A FUZZY APPROACH TO THRESHOLDING OF IMAGE DATA)

Jan Hrubeš
Doctoral Degree Programme (1), FEEC BUT
E-mail: jan.hrubes@phd.feec.vutbr.cz

Supervised by: Jiří Kozumplík
E-mail: kozumpli@feec.vutbr.cz

ABSTRACT

This paper is concerned with fuzzy entropy definition used for automatic image thresholding. The fuzzy entropy is applied in selection of the fuzzy region of membership function so that an image is able to be transformed into fuzzy domain with maximum entropy. Then we are able to divide the fuzzy region and establish the thresholds.

Genetic algorithm is used for selection of optimal membership function.

1. INTRODUCTION

In many applications in image processing, it is useful to be able to separate out the regions of the image corresponding to objects in which we are interested, from the regions of the image that correspond to background. Thresholding often provides an easy and convenient way to perform this segmentation on the basis of the differences between the foreground and background regions of an image.

The basis of this work is automatic estimation of fuzzy areas based on the principle of maximum fuzzy entropy and its application to match the membership function. For searching the best combination of coefficients a, b and c (see eq. (1)) of the membership function is used the genetic algorithm. Optimal membership function is used to threshold the fuzzificated picture to N partitions. In this case, the word optimal means, that the consequent picture will contain as much of original information as possible.

2. FUZZY SETS AND ENTROPY

In classic set is clear if some element belongs into it or not. The fuzzy sets represent generalization and we can establish the degree of membership of the element. Fuzzy set is defined as collection of ordered pairs of element and its degree of membership (from interval <0, 1>) to the set. The degree of membership denotes how much the element belongs to the set. In grayscale picture we can imagine it means how much the pixel with specific intensity (from interval <0, 255>) is bright.

For transform of the picture to fuzzy domain was used S membership function, defined as

$$\mu\left(x,a,b,c\right)=\begin{cases}0, & x\le a\\ \dfrac{\left(x-a\right)^2}{\left(b-a\right)\left(c-a\right)}, & a\le x\le b\\ 1-\dfrac{\left(x-c\right)^2}{\left(c-b\right)\left(c-a\right)}, & b\le x\le c\\ 1, & c\le x\end{cases} \tag{1}$$

where x is the variable in the intensity domain, and a, b and c are the parameters which determine the shape of S-function. The range [a, c] is called the fuzzy region.

For objective judgment of amount of information in the image after thresholding is used an entropy definition from [1], which respects both – a membership function and distribution, therefore, it can respond to the variety of input information.

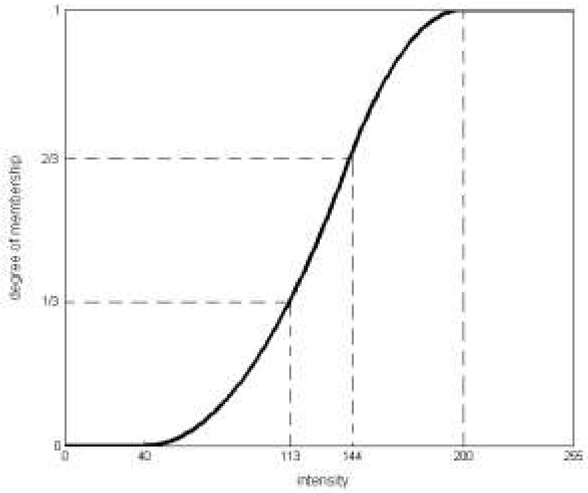

Fig. 1: Intensity domain divided into three partitions with S-function (40,140,200) and equal partitioning.

Let A be the fuzzy event, and N be the number of partitions of A in the fuzzy domain denoted as sub-events $A_1, \ldots, A_N$. The fuzzy domain partition method M can be equal partition or non-equal partition. $\mu_A(x)$ is the membership function (in this case it is the S-function) and $P(x)$ is the probability of x in the space domain. The entropy of a fuzzy event is defined as

$$\begin{aligned} H\left(A,N,M,\mu_A\right) &= \frac{-1}{\log N}\sum_{i=1}^{N} P_P\left(A_i\right)\log P_p\left(A_i\right), \\ P_p\left(A_i\right) &= \sum_{\mu_A\left(x\right)\in A_i} P\left(x\right) \end{aligned} \tag{2}$$

$P_P(A_i)$ means the probability summed in the space domain for the x (space domain) mapping into A_i (fuzzy domain) by the membership function $\mu_A(\ )$. Before entropy computation we must decide the number of partitions N and partition method M and notify that membership function $\mu_A(\ )$ is S-function with parameters (a,b,c). Then we can simplify $H(A,N,M,\mu_A)$ as $H(A,a,b,c)$. The problem is to find a combination of (a,b,c) such that $H(A,a,b,c)$ has the maximum value. To problem were solved by genetic algorithm.

3. GENETIC ALGORITHM

Genetic algorithm (GA) is search technique used for optimization. The problem is encoded into binary matrix, where one row is one set of parameters (one of solution). The quality of all solutions in one generation is evaluated by fitness function. Better solutions are selected with higher probability for crossover (random swap of part information between two solutions). Then is applied genetic operator mutation used to maintain genetic diversity and avoid local minima. Mutation operator inverts randomly selected bits in a genetic sequence. This set of genetic operators is repeated until we can qualify that the solution is optimal. There can be used different conditions for that. An example of flow diagram of genetic algorithm is on Fig. 2. More about GA you can read in [2] or [3].

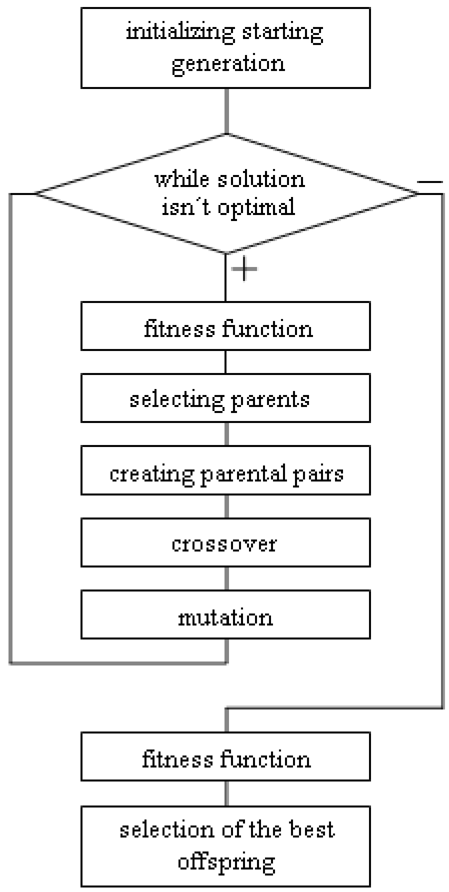

Fig. 2: Flow diagram of genetic algorithm.

In this case one solution consist of parameters (a,b,c) of membership function $\mu(a,b,c)$. All of these parameters are integers on interval from 0 to 255 (2^8=256), consequently one solution will be coded with 24 bits.

4. RESULTS

We implemented the algorithm for automatic optimal thresholding of image data in MATLAB to obtain some experimental results. The program can reduce the number of intensity levels from 256 to any lower, but interesting results are with less than 20 levels. For segmentation are important black & white (binary) outputs.

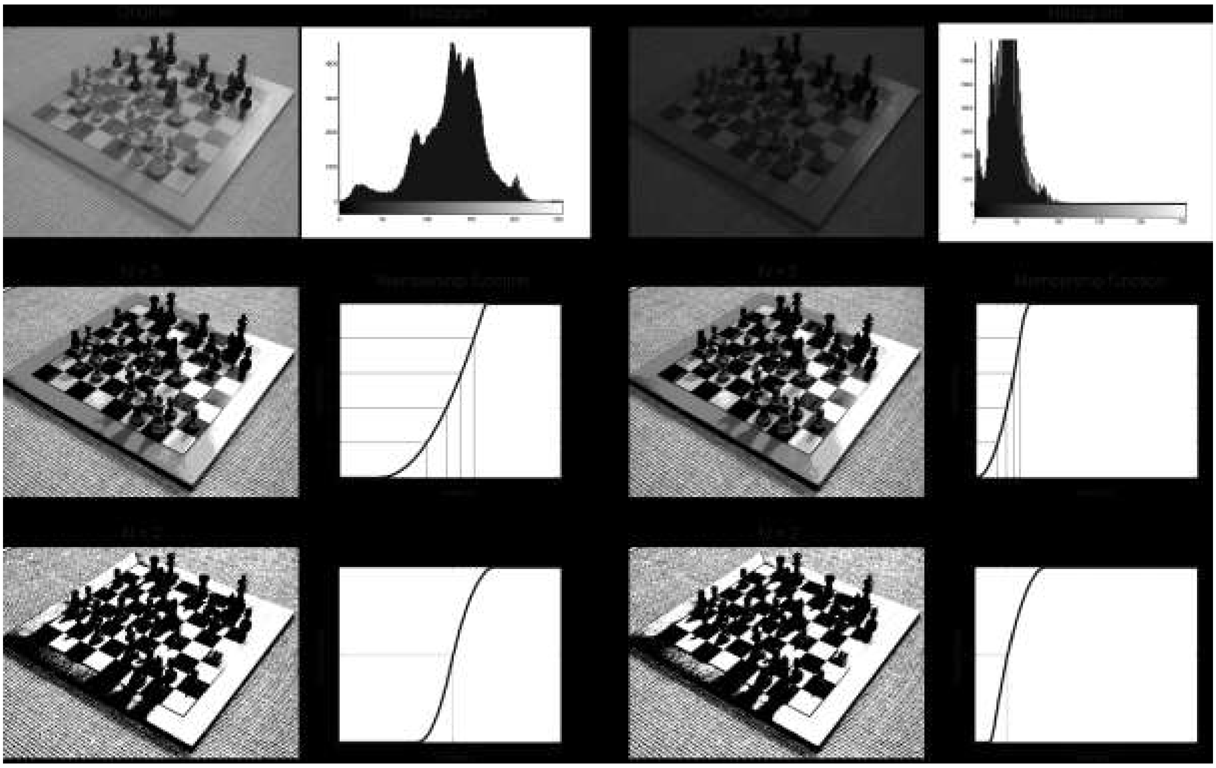

Fig. 3: Chess normal and dark and outputs with 2 and 5 levels of intensity.

On Fig. 3 is presented adaptability of this method for different contrast. The outputs are almost equal.

On next figure (Fig. 4) are shown results of optimal thresholding with various *N* (number of partitions). Parameters of S-function and corresponding thresholds used on the picture of Mona Lisa are summarized in Table 1. There is evident that thresholding with higher density is located near intensity 100.

N	*a*,*b*,*c*	thresholds
20	49,51,222	-
10	24,81,240	59,73,84,96,108,122,138,157,181
5	25,51,253	61,86,117,157
3	15,39,254	68,123
2	31,106,153	98

Table 1: Parameters of S-function and thresholds used in Fig. 4.

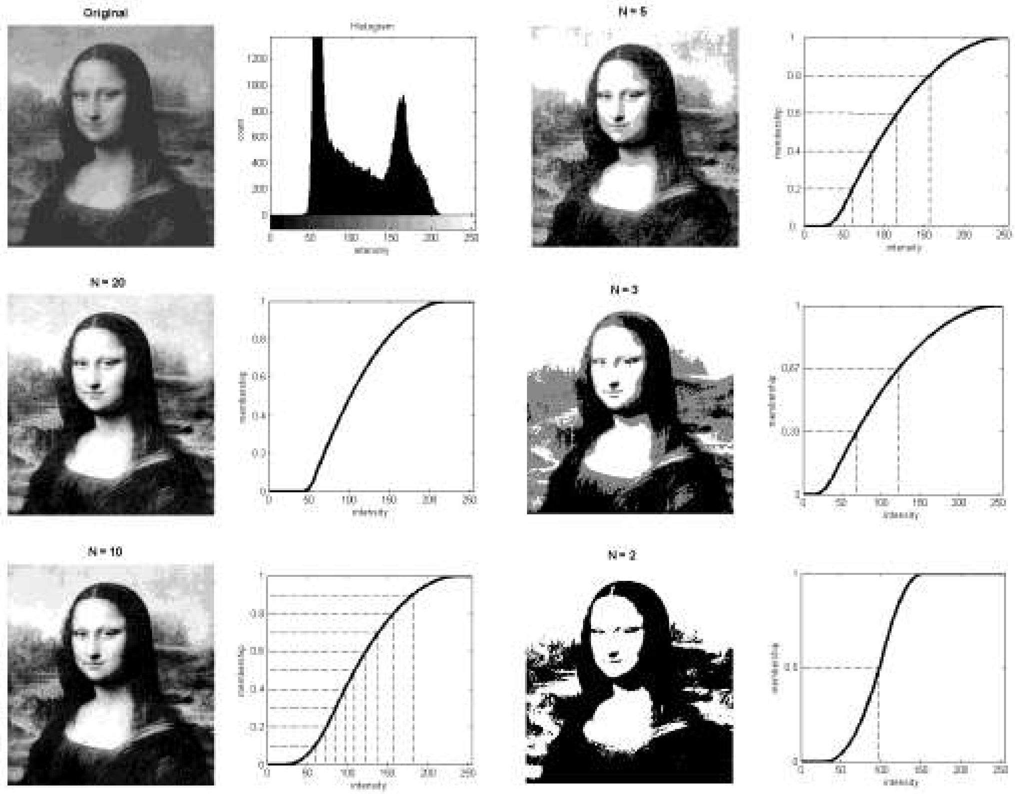

Fig. 4: Mona Lisa thresholded with different *N.*

5. CONCLUSIONS

Presented method is very effective for reducing the number of intensity levels. Problems may cause images with height amount of unwanted information which is saved to the expanse of subjective more important information.

ACKNOWLEDGEMENT

The project was supported by the Research Programme of Brno University of Technology MSM 0021630513.

REFERENCES

[1] CHENG, H.D., CHEN, Y.H., SUN, Y.: A Novel Fuzzy Approach to Image Enhancement and Theresholding. Signal Processing, Vol 75, No 3, June 1999, pp 277-301GUPTA, K. C., GARG, R., BAHL, I., BHARTIA, P. *Microstrip Lines and Slotlines*. 2nd ed. Norwood: Artech House, 1996.Orság, F.: Vision für die Zukunft. Biometrie, Kreutztal, DE, b-Quadrat, 2004, s. 131-145, ISBN 3-933609-02-X

[2] VOMLEL, J.: Genetické algoritmy, presentation, 2004, available from: <http://staff.utia.cas.cz/vomlel/slides/presentace-ga.pdf>

[3] WHITLEY, D.: A Genetic Algorithm Tutorial, available form: <http://samizdat.mines.edu/ga_tutorial/ga_tutorial.ps>