

# Automatic Corresponding Control Points Selection for Historical Document Image Registration

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

## Abstract

*Image registration is crucial for various image analysis tasks. In particular, most approaches to correction of bleed-through distortion on handwritten document images require the recto image and the verso image to be precisely registered. In this paper, we present a fully automatic method which detects specific number of corresponding control points from historical documents for the purpose of registration. First, candidate points are located by inspecting the gradient direction maps of document images. Corresponding control points are selected based on a dissimilarity metric that incorporates image intensity, gradient magnitude, gradient orientation and displacement. To improve the quality of the detected control points, median filters and consistency checking are applied to correct mismatches. Experiments on real historical document images have shown encouraging results and further improvements can be made by exploiting more sophisticated similarity metric tailored to historical documents' characteristics.*

## 1. Introduction

Historical document images are subject to various types of noise and degradation. In particular, double-sided handwritten historical documents are often affected by ink bleed-through as shown in Figure 1. To facilitate both human perception and machine recognition, these distortions must be corrected. Early work to correct bleed-through distortion on historical document images focused primarily on thresholding techniques such as multistage thresholding [8], local adaptive filters [4] and noise-based thresholding [2]. These methods however are limited as they assume distinct intensities for foreground text, bleed-through and background in a single image.

More sophisticated methods to bleed-through removal include independent component analysis (ICA) [12, 13],

wavelet decomposition/construction [11] and per-pixel classification [7]. In order to use these techniques, the input documents need to be first registered. Furthermore, the accuracy of bleed-through correction heavily relies on the registration precision. Perfect registration of the recto image and the verso image of a page is however difficult for several reasons. Firstly, as the registration is actually between foreground texts and their partially shown bleed-through interference, corresponding points may have extremely different gray levels. Secondly, different positioning of a page during image acquisition results in translational or rotational displacements between the recto image and the verso image. Thirdly, complicated local deformations such as uneven or warped surfaces could be caused by the bounding effect or due to the unevenness of the aging paper. Finally, background noise due to discolorization or stains can affect the registration result.

Due to the above difficulties, the selection of corresponding control points for historical document registration is often performed manually by a document analysis expert [11, 13]. This is time-consuming and tedious when a considerably large collection of documents are to be processed. Techniques for automatic registration are therefore demanded and have been proposed. For example, area-based techniques [15, 14, 7] perform registration using image patches and standard image similarity metrics such as normalized cross correlation or sum of absolute differences, but they are susceptible to matching errors. Control points (salient points or landmarks) selection techniques that have been proposed include corner detection [5, 1] and local extrema of wavelet transform detection [6, 3, 10]. These methods are not applicable in historical document registration due to efficiency reasons and the fact that corners or extreme points are often not present on bleed-through interference.

This paper presents a method to automatically select a specific number of corresponding control points for the registration of historical document images. The actual registration can be performed by using our detected control points to determine the transformation (e.g. affine transform [15]



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

or non-linear deformation [7]) between input images. To achieve this task, we first identify candidate points by analyzing the gradient direction maps and the labeled images of input documents. Then the correspondences between these points are established by locally searching for the point which minimizes the dissimilarity measure that considers intensity, gradient magnitude, gradient orientation and displacement. Median filters are then locally applied to detect and correct non-collectively occurring mismatches. To further improve the precision of the established correspondences, the roles of the target image and the reference image are switched so that consistency checking is conducted.

## 2. Proposed Approach

Our approach assumes that the input images are of the same spatial resolution and the two images of a document have been identified, paired and coarsely aligned. Such assumptions are more than reasonable for typical archival imaging. Before any subsequent processing, a mirror image of the verso image is acquired by flipping the verso image horizontally or vertically with respect to the layout of the document. Without loss of generality, from what follows, the flipped verso image is referred to as the target image and the recto image as the reference image.

### 2.1. Candidate Points Selection

As shown in Figure 1, for Western languages, most foreground texts are written with a notable slanting angle that ranges from  $0^\circ$  to  $90^\circ$ . More importantly, the strokes slanting at particular angles such as between  $30^\circ$  and  $60^\circ$  are much stronger than other portions of the foreground texts, which means they have higher probability of seeping through the page and causing bleed-through interference. In

other words, the original strokes which are responsible for most observed bleed-through interference usually slant at angles between  $30^\circ$  and  $60^\circ$ . Therefore, among pixels on these strokes, more common elements can be found for the purpose of registration. In addition, according to our experience in manually selecting corresponding control points, pixels which are darker and which lie at the locations where only one side of the document has foreground texts are easier to be accurately located and precisely matched. As a result, our method aims to select corresponding control points from points where the gradient directions are between certain range and where no foreground texts and bleed-through overlap.

To extract these points, the gradient magnitude map and the gradient direction map are first determined for the target image and the reference image respectively. Then the labeled images of the target image and the reference image are obtained using Otsu's binarization method [9]. The gradient direction maps and the labeled images are used to identify candidate points, resulting in:

$$\begin{cases} \theta_{can} = \theta_{xy} \otimes (\theta_{xy} > \theta_{low}) \otimes (\theta_{xy} < \theta_{high}) \otimes B_{xy} \\ M_{can} = M_{xy} \otimes (\theta_{xy} > \theta_{low}) \otimes (\theta_{xy} < \theta_{high}) \otimes B_{xy} \\ I_{can} = I_{xy} \otimes (\theta_{xy} > \theta_{low}) \otimes (\theta_{xy} < \theta_{high}) \otimes B_{xy} \end{cases}$$

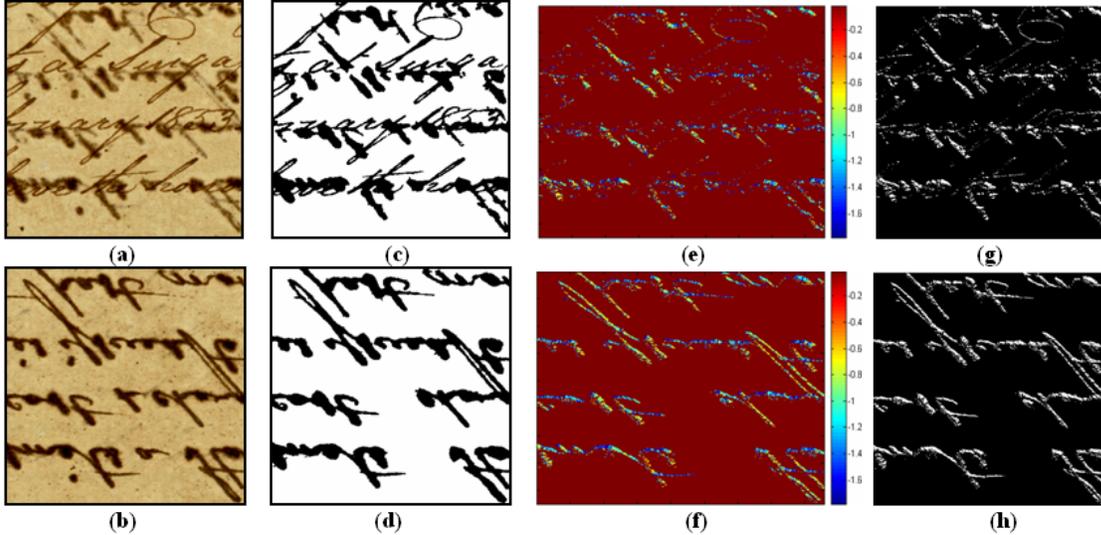
where  $\theta_{xy}$ ,  $M_{xy}$  and  $B_{xy}$  represent the gradient direction map, the gradient magnitude map and the labeled image of the original image  $I_{xy}$ .  $\theta_{can}$  and  $M_{can}$  are the gradient direction and magnitude maps of the identified candidate point sets  $I_{can}$ . Parameters  $\theta_{low}$  and  $\theta_{high}$  are used to threshold the gradient direction map and  $\otimes$  represents the Hadamard product. In our experiments,  $\theta_{low}$  and  $\theta_{high}$  are set as  $-1.8$ ,  $0.5$  and  $-0.5$ ,  $1.8$ . Figure 2 shows the two images of a sample document, the labeled images, the detected candidate points and the gradient direction map for candidate points.

### 2.2. Correspondence Establishment

Once two sets of candidate points have been detected from the reference image and the target image, the correspondences between them are established with a searching procedure which minimizes a dissimilarity metric defined as:

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

where  $I_{xy}$ ,  $M_{xy}$ ,  $\theta_{xy}$  are the intensity, the gradient magnitude and the gradient direction of the candidate point  $(x, y)$  in the target image;  $I'_{x'y'}$ ,  $M'_{x'y'}$ ,  $\theta'_{x'y'}$  are the corresponding values for point  $(x', y')$  in the reference image. The parameters  $w_i$ ,  $w_m$ ,  $w_\theta$  and  $w_d$  are the weights that specify the relative importance of intensity, gradient magnitude,



**Figure 2.** (a-b) recto image and horizontally flipped verso image (cropped from  $1628 \times 2480$ ); (c-d) labeled recto image and verso image; (e-f) band-pass filtered gradient direction maps of recto image and verso image; (g-h) identified candidate control points in recto image and verso image.

gradient direction and displacement when determining the correspondences. In our experiments, they are empirically chosen as  $(1, 3, 10, 20)$  corresponding to  $(w_i, w_m, w_\theta, w_d)$ .

As the differences in the intensities of matching points vary between 0 and 255, the values of gradient magnitude and gradient direction need to be normalized to 0-255. The displacements between matching points vary between 0 and the size of the searching window which is considerably smaller than the other three terms in Equation 1. Usually image intensity and gradient magnitude are used for image matching. However, in this application since the matching pairs are actually original foreground texts and their partially shown bleed-through strokes, the intensities and the gradient magnitudes of two corresponding points are intrinsically different. Gradient direction therefore aids in deciding the correspondence. If a unique correspondence cannot be established from image intensity and gradient, the fourth term selects the correspondence that results in the smallest displacement. Using this dissimilarity metric we find we obtain better results than using only a single term (e.g. intensity or gradient orientation only).

For each candidate point  $(x', y')$  in the reference image, all the candidate points in the square window which has a size of  $(2n + 1) \times (2n + 1)$  (in our experiment,  $n = 8$ ) and which centers at the point  $(x', y')$  in the target image are examined. For each point  $(x, y)$  in the window, the dissimilarity between it and  $(x', y')$  is computed and the point with minimum dissimilarity is selected as the point corresponding to  $(x', y')$ .

### 2.3. Mismatches Detection and Correction

As discussed in section 1, the two images to be registered are intrinsically dissimilar, which greatly increases the probability of mismatches. However, regardless of the type of the deformations, neighboring points in the target image should always map to neighboring points in the reference image. Therefore, if mismatches do not collectively occur, they can be detected and corrected with this continuity constraint. For this purpose, the validation of the detected control points are examined using  $3 \times 3$  windows centering at each control point in the reference image. In such a window in the reference image, for each point  $(x', y')$  whose correspondence is known, we compute its horizontal and vertical displacements as:

$$\begin{aligned} dx_i &= \mathbf{median}(x_i - x'_i) \\ dy_i &= \mathbf{median}(y_i - y'_i) \end{aligned} \quad (2)$$

where  $i = 1 \dots n$ .  $n$  is the total number of candidate control points in the  $3 \times 3$  window. The medians of these displacements are used as the actual displacements of point  $(x', y')$ . In this way, if the control point  $(x', y')$  in the reference image has a displacement value which is considerably larger or smaller than those of its neighbors, its correspondence to the point in the target image is reassigned. This process essentially enforces the continuity of the established correspondences and therefore is effective in correcting non-collectively occurring mismatches.

## 2.4. Consistency Checking

While median filters remove non-collectively occurring mismatches, collective mismatches are possible to occur. We therefore perform consistency checking to further improve the accuracy of the established correspondences. Traditional consistency checking is conducted between the correspondences resulting from different matching methods which are applied to the same dataset and the correspondences supported by both methods are considered as valid ones. In our method however the same matching procedure is applied but with the roles of the reference image and the target image switched. When the recto image is used as the reference image and the flipped verso image as the target image,  $\theta_{high}$  and  $\theta_{low}$  are set as  $-0.5$  and  $-1.8$ . The resultant control point pairs are mostly foreground texts on the verso image and their corresponding bleed-through points on the recto image. Then by switching the target image and the reference image and setting  $\theta_{high}$  and  $\theta_{low}$  as  $1.8$  and  $0.5$ , the resultant control point pairs are mostly foreground texts on the recto image and their bleed-through points on the verso image.

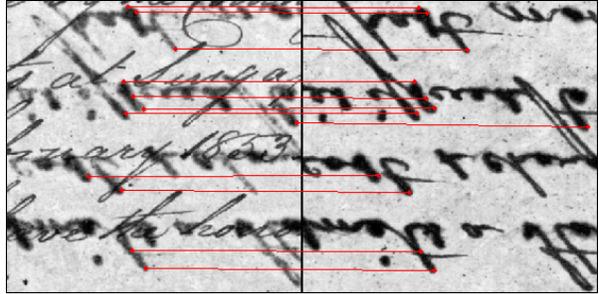
The two sets of resultant point pairs are combined to form the whole set of corresponding points. For each point pair  $P'_1 \leftrightarrow P_2$  in the combined set, if there are  $P'_3 \leftrightarrow P_4$  satisfying  $(P'_1 = P'_3) \wedge (P_2 \neq P_4)$  or  $(P'_1 \neq P'_3) \wedge (P_2 = P_4)$ , these pairs are inconsistent and deleted from the set. Then the reference image is divided into  $8 \times 8$  patches. For each point pair in each patch  $(x', y') \leftrightarrow (x, y)$ , we compute  $x' - x$  and  $y' - y$  which vote to determine the major directions of the displacements. The point pairs with different displacement directions are considered as inconsistent and removed.

Finally there are two ways to determine the number of point pairs returned by our method. The user can specify the desired number of points,  $N$ , and our method will select the  $N$  most similar pairs. Alternatively, a user-defined threshold  $dis_{th}$  can be used to select all control point pairs with dissimilarity less than the given value. Figure 3 shows some detected corresponding control points. For clarity, only part of pairs with high similarity are shown.

## 3. Experimental Results

The proposed method has been tested on 28 pairs of real historical document images obtained from the National Archives of Singapore. These images are mostly scanned at 150dpi and have a size of  $1800 \times 2800$ . To evaluate the proposed method, visual assessment is first performed by experts working on historical document image registration. For most tested images, the proposed method shows encouraging results.

For quantitative assessment, we manually select control points for 10 pairs of images using the matlab inter-



**Figure 3. Illustration of detected control point pairs from sample images in Figure 2.**

face ‘cpsselect’ and use the result as ground truth. We assume the manually selected point pairs  $S_{manu}$  are correct. For comparison, the proposed method returns point pairs with dissimilarity less than the median value of all detected point pairs. For a point pair  $(x', y') \leftrightarrow (x, y)$  detected by the proposed method, if  $(x', y') \leftrightarrow (x_1, y_1)$  exists in  $S_{manu}$ , this point pair is regarded as ‘detected’. We compute  $|x - x_1| + |y - y_1|$  and if the value is less than 2 (i.e. our detected point is within 2 pixels away from the manually selected point), the point pair  $(x', y') \leftrightarrow (x, y)$  is regarded as ‘correctly detected’. The proposed method is evaluated with precision and recall measures defined as follows:

$$Precision = \frac{N_{correct}}{N_{detect}}, \quad Recall = \frac{N_{correct}}{N_{manu}} \quad (3)$$

where  $N_{detect}$ ,  $N_{correct}$  and  $N_{manu}$  refer to the number of the ‘detected’, the ‘correctly detected’ and the manually selected point pairs.

Table 1 shows the precision and the recall of the proposed method on the tested images. As manual selection of control points is very subjective and time-consuming, the value of  $N_{manu}$  is small. Meanwhile, the number of point pairs returned by the proposed method is much larger than  $N_{manu}$ , therefore the recall is quite high. Note also that it takes us around 15 minutes to precisely select the corresponding control point pairs for each document. Thus, the ability to automate this procedure with reasonable precisions is greatly welcomed and we believe our results are promising.

## 4. Conclusion and Future Work

In this paper, we present a fully automatic control points selection method for historical document registration. Our approach makes use of the gradient orientation maps and the labeled images to identify candidate points. Corresponding control points are then detected and matched with a search procedure which optimizes the dissimilarity metric consid-

Document number	1	2	3	4	5	6	7	8	9	10	Average
No. of manually selected	61	50	48	68	46	43	55	60	76	64	57
No. of total returned	99	97	89	117	99	85	124	99	148	106	106
No. of 'detected'	60	50	48	66	46	42	53	57	73	63	56
Recall (%)	96	89	96	91	89	92	86	83	88	90	90
Precision (%)	98	89	96	94	89	94	89	87	92	91	92

**Table 1. The precision and recall of the proposed method on 10 documents using the results of manual selection as ground truth.**

ering image intensity, gradient orientation, gradient magnitude and displacement. Median filters and modified consistency checking are applied to correct mismatches. Using our empirically chosen weight parameters, the accuracy of the detected control points is quite comparative to that of manually selected points. In the future, we aim to explore the methods which can wrap the target image to the reference image using the corresponding control points detected by this method. In particular, mapping functions which can capture local distortions are demanded.

## 5. Acknowledgements

This research is supported in part by MOE ARC grant R252-000-349-112 and IDM R&D grant R252-000-325-279. The authors would like to thank the National Archives of Singapore for the use of their archival documents.

## References

- [1] D. Bhattacharya and S. Sinha. Invariance of stereo images via the theory of complex moments. *Pattern Recognition*, 30(9):1373–1386, September 1997.
- [2] H. Don. A noise attribute thresholding method for document image binarization. *International Journal on Document Analysis and Recognition*, 4(2):131–138, 2001.
- [3] L. Fonseca and M. Costa. Automatic registration of satellite images. In *Brazilian Symposium on Computer Graphics and Image Processing*, pages 219–226, 1997.
- [4] K. Franke and M. Koppen. A computer-based system to support forensic studies on handwritten documents. *International Journal on Document Analysis and Recognition*, 3(4):218–231, May 2001.
- [5] C. Harris and M. Stephens. A combined corner and edge detection. In *Proceedings of The Fourth Alvey Vision Conference*, pages 147–151, 1988.
- [6] J. Hsieh, H. Liao, K. Fan, and M. Ko. A fast algorithm for image registration without predetermining correspondences. In *Proceedings of the 1996 International Conference on Pattern Recognition (ICPR '96)*, page 765, Washington, DC, USA, 1996. IEEE Computer Society.
- [7] Y. Huang, M. Brown, and D. Xu. A Framework for Reducing Ink-Bleed in Old Document. In *IEEE Computer Vision and Pattern Recognition (CVPR'08)*, pages 1–7.
- [8] G. Leedham, S. Varma, A. Patankar, and V. Govindarayu. Separating text and background in degraded document images: A comparison of global thresholding techniques for multi-stage thresholding. In *Proceedings of the 8th International Workshop on Frontiers in Handwriting Recognition (IWFHR'02)*, page 244, Washington, DC, USA, 2002. IEEE Computer Society.
- [9] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1):62–66, January 1979.
- [10] N. Sebe, Q. Tian, E. Loupias, M. S. Lew, and T. S. Huang. Evaluation of salient point techniques. In *Proceedings of the International Conference on Image and Video Retrieval (CIVR'02)*, pages 367–377, London, UK, 2002. Springer-Verlag.
- [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, Oct. 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] J. Wang and C. Tan. Accurate Alignment of Double-sided Manuscripts for Bleed-through Removal. In *8th IAPR International Workshop on Document Analysis Systems (DAS'8)*, pages 69–75, Dec. 2008.
- [15] Q. Wang and C. L. Tan. Matching of double-sided document images to remove interference, pattern recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'01)*, pages 11–13, Hawaii, Dec. 2001.