# A Laplacian Energy for Document Binarization

Nicholas R. Howe
Department of Computer Science
Smith College
Northampton, MA, USA
E-mail: nhowe@smith.edu

***Abstract*—This paper describes a new algorithm for document binarization, building upon recent work in energy-based segmentation methods. It uses the Laplacian operator to assess the local likelihood of foreground and background labels, Canny edge detection to identify likely discontinuities, and a graph cut implementation to efficiently find the minimum energy solution of an objective function combining these concepts. The results of this algorithm place it near the top on both the DIBCO-09 and H-DIBCO assessments.**

***Index Terms*—I.4.0.b Image processing software, I.4.6.b Graph-theoretic methods, I.4.6.d Pixel classification.**

## I. Introduction

Binarized document images represent a miracle of efficiency: by recording each pixel as a simple on-or-off value they require a fraction of the storage space as compared to a full-color scan, yet they can preserve most if not all of the significant information content of a document. Many useful computations on document images rely on binarization as an initial step, and a high-quality binarization simplifies most subsequent processing. By contrast, a low-quality binarization that omits significant detail or contains other errors can hinder the success of any methods that rely upon it.

For these reasons, researchers have exerted considerable effort toward improving binarization algorithms. In particular, commonly cited milestones include the work of Otsu [11], Niblack [10], and Sauvola et. al. [14]. Each improves upon its predecessors, but still falls short of perfection, particularly for documents with challenging attributes such as fading and bleed-through. Indeed, research continues on various binarization methods, and a recent contest comparing different techniques attracted 43 distinct entries.

Perhaps not surprisingly, document binarization is a subjective and ill-posed problem. The configuration of intensities that form the dot of a letter 'i' in one case may represent a simple inkstain in another. The presence of pathological examples such as this does not diminish the usefulness of binarization in the vast majority of instances. Nevertheless subjectivity poses a problem for the evaluation and comparison of different binarization methods. One way around this difficulty is to look at how well binarization supports some end application, such as optical character recognition [5], [9]. A second approach simply ignores the ambiguities, and evaluates binarization performance in terms of a chosen ground truth [3]. This paper employs the latter method, using the data set created for the 2009 Document Image Binarization Contest (DIBCO-09) and the sequel contest that focused on handwriting (H-DIBCO).

Reading the brief descriptions of the techniques entered in the first contest reveals that most employ some combination of background estimation and compensation, adaptive thresholds, and heuristic filtering techniques [3]. A few rely on Markov random field models, like the method presented herein. Two make use of the Laplacian in some way, while a third detects "domes", perhaps performing an analogous function. The different techniques vary in complexity, but many involve a series of pre- and post-processing steps. At least one other binarization method has employed a graph cut implementation similar to the one used here [7], [6], but does not report DIBCO results.

## II. Motivation and Method

The approach explored in this paper relies on a combination of several simple concepts. First, it employs a well-known vector property to achieve illumination invariance: the Laplacian of the image intensity, $\nabla^2 I$. The Laplacian measures the divergence of the intensity gradient, and therefore has greatest magnitude in areas that are local peaks and valleys of intensity – negative in sign for the former and positive for the latter. Thus it naturally separates concentrations of darkness and lightness, independent of the absolute local intensity level.

Second, the cues from the Laplacian operator are aggregated across the entire image by finding the optimal solution to a global fitness function. This enforces long-range consistency in the final solution, and discourages local choices that are incompatible with one another. The global fitness function chosen is easily solved with graph cut (maximum flow) methods, which efficiently compute the optimal binarization [1]. Similar approaches have proven effective in foreground segmentation for videos [15].

Third, the results of Canny edge detection inform the binarization, so that discontinuities in the output binarization coincide with detected edges where possible. The graph cut formulation provides the mechanism behind this linkage: one can omit graph edges between neighboring pixels wherever an image edge appears. The next section gives further details on the graph construction. Edge detection has been used in some prior work, but not in combination with a global energy function [12], [4].

The method treats dark and light areas asymmetrically in one respect. The Laplacian operator approaches zero in areas of near uniform intensity. Thus the binarization of large uniform areas depends entirely on the Laplacian at the boundaries, and in the absence of any strong signal some low-texture areas may receive incorrect labels on the basis of small fluctuations at their border. To guard against this possibility, the labeling fitness expression incorporates a strong bias towards the background label for a small set of bright outlier pixels (i.e., those 2 standard deviations above their local mean). This breaks the symmetry in favor of designating uniform areas as background, corresponding to an assumption that most text documents will not contain large areas of pure ink. For the ten DIBCO-09 test images, the heuristic described above misidentifies only five pixels on one single image that happens to contain a large inked region.

### A. Problem Formulation

Assume an $m \times n$ grayscale image $I$ where the pixel intensities $I_{ij}$ lie within the range from 0 (black) to 1 (white). The goal of binarization is to produce a new image $B$ of the same dimensions, composed exclusively of black and white pixels, i.e., $B_{ij} \in \{0, 1\}$. The binarized image should be perceptually similar to the original image, a notion we can formalize by defining an energy function $\mathcal{E}_I(B)$ minimized by the ideal binarization $B^*$.

$$B^* = \arg\min_B \mathcal{E}_I(B) \tag{1}$$

Markov random field modeling suggests the use of energy functions comprising a sum of individual label penalty terms (meant to capture the affinity of a particular pixel for a particular label) and pairwise label mismatch terms (meant to capture the tendency of neighboring pixels to share a label, for example). Such functions can be efficiently solved via algorithms based upon graph cuts, among other means [1]. For the average DIBCO image, the solution takes less than two seconds to compute on a 2.2 GHz laptop.

$$\begin{aligned} \mathcal{E}_I(B) &= \sum_{i,j:B_{ij}=0} L^0_{ij} + \sum_{i,j:B_{ij}=1} L^1_{ij} \\ &\quad + \sum_{i,j,i',j':B_{ij}\neq B_{i'j'}} C(i,j,i',j') \end{aligned} \tag{2}$$

Here $L^0_{ij}$ is the cost of assigning label 0 to the pixel at $(i, j)$, which intuitively should be lowest for intensity valleys. Likewise $L^1_{ij}$ is the cost of assigning the label 1. $C(i, j, i', j')$ is the cost of assigning different labels to the pixel at $(i, j)$ compared to the pixel at $(i', j')$. $C$ will be zero for non-neighboring pixels; this formulation corresponds to a Markov random field and allows a more specific expression for the energy.

$$\begin{aligned} \mathcal{E}_I(B) &= \sum_{i=0}^{m}\sum_{j=0}^{n} \left[L^0_{ij}(1 - B_{ij}) + L^1_{ij} B_{ij}\right] \\ &\quad + \sum_{i=0}^{m-1}\sum_{j=0}^{n} C^h_{ij}(B_{ij} \neq B_{i+1,j}) \\ &\quad + \sum_{i=0}^{m}\sum_{j=0}^{n-1} C^v_{ij}(B_{ij} \neq B_{i,j+1}) \end{aligned} \tag{3}$$

Here $C^h_{ij}$ and $C^v_{ij}$ represent the costs of a label mismatch between $B_{ij}$ and its neighbor to the south or east, respectively. They take on either a constant value $c$ or zero, as described below. The boolean inequality expression converts to either 0 or 1 in the standard manner.

The easiest choice would simply set all $C^h_{ij}$ and $C^v_{ij}$ to a positive constant $c$. This approach enforces smoothness in the binarized solution by penalizing any discontinuities. However, some discontinuities must be tolerated in an accurate binarization. Specifically, the edges of the inked regions are discontinuous. Thus the energy function must not penalize discontinuities between neighbors if an edge separates them.

Standard edge detectors identify individual edge pixels, but the formulation in Equation 3 requires knowing which side of the pixel should be discontinuous with its neighbors. Fortunately, the gradient direction provides an appropriate cue, with two possible choices. If one places the discontinuity on the high-gradient side, inked areas will tend to include edge pixels, whereas placing discontinuities on the low-gradient side will tend to group edge pixels with the background. The experimental implementation here chooses the former policy. Assuming that $E_{ij}$ represents the presence or absence of a Canny-detected edge at pixel $(i, j)$ [2], the final expressions for $C^h_{ij}$ and $C^v_{ij}$ appear below.

$$C^h_{ij} = \begin{cases} 0 & \text{if } E_{ij} \wedge (I_{ij} < I_{i+1,j}) \\ 0 & \text{if } E_{i+1,j} \wedge (I_{ij} \geq I_{i+1,j}) \\ c & \text{otherwise} \end{cases} \tag{4}$$

$$C^v_{ij} = \begin{cases} 0 & E_{ij} \wedge I_{ij} < I_{i,j+1} \\ 0 & E_{i,j+1} \wedge I_{ij} \geq I_{i,j+1} \\ c & \text{otherwise} \end{cases} \tag{5}$$

As mentioned, the Laplacian of the intensity provides a useful starting point for the label costs because it identifies areas of converging and diverging gradients (which indicate heights and depressions respectively):

$$L^0_{ij} = \nabla^2 I_{ij} \tag{6}$$

$$L^1_{ij} = -\nabla^2 I_{ij} \tag{7}$$

Note that this formulation is independent of absolute intensity and perfectly symmetric with respect to light and dark pixels. The asymmetric bias mentioned previously is then applied by setting $L^1_{ij} = \tau$ for certain bright outlier

pixels. With some abuse of notation, in the formulas below let $G_r(I_{ij})$ indicate the intensity at pixel $(i, j)$ after the image $I$ has undergone smoothing by convolution with a Gaussian kernel of radius $r$, representing the extent of the search for outliers.

$$L^1_{ij} = \begin{cases} -\nabla^2 I_{ij} & H_{ij} \leq 2S_{ij} \\ \tau & H_{ij} > 2S_{ij} \end{cases} \quad (8)$$

where

$$H_{ij} = I_{ij} - G_r(I_{ij}) \quad (9)$$

$$S_{ij} = \sqrt{G_r(H^2_{ij})} \quad (10)$$

The computed image $H$ resembles the original image $I$ adjusted to have local mean intensity of zero. (In this case, "local" means the contributions of neighboring pixels are weighted by a Gaussian of radius $r$.) The computed image $S$ represents the local standard deviation of $H$.

### B. Parameters

The algorithm just described includes five important parameters: $\tau$, $c$, $r$, and two Canny thresholds $t_{lo}$ and $t_{hi}$. Actually, the specific value of $\tau$ matters little so long as it is large enough to force pixels to take a foreground label. The expected size of ink components in the document should guide the choice of $r$: it should be at least several multiples of the expected ink stroke width. The remaining three parameters interact more strongly. With low values of $c$, the edge locations matter less and local sign changes in the value of the Laplacian dominate the discontinuities in $B$. Alternatively, with high values of $c$, discontinuities in $B$ will increase the overall energy unless they align with detected edges, and the choice of $t_{lo}$ and $t_{hi}$ becomes critical because it determines which edges appear and thus the components included in $B$. In general, the edges of ground-truth ink boundaries often have higher contrast than noise sources such as stains, smudges, and bleed-through from the opposite side of a paper document, etc. However, with high values for $c$, edges must be detected as completely as possible to minimize the discontinuity costs in Equation 3. These considerations motivate a high value for $t_{hi}$ and low value for $t_{lo}$.

Empirically the following choices are effective parameter settings for a range of documents and may be used as sensible defaults: $\tau = -2$, $c = 0.8$, $r = 20$, $t_{lo} = 0$ and $t_{hi} = 0.4$. (The latter two are specified as a fraction of the maximum observed edge gradient.) On the other hand, given a particular document or set of documents with ground truth, one can optimize the parameters for greater performance on those documents and others like them. This strategy was adopted for the DIBCO contests. For example, with clean documents the values of $c$ and $t_{hi}$ may be reduced for better recall without loss of precision. In cases where a training set can be used for explicit optimization, the experiments in the next section begin with the values above and execute a derivative-free unconstrained minimization on the F-measure using Matlab's *fminsearch* function [8]. It is worth noting that such tuning only changes the final performance by a few percent: in other words, a wide range of parameter values (including the defaults previously mentioned) still give acceptable results.

### C. Post Processing

One innovation that can slightly improve binarization quality is to repeat the energy minimization step after adding additional low-strength edges located within inked areas of $B$ after the first pass. (Trying to include these faint edges in one pass would also pick up unwanted noise in background areas, but restricting their inclusion to areas likely to contain ink avoids this problem.) If $E^1_{ij}$ are the Canny edges with the original value of $t_{hi}$, and $E^2_{ij}$ are the Canny edges with $t_{lo} = t_{hi} = 0$, then generate a new combined edge map:

$$E_{ij} = E^1_{ij} \vee (E^2_{ij} \wedge B_{ij}) \quad (11)$$

Binarization with the new edge map introduces no new false positive labels, although it can produce some false negatives particularly on print documents (examples visible in Figure 1). However, it also lowers the false positive rate by hollowing out letter loops that were mistakenly filled in the original binarization. Because the net effect is usually beneficial, the experiments adopt this procedure.

A final wrinkle may reflect a quirk of the ground truth data used in the experiments. A visual examination of the discrepancies between initial computed binarizations and the ground truth provided shows a disproportionate number of false positives positioned on the northwest border of each ink component. A single erosion of the binarization output to remove all northwest corner pixels (those with background to north and west of them) restores isotropy to the error profile and consistently improves the quantitative results. Since the algorithm development provides no justification for such an operation, it may reflect the way in which the ground truth was developed. The experiment section reports numbers both with and without this adjustment.

## III. Experiments

This paper adopts its experimental framework from the two DIBCO contests [3], [13], using a small set of documents for training and a separate set for testing. DIBCO-09 used a selection of ten test documents with ground truth binarization; five were handwritten and five printed. All contain one or more features known to hinder standard binarization algorithms: stains, bleed-through, colored text, large areas of background, and unusual fonts. Following the contests' conclusion the document images and ground truth were released as public data sets.

Results in the contest were evaluated on four measures of binarization quality: F-measure, peak signal-to-noise ratio, negative metric rate, and misclassification metric penalty. Good binarizations maximize the first two of these and minimize the latter two. Formulas for the four quantities appear below, assuming the following definitions: $N_{TP}$,

$N_{FP}$, $N_{TN}$, $N_{FN}$ are respectively the number of true positive, false positive, true negative, and false negative identifications of ink pixels; $T_{ij}$ is the ground truth labeling, and $D_{ij}$ is the distance of each pixel to the boundary contours of the ground truth.

$$F = \frac{2 \cdot R \cdot P}{R + P} \tag{12}$$

where

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \tag{13}$$

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}} \tag{14}$$

$$PSNR = 10 \log \left( \frac{1}{MSE} \right) \tag{15}$$

where

$$MSE = \frac{1}{n \cdot m} \sum_{i=1}^{m} \sum_{j=1}^{n} (B_{ij} - T_{ij})^2 \tag{16}$$

$$NRM = \frac{R_{FN} + R_{FP}}{2} \tag{17}$$

where

$$R_{FN} = \frac{N_{FN}}{N_{FN} + N_{TP}} \tag{18}$$

$$R_{FP} = \frac{N_{FP}}{N_{FP} + N_{TN}} \tag{19}$$

$$MPM = \frac{MP_{FN} + MP_{FP}}{2} \tag{20}$$

where

$$MP_{FN} = \frac{1}{SD} \sum_{i=1}^{m} \sum_{j=1}^{n} D_{ij} \cdot (\neg B_{ij} \wedge T_{ij}) \tag{21}$$

$$MP_{FP} = \frac{1}{SD} \sum_{i=1}^{m} \sum_{j=1}^{n} D_{ij} \cdot (B_{ij} \wedge \neg T_{ij}) \tag{22}$$

$$SD = \sum_{i=1}^{m} \sum_{j=1}^{n} D_{ij} \tag{23}$$

Table I summarizes the results of the method described in this paper for several parameter settings on the DIBCO-09 test set. The reported results use parameters tuned on a training set provided to all entrants in the contest, consisting of two handwritten images and two printed images, all with ground truth. Despite the small size of this training set, the method still performs strongly on the test images. It beats 42 of 43 contestants, scoring significantly above the median, and falls an insignificant fraction short of the top method's results. Figure 1 shows the results for a printed document from the test set.

The table also shows several results for comparison. *Default* uses the standard parameter values from Section II-B and omits the post-processing described in Section II-C. Comparative results from DIBCO-09 appear below the double line: Lu & Tan's unpublished method was the highest-rated in the competition, and Fabrizio & Marcotegui was second highest [3].

| Method | F (%) | PSNR | NRM ($\times 10^{-2}$) | MPM ($\times 10^{-3}$) |
|---|---|---|---|---|
| All test documents | 91.07 | 18.51 | 4.39 | 0.67 |
| Print documents | 94.30 | 18.95 | 2.87 | 0.54 |
| Hand documents | 87.31 | 19.66 | 4.92 | 0.70 |
| Default | 90.02 | 17.91 | 3.12 | 1.25 |
| DIBCO-09 first | 91.24 | 18.66 | 4.31 | 0.55 |
| DIBCO-09 second | 90.06 | 18.23 | 4.75 | 0.89 |
| DIBCO-09 median | 83.98 | 15.81 | 4.51 | 5.48 |

TABLE I
RESULTS ON DIBCO-09 TEST IMAGES.

| Tuning | $c$ | $r$ | $t_{lo}$ | $t_{hi}$ |
|---|---|---|---|---|
| Training set | 0.48 | 22.2 | 0.0001 | 0.47 |
| Print (median) | 0.78 | 16.8 | 0 | 0.56 |
| Hand (median) | 0.48 | 21.8 | 0.0005 | 0.39 |

TABLE II
PARAMETER SETTINGS FOUND FROM TRAINING DATA, USED TO GENERATE THE RESULTS IN TABLE I.

The table also shows results achieved for print and handwritten documents separately, although no prior results have been reported in these subcategories for the DIBCO-09 images. Since the number of documents is so small, these experiments combine the training and test images and adopt a leave-one-out methodology. In this framework parameter tuning uses all the documents except one, which is tested using the resulting parameter set. The reported numbers average the results for all documents together. The tuned parameters for each group of documents mostly resemble each other, but differ somewhat between the two groups. The median parameter values found appear in Table II.

Table III shows blind results from the H-DIBCO competition [13], using parameter values trained from the DIBCO-09 handwriting samples. The method does comparatively well: only four of the seventeen entrants placed better. Without a detailed description of the other methods in the contest, the reasons for the differing levels of performance are unclear. Visual inspection suggests that the algorithm was too conservative in identifying faint pen strokes under the chosen parameters; Figure 2 shows one example where thin connecting lines disappear in the result.

| Method | F (%) | PSNR | NRM ($\times 10^{-2}$) | MPM ($\times 10^{-3}$) |
|---|---|---|---|---|
| Tuned on DIBCO-09 | 89.73 | 18.90 | 5.78 | 0.41 |
| H-DIBCO-09 best | 91.78 | 19.78 | 8.180 | 0.231 |
| H-DIBCO-09 median | 85.06 | 17.56 | 10.42 | 0.95 |

TABLE III
RESULTS FROM H-DIBCO [13].


Fig. 1. Binarization of a print document from the DIBCO-09 test set.

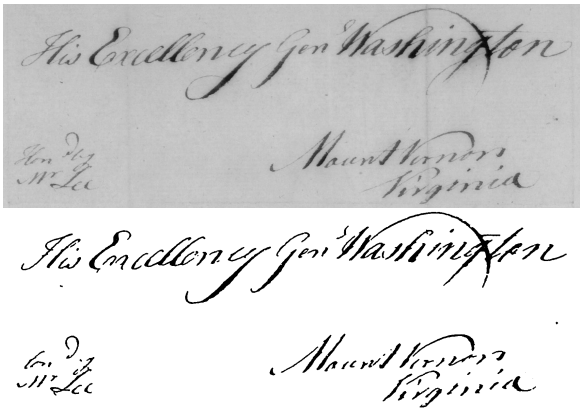

Fig. 2. Binarization of a handwritten document from the H-DIBCO test set (third worst of the ten test images).

## IV. Conclusion

This paper presents a document binarization algorithm based on the Laplacian of the image intensity, with an energy function minimized efficiently via a graph-cut computation. It incorporates Canny edge information in the graph construction to encourage solutions where discontinuities align with detected edges. Graph cut methods have proven successful for other sorts of segmentation but have received fairly little attention to date for document binarization. These results show that they should be taken seriously. A reference implementation of the algorithm in Matlab is available from the author's web site.

Aside from its excellent performance on challenging data sets as compared to state-of-the-art competitors, the algorithm also retains an attractive simplicity. It seems likely that some of the more complicated techniques developed by others to solve specific problems in binarization might prove complementary to the basic approach. For example, parameter tuning in the edge detector currently provides the main mechanism for ignoring marks bleeding through from the reverse side of the paper. Others have developed techniques that explicitly recognize such marks and remove them from the final binarization [16]. Such methods might prove even more effective when used in combination with the basic algorithm in this paper.

## References

[1] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26(9):1124–1137, September 2004.

[2] J. Canny. A computational approach to edge detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8(6):679–714, 1986.

[3] B. Gatos, K. Ntirogiannis, and I. Pratikakis. Icdar 2009 document image binarization contest (dibco 2009). In *Int. Conf. on Document Analysis and Recognition*, pages 1375–1382, 2009.

[4] B. Gatos, I. Pratikakis, and S. J. Perantonis. Efficient binarization of historical and degraded document images. In *International Workshop on Document Analysis Systems*, pages 447–454, 2008.

[5] J. He, Q. D. M. Do, A. C. Downton, and J. H. Kim. A comparison of binarization methods for historical archive documents. In *Int. Conf. on Document Analysis and Recognition*, pages 538–542, 2005.

[6] J. G. Kuk and N. I. Cho. Feature based binarization of document images degraded by uneven light condition. In *10th International Conference on Document Analysis and Recognition*, pages 748 – 752, 2009.

[7] J. G. Kuk, N. I. Cho, and K. M. Lee. MAP-MRF approach for binarization of degraded document image. In *Int. Conf. on Pattern Recognition*, pages 2612–2615, 2008.

[8] J.C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright. Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM Journal of Optimization*, 9(1):112–147, 1998.

[9] R. Milewski and V. Govindaraju. Binarization and cleanup of handwritten text from carbon copy medical form images. *Pattern Recognition*, 41(4):1308–1315, 2008.

[10] W. Niblack. *An Introduction to Digital Image Processing*. Prentice-Hall, Englewood Cliffs, New Jersey, 1986.

[11] N. Otsu. A threshold selection method from graylevel histogram. *IEEE Trans. on System, Man, Cybernetics*, 19(1):62–66, January 1978.

[12] P. Palumbo, P. Swaminathan, and S. Srihari. Document image binarization: Evaluation of algorithms. *SPIE Applications of Digital Image Processing IX*, 697:278–285, 1986.

[13] I. Pratikakis, B. Gatos, and K. Nirogiannis. H-DIBCO 2010 - handwritten document image binarization competition. In *12th International COnference on Frontiers in Handwriting Recognition*, pages 727–732, 2010.

[14] N. Sauvola and M. Pietikainen. Adaptive document image binarization. *Pattern Recognition*, 33(2):225–236, January 2000.

[15] Y. Sun, B. Yuan, Z. Miao, and C. Wan. Better foreground segmentation for static cameras via new energy form and dynamic graph-cut. In *ICPR (4)*, pages 49–52, 2006.

[16] J. Wang, M. S. Brown, and Chew Lim Tan. Automatic corresponding control points selection for historical document image registration. In *Int. Conf. on Document Analysis and Recognition*, pages 1176–1180, 2009.