical, experimental and computational evidences and connections," *Brazilian Journal of Physics*, vol. 29, pp. 1 – 35, 03 1999.

[14] H. Erdmann, L. Lopes, G. Wachs-Lopes, M. Ribeiro, and P. S. Rodrigues, "A study of a firefly meta-heuristics for multithreshold image segmantation," *Proceedings of Computational Vision and Medical Image Processing IV*, pp. 211–217, 2013.