

Directed Assistance for Ink-Bleed Reduction in Old Documents

Zheng Lu Zheng Wu Michael S. Brown
School of Computing; National University of Singapore
luzheng | wuz | brown@comp.nus.edu.sg

Abstract

Ink-bleed interference is a serious problem that affects the legibility of old documents. Ink-bleed can be reduced using pixel classification based on user-supplied markup that labels examples of ink-bleed, foreground-ink, and background. The main challenge is ensuring that the user's markup sufficiently captures the characteristics of the document. This is particularly troublesome for old documents that can exhibit significant change within the same page.

In this paper, we address this markup problem using a "directed assistance" approach in which the user provides a small amount of initial markup. The image is then classified and regions with low classification confidence are grouped and displayed to the user for another round of markup. The key idea is to direct the user to where markup is needed. In addition, local markup can be weighted in the classification algorithm to produce better results.

1. Introduction and Motivation

Ink-bleed in old handwritten documents is a serious problem faced by archives and libraries. Documents suffering from ink-bleed can exhibit a wide range of visual characteristics making it difficult to develop a universal model for ink-bleed reduction. As a result, the vast majority of existing ink-bleed reduction approaches make assumptions that limit their application to specific cases (e.g. ink-bleed is assumed to be lighter than foreground-ink).

In our previous work [5], we introduced a user-assisted approach that avoids prior assumptions by having the user label examples of ink-bleed, foreground-ink, and background on the input image. This training-data is used to compute per-pixel data-costs for an MRF framework that determines the ink-bleed interference. The only drawback of this approach is if the user markup does not sufficiently capture the characteristics of the entire document then the results will be unsatisfactory. We refer to this style of interactive processing as a "try and see" approach, i.e. the user tries an initial markup and then visually validates the result to see if the markup was sufficient; if not, the user amends the markup and tries again.



Figure 1. (Top): The user supplies a small amount of markup to produce an initial classification that is not optimal. (Bottom): Based on the initial classification, pixels classified with low confidence are grouped (color coded) according to similar image characteristics. This pixel grouping visually directs the user to where more markup is potentially needed.

Our Contribution We introduce a *directed assistance* approach for reducing ink-bleed. Our aim is to direct the user to provide markup where it is needed. To do this, our approach first collects a small amount of user markup and classifies the image. Pixels classified with low-confidence are grouped together into regions based on similar image characteristics. The user can then provide more markup in these segmented regions as needed for another round of classification. This new directed markup should better represent the characteristics of the document and improve results. Figure 1 shows an example. The procedure can be performed iteratively until a desired confidence is achieved. Our directed assistance approach serves two purposes: 1) it provides a convenient way for the user to supply additional markup in potentially erroneous regions (i.e. low classification confidence) and; 2) it allows the region-based markup to be weighted to improve subsequent classification.

The remainder of this paper is organized as follows: section 2 discusses related work; section 3 and section 4 discuss the overall framework and directed assistance implementation; section 5 shows several results; section 6 concludes our work.

2. Related Work

Related work pertaining to ink-bleed reduction and interactive computer vision are discussed.

Ink-Bleed Reduction A great deal of work addressing ink-bleed involves some form of local, global, or adaptive thresholding (e.g. [8, 10, 3]). Other approaches perform signal separation to extract the ink-bleed, foreground-ink, and background from the RGB input (e.g. [12, 13]). Tan *et al.* [11] and Wang *et al.* [14] enhanced the input image first via selective sharpening and dampening in the wavelet domain to help improve thresholding of ink-bleed pixels. Work by Wolf [15] used a two-layer MRF to separate the ink contribution from the front and back page using a single image of the front. With the exception of [15] these approaches all assume that ink-bleed is *always* lighter in intensity than foreground-ink; an assumption that is not always valid (see Figure 1). In the case of [15], the ink-bleed and foreground-ink characteristics are assumed to be similar over the whole document. This assumption is contrary to our examples that have spatially varying characteristics over the whole document (again see Figure 1). Another problem with these previous approaches is the need for complex parameter tuning. For example, Tan *et al.* [11] tunes six parameters per example.

Our previous work, Huang *et al.* [5], avoided prior assumptions and parameter tuning by instead having the user provide examples of the ink-bleed, foreground-ink, and background. This work however, placed the burden on the user to provide sufficient markup in a single attempt. Insufficient markup must be verified and amended by the user. The impetus of the work in this paper is to make markup easier and more controlled.

Interactive Computer Vision Interactive computer vision applications and algorithms enlist user assistance to provide high-level visual information that is otherwise difficult to derive automatically. In many cases, the user simply denotes regions for processing (e.g. image inpainting [1], natural image matting [2]), however, user-assistance can also take the form of training-data collection, such as with Lazy Snapping [7].

Our approach is similar to Lazy Snapping [7] which segments out foreground objects from an image based on user-supplied training examples. Lazy Snapping is also a “try and see” approach, where markup is performed and the results are visually evaluated. This is reasonable for object segmentation since visually validating a large segmented foreground object is simple for a user. Visually validating

handwritten foreground strokes, on the other hand, is significantly more tedious. Our directed assistance procedure provides a way to help the user to see potential mistakes as well as control the markup at the same time.

3. System Overview

Our framework is fashioned after the ink-bleed reduction approach in Huang *et al.* [5]. Relevant components of that work are briefly described here.

3.1. Input and Initial Markup

The input is two images of the front and back page of a single document. These images are aligned (the back image is mirrored) such that a pixel, p , in the front and back image represents the same physical location on the front and back of the document. Details to how this alignment can be performed can be found in [5].

Given two aligned images, the user supplies markup on the front image (in the form of sparse strokes and points) to specify foreground-ink, ink-bleed, and background. For each pixel, p , in the markup, a 2D feature vector is defined as $\rho_p = [C_f(p), C_b(p)]$, where $C_f(p)$ and $C_b(p)$ are the intensities of the front and back image at pixel location p respectively.

3.2. KNN Classification

Given the training-data from the user-supplied markup, K-Nearest Neighbor (KNN) classification is performed as follows. Training features are divided into three sets: $\{\rho_i^{\mathcal{F}}\}_{i=1}^L$, $\{\rho_j^{\mathcal{I}}\}_{j=1}^M$ and $\{\rho_k^{\mathcal{B}}\}_{k=1}^N$, which represent foreground-ink, ink-bleed and background respectively. Each element ρ in the sets is the 2D feature vector described above and L , M and N are the respective set sizes.

To classify an unlabeled pixel in the front image, we extract its feature vector ρ_p and compute the Euclidean distance between ρ_p and all the $L + M + N$ training vectors. Based on these distances, we select the top- K closest vectors where $K = 8$. The top- K vectors are denoted as $\{\rho_m\}_{m=1}^K$ and are further divided into three index sets $\pi^{\mathcal{F}}$, $\pi^{\mathcal{I}}$ and $\pi^{\mathcal{B}}$ according to their labels. The pixel p is classified by the majority class of $\{\pi^{\mathcal{F}}, \pi^{\mathcal{I}}, \pi^{\mathcal{B}}\}$ expressed as:

$$l_p = \arg \max_{\theta} \|\pi^{\theta}\|, \quad (1)$$

where θ is a label in $\{\mathcal{F}, \mathcal{I}, \mathcal{B}\}$, and $\|\pi^{\theta}\|$ is the number of elements in set π^{θ} . In the event of a tie in Eq. 1, l_p is assigned with the following label priority $\{\mathcal{F}, \mathcal{I}, \mathcal{B}\}$.

This feature collection procedure with KNN classification serves as the basis for our application to reduce ink-bleed. The first time this is performed, we refer to this procedure (and the labeled data) as the *global markup* and

global classification. The following section describes the steps that will direct the user to regions that require further assistance. Note that the KNN algorithm described above will be changed to weight local training-data higher in the final classification decision.

4. Directed Assistance Framework

Step-by-step details of the directed assistance framework are described here.

4.1. Determining Low-Confidence Pixels

In our framework, global markup is assumed to be only a few strokes and points. This initial markup is used to classify the entire document. Afterwards, low-confidence pixels will be determined and grouped into similar regions based on image characteristics. While the per-pixel KNN results can be converted into confidence measurements, we find the following procedure produces a better confidence map for finding potential errors. To obtain the confidence score c_ρ for a feature vector ρ , we find its similarity S_ρ^θ to each class θ as the log conditional probability $\ln p(\theta|\rho)$. To estimate $p(\theta|\rho)$, we use a Gaussian mixture model (GMM) and assume that the data of each class follows a Gaussian distribution. For each class obtained by the KNN classification, the prior probability P_θ , the mean vector μ_θ and the covariance matrix Σ_θ are calculated, and S_ρ^θ is given by:

$$S_\rho^\theta = -\frac{1}{2}(\rho - \mu_\theta)^T \Sigma_\theta^{-1} (\rho - \mu_\theta) - \frac{1}{2} \ln |\Sigma_\theta| + \ln P_\theta, \quad (2)$$

where the constant item is omitted.

To compute the confidence of a feature point ρ , we compare its similarity values among the three classes. If the differences among the three similarity values are small, ρ will lie on a decision boundary between classes and can be considered to be of low-confidence; if one of the similarity values is much greater than the other two, ρ is considered confident. The confidence score c_ρ is therefore defined as:

$$c_\rho = |S_\rho^{\mathcal{F}} - \max(S_\rho^{\mathcal{I}}, S_\rho^{\mathcal{B}})|, \quad (3)$$

where only the greater value between $S_\rho^{\mathcal{I}}$ and $S_\rho^{\mathcal{B}}$ is compared with $S_\rho^{\mathcal{F}}$. This formulation considers the decision boundaries between foreground-ink and ink-bleed, and foreground-ink and background. Note that we are not concerned when ink-bleed is mistaken as background and its decision boundary is therefore ignored. With c_ρ calculated, a confidence map can be formed by assigning each pixel p 's intensity value as c_{ρ_p} .

To identify low-confidence regions, the confidence map is binarized using standard two-class threshold selection based on Otsu's method [9]. To remove noise-like regions in the binary map, the confidence map is down-sampled,

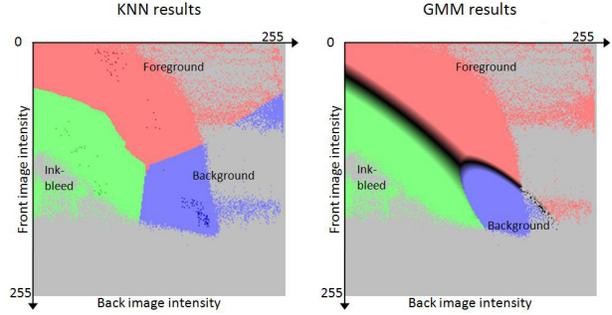


Figure 2. (Left) training-data (shown slightly darker) and the resulting KNN classification decision boundary plotted in 2D feature space. (Right) GMM classification result and the low-confidence points (shaded in black) located along the decision boundary.

then binarized and finally restored to the original size. The sampling ratio is set to the estimated stroke width in the document. This stroke width is estimated using a 50% trimmed mean (i.e. the first 25% and last 25% are ignored) of the lengths of sequences of consecutive foreground pixels along each row and column of the initial global classified image.

Figure 2 shows an example of the decision boundary created by KNN classification with an initial markup, and the low-confidence feature points that are obtained by our procedure. Note that the decision boundary is changed with the calculation of the feature points' similarities to the three classes. The new label of pixel p is decided by the Bayesian classification rule based on the GMM estimated above, i.e.

$$l_p = \arg \max_{\theta} S_{\rho_p}^\theta. \quad (4)$$

Compared with the KNN result, the new boundary is smoother and the range of the background class shrinks, which reflects the true background class distribution which densely populates this region of the feature space. Figure 3 gives an example of a confidence map and its corresponding map of low-confidence regions, in which most of the pixels with erroneous classification results have been marked. In the confidence map, the pixels corresponding to foreground-ink are often circled by dark outliers, which are caused by ink diffusion. By first down-sampling by the estimated stroke width and then thresholding the confidence map, outliers around high-confidence pixels are removed.

4.2. Region Grouping and Directed Markup

The identified low-confidence pixels are grouped into regions that share similar visual characteristic in the document. The user can then provide new markup inside these regions. Local markup performed in any of these directed regions will be applied to all pixels that belong to the same group.

The regions can be effectively grouped using the the local intensity histogram as follows. For each low-confidence

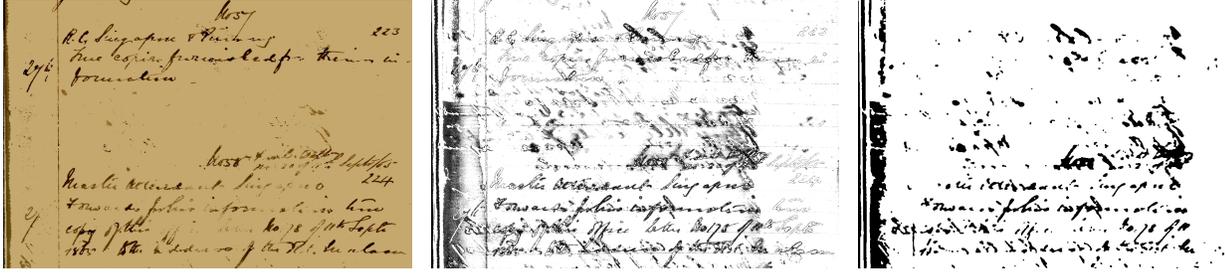
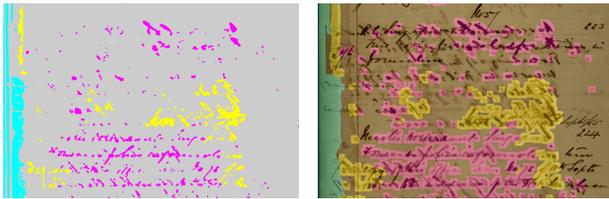


Figure 3. (Left) Classified image where the foreground-ink pixels are displayed in their original colors, and the ink-bleed and background pixels are displayed by the average color of the background pixels. (Middle) The confidence map related to the classified image, where dark pixels are of low confidence, while light pixels are of high confidence. (Right) The map of low-confidence regions, i.e. the binarized confidence map, where black pixels are low-confidence pixels.



(a) Grouped regions (b) Enlarged region map

Figure 4. (Left) Grouped regions. (Right) The enlarged region overlaid on the input image.

pixel, we collect pixel intensities within an image window that 2 times the estimated stroke width. The histogram is discretized into 16 bins to produce a 16D feature vector. The data is then clustered using K-means (see [4]) to find several results with increasing number of clusters. The silhouette statistic [6] is used to determine the best cluster size (i.e. K) from these results. In our implementation, we set the maximum number of clusters to $K = 3$, which appears to be well suited to the documents we process. A larger maximum number of possible regions can be easily increased if desired.

After the regions have been determined, region growing is used to expand their overall size for clearer visualization and easier marking by the user. Figure 4 shows an example of such grouped regions and enlarged region map, where different colors represent the different regions. Figure 4 also shows an example of how these regions will appear to the user for markup.

While global training-data is collected by the initial global markup, local training-data is collected in subsequent rounds. Any markup made in a low-confidence region is collected as local training-data and used for *all* pixels grouped into that region, i.e. all pixels sharing the same color coding. The user does not need to provide all three types of markup in a low-confidence region, instead they only need to provide additional labels (i.e. ink-bleed, foreground-ink, or background) that they feel is necessary to help that region. Furthermore, these displayed regions

only represent where low-confidence results have been detected. Low-confidence does not always mean there is an error. As such, some regions may require no further markup and can be left alone.

When new markup is added to a region, all pixels within that region will be classified again using a KNN algorithm similar to the one described in Section 3. However, the KNN algorithm used with the low-confidence regions will favor the local training-data more than initial global training-data. This is achieved by assigning different weights to each of top- K nearest neighbors according to which set of training data it belongs to. The original KNN decision rule, expressed earlier in Eq. 1, is modified as follows:

$$lp = \arg \max_{\theta} (w_g \|\pi_g^{\theta}\| + w_l \|\pi_l^{\theta}\|), \quad (5)$$

where w_g and w_l are weights for global and local training data π_g^{θ} and π_l^{θ} are subsets of π^{θ} which correspond to global or local training data, $\|\pi_g^{\theta}\|$ is the number of elements in set π_g^{θ} , and $\|\pi_l^{\theta}\|$ is the number of elements in set π_l^{θ} . We set $w_g = 1$ and $w_l = 2$ to weight local training data labeled by directed markup more than global training data.

All other pixels not within a low-confidence region, or whose regions had no additional local markup made, are not changed.

4.3. Iterations

Our approach is intended to be performed in an iterative fashion. The procedure is stopped if the confidence threshold described in 4.1 or the number of low-confidence pixels drops below a fixed value, or the user provides no more markup. In general, we find that in most cases only one iteration is needed for our task of ink-bleed reduction. In the following, we show an example that uses two rounds of directed markup.



Figure 5. (Top) First image shows the input image (the global markup is out of the region). Confidence scores are shown in the second image. Third image shows the classification result. (Bottom) Three detected regions (color coded) from the initial classification are shown. The new local markup made in the detected regions, new confidence map, and new results with zoomed comparisons are shown on the right.

5. Results

Several examples are used to demonstrate the ability of our approach to direct the user’s markup. Note that we do not compare results with automatic ink-bleed reduction approaches (e.g. [3, 11, 15]). Comparisons previously performed in [5] demonstrate that a user-assisted approach produces better results.

Figure 5 shows a portion of a document that has dark ink-bleed and a water-stain. Figure 5(top) shows the initial classification in which several errors occur, including background misclassified as foreground-ink in the water-stain area as well as dark ink-bleed misclassified as foreground-ink. The global markup is made outside the region shown and consists of only a few strokes. Per-pixel confidences are shown. The clustered low-confidence regions are shown in Figure 5(bottom) where new markup is applied. The new result shows the pixels’ confidences significantly improve with the directed markup.

Figure 6 compares the results of using the training-data in a global versus local manner with the KNN classifier. In this example, the markup drawn in the directed regions are used for *global classification* as described in Section 3. The local KNN produces better results.

Figure 7 shows a full page result where the initial global markup is shown. After markup, three regions are found and presented to the user. Only a small amount of additional markup is used to improve the overall classification. The

“try and see” approach in our initial work [5] can potentially take several tries. In addition, mistakes would need to be verified by the user without the aid of highlighting.

Figure 8 shows three rounds of markup applied on a difficult document. The first global markup results in a large amount of low-confidence pixels. The first round of directed markup improves the overall classification. The second round of directed markup helps to reduce the remaining noise.

6. Discussion and Summary

We have described a user-assisted framework that effectively directs users in their efforts to supply training-data for the task of ink-bleed reduction. Our approach provides a mechanism to incrementally train a classifier in a directed manner. Several issues are discussed here.

Too Obvious It is obvious that adding more quality training-data will produce a better result. However, from an application standpoint, the traditional “try and see” approach used by many supervised learning tasks is not well-suited for use on old and deteriorated documents. Visually determining image regions that warrant markup is not trivial, and visually verifying subsequent mistakes is equally tedious. Directed assistance provides a reasonable way to address both of these issues.

Segment First An alternative strategy to directing user markup would be to segment the entire image first, then

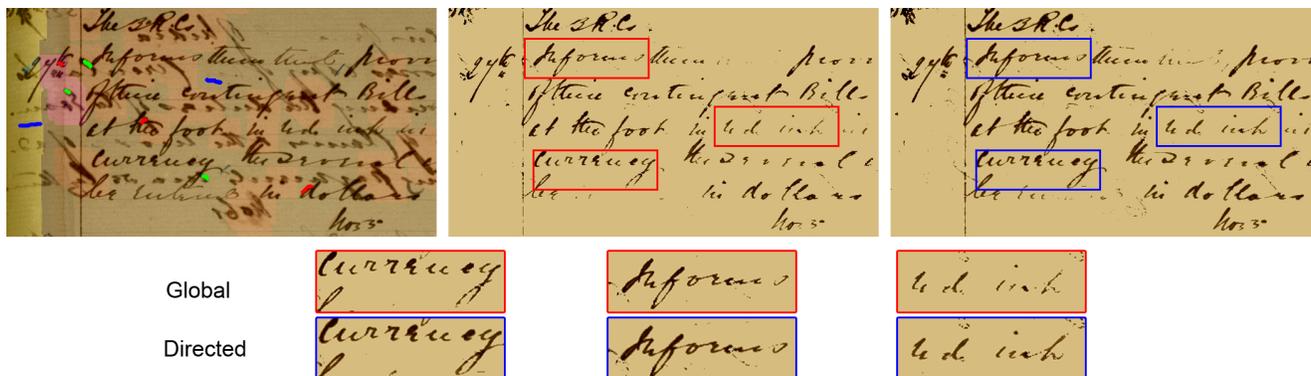


Figure 6. [Global vs. Local] (First) Image with markup drawn in low-confidence regions, (Middle) KNN result treating *all* markup as global markup, (Last) Locally weighted KNN result using our method with initial global classification followed by local weighting of training-data applied to low-confidence regions only.

present the segmented regions to the user for markup. While this would help the user visually determine regions that share similar characteristics, it may also overwhelm the user from the start. Pre-segmentation also does little to help the user verify the results afterwards. Our approach can be considered a lazy-evaluation strategy that segments the image only as needed when classification confidence is low. It also helps to visually direct the user to where markup is needed.

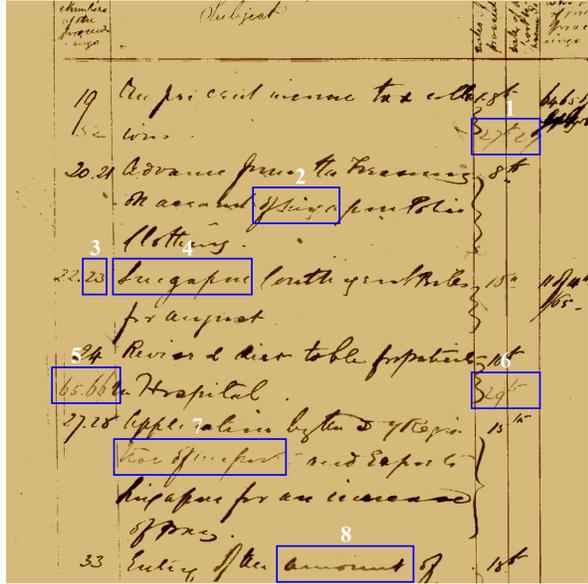
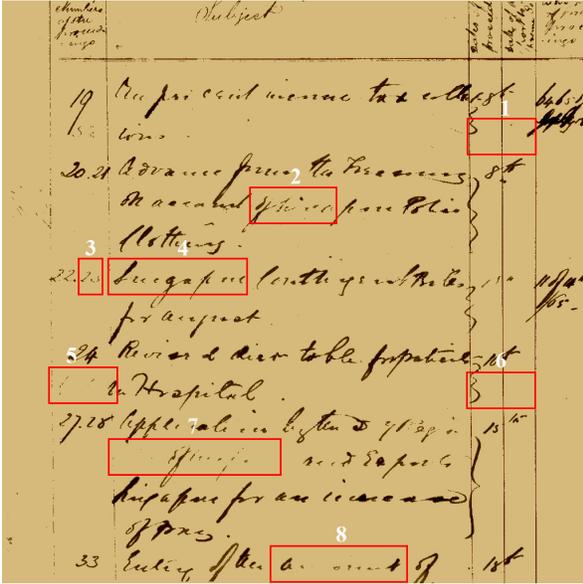
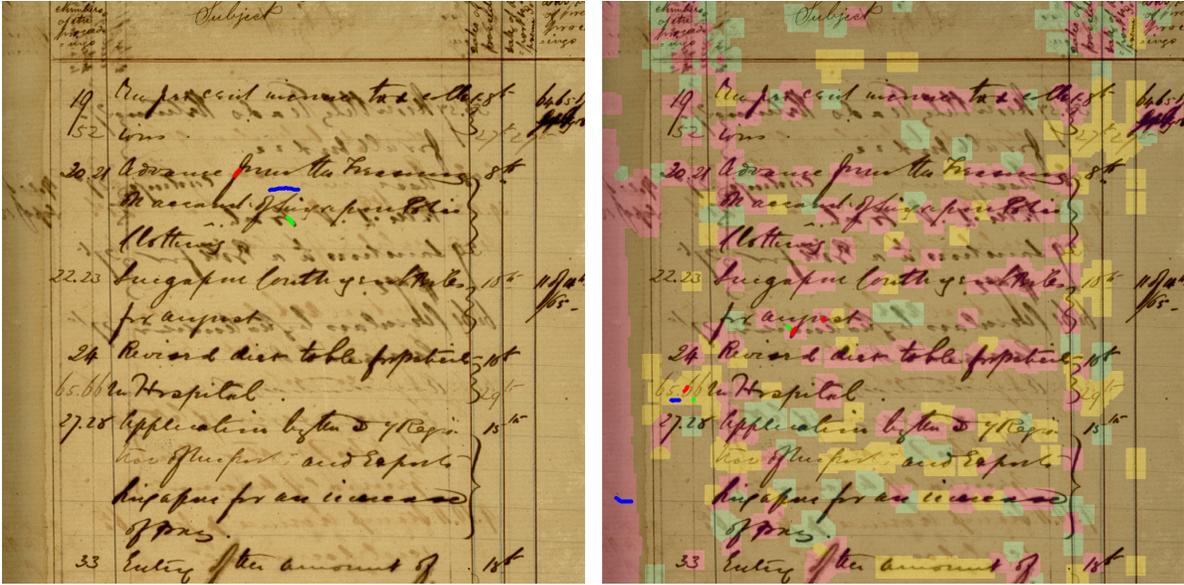
Convergence and Overall Quality As shown in Figure 8, there is a diminishing return as the number of iterations of markup is performed. In our application the main benefits are gained after a single iteration. The quality of our output is only as good as the classifier and features used. Figure 8 is a difficult example where some ink-bleed and foreground-ink cannot be distinguished by our classifier. At some point, we will hit a steady state where more markup will not help improve classification confidence. However, since the user is actively involved in the iterations they can easily determine when to stop. Lastly, there is still room to find better features and classifiers that can improve the ink-bleed reduction results, however, we believe directed markup can still play an important role in future work.

In summary, we have presented a novel approach to ink-bleed reduction by directing the user’s markup. Our directed assistance approach provides a mechanism to show where markup is needed, as well as improve classification via local weighting of the classifier.

References

- [1] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. In *SIGGRAPH*, 2000.
- [2] Y. Chuang, B. Curless, D. H. Salesin, and R. Szeliski. A bayesian approach to digital matting. In *CVPR*, 2001.
- [3] F. Drira, F. L. Bourgeois, and H. Emptoz. Restoring ink bleed-through degraded document images using a recursive unsupervised classification technique. In *Document Analysis Systems (DAS’06)*, 2006.

- [4] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience Publication, 2000.
- [5] Y. Huang, D. Xu, and M. S. Brown. A framework for removing ink-bleed in old documents. In *CVPR*, 2008.
- [6] L. Kaufman and P. J. Rousseeuw. *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley-Interscience Publication, 1990.
- [7] Y. Li, J. Sun, C.-K. Tang, and H. Shum. Lazy snapping. *ACM Transactions on Graphics (SIGGRAPH)*, 23(3):303–308, 2004.
- [8] F. C. Mintzer et al. Toward on-line, worldwide access to vatican library materials. *IBM Journal of Research and Development*, 40(2):139–162, 1996.
- [9] N. Otsu. A threshold selection method from gray level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9:62–66, 1979.
- [10] Z. Shi and V. Govindaraju. Historical document image enhancement using background light intensity normalization. In *ICPR*, 2004.
- [11] C. L. Tan, R. Cao, and P. Shen. Restoration of archival documents using a wavelet technique. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(10):1399–1404, 2002.
- [12] A. Tonazzini, L. Bedini, and E. Salerno. Independent component analysis for document restoration. *International Journal on Document Analysis and Recognition*, 7(1):17–27, 2004.
- [13] A. Tonazzini, E. Salerno, and L. Bedini. Fast correction of bleed-through distortion in grayscale documents by a blind source separation technique. *International Journal on Document Analysis and Recognition*, 10(1):17–25, 2007.
- [14] Q. Wang, T. Xia, L. Li, and C. Tan. Document image enhancement using directional wavelet. In *CVPR*, 2003.
- [15] C. Wolf. Document ink bleed-through removal with two hidden markov random fields and a single observation field. In *Technical Report RR-LIRIS-2006-019*, 2006/2007.



Global

Directed Markup

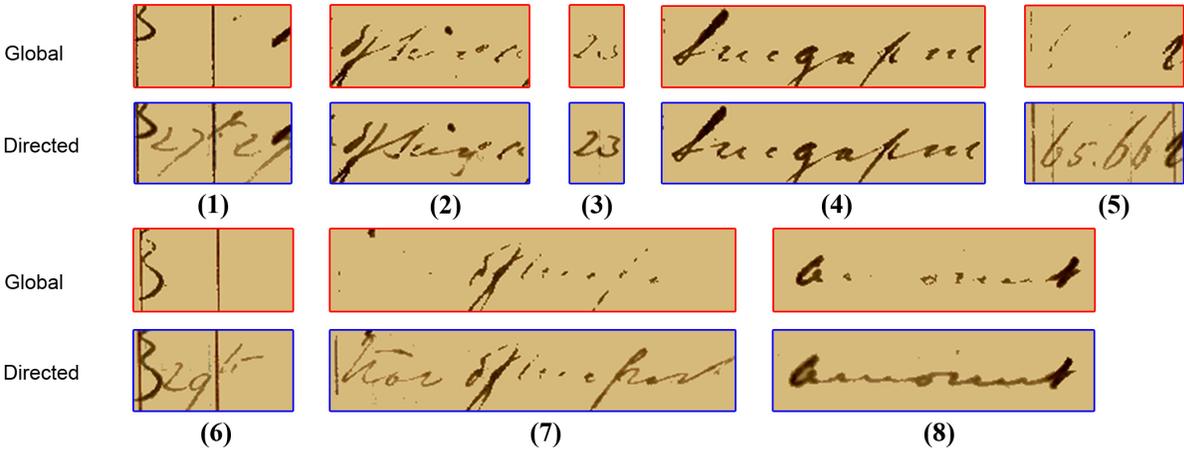
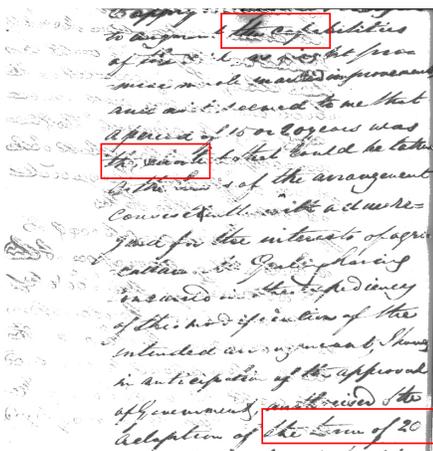
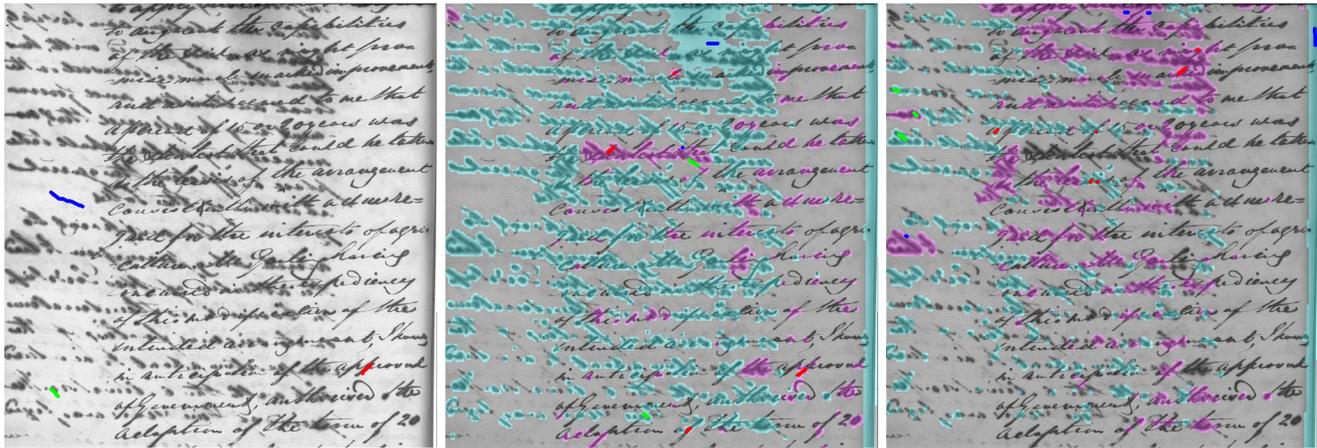
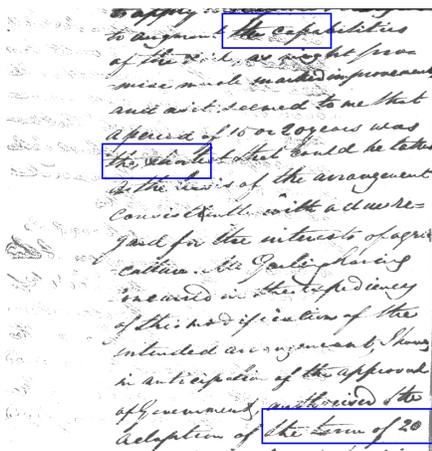


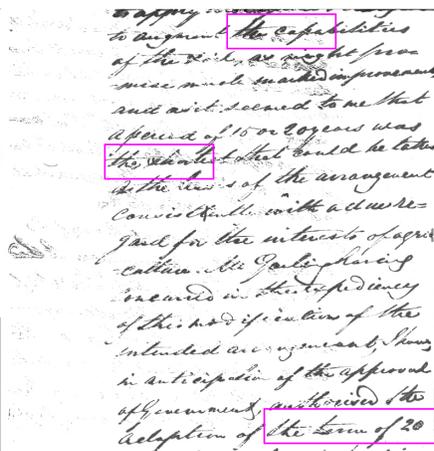
Figure 7. [(Top) Initial global markup and markup in local directed regions. (Middle) Results of the global and local markup (three regions color coded). (Bottom) Eight selected regions zoomed for comparison.



Global



Directed Markup (first round)



Directed Markup (second round)



Figure 8. Three rounds of markup applied on a difficult example with strong ink-bleed. The first global markup results in a large amount of low-confidence pixels. The image is grouped into two regions (color coded). After one round of directed markup low-confidence pixels are reduced, but not completely removed. Another round improves the results slightly (removing some noise). More markup will most likely not help. Zoomed regions are shown for comparison.