Non-rigid Registration and Restoration of Double-sided Historical Manuscripts

Jie Wang and Chew Lim Tan
School of Computing
National University of Singapore
{wangjie,tancl}@comp.nus.edu.sg

Abstract

This paper presents a fully automatic framework for the restoration of double-sided historical manuscripts which are impaired by ink bleed-through distortions. First, the recto side and the verso side of a manuscript are precisely registered with a non-rigid registration method. The registration procedure selects salient control points as features, employs a free-form transformation model as mapping function and uses residual complexity as similarity measure. Next, we extract prominent foreground feature images with the two registered images. By referring to the feature images, we iteratively perform wavelet decomposition and construction to smear bleed-through interferences and to enhance foreground texts on the manuscript. We test the proposed framework with real manuscripts from the archives and the experimental results are encouraging.

1. Introduction

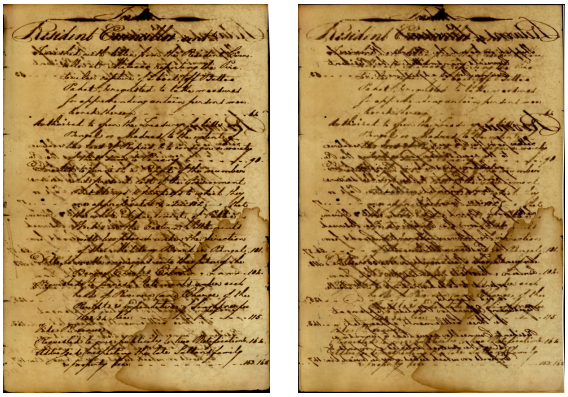
Historical document images which are preserved in the form of scanned images or microfilms are usually degraded by various distortions. On the one hand, due to aging or thick binding effects, these documents are difficult to be presented flat to the image capturing plane, which causes photometric distortions such as uneven surfaces, optical blur, etc. On the other hand, geometric distortions such as translation, rotation and scale difference are inevitably introduced due to imprecise positioning of pages and different scale settings. In particular, the foreground texts on double-sided manuscripts are additionally impaired by ink seeping from the reverse side of the document due to the low quality of carrying medium. This type of distortion is usually referred to as bleed-through distortion as shown in Figure 1. All these distortions should be corrected to facilitate human perception and machine recognition.

Early researches to the problem of bleed-through correction focused on threshold-like techniques such as multistage thresholding [5], local adaptive filters [3] and noise-based thresholding [2] to extract main text. In common, these approaches process the two sides of a document independently and are limited by their assumption that the foreground text layer and the background layer of a document are separable in the term of intensity.

Considering the fact that bleed-through strokes on one side is caused by the foreground texts on the other side of the document, complicated methods were proposed to simultaneously process the two sides of a document. These methods require the two sides of a document being available and being precisely registered. Then they can separate foreground texts and bleed-through interferences with independent component analysis techniques [11, 12], wavelet decomposition/composition [9] and per-pixel classification [4]. As more information is employed for separation, this line of methods are supposed to be more robust especially with severely degraded manuscripts where the foreground texts and bleed-through strokes are heavily overlapped. The problem however is that precisely registering the two sides of a document is really difficult. Therefore, currently the registration task for this application is mostly done manually, which is error-prone, time consuming and tedious. To automate this registration task, [15] proposed point pattern matching strategy and [14] also tried block-based rigid matching using correlation coefficients. These methods are not satisfying as they are either computational intensive or susceptible to matching errors.

This paper presents a fully automatic framework that restores historical manuscripts impaired by ink bleed-through distortions. First, we extract candidate control points from the two images of a document by filtering their gradient and intensity maps. The correspondences between these points are then established by minimizing a distance consisting of gradient, intensity and displacements. We model the spatial relationship between the two images with a free-form transformation which is based on a B-splines. The estimation of the transformation model takes advantage of residual complexity similarity measure and derivatives regularization term. With the registered images, we extract prominent foreground strokes from the two images and iteratively per-

1520-5363/11 $26.00 © 2011 IEEE
DOI 10.1109/ICDAR.2011.276



(a) recto image (b) flipped verso image

Figure 1. The two sides of a historical document degraded by various distortions

form wavelet decomposition and construction so that foreground texts are enhanced and bleed-through strokes are simultaneously smeared. Compared with our previous work, this framework is more general, accurate and robust to deal with severely degraded historical document images.

2. Non-rigid Registration

At the beginning, we assume the two side images of a manuscript are of the same spatial resolution and have been identified and paired. Such assumptions are more than reasonable for typical archival imaging. Before any subsequent processing, a mirror image of the verso side is acquired by flipping the verso image horizontally or vertically with respect to the layout of the document. Without loss of generality, the flipped verso image is referred to as the target image and the recto image as the reference image.

2.1. Feature Points Selection

As high registration accuracy is required and sophisticated image features are difficult to be extracted from historical document images, we register the two images on pixel level with salient control points detected by following the below steps:

- Select candidate points (from both two images) which satisfy two conditions: their gradient directions are in the range of -1.8-0.5 (reference image) and -0.5-1.8 (target image); their gray intensities are higher than T_i which is determined by an adaptive binarization procedure [7]. According to our experience and experiments, points satisfying these conditions are most possible to cause strong bleed-through interference. Therefore, in this way, we make sure that most selected control points on the reference image have a corresponding point on the target image.

- Establish the correspondences between the points on the reference image and the points on the target image by minimizing a distance measure combining gray intensity I, gradient magnitude M, gradient direction θ and displacements $x - x'$, $y - y'$:

$$\mathbf{dis}(x, y, x', y') = w_i(I_{xy} - I'_{x'y'}) + w_m(M_{xy} - M'_{x'y'}) + w_\theta(\theta_{xy} - \theta'_{x'y'}) + w_d\sqrt{(x - x')^2 + (y - y')^2} \quad (1)$$

  where w_i, w_m, w_θ and w_d are the weights that specify the relative importance of the four components and are empirically set as (1, 3, 10, 20) in our experiments.

- Check the displacements of each point pair block by block to correct non-collectively occurring mismatches. As the established correspondences are supposed to be locally continuous, point pairs with a displacement larger than the median displacement of the local block are regarded as mismatches and discarded.

- Switch the roles of recto image and verso image to conduct consistency checking. This step is useful for documents with severe bleed-through distortions on both sides and could further refine detected control point pairs. Feature 2 shows some of the extracted corresponding points.

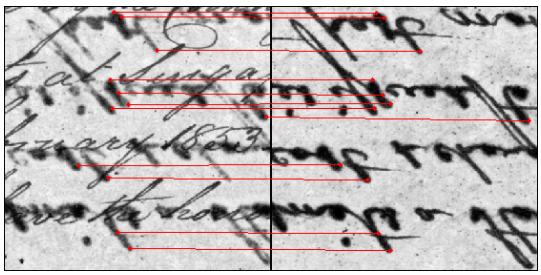


Figure 2. Samples of detected corresponding point pairs.

2.2. Free-form Transformation Model

We model the spatial relationship between the two images to be registered with a free-form transformation which is based on a 3 hierarchical levels of B-spline [8]. It consists of two components which are optimized sequently in a hierarchical structure (coarse to fine).

$$T(x, y) = T_{global}(x, y) + T_{local}(x, y) \quad (2)$$

T_{global} is an affine transformation (as in equation 3) which is supposed to capture the global rotation, translation and non-isotropic scaling deformation between the two images.

$$T_{global}(x,y) = \begin{pmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (3)$$

T_{local} is defined as the tensor product of the cubic B-splines (as in equation 4), which describes the local deformation. This transformation model is chosen as it usually produces smooth and continuous transformation. Besides, B-splines are locally controlled, which makes them computationally efficient for a large number of control points.

$$T_{local}(x,y) = \sum_{m=0}^{3}\sum_{n=0}^{3} B_n(u)B_m(v)\phi_{i+n,j+m} \quad (4)$$

where $i = \lfloor x/n_x \rfloor - 1$, $j = \lfloor y/n_y \rfloor - 1$, $u = x/n_x - \lfloor x/n_x \rfloor$, $v = y/n_y - \lfloor y/n_y \rfloor$ and $[n_x, n_y]$ is the size of the control mesh. ϕ denotes the mesh of control points and B_i refers to i^{th} basis function as follows.

$$\begin{cases} B_0(u) = (1-u)^3/6 \\ B_1(u) = (3u^3 - 6u^2 + 4)/6 \\ B_2(u) = (-3u^3 + 3u^2 + 3u + 1)/6 \\ B_3(u) = u^3/6 \end{cases} \quad (5)$$

2.3. Cost Function Optimization

To estimate the parameters of the transformation model, we define a cost function consisting of two parts.

$$C = C_{sim}(I_{xy}, T(I_{xy})) + \lambda C_{smo}(T) \quad (6)$$

C_{sim} refers to the similarity of the registered images, which ensures an accurate transformation model. C_{smo} is a regularization term to make sure the estimated transformation is smooth. λ weights the relative importance of the two constrains and is set to 0.02.

The similarity measure used in this work is residual complexity which is the compression complexity of the residual images between the two registered images [6]. This measure is used as the intensities of matching components (foreground strokes vs bleed-through interferences) are significantly different, dependent and non-stationarily distorted.

$$\begin{aligned} &C_{sim} = \textstyle\sum_{n=1}^{N} log((\mathbf{q}_n^T\mathbf{r})^2/\alpha + 1) \\ &\mathbf{q}_n = dctn(\mathbf{r}) \\ &\mathbf{r} = I_{ref} - T(I_{tar}) \end{aligned} \quad (7)$$

where **r** refers to the residual image and $dctn$ is multidimensional **DCTs**. α is a trade-off parameter and is set to 0.05 in our experiments.

For implementation simplicity and time efficiency, C_{smo} is chosen as the space integral of the square of the second order derivatives:

$$C_{smo} = \frac{1}{s}\int\int[(\frac{\partial^2 T}{\partial x^2})^2 + (\frac{\partial^2 T}{\partial y^2})^2 + 2(\frac{\partial^2 T}{\partial xy})^2]dxdy \quad (8)$$

where s denotes the area of the image domain and T represents the estimated transformation. For the purpose of function optimization, gradient descent method is used and the transformation model is iteratively refined with increasing size of control points. In our implementation, the optimization procedure terminates when the relative function difference is larger than a threshold or the maximum number of iterations has been reached. We avoid finding some local minima by reducing the optimization step size and slightly increasing the value of the objective function. Figure 3 shows an example of the registration result.

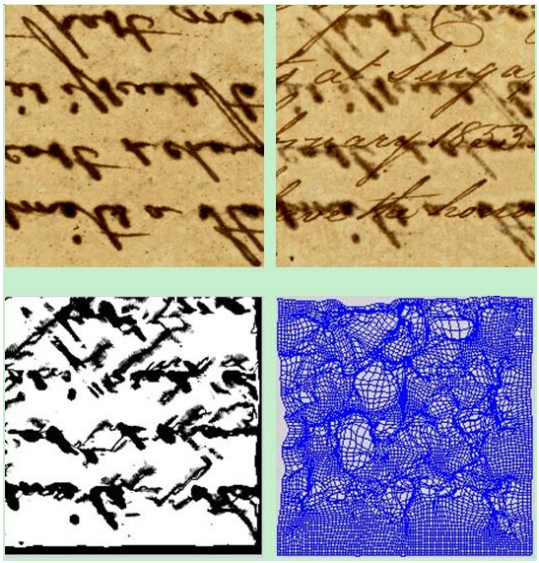


Figure 3. Illustration of registration procedure. Top to bottom and left to right are: reference image, target image, registered image and estimated transformation

3. Ink Bleed-through Correction

Once the two images of a document have been registered, we adopt the wavelet-based technique proposed in [9] to correct the bleed-through distortions on this document. The detailed steps for this approach can be described as follows:

- Compute the foreground overlay image $a(x,y)$ for recto image $f(x,y)$ by:

$$a(x,y) = flip(invert(b(x,y))) + f(x,y) \quad (9)$$

where $b(x,y)$ refers to the registered verso image.

- Weaken the suspected bleed-through interference on the foreground overlay image by scaling it with a non-linear transformation:

$$curve(x) = 2^k - 1 - ((2^k - 1)^2 - x^2)^{0.5} \quad (10)$$

- Detect the foreground strokes on the foreground overlay image using a modified canny edge detector with orientation filters and constraints [1, 10] to form the "enhancement feature image", $E(x, y)$.

- Conduct the above steps on the registered verso image $b(x, y)$ to get the "smearing feature image", $S(x, y)$.

- Decompose an original image into wavelet domain while retaining the size of the image as:

$$wf(x, y) = \{C_j(x, y), D_j^k(x, y), j = 0, \cdots, k = 1, 2, 3\}$$

  where $C_j(x, y)$ is the wavelet approximation coefficient and $D_j^k(x, y), (k = 1, 2, 3)$ are the wavelet detail coefficients at scale j of the wavelet decomposition.

- Modify the wavelet detail coefficients by referring to $E(x, y)$ and $S(x, y)$:

$$D_j^k(x, y) = \begin{cases} e_j^k D_j^k(x, y) & \text{if } E(x, y) == 1 \\ s_j^k D_j^k(x, y) & \text{if } S(x, y) == 1 \end{cases} \quad (11)$$

  where $j = 0, \cdots, J; k = 1, 2, 3$.

- Reconstruct the wavelet representation of the original image with the modified wavelet detail coefficients.

- Iteratively repeat the above decomposition and construction procedure at most 15 times to get restored images with foreground texts enhanced and bleed-through distortion reduced.

- Apply the same edge detector to the enhanced images to obtain final output images. Figure 4 shows a sample image and the restored image. As we can see from the figure, although some foreground strokes are mistakenly broke, the restored image is much better for human perception than the original one. With proper recovering methods to connect the broke foreground strokes, the quality of the resultant images can be further improved.

4. Experiments

The proposed framework has been tested with 28 double-sided historical documents (56 images) from the national archives and the experimental results are encouraging. Our experiments were conducted on a Pentium 4GHZ CPU with

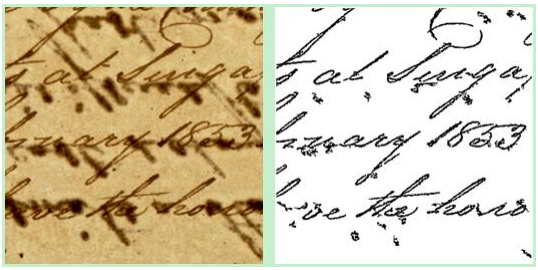


Figure 4. A degraded image (cropped from a large image) and the result after restoration

MATLAB. Typically, it takes 5 minutes to process a pair of images scanned at 150dpi and with a size of 1800×2800. This speed is acceptable since most document restoration tasks are conducted off-line and with more powerful machines and parallel strategies, this time cost can be further reduced. As usual, visual assessment was first performed by the experts working at the achieves to qualitatively evaluate this new restoration framework. In particular, we compared the results from this framework with those of other three methods [10, 13, 15]. As shown in figure 5, this new framework produced better results for most testing images.

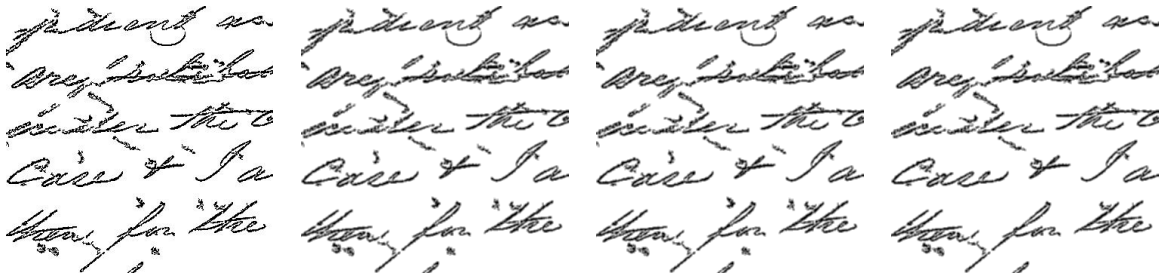
For quantitative assessment, we manually labeled 12 documents (4 slightly degraded, 4 moderately affected and 4 severely impaired as shown in Figure 1) with "ground truth". To be specific, we counted the following numbers: the total number of foreground words N_{fgd} on the image; the number of words detected after applying this restoration method N_{detect}; the number of words correctly detected after restoration $N_{correct}$. When calculating these numbers, connected words were counted separately. N_{detect} included fully detected foreground words and partially or fully detected interfering words. However, only fully detected foreground words were counted in $N_{correct}$. With these numbers, the proposed framework was evaluated with the traditional document analysis metrics as:

$$Precision = \frac{N_{correct}}{N_{detect}}, \quad Recall = \frac{N_{correct}}{N_{fgd}} \quad (12)$$

Table 1 lists the results of 12 degraded documents from the new framework and the other two methods in [13, 15]. As we can see, the average restoration precision and recall of the proposed framework are 89 and 90, respectively, which are higher than those of the other methods. Moreover, experiments show that when working on severely degraded manuscripts, the new framework is more robust.

5. Conclusion and Future Work

In this paper, we propose a fully automatic framework for the restoration of double-sided historical documents. In



(a) Tan [9] (b) Q Wang [15] (c) J Wang [13] (d) Proposed

Figure 5. Comparison of images resulted from different methods

No. of words		47	241	134	79	83	257	209	189	217	145	97	89	Average
Precision (%)	Q Wang [14]	87	57	63	76	82	47	58	76	62	64	71	68	62.4
	J Wang [12]	92	79	84	89	88	73	81	79	78	86	84	90	83.6
	Proposed	97	88	89	95	93	81	87	90	87	84	88	89	89
Recall (%)	Q Wang [14]	91	65	80	82	85	72	69	77	79	68	83	81	77.7
	J Wang [12]	96	76	87	91	90	79	73	87	81	79	91	86	84.7
	Proposed	95	83	93	97	89	86	84	92	87	84	96	94	90

Table 1. Quantitative evaluation and comparison of the proposed method

order to completely remove bleed-through distortions without impairing foreground contents, we first precisely register the two sides of a document. For registration, we select salient control points as matching features and model the spatial relationship between the two images with a B-spline based free-form transformation. We estimate the transformation with a similarity measure called residual complexity and regularize the estimated transformation with smoothness criteria. With the registered images, we then extract two prominent foreground maps and then conduct directional wavelet decomposition and construction to enhance foreground texts and simultaneously smear bleed-through interference. In future, we will exploit post-processing methods to recover mistakenly broke foreground strokes after this restoration framework.

6. Acknowledgements

This research is supported in part by Ministry of Education, Singapore, under research grant R252-000-349-112.

References

[1] R. Cao, C. L. Tan, Q. Wang, and P. Shen. Segmentation and analysis of double-sided handwritten archival documents. In *DAS*, 2000.

[2] H. Don. A noise attribute thresholding method for document image binarization. *IJDAR*, 4(2):231–234, 2001.

[3] K. Franke and M. Koppen. A computer-based system to support forensic studies on handwritten documents. *IJDAR*, 3(4):218–231, 2001.

[4] Y. Huang, M. Brown, and D. Xu. A Framework for Reducing Ink-Bleed in Old Document. In *CVPR*, 2008.

[5] G. Leedham, S. Varma, A. Patankar, and V. Govindarayu. Separating text and background in degraded document images: A comparison of global threshholding techniques for multi-stage threshholding. In *PIWFHR*, 2002.

[6] A. Myronenko and X. B. Song. Image registration by minimization of residual complexity. In *CVPR*, 2009.

[7] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1):62–66, 1979.

[8] D. Rueckert, L. Sonoda, I. Hayes, D. Hill, M. Leach, and D. Hawkes. Nonrigid registration using free-form deformations: Application to breast mr images. In *IEEE TMI*, 1999.

[9] C. L. Tan, R. Cao, and P. Shen. Restoration of archival documents using a wavelet technique. *TPAMI*, 24(10):1399–1404, 2002.

[10] C. L. Tan, R. Cao, P. Shen, Q. Wang, J. Chee, and J. Chang. Removal of interfering strokes in double-sided document images. In *IEEE WACV*, 2000.

[11] A. Tonazzini, L. Bedini, and E. Salerno. Independent component analysis for document restoration. *IJDAR*, 7(1):17–27, 2004.

[12] A. Tonazzini, E. Salerno, and L. Bedini. Fast correction of bleed-through distortion in grayscale documents by a blind source separation technique. *IJDAR*, 10(1):17–25, 2007.

[13] J. Wang, M. Brown, and C. L. Tan. A fully automatic system for restoration of historical document images. In *IAAI*, 2009.

[14] J. Wang and C. Tan. Accurate Alignment of Double-sided Manuscripts for Bleed-through Removal. In *DAS*, 2008.

[15] Q. Wang and C. L. Tan. Matching of double-sided document images to remove interference, pattern recognition. In *CVPR*, 2001.