A LOCALLY ADAPTIVE BINARY-TREE METHOD FOR BINARIZATION OF TEXT IMAGES

D. T. DIMOV, G. Y. GLUHCHEV
Institute of Information Technologies, Bulgarian Academy of Sciences,
Acad. G. Bontshev Str., Block 29-A, 1113 Sofia, Bulgaria.
E-mail: dtdim@bgcict.acad.bg , gluhchev@iinf.bg

A heuristic method is proposed, locally adaptive to noise in the recognized image (text and/or graphics). A modified image quality measure is defined. Using this measure, at every step of the image division by a B-tree approach, a decision is made either for continuation or for straight-forward binarization of the current subimage by a given method (currently the Otsu's discriminant procedure is employed). Experimental results are illustrated comparatively. Some improvements of the method under investigation are also reported.

1 Introduction

Binarization, i.e. discrimination of a halftone image into two classes, e.g. black (*Object*) and white (*Background*), has always been an important stage in image pre-processing, especially in text-and-graphics image recognition [1÷12]. Most commonly, the binarization over an area *D* from the image can be done in two ways: (i) by a global threshold t , t Î $[L,R]$, where t is an integer and *L* and *R* are the leftmost and rightmost bounds of the image greys, respectively, and (ii) by local thresholds t_i , acting over non-overlapping subregions D_i , *i=1,2,¼* , covering *D*. Usually, a reduced information about the image is processed. According to the amount of this information, the methods known [1, 2, 3] may be differentiated as:

- Non-contextual: by *histograms* of the grey scale distribution, cf. [4, 5, 6];
- Contextual: by second order statistics for neighbour pixel intensities, cf. [7];
- Other contextual ones: by gradient approaches [8], morphological ones, etc.

In the present work, a locally adaptive method of histogram type is considered for the intuitive reason that the treatment of the image by parts will make against the increased information loss.

2 Locally Adaptive Approaches to Binarization

Adaptation to image noise depends on the image decomposition approach to a great extent. The most preferred way is to divide the image into congruent squares of size (*m´m*), while the local threshold is computed over a larger en-compassing square *(n´n)*, *n>m*, to account for neighbour pixel correlation [2, 4, 9]. Also, the order of scanning with windows is of crucial importance as shown in [4]. The dimensions *n, m* are subject to experimental adjustment for different types of images as well, except in the special case of working on separate pixels ($n>m=1$). An additional smoothing (when $m>1$) can be applied to achieve more 'even' threshold surface [3, 9]. Gradient methods [8] for 'cleaning up possible ghosts' after binarization, are an efficient auxiliary procedure in many cases [2].

Decomposition into equal parts is convenient for interpretation but poses certain problems like: (i) the optimal choice of window dimensions and the scanning order at the same time [4], (ii) the problem of adaptation to areas occupied entirely with background or object, i.e. when only one of the two classes is practically present, etc. Thus in [4] a parameter adjustment is proposed (but (?) if the average intensity steadily approaches the background or the object intensity). In [9], a neighbour extrapolation is considered for the case of no threshold decision (but (?) if the neighbours are of the same type of ambiguity).

An uneven decomposition may solve these problems. There are known well-structured approaches to obtain an uneven division, for example by *Q-tree* [10] or by *B-tree* [11, 12]. But few data is available on their use for binarization, probably due to possible complexity of algorithms or due to the assumption that a global binarization method could be applied almost directly.

Using a B-tree approach to histogram binarization we search for an algorithm *data-driven* to the local particularities of text images as well as the noise over them.

3 The Binary-Tree Histogram Approach to Binarization

It can be shown that considering a Q-tree instead of the simpler B-tree approach for image decomposition does not lead to more favourable conditions for binarization. We shall consider the following

3.1 B-tree Based Algorithm

A1: (Initialization): Make the initial image current.

A2: (*Base Recursion*): Split alternatively (horizontally or vertically) the current image (*parent*) into two equal parts (*daughters*), and decide: (i) make the daughter a new parent and proceed with its division, or (ii) binarize it, and proceed with the rest of daughters if exist, otherwise end.

3.2 Choice of a Global Binarization Method for Application

The binarization of the daughters by their histogram can be performed by a known method [1, 2, 3]. Most methods expect a Gaussian distribution of the two classes O and B of binarization, with parameters: m_O and m_B (mean values), s_O and s_B (variances); w_O and w_B , $w_O+w_B=1$ (normalized class areas), cf. [3, 4, 6, 9]. Entropy approaches are viewed as optimal (in a Bayesian sense). Modelling by mixed populations yields the same best result [6]. Also, the authors proved there that the Otsu's method [5] was optimal only for the case: $w_O = w_B = 1/2$ and $s_O = s_B$. Besides, the iterative approaches [3] are rather “clumsy” to be used in our case.

Like in [4], we think that the Otsu's method performance is good enough, and, what is more important – is apt to determine two classes (except when $s_O = s_B = 0$, and $m_O = m_B$). The optimal threshold t , $t \in [L,R]$, is determined using the maximum of the h criterion, cf. [1, 2, 5, 6] :

$$h \;=\; w_O w_B (m_O - m_B)^2/s^2 \;=\; (w_O(m_O - m)^2 + w_B(m_B - m)^2)/s^2 \;; \quad (1)$$

where m and s are the total mean and variance of the image.

As Otsu notes in [5], the maximum of h can be considered a measure of “goodness” of the image. Still, h does not distinguish between “pure” cases ($s_O = s_B = 0$, $m_O \neq m_B$ when $h \equiv 1$). So, we introduce the following *modified quality measure*:

$$x \;=\; h.Dm/(R-L) \;=\; \frac{1}{(R-L)} \cdot \frac{(Dm)^3}{(s_O)^2/w_B + (s_B)^2/w_O + (Dm)^2}\,, \quad Dm = |m_O - m_B|. \qquad (2)$$

Clearly, the new measure x is normalized ($x \in [0,1]$) and is monotonically varying with Dm, s_O, s_B, and with $w_O \in (0, s_B/s_O]$, $w_O \in [s_B/s_O, 1)$, like h.

3.3 Description of the Proposed Algorithm. Experiments

The comparative study of experiments on 8 text images (256×256×256), with different amount of noise, leads us to the following algorithm for *decision making* in the base recursion (BR):

D1: Compute parameters (w_O, m_O, s_O ; w_B, m_B, s_B ; t, m, s) for the current parent $\langle O_0, B_0\rangle$ and its two daughters $\langle O_1, B_1\rangle$, $\langle O_2, B_2\rangle$ according to Otsu [5], and

if (the dimensions of $\langle O_0, B_0\rangle$ have reached a predetermined low bound)

then (binarize $\langle O_1, B_1\rangle$ and $\langle O_2, B_2\rangle$ by t_0 and **goto** D4); **else** continue;

D2: Compute the measures $x_i, i=0,1,2$ (for the three subimages) by (2). Declare the threshold t_i, $i \in \{0,1,2\}$ of the image with highest x, a *dominant threshold* t_{Dom}.

D3: Make a choice for each daughter, i.e.:

for ($i = 1, 2$) do

if (the daughter $\langle O_i, B_i\rangle$ has got the dominating threshold)

then call BR for $\langle O_i, B_i\rangle$, i.e. make it current parent and proceed it;

else if ($((t_{Dom}+t_i)/2 \leq m_{Oi} + s_{Oi})$ or $(m_{Bi} + s_{Bi} \leq (t_{Dom}+t_i)/2\,)$)

then binarize $\langle O_i, B_i\rangle$ by the threshold $(t_{Dom}+t_i)/2$, and **goto** D4;

else call BR for $\langle O_i, B_i\rangle$ (i.e. forwards by the BR-tree);

D4: **return** (i.e. backwards by the BR-tree).

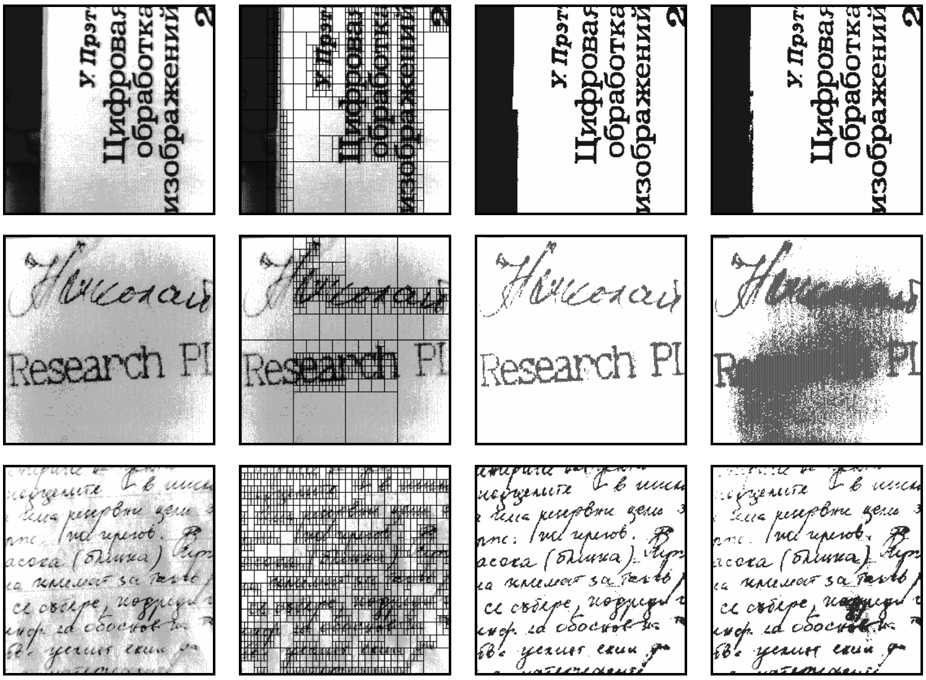

Figure 1. Three text images binarization results. From the left to the right side – the originals; their division; their binarization by the proposed method, and by the Otsu's method.

The results of this algorithm on 3 test images, compared to the global method of Otsu, are shown in Fig. 1. Subimages with a lower x (including the extreme case: $s_O = s_B = 0$, and $m_O = m_B$) are prone to binarization, while these with a higher x (including the extreme case: $s_O = s_B = 0$, and $m_O < m_B$) are rather inclined to further division. The obtained layout of "minimal subimages" tends to areas with maximal gradient (i.e. to cover the text object). This effect, though positive, actually leads to lengthening the processing of "clear" images. The latter phenomenon can be limited by setting a bound x_{max} for x.

4 Improvements to the Proposed Algorithm. A Future Work

Further experiments show that the proposed algorithm meets difficulties when processing rather noisy images. Similar images are usually obtained by TV-cameras working in poor lighting environment. Several approaches considered to help in solving this problem has been investigated. Three of them are briefly reported below:

A1: After a histogram smoothing (i.e. computing a moving average on a segment d, $d=0,1,...(R-L)/2$), the current histograms (of the parent and its two daughters, at the current node of B-tree) would fall in two classes: *bimodal* and *unimodal*. Bimodal histograms, especially with $d=0$, are intuitively preferred for binarization to unimodal ones which "tend" (except in the case: $x=0$) to further division. Besides, as the process is going forwards the tree, w_O is inclined to increase and w_B is inclined to decrease or vice versa,i.e. an *intensity mean inversion* during division. Comparative experiments (not listed) show no better results.

A2: After each division (at B-tree node), the following conditions must hold:

$$|w_{C0} - lw_{C1} - (1-l)w_{C2}| = Dw(t_1,t_0,t_2) » 0,$$
$$|w_{C0}m_{C0} - lw_{C1}m_{C1} - (1-l)w_{C2}m_{C2}| = Dm(t_1,t_0,t_2) » 0 ; \quad (3)$$

where $w_{Ci}=w_{Ci}(t_i)$, $m_{Ci}=m_{Ci}(t_i)$, $C Î \{O,B\}$, $i Î \{0,1,2\}$; and $0£l£1$ is the coefficient of splitting (in our case $l»1/2$). These conditions, together with the correlation between the measures $x_i=x_i(t_i)$, $i Î \{0,1,2\}$, taken for all splitting following the algorithm, leads to an optimis ation task for correction of the correspondent thresholds t_i, $i Î \{0,1,2\}$. Because of the task complexity, an iterative (top-down and vice versa) approach has been experimented to evaluate optimal thresholds.

A3: Proceeding the whole B-tree, from an initial image to minimal subimages, we can observe the trend of subimages to cluster, see Fig. 2. This phenomenon is mostly manifested in the case of field (m_0, s_0), and field (Dm, j) as well where:

$$j = j_0 = l_1 l_2 (Dm)^2 / (l_1(s_1)^2 + l_2(s_2)^2 + l_1 l_2 (Dm)^2) , \quad Dm = |m_1 - m_2|. \quad (4)$$

by analogy with (1) and (2), $l_1 » l_2 » 1/2$, cf. (3), and m_i, s_i, $i Î \{0,1,2\}$ are the total means and variances of the subimages. Obviously, the latter allows a classification of subimages into two classes to be done. For instance, Fig. 2a, the first class ($s£s_{limit}$) consists of single-coloured (O or B) subimages, while the second class ($s>s_{limit}$) consists of parti-coloured (O and B) ones. Instead of discussing methods of more sophisticated classification, we represent only the results of an improved algorithm for binarization, Fig. 3. Its resemblance to the method for a stepwise down-top merging by Beaulieu and Goldberg [12] can be lightly proved. And this improvement of the proposed algorithm has to be considered the most promising one in our research so far.

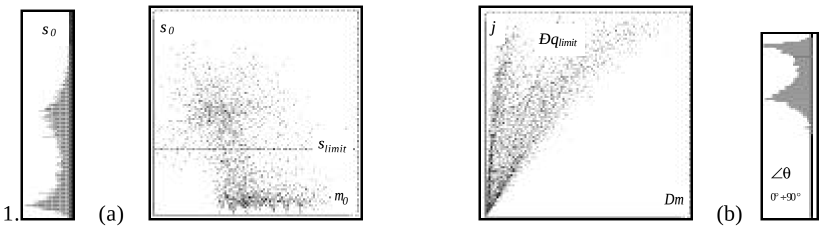

Figure 2. Trends to clustering for the image of Fig. 3: (a) the field (m_0, s_0), and (b) the field (Dm, j)

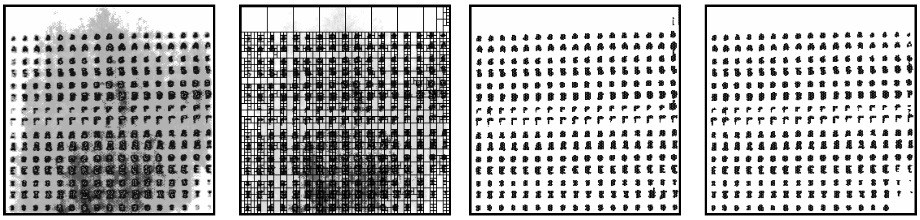

Figure 3. Binarization results for a rather 'dirty' image by the improved algorithm, cf. A3. From the left to the right side – the original; their division; their binarization by the improved as well as by the base algorithm, cf. 3.3 (instead of the global Otsu's which gives nothing in the case).

5 Conclusion

A modified measure of the quality of text-and-graphics images is proposed. A heuristic B-tree algorithm for data-driven binarization is developed. Improvements of the above described technique are under investigation. It can be stated that the method efficiency is comparable to the efficiency of known methods [2, 4], but possesses to a higher extent the advantage of data driven processing. The approach can be applied for processing of images captured by a TV-camera.

Acknowledgements

This work is sponsored by Bulgarian National Science Fund, grant No. I-524/95.

References

1. P. Sahoo, S. Soltani, A. Wong and Y. Chen, *Comp. Vision, Graphics, and Image Proc.*, **41**, 233 (1988).
2. ∅. Trier and T. Taxt, *IEEE Trans. on PAMI*, **17**, 312 (1995).
3. K. Mardia and T. Hainsworth, *IEEE Trans. on PAMI*, **10**, 919 (1988).
4. L. Eikvil, T. Taxt and K. Moen, in *Proceed. of ICDAR'91* (Saint-Malo, France, 1991), 435
5. N. Otsu, *IEEE Trans. on Systems, Man, and Cybernetics*, **9**, 62 (1979).
6. T. Kurita, N. Otsu and N. Abdelmalek, *Pattern Rec.*, **25**, 1231 (1992).
7. A. Brick, *Pattern Rec.*, **25**, 803 (1992).
8. S. Yanowitz and A. Bruckstein, *Comp. Vision, Graphics, and Image Proc.*, **46**, 82 (1989).
9. Y. Nakagawa and A. Rosenfeld, *Pattern Rec.*, **11**, 191 (1979).
10. P. Strobach, *IEEE Trans. on PAMI*, **11**, 742 (1989).
11. N. Samet and M. Tamminen, *IEEE Trans. on PAMI*, **10**, 579 (1988).
12. J.-M. Beaulieu and M. Goldberg, *IEEE Trans. on PAMI*, **11**, 150 (1989).