Project Report

Rahul Anand Sharma rahul.anand@research.iiit.ac.in

Mehak Sood mehak.sood@studentsiiit.ac.in

Abstract:

The objective of work is to solve the text binarization problem using the TRW and BP graph cut algorithm. Our code implements the TRW (Kolmogorov, 2006) and BP algorithm (S.Yedidia) for Binary Images. It also includes the evaluation measures as well as the performance of the methods along with a short description of each method and comparison of both the algorithms with the ground truth dataset.

1 Introduction:

Document image binarization is an important step in the document image analysis and recognition pipeline since it affects further stages of the recognition process. The evaluation of a binarization method aids in verifying its effectiveness and studying its algorithmic behaviour. Therefore, it is imperative to have a benchmarking dataset along with an objective evaluation methodology in order to capture the efficiency of current document image binarization practices. We focused on the evaluation of document image binarization methods using a variety of scanned machine-printed and handwritten documents for which the binary image ground truth was given. The problem description is explained in section 1.1 and the two energy minimization techniques are explained in section 1.2 below.

1.1 Problem Description:

Binarization is one of the key pre-processing steps in any document image analysis system. The performance of the subsequent steps like character segmentation and recognition are highly dependent on the success of binarization. Document image binarization is an active area of research for many years.

We focused on binarization of DIBCO 2011 Dataset and formulated the binarization as text is foreground and anything else is background, and define a novel energy (cost) function such that the quality of the binarization is determined by the energy value. We minimize this energy function to find the optimal binarization using TRW and BP graph cut scheme. The graph cut method needs to be initialized with foreground/background seeds and also the edge weights between the neighbours have to be initialized.

The binarization problem is formulated as a labelling problem, where we define an energy function such that its minimum corresponds to the target binary image

Section 1.3 describes experiments and results based on the challenging DIBCO 2011 word dataset. Some sample images of this dataset are shown in Figure 1.

We formulate the binarization problem as a labelling problem as follows:

Input: Graph $G = (V, E)$
Discrete label set L = (0: background, 1: foreground)
Cost of a labelling $Q(f)$ = Unary Cost + Pairwise Cost
Problem: Find $f = \text{argmin}Q(f)$

In General

$$E(x) = \sum_{i} CiXi + \sum_{i,j} Ci,j\, Xi(1 - Xj)$$

C_I and $C_{i,j}$ are set such that:

- The higher a pixel looks like the object, it is preferred to be labelled as object. (Unary term)
- If all the neighbours of a pixel are labelled as object, then this pixel also may be preferred to be labelled as object. (Pairwise term)

1.2 EM Technique Used

Energy Function Used

$$E(x) = \sum_{i=1}^{N} \sqrt{(xi - L)(xi - L)} + \sum_{i,j} K\sqrt{(xi - xj)(xi - xj)}$$

X $_I$ are grayscale intensity values.

L are labels.

K is smoothness factor

Given an input image of size m x n.

$$\boldsymbol{N} = m\ x\ n$$

We first create **N** nodes.

Initialize labels foreground (L = 255) and background (L = 0)

Compute weights between nodes and labels as

Unary Term: (Xi – L)

Compute weights between the neighbours as

Pairwise Term: (Xi-Xj)

First Order Neighbourhood (4-nbd):

	1	
1	X	1
	1	

1.3 Evaluation Measures:

1.) F-Measure

$$Measure = 2 * \frac{Recall * Precision}{Recall + Precision}$$

Where

$$Recall = \frac{TP}{TP + FN} \quad Precision = \frac{TP}{TP + FP}$$

TP, FP, FN denote the True positive, False positive and False Negative values, respectively.

2.) PSNR

$$PSNR = 10 * \log(C^2/MSE)$$

Where

$$MSE = \sum_{x=1}^{M}\sum_{y=1}^{N}(I(x,y) - I'(x,y))^2/MN$$

PSNR is a measure of how close is an image to another. Therefore, the higher the value of PSNR, the higher the similarity of the two images. We consider that the difference between foreground and background equals to C.

3.) Accuracy

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn}$$

4.) Precision

$$Precision = \frac{tp}{tp + fp}$$

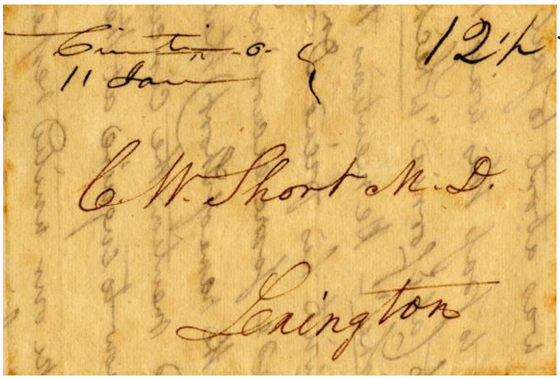
5.) Recall

$$Recall = \frac{tp}{tp + fn}$$

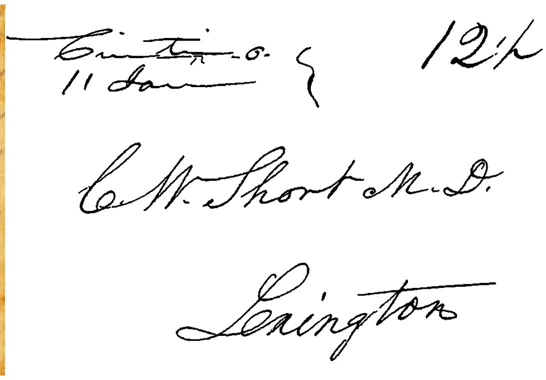
1.3 Experimental Analysis:

The DIBCO 2011 testing dataset consists of 8 machine-printed and 8 handwritten images resulting in a total of 16 images for which the associated ground truth was used for the evaluation.

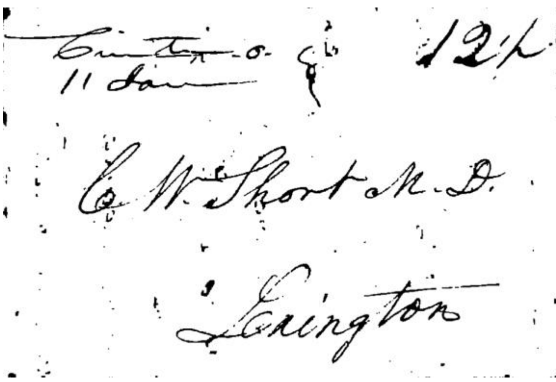
		HW1	HW2	HW3	HW4	HW5	HW6	HW7	HW8	PR1	PR2	PR3	PR4	PR5	PR6	PR7	PR8
TRW	Precision	88.34	95.68	93.22	88.82	90.66	96.05	96.71	100	79.62	91.09	81.51	92.85	86.02	99.8	99.04	91.72
	Recall	80.46	97.11	96.01	65.57	90.97	86.56	95.54	0.78	88.98	82.92	89.29	89.4	77.25	0.6	71.66	94.46
	Accuracy	73.61	93.06	89.79	62.23	83.51	83.69	92.56	3.94	74.34	77.43	75.41	83.99	70.97	4.34	71.3	87.18
	F Measure	84.22	96.39	94.6	75.44	90.82	91.06	96.12	1.55	84.04	86.82	85.22	91.09	81.4	1.19	83.16	93.07
	PSNR	5.79	11.59	9.91	4.23	7.83	7.88	11.28	0.17	5.91	6.46	6.09	7.96	5.37	0.19	5.42	8.92
BP	Precision	88.32	95.68	93.21	88.91	90.66	95.99	96.7	100	79.63	91.16	81.54	92.89	86.29	99.84	98.95	91.73
	Recall	80.43	97.15	96.13	65.49	91	85.55	95.53	0.01	89.01	82.17	89.35	89.27	75.4	0.7	72.46	94.4
	Accuracy	73.57	93.1	89.88	62.23	83.54	82.71	92.55	3.18	74.37	76.88	75.49	83.92	69.93	4.44	72.03	87.14
	F Measure	84.19	96.41	94.65	75.42	90.83	90.47	96.11	0.01	94.06	86.43	85.27	91.04	80.48	1.38	83.66	93.04
	PSNR	5.78	11.61	9.95	4.23	7.83	7.62	11.28	0.14	5.91	6.36	6.11	7.94	5.22	0.2	5.53	8.91



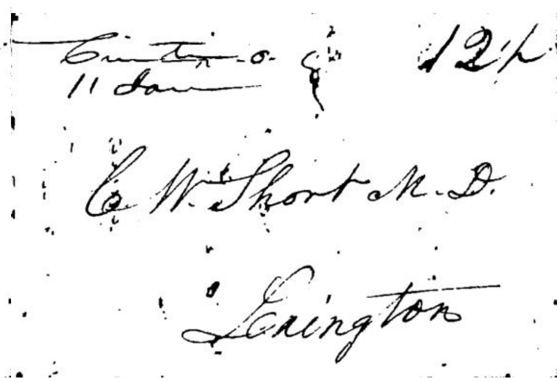
(a)



(b)



(c)



(d)

Figure 1. Representative samples and corresponding binarization results of the HW7.png from the DIBCO 2011 Dataset (a) Original image; (b) Binarized ground truth image; (c) Binarized BP image; (d) Binarized TRW image.

1.5 Conclusion

Our Implementation of using Graph cut based technique for binary segmentation yields comparable results as mentioned in DIBCO 2011 paper.
For Several images our implementation outperforms the standard techniques as mentioned in the paper on the basis of F Measure evaluation technique (3 Hand Written and 3 Printed).

1.6 Future work:

Till now we have implemented a simple technique for providing the weights between the nodes and the edges for the graph cut problem. Our method is powerful than one shot graph cut based binarization because it refines the seeds and binarization output at each iteration and produces clean binary images. However, a good binarization method also needs to be robust to variations in foreground -background colours. So, we propose to improve our method using GMM for modelling colour distributions. This will help us reduce the undesired prints in the results.

References

Kolmogorov, V. (2006). Convergent Tree-reweighted Message Passing. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*.

Mishra, A. (2011). An MRF Model for Binarization of Natural Scene Text. *Proceedings of International Conference on Document Analysis and Recognition*.

S.Yedidia, J. (n.d.). Understanding Belief Propagation and its Generalizations.