Scientific Journal Impact Factor (SJIF): 1.711
International Journal of Modern Trends in Engineering and Research
ISSN (ONLINE) : 2349 - 9745 ISSN (PRINT) : 2393 - 8161

BINARIZATION OF TEXTUAL IMAGES USING A NETWORK FLOW TERMINOLOGY

Swagat Barot[1], Prof. S. Rama Mohan[2] .
[1,2] *Dept. of Applied Mathematics,Faculty of Tech. & Engineering,*
Maharaja Sayajirao University of Baroda,Vadodara, India

Abstract - The problem of Optical Character Recognition (OCR) needs, as a first prerequisite, a good and powerful algorithm for binarization of scanned text. Most of the available algorithms used for binarization, uses the concept of thresholding. Such algorithms, in spite of being powerful, simple and the one with very small computational cost, are sometimes less efficient in case of scanned text of low quality. This leads to the requirement of other mathematical model for binarization. In this paper, we present a new mathematical model for binarization problem. We transform the problem in to an equivalent optimization problem and attempt to solve it through network flow terminology. Max-flow min- cut theorem given by Ford and Fulkerson plays a very crucial role in the model.

Keywords – Binarization of image, Network flow problem, Max-flow min-cut theorem, Optimization, Ford – Fulkerson algorithm

I. INTRODUCTION

Optical Character Recognition (OCR) is a technique to convert text image in to computer readable text. It is a problem of machine learning whose solution requires binarization of the scanned text images as a primary step. There are variety of algorithms used for binarization, most of which relies on the technique of thresholding. Despite of numerous advantages of the technique, there are times, when the shortcomings of the technique overshadow the positive side of the technique and force the user for the search of some new method for binarization. We have tried to address the problem by means of mathematical model involving network flow terminology.

A digital image of text document (i.e. scanned text) is composed of rectangular arrangement of pixels, each of which represents intensity. Binarization is defined as a process of formulating a function $l : X \to \{t,b\}$ which reassigns a binary value *t* (Text) or *b* (Background) to every pixel *x* of the image, subject to the given data. The choice of the function depends on two constraints. 1) The neighboring pixels should be assigned similar value by the function almost everywhere with the exception of the pixels representing boundary of the text. 2) The assignment should be made in light of the data given by scanned image. We have used the approach of Energy minimization in order to construct the function. The first step includes construction of an objective function in the light of constraints of the problem. Let X be a set of all pixels of the image and $\{g_x : x \in X\}$ be the set of grey values of pixels. The objective function $O : L \to \mathbb{R}$ provides the measure of inappropriateness for every possible function $l \in L$ and plays a crucial role in the selection of most suitable function. It is defined as,

$$O(l) = \sum_{\substack{(x,y)\in N \\ x,y\in X}} kp_{xy} + \sum_{x\in X} |l_x - g_x| \qquad (1.1)$$

Where, *N* denotes the set of all pairs of neighboring pixels of *X*, *k* is a nonzero constant, p_{xy} denotes the penalty imposed by the objective function to *l* for assigning different binary values to the neighboring pixels *x* and *y* of *X* , and is defined as

$$p_{xy} = |l_x - l_y| \tag{1.2}$$

and l_x is defined as,

$$l_x = \begin{cases} 0, & \text{if } l(x) = t \\ 255, & \text{if } l(x) = b \end{cases} \tag{1.3}$$

The objective function takes care of both the constraints imposed on the required binary function *l*. Hence our reduced problem is to find a binary function *l* that minimizes the objective function. This optimization problem is computationally expensive as the space all possible binary functions has dimension $|X|$, which is in thousands. Making the situation worst, the function being non-convex may have many local minima. Our model uses the approach of network flow for the minimization of objective function. We construct a network flow using the objective function and the scanned image. Minimum cut of the network flow plays a decisive role in the evaluation of the required binary function.

II. GRAPH THEORETIC MODEL

Given a scanned textual image with pixel set *X*, our model considers its gray value set $\{g_x : x \in X\}$ as an input data. A network *G (V, E, W)* using the input data is constructed with vertex set *V* containing vertices v_x corresponding to each member *x* of X and a pair of distinguished vertices *s* and *t* for source and sink resp. Every such vertex has a pair of edges e_{sx} and e_{xt} joining it with source *s* and sink *t* resp. Their edge weights are $|l_x - g_x|$. For every pair of neighbors *x* and *y*, an edge e_{xy} with weight $k.p_{xy}$ is created in the network flow. The cut of the network flow is a minimal collection of its edges, whose removal from the network flow makes the terminal vertices *s* and *t* disconnected. A minimum cut of the network flow is a cut with minimum cost. The minimum cut of the network flow gives the partition of the vertices v_x in to two sets *S* and *T*. This also gives the most suitable choice of binary function for the image. All pixels *x* corresponding to vertices v_x which are part of the set *S* should be assigned binary value *t*, whereas the remaining pixels of the image should be assigned value *b*. In the next session, we present the results showing that, the model gives the binary function which minimizes the objective function.

III. SINGNIFICANCE OF THE MODEL

In this section, we will prove that, our model efficiently minimizes the objective function theoretically, and as a result gives a binary function which is the best solution of the problem in the light of available constraints.

Theorem 3.1 *For any scanned textual image, the set L of all corresponding binary functions and the set C of all cuts on the network corresponding to the image are in one to one correspondence.*

Proof: Let *L: X ⟶ {t, b}* be a binary function corresponding to the given scanned image. Then, there exists a cut c = *{S, T}* on the network for the image corresponding to the function *l* defined as follows:

S = $\{v_x / l_x = 0\}$ and *T* = $\{v_x / l_x = 255\}$

Note that, $l_x = 0$, when $l(x) = t$ and $l_x = 255$ when $l(x) = b$.

This proves that, every binary function finds an equivalent cut in the set of all cuts.

Conversely, let *C = {S, T}* be a cut on the network for the given textual image. Then, we can define *l* as follows:

$l(x) = t$, if $v_x \in S$

And $l(x) = b$, if $v_x \in T$

This proves the theorem.

Theorem 3.2 *The minimum cut for the network flow under consideration gives binarization which minimizes the objective function.*

Proof: let l be the binarization corresponding to any cut C of the network flow. First, we show that, cost of the cut C is $O(l)$.

Note that, cost of any cut is sum of weights of the edges which are member of the cut set. In case of our network flow, there are two types of edges: (i) edges joining vertices corresponding to pixels of the image. i.e. edge e_{xy} with $(v_x, v_y) \in N$ (ii) edges between terminal vertex and a vertex corresponding to pixel, i.e. edges e_{sx} or e_{xt} for some vertex v_x. Thus, the cost of C is given by,

$$\text{Cost}(C) = \sum_{\substack{(v_x,v_y)\in N \\ e_{xy}\in C}} w(e_{xy}) + \sum_{e_{sx}\in C} w(e_{sx}) + \sum_{e_{xt}\in C} w(e_{xt}) \quad (3.1)$$

Claim 1: For every vertex v_x, exactly one of the edges e_{sx} and e_{xt} can be part of the cut C.

If both of the edges are part of the cut C, the cut C' obtained by removing one of the edges e_{sx} and e_{xt} is again a cut, which is contradiction with the fact that C is a cut, as no proper subset of a cut can be a cut. This proves that, at most one of the edges e_{sx} and e_{xt} can be in C.

If the cut C contains none of the edges e_{sx} and e_{xt} then, the terminals s and t stays connected even after removal of all edges of C, (because of the existing path s-e_{sx}-v_x-e_{xt}-t in $G \setminus C$) which contradicts with the fact that, C is a cut.

This proves *claim 1*.

As a result, every vertex v_x of the graph contributes to exactly one of the last two terms of (3.1). Thus, (3.1) becomes,

$$\text{Cost}(C) = \sum_{\substack{(v_x,v_y)\in N \\ e_{xy}\in C}} w(e_{xy}) + \sum_{x\in X} |l_x - g_x| \quad (3.2)$$

As $w(e_{xy}) = k.p_{xy}$, (3.2) becomes,

$$\text{Cost}(C) = \sum_{\substack{(v_x,v_y)\in N \\ e_{xy}\in C}} k.p_{xy} + \sum_{x\in X} |l_x - g_x| \quad (3.3)$$

Claim 2: For every pair of neighboring vertices v_x and v_y, the edge e_{xy} is part of C iff either e_{sx} and e_{yt} are both in C or e_{xt} and e_{sy} are both in C.

Assume that, e_{sx} and e_{yt} are in C but e_{xt} and e_{sy} are not. If possible, assume that, $e_{xy} \notin C$. Then there is a path s-e_{sy}-v_y-e_{xy}-v_x- t joining s and t in $G–C$, which is a contradiction with the fact that C is a cut. Hence, in this case e_{xy} must be in C.

Assume that, e_{sx} and e_{sy} are in C but e_{xt} and e_{yt} are not. If possible, assume that, $e_{xy} \in C$. Then, $C - \{e_{xy}\}$ is still a cut, which is a contradiction with the fact that, C is a cut and no proper subset of C can be a cut.

Remaining two cases can be addressed by analogous argument. Thus, claim 2 is proved.

The first term in the expression of cost(C) mentioned in (3.3) presents sum of weights of all e_{xy} corresponding to neighboring vertices v_x and v_y which are left connected to different terminals in

$G \backslash C$. The edges e_{xy} not contained in C, are the ones, for which corresponding v_x and v_y are assigned same binary value by l. Hence, weights of such e_{xy} will be zero.

Thus, equation (3.3) becomes,

$$\text{Cost}(C) = \sum_{\substack{(x,y)\in N \\ x,y\in X}} k.p_{xy} + \sum_{x\in X} |l_x - g_x| \qquad (3.4)$$

But, by (1.1),

Cost(C) = O(l).

This proves that, the cost of the minimum cut C minimizes the objective function O(l).

IV. DISCUSSION AND RESULTS

We have implemented the model using a code written in java programming language. Although the model gives sufficiently good results, the code still needs refinement as it takes considerably long computational time for binarization of large images. The results of experimentation are presented in figure 4.1.

(a) (b) (c) (d) (e) (f) (g) (h)

Figure 4.1: (a) ,(c), (e) and (g) are scanned images whereas, (b),(c),(f) and (h) are their binarization produced by our code

We have carried out time analysis on the data of implementation of our code on textual images of various sizes. The results are discussed in table 4.1 and figure 4.2. From the theoretical results, it is evident that, the model gives one of the best results in light of the constraints. Still, the exponential growth of computational time with respect to the image size appears as a big hurdle for real acceptability of the method and needs to be addressed.

Dimension of the scanned image	No. of pixels in the image	Time taken (in Seconds)
91 X 33	3003	2
91 X 35	3185	2
89 X 45	4005	3
100 X 63	6300	5
92 X 92	8464	10
120 X 95	11400	15
320 X 200	64000	75
480 X 216	103680	2258

Table 4.1: Computational time for images of various sizes

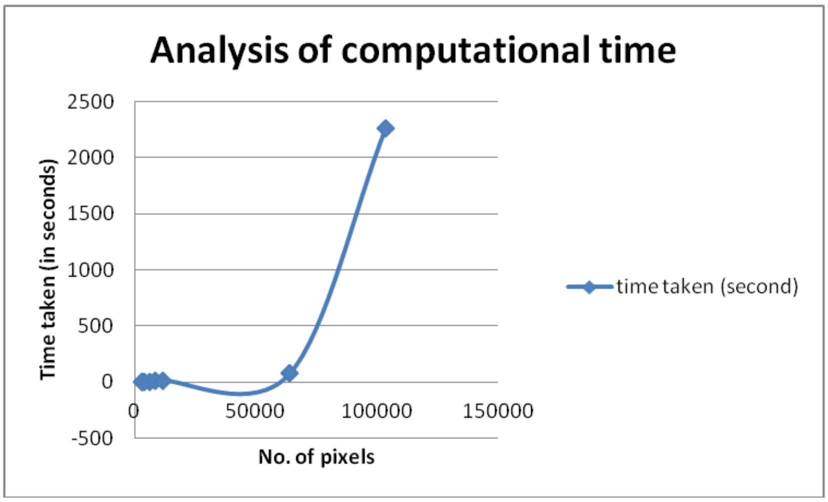

Figure 4.2: Scatter plot for data of table 4.1

V. CONCLUSION

Our mathematical model efficiently optimizes the objective function theoretically and hence gives the best possible results. However, the computational complexity of the model needs to be addressed in a better way at the time of real implementation by encoding the model by means of more efficient programming for implementation.

REFERENCES

[1] Adrian Burian, Markku Vehnilainen, Mejdi Trimeche, and Jukka Saarinen “Document Image Binarizatzon Using Camera Device in Mobile Phones”, Proc. Intl. Conf. on Image Processing, 2005, (ICIP 2005), Volume 2, pp: 546-548.

[2] B. Gatos, I. Pratikakis and S.J. Perantonis,’ Improved Document Image Binarization by Using a Combination of Multiple Binarization Techniques and Adapted Edge Information’, 978-1-4244-2175-6/08/$25.00 ©2008 IEEE

[3] B. Gatos, I. Pratikakis, and S. J. Perantonis, “Adaptive Degraded Document Image Binarization,” Pattern Recognition, Vol. 39, No. 3, pp. 317–327, March 2006.

[4] B. Su, S. Lu, and C. L. Tan, “Binarization of Historical Document Images Using The Local Maximum and Minimum,” Proc. Intl. Workshop on Document Analysis Systems, pp. 159–165, June 2010.

[5] B. Su, S. Lu, and C. L. Tan, “Document Image Binarization Using Background Estimation and Stroke Edges,” *Proc. Intl. Journal on Document Analysis &Recognition*, Vol. 13, No. 4, pp. 303-314, 2010.

Volume 02, Issue 08, [August– 2015] **ISSN (Online):2349–9745 ; ISSN (Print):2393-8161**

[6] B.K.P. Horn and B. Schunk, "Determining Optical Flow," Artificial Intelligence, vol. 17, pp. 185-203, 1981.

[7] Costin-Anton Boiangiu, Alexandra Olteanu, Alexandru Stefanescu, Daniel Rosner, Nicolae Tapus, Mugurel Andreica "Local Thresholding Algorithm Based on Variable Window Size Statistics" Proceedings CSCS-18, The 18-th International Conference on Control Systems and Computer Science, May 24-27 2011, Bucharest, Romania, Volume 2, Pag. 647-652, Politehnica Press, ISSN: 2066-4451.

[8] Costin-Anton Boiangiu, Andrei Iulian Dvornic, Dan Cristian Cananau, „Binarization for Digitization Projects Using Hybrid Foreground-Reconstruction", Proceedings of the 2009 IEEE 5th International Conference on Intelligent Computer Communication and Processing", Cluj-Napoca, Editor Ioan Alfred Letia, August 27-29, pp.141-144, ISBN: 9781-4244-5007-7, Library of Congress Number:200906791, IEEE Catalog Number: CFP0909D-PRT, 2009.

[9] Costin-Anton Boiangiu, Andrei Iulian Dvornic. "Methods of Bitonal Image Conversion for Modern and Classic Documents". WSEAS Transactions on Computers, Issue 7, Volume 7, pp. 1081 1090, July 2008, ISSN: 1109-2750.

[10] Costin-Anton Boiangiu; Olteanu, A.; Stefanescu, A. V.; Rosner, D. & Egner, A. I. (2010). Local Thresholding Image Binarization using Variable-Window Standard Deviation Response (2010), Annals of DAAAM for 2010 & Proceedings of the 21st International DAAAM Symposium, 20-23 October 2010, Zadar, Croatia, pag. 133-134, ISBN 978-3901509-73-5, ISSN 1726-9687, Editor B. Katalinic, Published by DAAAM International, Vienna, Austria 2010.

[11] D. Greig, B. Porteous, and A. Seheult, "Exact Maximum A Posteriori Estimation for Binary Images," J. Royal Statistical Soc.,Series B, vol. 51, no. 2, pp. 271-279, 1989.

[12] F. Shafait, D. Keysers, and T. M. Breuel, "Performance comparison of six algorithms for page segmentation," in 7th IAPR Workshop on Document Analysis Systems, pp. 368–379, (Nelson, New Zealand), Feb. 2006.

[13] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. 3, G. T. Rado and H. Suhl, Eds. New York: Academic, pp. 271–350, 1963.

[14] J. Sauvola and M. Pietikainen, "Adaptive document image binarization," Pattern Recognition 33(2), pp. 225–236,2000.

[15] O. H. Hwa, L. T. Kil, and C. I. Sung, "An Improved Binarization Algorithm Based on a Water Flow Model for Document Image with Inhomogeneous Backgrounds," *Pattern Recognition*, Vol. 38, pp. 2612 –2625, 2005.

[16] P. Viola and M. J. Jones, "Robust real-time face detection," Int. Journal of Computer Vision 57(2), pp.137– 154, 2004.

[17] S. Birchfield and C. Tomasi, "A Pixel Dissimilarity Measure that Is Insensitive to Image Sampling," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, no. 4, pp. 401-406, Apr. 1998.

[18] S. Birchfield and C. Tomasi, "Depth Discontinuities by Pixel-to-Pixel Stereo," Int'l J. Computer Vision, vol. 35, no. 3, pp. 1-25, Dec. 1999.

[19] S.A. Barker and P.J.W. Rayner, "Unsupervised Image Segmentation," Proc. IEEE Int'l Conf. Acoustics, Speech and Signal Processing, vol. 5, pp. 2757-2760, 1998.

[20] Sauvola, J., Seppanen, T., Haapakoski, S., and Pietikainen, M.: 'Adaptive document binarization'. Proc. 4th Int. Conf. on Document Analysis and Recognition, Ulm Germany, 1997, pp. 147–152

[21] T.Hoang Ngan Le, Tien D. Bui, and Ching Y. suen, "Ternary Entropy – based Binarisation of Degraded Document Images Using Morphological Operators" 15205363/11$26.00 ©2011 IEEE.

[22] Ulichney, Robert A (1994). "Halftone Characterization in the Frequency Domain". Retrieved 2012-7-20.

[23] V. Kolmogorov and R. Zabih, "Computing Visual Correspondence with Occlusions via Graph Cuts," Proc. Int'l Conf. Computer Vision, vol. 2, pp. 508-515, 2001.

[24] W. Niblack, "An Introduction to Digital Image Processing ", pgs: 115116, Englewood Cliffs, NJ, Prentice Hall, 1986.

[25] W.E.L. Grimson and T. Pavlidis, "Discontinuity Detection for Visual Surface Reconstruction," Computer Vision, Graphics and Image Processing, vol. 30, pp. 316-330, 1985.

[26] Y. Boykov, O. Veksler, and R. Zabih, "Fast Approximate Energy Minimization via Graph Cuts", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 11, pp. 1222-1229, 2001.

[27] Ying Liu, Richard Fenrich, and Sargur N Srihari "An Object Attribute Thresholding Algorithm for Document Image Binarization"Proc. 2 Intl. Conf. on Document Analysis and Recognition, Tsukuba Science City, Japan, pp. 278-281 October, 1993.

