A RBFN Perceptive Model for Image Thresholding

Fabrício Martins Lopes[1], Luís Augusto Consularo[2]
[1]*CEFET-PR - Centro Federal de Educação Tecnológica do Paraná*
Av. Alberto Carazzai, 1640, 86300-000, Cornélio Procópio, PR, Brasil.
fabricio@cp.cefetpr.br
[2]*UNIMEP - Universidade Metodista de Piracicaba*
Rodovia do Açúcar, Km 156, 13400-911, Piracicaba, SP, Brasil.
laconsul@unimep.br

Abstract

The digital image segmentation challenge has demanded the development of a plethora of methods and approaches. A quite simple approach, the thresholding, has still been intensively applied mainly for real-time vision applications. However, the threshold criteria often depend on entropic or statistical image features. This work searches a relationship between these features and subjective human threshold decisions. Then, an image thresholding model based on these subjective decisions and global statistical features was developed by training a Radial Basis Functions Network (RBFN).

This work also compares the automatic thresholding methods to the human responses. Furthermore, the RBFN-modeled answers were compared to the automatic thresholding. The results show that entropic-based method was closer to RBFN-modeled thresholding than variance-based method. It was also found that another automatic method which combines global and local criteria presented higher correlation with human responses.

1. Introduction

Although simple, the image segmentation thresholding approach is still intensively explored to distinct background from objects in digital images. A number of criteria have been proposed to determine the threshold gray level performance, but in general, they do not take into account the human vision perception [1].

This work presents an innovative approach to threshold gray level images based on perceptive responses. These responses were given by human subjects who answered a psychophysical test. The threshold human decision was modeled by RBFN (Radial Basis Functions Network). The RBFN approach allowed us to set smoothing parameters to change the correlation between human and modeled responses of new sample tests. Hence, the correlation coefficient between human responses and, for instance, automatic entropic-based thresholding responses, revealed a quantitative similarity between human and automatic responses.

The next section (Sec. 2) explains the subjective thresholding modeling technique proposed here. Then, in the section 3, the psychophysical test is described to show how the collected data is related to the modeling process. In the section 4, the RBFN modeling technique is presented, and, in section 5, the validation process is reported. The section 6 compares the proposed subjective model method and three others representative automatic thresholding methods.

2. The Perceptive Thresholding Modeling

Automatic thresholding methods often require one or a couple of global or local criteria to select the best level threshold of a digital gray level image. These criteria are good choices when to classify pixels as belonging to background or object is fairly easy, i.e. images that has bimodal histograms. However, it is not quite easy to perform this classification when the image was acquired in environments with irregular illumination, or when the images whose interested object is not present in the main focus. Furthermore, some applications claim some human interference to define “the right” decisions.

The thresholding problem is to choice a level, which could best classify pixels as belonging to background or object. This work proposes to train this choice based on human decisions. Then, with a slide bar, people pointed the gray level which could better separate the background pixels from the object. These decisions were collected in a two columns table holding the chosen gray level at first column and a global image feature such as entropy, gray-level standard deviation, or other feature, at the second column.

A space was spanned by feature and gray level where in each subject answer was a sample. These answers were interpolated with Radial Basis Functions (RBF) whose responses were evaluated by RMS errors [2]. In fact, some images formed a training set, and others a test set, according to the validation method (see Sec. 5). Since the RBF generalization is constrained to the training set, it was performed a regression by searching the centers for each basis functions in the RBFN (RBF Network). This approach detaches the features from the model, so it is possible to remove or include new features just applying a transformation to the model. Hence, blend of features could be selected, tested and validated.

The automatic thresholding methods [3, 4, 5, 1, 6, 7], in general, apply statistical properties of pixels gray-level distribution such as variance, entropy, average, median or even quartiles as criterion to decide the best level to threshold. Therefore, a thresholding model based on these features could provide some cue about what are the human vision criteria to threshold a gray level image.

3. Data source: psychophysical experiment

The human vision perception modeling demands human choices or opinions. Hence, it was developed a psychophysical experimentation where gray level images (the stimulus) were presented to subjects who were asked to choice a threshold level (dependent variable). Before, they were instructed to choose the black-and-white image that showed the better distinction. It was allowed to the subjects to slide a bar among the gray levels to see, in real-time, the thresholded image for the selected intensity.

In the experiment each one of the 137 voluntary subjects (undergraduate students) were exposed to 12 distinct images, randomly selected from 110 images (Table 1). Each group of 12 images was also randomly combined from the 110 before they would be shown to the subjects. Although some subjects had assigned different threshold values to identical images, the difference among their answers were not more than 10 pixels.

In the homogeneous group of images, the experiments were carried out with only one subject, applying the same methodology.

The collected data (features and human threshold) were stored in a database table so that the model could be evaluated. Afterward, this table was joined to the automatic thresholding results for the same image. These results will be compared later in the Sec. 6.

This psychophysical test was implemented as a two-tier client-server application, i.e. it has two parts: the Java applet (the client) and the servlet (the server). This design provided a flexible version control, and enabled the experimentation to reach, at same time, multiple and sparse audiences. The Fig. 1 shows a snapshot for the client.

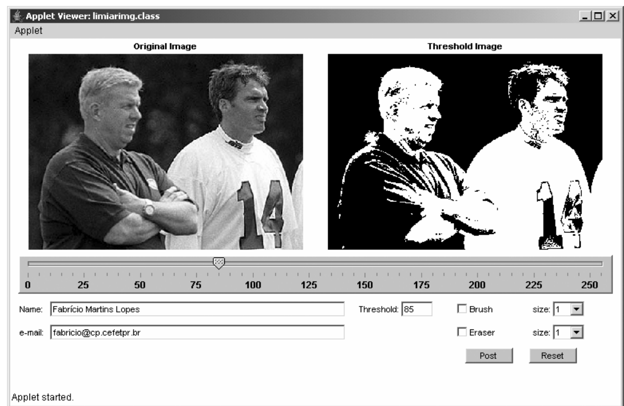


Figure 1 - A snapshot for the psychophysical experimentation client part.

There were two groups of images: (1) a heterogeneous group, whose image content is photographic motivations such as cars, houses, people and landscapes; (2) a homogeneous group, only with face images.

Table 1 - Number of images and ranges for entropy and standard deviation of heterogeneous (Het.) and homogeneous (Hom.) groups of images.

Group	*#Img*	*Entropy*	*Std Dev*
Het	110	5.84 to 7.8	23.38 to 92.27
Hom	59	6.13 to 7.37	50.32 to 98.81

These images were picked out to try spans the gray level entropy and the standard deviation theoretical ranges: [0:8] for the entropy and [0:255] for the standard deviation, as well as the other statistical parameters. However, real images hardly would span all these ranges. So, the Table 1 shows the number of images and the ranges for entropy and standard deviation used in this work.

4. The RBFN to train global thresholds

The RBFs have been intensively applied to implement regressive models, to interpolate surfaces or volumes in computer graphics applications, to smooth signals, just to mention a few. This work applies the RBF as a regressive modeling to generalize human responses to threshold gray level choices. Two steps are required: (a) the training and (b) the test. At the training step, some selected points, i.e. some pairs (features, threshold) approximate a function (1) whose components are Radial Basis Functions (RBFs) [8]. In fact, the approximation result is a linear system solution for (2).

$$W = \Phi(S) \cdot F(S) \tag{1}$$

The RBF ϕ is evaluated for distances ||.|| between each feature point and each function center $\bar{\mu}_i$. The center choice depends on the learning strategy. This work adopted a solution suggested in [9], where the centers were defined by k-means algorithm.

$$\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_k \\ b \end{bmatrix} = \begin{bmatrix} \phi_1(\|\bar{s}_1 - \bar{\mu}_1\|) & \phi_1(\|\bar{s}_2 - \bar{\mu}_1\|) & \cdots & \phi_1(\|\bar{s}_n - \bar{\mu}_1\|) \\ \phi_2(\|\bar{s}_1 - \bar{\mu}_2\|) & \phi_2(\|\bar{s}_2 - \bar{\mu}_2\|) & \cdots & \phi_2(\|\bar{s}_n - \bar{\mu}_2\|) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_k(\|\bar{s}_1 - \bar{\mu}_k\|) & \phi_k(\|\bar{s}_2 - \bar{\mu}_k\|) & \cdots & \phi_k(\|\bar{s}_n - \bar{\mu}_k\|) \\ 1 & 1 & \cdots & 1 \end{bmatrix}^{+}_{k+1 \times n} \times \begin{bmatrix} f(\bar{s}_1) \\ f(\bar{s}_2) \\ \vdots \\ f(\bar{s}_n) \end{bmatrix} \tag{2}$$

The W vector elements are weights for each k RBF centered in $\bar{\mu}_i$. The S vectors hold the features for each gray level image. The S dimension, n, is the number of global features to generalize the threshold human decision, denoted here as F. Each gray level image was thresholded by several subjects. This F is actually the average among the thresholds assigned by subjects when they were doing the psychophysical experiment.

The matrix Φ is the interpolation matrix, where each element is the RBF function value of the Euclidean norm ||.|| of the difference vector between the each pair of input sample and function center.

The vector W, i.e. the model, can now generalize a response for new images, for instance, a set of test images (3).

$$F(X) = \Phi(X) \cdot W \tag{3}$$

Now, the same trained features must be evaluated for test images. The X vector holds these test features, and similarly to the training step, the RBF function must be evaluated for Euclidean norm of difference vector between each pair of test sample and function center (4).

$$\begin{bmatrix} f(\bar{x}_1) \\ f(\bar{x}_2) \\ \vdots \\ f(\bar{x}_m) \end{bmatrix} = \begin{bmatrix} \phi_1(\|\bar{x}_1 - \bar{\mu}_1\|) & \phi_2(\|\bar{x}_1 - \bar{\mu}_2\|) & \cdots & \phi_k(\|\bar{x}_1 - \bar{\mu}_k\|) & 1 \\ \phi_1(\|\bar{x}_2 - \bar{\mu}_1\|) & \phi_2(\|\bar{x}_2 - \bar{\mu}_2\|) & \cdots & \phi_k(\|\bar{x}_2 - \bar{\mu}_k\|) & 1 \\ \vdots & & \ddots & \vdots & \vdots \\ \phi_1(\|\bar{x}_m - \bar{\mu}_1\|) & \phi_2(\|\bar{x}_m - \bar{\mu}_2\|) & \cdots & \phi_k(\|\bar{x}_m - \bar{\mu}_k\|) & 1 \end{bmatrix}_{m \times k+1} \times \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_k \\ b \end{bmatrix} \tag{4}$$

There exist some requirements to choice a RBF function. This work tried the TPS (thin-plate spline) $f(r) = r^2 log(r)$ [10]. Gaussians RBFs were also tested, but they presented greater errors than the TPSs.

5. Validation: methods and results

Although the thresholds for test images $F(X)$ had been assessed, it is not enough to assert that W is a good model. The validation searches the best or better train/test images arrangement to model the threshold decision.

It was applied three validation methods: the Generalized Cross-Validation (GCV), also known as k-fold; the Leave-One-Out (LOO); and the Bootstrap. The GCV tests the modeling stability, as well as the generalization efficiency, for different arrangements [11], but it is not suitable to validate a model from a few samples. The LOO is an extreme GCV variant, where only one sample is leaved out of training and so tested [12]. This approach is conservative, i.e. the errors are greater than others. In contrast, the Bootstrap validation method [13] is more appropriate to scarce samples, since its random nature to arrange train and test samples.

Initially the perceptive model was tested with data generated from functions with known behaviors such as linear, logarithmic, exponential, polynomial and furthermore a randomly uniform distribution. For each one of these known behaviors, it was generated two data groups, one holding 100 pairs of values and other 50 pairs. The pairs of values simulated the threshold of a hypothetical attribute. This test was developed to verify the modeler ability to deal with more certain data and to test the validation approaches.

In that way, each group was randomly divided in two other groups, training and test, with the same number of elements.

The performance test of the RBFN perceptive model applied to the mentioned groups consisted of three validation types: Generalized Cross Validation (GCV), Leave-One-Out (LOO) and the Bootstrap.

A crucial parameter in generalizations with RBFN was the number of centers, i.e. the number of RBFs that would compose the generalized function at training and test steps. The Table 2 shows the results for the generalization of the named known behaviors as well as the number of centers (#RBF column) for their evaluation.

Table 2 - Results of model checking with functions or distributions whose behavior were well known. The RMS Errors column is resultant of the average of 96 executions for GCV and LOO; and 40 executions for Bootstrap validation.

Experiment	#Training	#Test	LOO		GCV		Bootstrap	
			RMS Error	#RBF	RMS Error	#RBF	RMS Error	#RBF
random	80	20	57,28	3	15,51	1	15,79	3
random	40	10	55,78	1	20,26	1	20,76	1
exponential	80	20	2,66	74	2,06	59	1,51	52
exponential	40	10	4,30	40	5,07	31	2,84	35
linear	80	20	0,70	80	0,75	48	0,35	60
linear	40	10	1,98	40	2,77	34	1,31	30
logarithmic	80	20	0,65	72	1,01	32	0,27	57
logarithmic	40	10	0,80	40	1,03	29	0,42	32
polynomial	80	20	1,19	63	1,97	64	0,76	60
polynomial	40	10	2,11	39	2,38	32	1,65	30

The Table 2 results deserve a remark. All results but not uniform random distribution, presented low errors for at least three validation methods. This result shows that the RBFN method can model certain data with low errors, and uncertain data with high errors, as would be expected for a modeling tool.

An important modeling parameter is the number of centers (#RBF) that could be thought as an uncertainty factor for the trained data. This behavior was observed when the random distribution model found better results as the number of centers decreased. On the other hand, the linear model demanded much more centers, revealing more model specificity. Next, this section presents the thresholding modeling results for data collected in the experimentation.

The Table 3 shows the modeling results for the heterogeneous group of images. Here, the best model response was obtained by modeling human subjective responses as a 2D surface function of entropy and standard deviation. The number of centers was small (11 and 13), pointing that this surface is smooth, since it could reach 80 centers.

Table 3 - Validation model results for the known models. The RMS Errors is resultant of the average among 88 executions for GCV and LOO, and 40 for the Bootstrap.

#Training	#Test	LOO		GCV		Bootstrap	
		RMS Error	#RBF	RMS Error	#RBF	RMS Error	#RBF
88	22	16,61	13	4,44	11	4,32	13

The Figure 2 illustrates the generated surface that represents graphically one execution of the GCV. There, the circles mean the centers, and the triangles mean the average of human thresholds. The space was spanned by two features (entropy and standard deviation) and the human threshold. The crosses mean the samples used by training with respective entropy and the standard deviation.

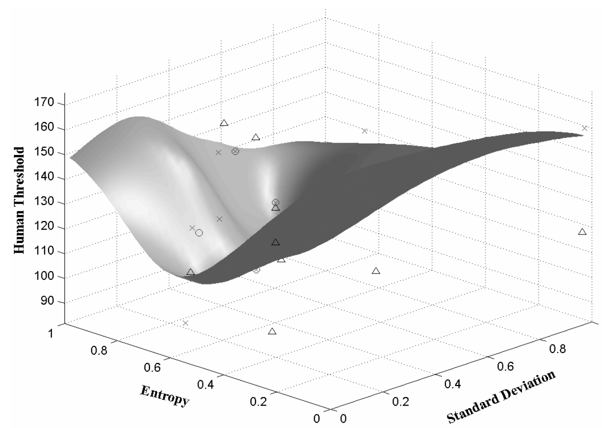


Figure 2 - Surface to illustrate the threshold human modeled as function of two attributes: entropy and standard deviation. This surface is one execution of GCV that use 10 RBF.

The work proceeds exploring the RBFN modeling approach as a feature selector. New image features were included in the perceptive threshold model. Having the validation as reference, it allowed us to measure the inclusion or exclusion of features from the model. Hence, three questions should be answered to characterize the models: *i*) what is the appropriate number of centers? *ii*) how many features would be enough to represent the human subjective answers? *iii*) which attributes would be the most representative?. In consequence, it was tested for errors, the features combinations generalizations applied to the group of heterogeneous images (Table 4). In the GCV and Bootstrap validations, the group was randomly partitioned in 88 images for training and 22 for test.

Models with just one feature combination had the smallest error. The modeled feature was the luminance average generalized with 6 centers (on average). On the other side, the contrast average presented the worst result with the maximum smoothness (1 center), that reveals an uncertain feature.

The combination of two attributes had its smallest error combining standard deviation and average of luminance, for intermediate smoothness.

The combination of three attributes models had its best result combining the standard deviation, the average of contrast, and the average of luminance, also with an intermediate smoothness.

Table 4 - Results of the validation applied to combined attributes to model the human thresholding of heterogeneous images. The RMS Errors are resultant of the average among 88 executions for GCV and LOO and 40 for the bootstrap. The attributes are codified with the first numerical character meaning the number of combined attributes and followed by the following abbreviations: e(entropy), sd(standard deviation), al(average of luminance), ac(average of the contrast), fc (first quartile of the contrast), tc (third quartile of the contrast), and mc (maximum of the contrast).

Features	LOO		GCV		Bootstrap	
	RMS Error	#RBF	RMS Error	#RBF	RMS Error	#RBF
1sd	17,69	10	4,51	1	4,59	3
1e	16,25	2	4,38	3	4,21	4
1al	14,03	9	3,82	4	3,64	5
1ac	18,28	1	4,54	1	4,76	1
2sd-al	13,54	41	3,82	34	3,76	29
3al-fc-tc	13,56	38	3,87	41	3,76	33
4sd-al-fc-tc	13,15	76	3,88	48	3,66	46
5e-sd-al-ac-mc	14,46	30	4,39	24	4,03	29
6e-sd-al-ac-fc-tc	13,51	37	4,17	39	3,85	36
7e-sd-al-ac-mc-fc-tc	13,95	31	4,20	33	3,82	38

The four attributes combination presented the smallest average error for the three validation methods, which was achieved combining the standard deviation, the average of luminance, the first and the third quartiles of the contrast.

The approach of RBFN was also applied to homogeneous images. A set of 59 distinct face images was used as training/testing set, all with the same size (71 x 92 pixels) and with faces in the same position.

The homogeneous images set was randomly divided in training set (48 images) and testing set (11 images) for GCV and Bootstrap validations. The results can be viewed in Table 5.

The Bootstrap validation was the most appropriate to assess the model performance for the face images.

It was observed (Table 5) that, again, the average of luminance presented the best result (the smallest error) with 6 centers (on average) models. The best result was achieved when two features were combined: the average of luminance and the average of contrast.

The combination of three attributes (entropy, luminance standard deviation, and the average of intensity) had the best result. It was achieved with 11 centers (on average) models.

Table 5 - Results of the validation applied to combined attributes to model the human thresholding of homogeneous images. The RMS Errors are resultant of the average among 88 executions for GCV and LOO and 40 for the bootstrap. The attributes are codified with the first numerical character meaning the number of combined attributes and followed by the following abbreviations: e(entropy), sd(standard deviation), al(average of luminance), ac(average of the contrast), fc(first quartile of the contrast), tc(third quartile of the contrast), and mc(maximum of the contrast).

Features	LOO		GCV		Bootstrap	
	RMS Error	#RBF	RMS Error	#RBF	RMS Error	#RBF
1sd	17,62	2	6,44	1	5,98	1
1e	17,76	1	6,71	3	6,19	4
1al	15,14	9	5,81	2	5,55	7
1ac	17,78	6	6,43	1	6,44	2
2al-mc	16,32	2	5,96	3	5,88	3
3al-fc-tc	15,91	12	6,34	4	5,94	5
4e-sd-al-mc	15,41	10	6,24	10	5,97	2
5e-sd-al-ac-mc	15,96	24	6,37	13	6,45	2
6e-sd-al-ac-fc-tc	16,46	14	6,31	1	6,16	1
7e-sd-al-ac-mc-fc-tc	15,88	12	6,07	1	5,97	1

The combination of four attributes presented the best result with entropy, luminance standard deviation, average of luminance, and the maximum of contrast, for 7 centers (on average) models.

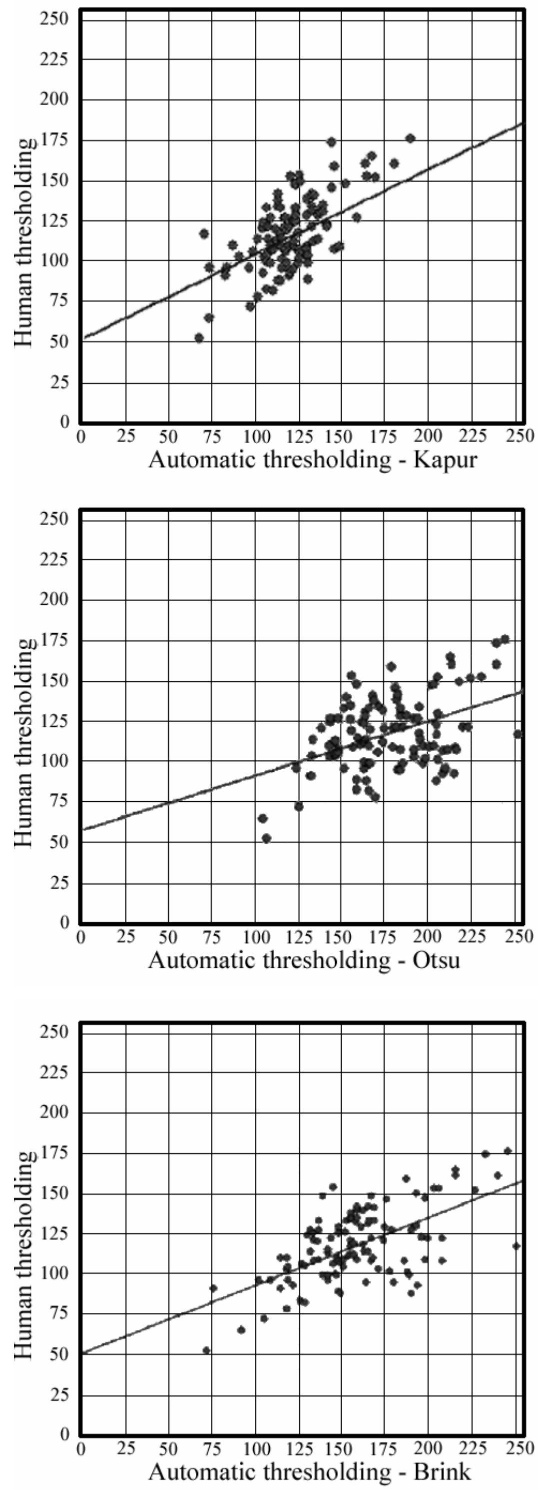
6. Automatic methods evaluation

Automatic thresholding methods are, in general, based on statistical pixel gray level properties. But, how similar are the automatic result and the human model result? This section searches to answer this question. So, the correlation coefficients between the automatic and the human decisions were computed to compare them. Three different methods: Otsu [5], based on gray level variance; Kapur et al [4], based on entropy; and Brink [3] that applied both, variance locally and entropy globally.

The Figure 3 illustrates these results for heterogeneous images with a dispersion diagram, and also the respective correlation coefficient (ρ). These results suggest that improvements performed in the method of Brink (e.g. to take into account the gray level variance for neighbor pixels) caused its good performance ($\rho = 0.71$). Furthermore, the Otsu variance-based method performance revealed low correlation with human decision ($\rho = 0.456$). In the other hand, the Kapur's entropic method presented a good correlation ($\rho = 0.6671$). This order was kept for homogenous face images, but the coefficient diminished for Kapur ($\rho = 0.5066$) and Brink ($\rho = 0.512$) methods.

The modeled threshold is more flexible than automatic methods. Each feature combination may present high and low correlation coefficients. For instance, the variance modeled thresholding reached the maximum $\rho = 0.9167$, but an uncorrelated model for variance caused a $\rho = 0.0937$. The number of centers, i.e. the number of RBFs can also improve or make worse the modeled correlation.

The Figure 4 shows a comparison between the modeled thresholding using the same feature than the automatic method. Therefore, Fig. 4 (a) denotes the mean of errors for modeled entropic thresholding (for bootstrap validation) related to each correlation coefficient. The solid square (the Kapur method) is the closest to the entropy-based model thresholding curve. The Fig 4 (b), differently, the Otsu method (circle) and the variance modeled thresholding were the most distant. Besides, the automatic method presented the greatest errors and the lowest correlation. The Brink method had to be compared with two features model: entropy and variance. These models did not present errors greater than 48. Although presented some correlation, the Brink method error was also considerable.

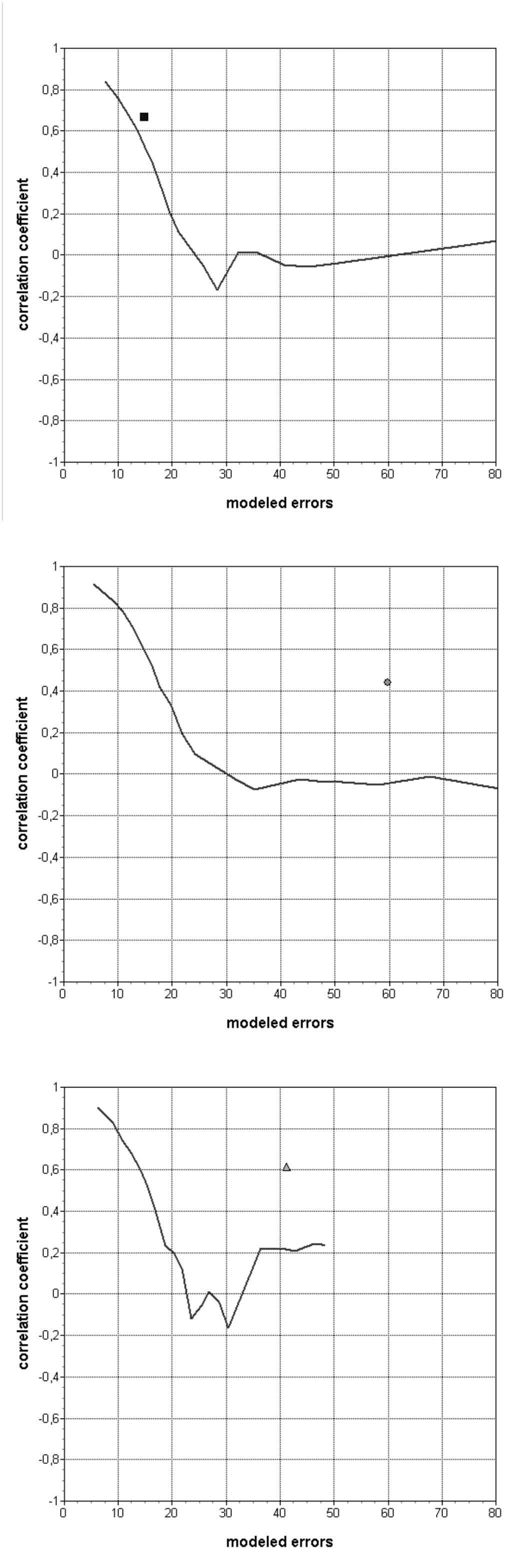


(a)

(b)

(c)

Figure 3 – Dispersion diagram for three representative automatic thresholding methods and Subjective Human Responses. (a) Kapur x Human $\rho = 0.6671$; (b) Otsu x Human. . $\rho = 0.456$.; Brink x Human. . $\rho = 0.71$.



(a)

(b)

(c)

Figure 4 - Comparison among modeled thresholding (lines) for different features and automatic thresholding. (a) Human x Kapur; (b) Human x Brink; (c) Human x Otsu.

A remarkable observation is that all the three modeled thresholding have shown a curve with high correlation coefficients and low errors. Even predictable, this result suggests that automatic thresholding methods could be classified according theirs human decision affinity.

7. Conclusion and Final Remarks

This work presented an innovative approach to implement gray level thresholding. It was built a modeler for subjective human decisions with two gray level image sets: one heterogeneous and other homogeneous. The modeling was based on global image features such as entropy and standard deviation, similarly to that used by automatic thresholding methods.

Firstly, the human responses and the automatic thresholding were compared by correlation coefficients. It was found a higher correlation for automatic methods which took into account a local variance.

Secondly, a RBFN generalization was applied to model the subject decisions, which had been collected with a psychophysical test.

The perceptive modeling results were compared to three representative automatic thresholding methods based on entropy and variance global features. These results revealed the flexibility of perceptive approach. It can model thresholding whose results are high or low correlated with human decisions. Besides, the entropic automatic thresholding presented high correlation and low errors when confronted to human responses for heterogeneous images, which could be visually observed.

This work also contributes with a test experiment to collect human thresholds, designed to works in the web environment, which enables distributed and sparse audiences.

The perspective for this work is to apply the models to threshold real-time video acquisition, since it would require just a global feature evaluation and the model application to perform early image segmentation.

Acknowledgements

Luís Augusto Consularo is grateful to CNPq (68.0053/01-8) for financial support.

8. References

[1] Y. J. Zhang, "A Survey on Evaluation Methods for Image Segmentation". *Pattern Recognition*, (31), 1996, pp. 1335-1346.

[2] Powell, M.J.D. *Radial basis functions for multivariate interpolation: A review*, Technical report 1985/NA12. Department of Applied Mathematics and Theoretical Physics, Cambridge University, 1985.

[3] A.D. Brink, "Using spatial information as an aid to maximum entropy image threshold selection". *Pattern Recognition Letters*, (17), 1996, pp. 29-36.

[4] J.N. Kapur, P.K. Sahoo, and K.C. Wong, "A new method for gray-level picture thresholding using the entropy of the histogram". *CVGIP*, (29), 1985, pp. 273-285.

[5] N. Otsu, "A threshold selection method from gray-level histogram". *IEEE Transactions on Systems, Man, and Cybernetics*, 1978, pp. 62-66.

[6] K.S. Fu, J.K. Mui, "A Survey on Image Segmentation". *Pattern Recognition*, (13), 1981, pp. 3-16.

[7] N.R. PAL, S.K. PAL, "A Review on Image Segmentation Techniques". *Pattern Recognition*, (26), 1993, pp.1277-1294.

[8] Ghosh, J., Nag, A. "An Overview of Radial Basis Function Networks". In: Howlett, R.J. & Jain, L.C. (Eds.) *Radial Basis Function Neural Network Theory and Applications*, Series Studies in Fuzziness and Soft Computing, Physica-Verlag, Heidelberg, Germany, 2001, pp. 1-36.

[9] Haykin, S. *Neural Networks: A Comprehensive Foundation*. Prentice Hall, New Jersey – USA, 1999.

[10] A.I. Mess, M.F. Jackson, and L.O. Chua, "Device Modeling by Radial Basis Functions", *IEEE Transactions on Circuits and Systems-I,* (39)1, 1992, pp. 19-27.

[11] R. A. Kohavi, "Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection". In: *Proceedings of IJCAI - International Joint Conference on Artificial Intelligence*, 1995.

[12] Bishop, C.M. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.

[13] B. Efron, R. Tibshirani, *An introduction to the bootstrap*. Chapman & Hall, 1993.