Use of Shape Deformation to Seamlessly Stitch Historical Document Images

● Wei Liu ● Wei Fan ● Li Chen ● Jun Sun ● Satoshi Naoi

In China, efforts are being made to preserve historical documents in the form of digital data so that they can be effectively used while being protected. Stitching technology is expected to play a role in these efforts as it can be used to divide a large historical document into multiple areas, scan those areas using a compact contactless scanner, and seamlessly connect the resulting images. Since historical documents are often on paper of low quality and have an uneven surface, there can be local distortion in the scanned images, making it difficult to stitch them together. To solve this problem, Fujitsu Research and Development Center Co., Ltd. has developed a seamless image stitching method using shape deformation. This method first deforms one of two images to approximate the other image and then estimates single or dual optimal seams to minimize deformation and maximize consistency. With this method, high-quality images of historical documents can be obtained. This paper describes the problems associated with the stitching of historical document images, presents a method for solving them, and evaluates its performance.

1. Introduction

In China, progress is being made on preserving historical documents in the form of digital data to ensure their survival and facilitate their use. Large-scale specialized scanners have so far been used to create digital archives for large documents such as old newspapers or works of calligraphy, but this approach can incur significant costs. Attention has therefore turned to methods that partition a document into multiple areas, scan each of those areas using an inexpensive compact scanner, and apply an image stitching method to combine the scanned images into a single image. Fujitsu Research and Development Center Co., Ltd. (FRDC) has developed technology for dividing a large-size historical document into multiple areas, scanning each area using a compact contactless scanner, and seamlessly stitching together the scanned images. This technology can be used to seamlessly stitch multiple overlapping images into a single high-resolution image. It can be applied to the creation of panoramic photographs, images of road scenery, etc. in addition to the digital archiving of large historical documents.

Requirements for stitching images of historical documents are more severe than those for conventional image stitching since any misalignment of the image structure or of the image artifacts would be obvious. Historical documents present particular problems for two main reasons: the lack of a flat document surface can distort the content of the stitched images, and the low-quality paper of historical documents can generate local distortion in scanned images. Stitching images in a way that preserves continuity in image structure as in the case of straight lines is an important issue here.

2. Conventional technology and associated issues

Conventional stitching methods can be classified into those that use a single optimal seam and those that locate corresponding seams in two images and then deform and stitch the contents of the two images along those seams. The general procedure followed by the first type is to roughly align the two images by estimating a global transformation[1] between them and then stitch those two parts together along an optimal seam to create a single image. For example, the method proposed by Kwatra et al.[2] locates the position of the optimal seam by minimizing the differences in intensity and texture in the overlap area. The method proposed by Veena[3)] controls both photometric inconsistencies and geometric misalignments between the two images by minimizing ghost artifacts along the seam.

However, in non-rigid image stitching, a number of local distortions generally arise, so using only one seam can result in misalignment of the image structure (characters, figures, etc.). To solve this problem, methods have been developed to find two corresponding seams in the two images targeted for stitching. For example, Jia et al. developed a method[4)] that estimates two corresponding seams on the basis of intensity coherence and structure continuity. These seams are used to propagate deformation in the two images and form a complete image. In addition, Fang et al. developed a deformation function[5)] that maximizes color matching while minimizing deformation distortion in the overlap area. Fan et al. developed a method[6)] that uses dynamic programming[7)] to find an optimal seam, i.e., one that minimizes the crossing of foreground characters, and then uses feature matching to estimate the second seam. These methods stitch together image content on either side of a seam after deformation to favorably preserve image content. However, geometrically continuous elements such as lines may no longer be straight after the transformation.

3. Overview of proposed method

To overcome the problems described in the previous section, we developed a seamless image stitching method that uses shape deformation. This method first stitches together two images using a single optimal seam that minimizes image deformation. Then, to overcome the problem of geometrical inconsistency, it uses dual seams that maximize deformation consistency. With this method, deformation in image content along each seam is consistent, and geometrically continuous elements such as lines are well preserved.

This seamless image stitching method follows the procedure shown in **Figure 1** and described below.

1) Scan upper half and lower half of document using two scanners
2) Determine overlap area by rough alignment
3) Extract robust control points
4) Deform one of the two images and use the image deformation to estimate optimal seams
5) Stitch together the two images using those optimal seams

The following sections describe image deformation and estimation of optimal seams, the key features of the proposed method.

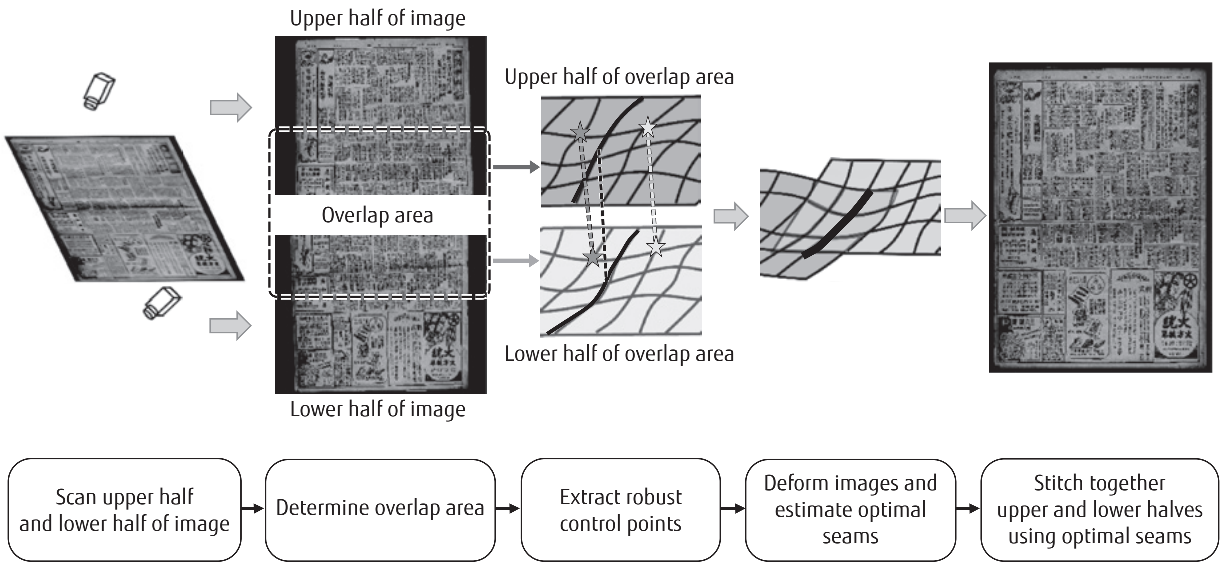
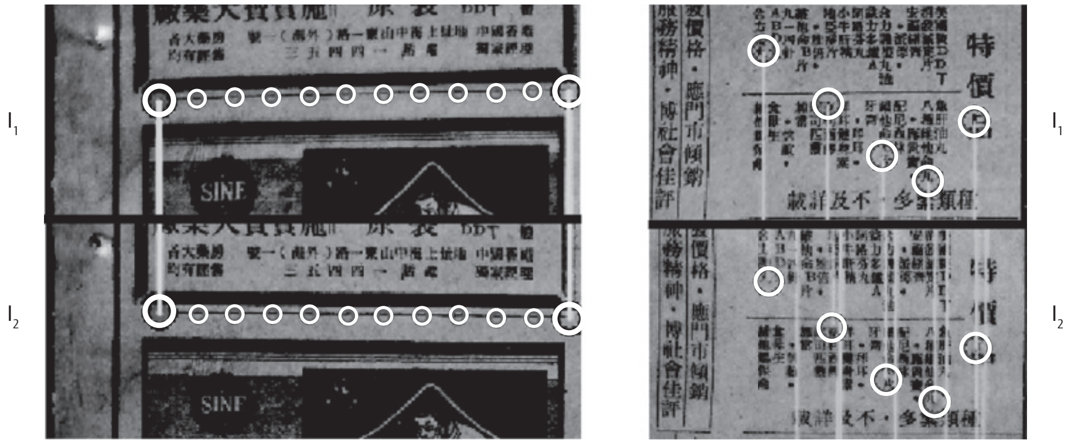


Figure 1
Process flow of proposed method.



(a) Lines (b) Feature points

Figure 2
Examples of extracted control points.

4. Image deformation

4.1 Extraction of robust control points

Since historical documents may lack a flat surface, differences may arise between two scanned images of a document. Nevertheless, several robust features should still exist in such images. Even in historical documents, the document surface will be locally continuous within a narrow range, so a corresponding relationship will exist between the two images in such local areas. Thus, if corresponding points in such an area were to be extracted from the two images, they would constitute a partial set of robust control points.

The image of a historical document may contain geometric elements such as straight lines. The correlation between such elements is invariant to image deformation, so they can be treated as robust features. Now, if corresponding lines in two images were to be extracted as a pair, the points on those lines would constitute another partial set of robust control points. Two examples of extracted control points are shown in **Figure 2**.

4.2 Image deformation by moving least squares

Given that the two images have been roughly aligned to obtain the overlap area, a global transformation between them would not be applicable due to local distortions. An alternative approach is to represent local deformation using a deformation field. Here, we use the moving least squares (MLS) method[8] to model the deformation field. The purpose of using MLS is to approximate the affine space[9] and minimize the approximation error in control points.

Let the positions of corresponding control points in the two images be denoted as p_i and q_i. Now, denoting the coordinates of an arbitrary pixel in the image as v, we derive the best affine transformation[note)] $l_v(x)$ that minimizes the value of the following expression by MLS.

$$\sum_i w_i \left| l_v(p_i) - q_i \right|^2,$$

where w_i denotes the weights of the control points with respect to the current pixel.

We solve for the affine space denoted as M by direct differentiation of the above expression. After determining the affine space, we deform the image by converting each and every pixel in this way.

5. Seam estimation by shape deformation

Achieving seamless image stitching requires that a single optimal seam or dual matchable seams be aligned between the two images. In this section, we introduce two methods for this purpose: Single Seam

note) A method of deforming figures and shapes that combines linear transformation and parallel shift. The transformation result preserves the original geometrical properties. For example, points on a straight line in the original figure will remain on a straight line, and parallel lines will remain parallel.

with Minimum Deformation (SSMD) and Dual Seam with Maximum Deformation Consistency (DSMDC). Both methods are used in the seamless image stitching method that we propose.

5.1 SSMD

We determine an affine space with respect to the image shown in Figure 2 (a) in the manner described above. To stitch in a seamless manner, misalignment near the seam must be minimized. Conventional methods use the graph cut method[2] for finding an optimal seam. This method maximizes the "cost" between adjacent pixels, which is defined as gradient smoothness or texture similarity. However, there is no clear standard for quantifying gradient smoothness and texture similarity, so we decided to adopt another method that estimates an optimal seam on the basis of shape deformation.

The symbols I_1 and I_2 in Figure 2 denote two images to be stitched together. In this case, I_1 and I_2 overlap vertically. Denoting primary pixel coordinates as v, deformation image V is defined by the following equation.

$$V = \| vM - v \|,$$

where M is the affine space.

The optimal row that minimizes the amount of deformation can be calculated using the following equation.

$$i^* = \arg\min_{1 \le i \le h} \sum_{j=1}^{w} V(i, j),$$

where w and h denote the width and height, respectively, of I_1. The rough seam is taken to be the point group on this optimal row. In the case of dynamic programming, the optimal seam is a traversing path with a minimum amount of deformation near the optimal row.

5.2 DSMDC

An experiment we performed showed that SSMD produced better results than dynamic programming, but there were cases in which nonexistent lines (ghost artifacts) appeared. An example of such artifacts is shown in **Figure 3**. These ghost-artifact features are clearly apparent in the stitched image.

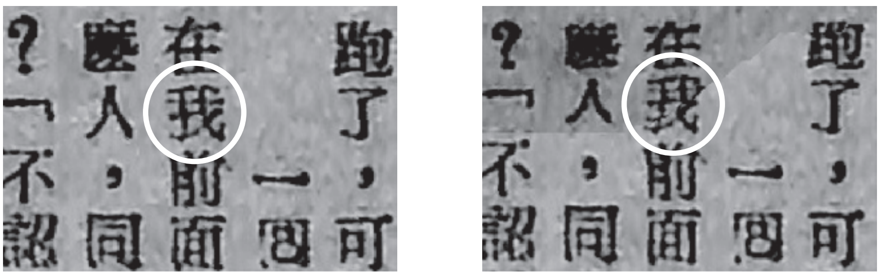
We developed a dual seam method to solve this problem. This method estimates and joins two matchable seams and deforms the images on the opposite sides of the two seams to generate the final result. An example of image stitching using dual seams is shown in **Figure 4**. As shown by the magnified portion on the right side of the photograph, the ghost artifacts have disappeared. However, it can be seen at the left of the photograph that, while the content is well preserved, a line is no longer straight due to distortion.

To deal with this problem, we developed the DSMDC method as an enhancement of the above dual seam method. With this method, the purpose of optimization is not simply to minimize deformation but also to maximize deformation consistency in content lying along the seam.

First, letting v denote pixel coordinates in one of the images, we construct deformation image D from each of the pixels corresponding to the vector

$$D = vM - v .$$

We need to find an optimal seam along which such deformation vectors are nearly the same. The objective function for obtaining coordinates (i, j) of the optimal seam is



(a) Original image (b) After stitching

Figure 3
Example of ghost artifacts after stitching.

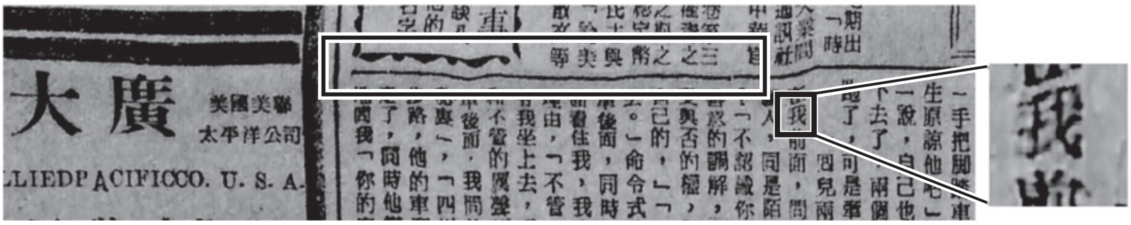


Figure 4
Image stitching using dual seams.

$$\min \sum_{j=1}^{w} \left\| D(i,j) \right\| + \lambda_1 \left\| X_{i,j+1} - X_{i,j} \right\| + \lambda_2 \left\| y_{j+1} - y_j \right\|,$$

where λ_1 and λ_2 are normalization coefficients, $X_{i,j}=(i,j)+D(i,j)$ are coordinates after deformation, and y_j is the y coordinate of $X_{i,j}$. The three terms in the above expression correspond to three objectives:

1) Minimize shape deformation
2) Minimize difference in amount of deformation between adjacent pixels (i.e., maximize deformation consistency between adjacent pixels)
3) Maximize smoothness in final seam.

Minimizing the above objective function tends to result in similar transformations between adjacent pixels, which means that the consistency of geometrically continuous elements can be well preserved. The Viterbi algorithm[10)] is used to optimize this objective function.

6. Performance evaluation

We evaluated performance using three types of historical documents: newspapers (10 items), drawings (10 items), and maps (1 item). The criterion used for judging performance was the degree of content preservation. We evaluated three methods for comparison purposes: conventional dynamic programming, SSMD, and DSMDC.

First, we evaluated image stitching using a single seam (**Figure 5**). Some loss in content can occur at the optimal seam with the graph cut method. For example, the black line simply disappeared, as shown in Figure 5 (b). In contrast, good stitching results were obtained with SSMD, as shown in Figure 5 (c).

Next, we evaluated image stitching using dual seams (**Figure 6**). Conventional methods usually estimate the first seam using minimized deformation (Figure 4) or minimized content crossing (method that minimizes image crossing) (Figure 6 (b)) as a criterion and then estimate the corresponding seam (second seam). Finally, the opposite sides of the two seams are transformed to form the final stitching result. As shown in Figure 6 (c), image stitching with the proposed DSMDC method obtained results closest to the original image shown in Figure 6 (a).

Differences between the stitched image and the original image (obtained by image subtraction) are shown in **Figure 7** for dynamic programming, SSMD, and DSMDC. A smaller number of white points corresponds to fewer differences. As shown, the transformed image generated with the DSMDC method was closest to the original image.

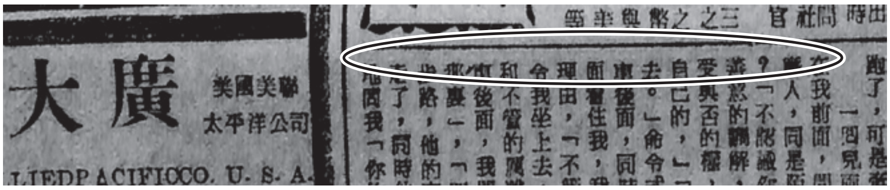
The recall percentages are shown in **Table 1**. We obtained these values by calculating the number of pixels for which a difference occurred in each of the images in Figure 7, dividing that number by the total number of pixels, and subtracting that result from 1. It can be seen from these results that the transformed images generated using the proposed SSMD and DSMDC methods were closest to the original image.

7. Conclusion

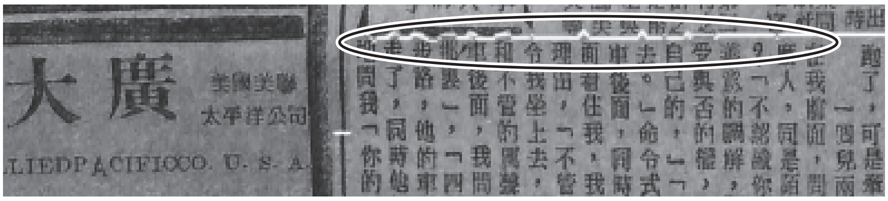
The seamless image stitching method proposed in this paper uses shape deformation to overcome the problem of geometrical inconsistency in which lines in a stitched image are no longer straight. This method first deforms one of the two images targeted for stitching to approximate the other image and then estimates single or dual seams that maximize consistency with

Table 1
Recall percentages.

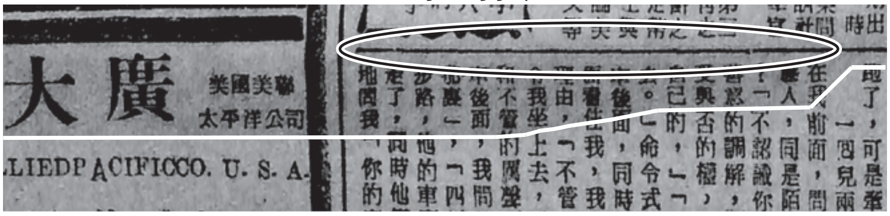
	Dynamic programming	SSMD	DSMDC
Newspapers	78.58%	80.17%	80.45%
Drawings	82.08%	83.33%	87.89%
Map	80.86%	81.10%	80.90%



(a) Original image

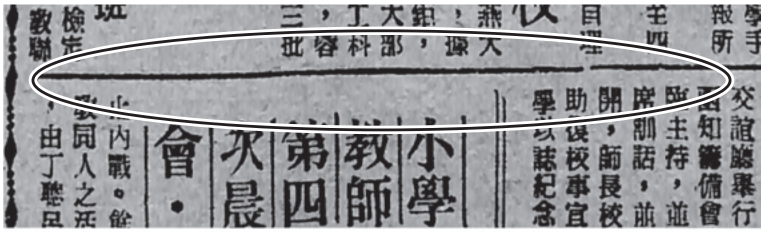


(b) Results of stitching using graph cut method

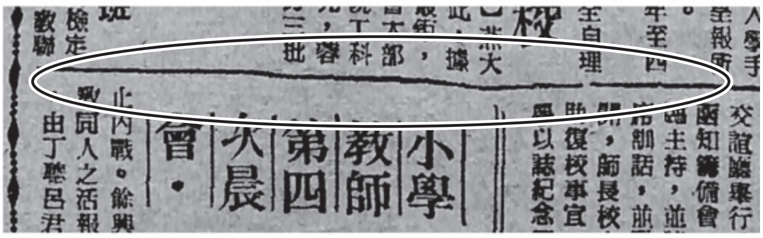


(c) Results of stitching using SSMD method

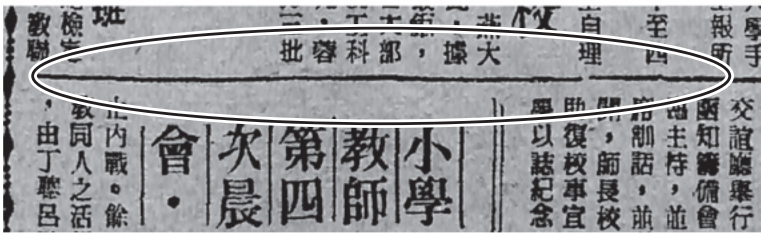
Figure 5
Evaluation of image stitching using single seam.



(a) Original image



(b) Results of stitching by minimized content crossing



(c) Results of stitching by DSMDC method

Figure 6
Evaluation of image stitching using dual seams.

(a) Conventional dynamic programming method (b) SSMD method (c) DSMDC method

Figure 7
Differences between stitched image and original image.

minimal deformation. With this method, the degree of deformation in image content along each seam is consistent, and geometrically continuous elements such as lines are well preserved. Fujitsu Research and Development Center Co., Ltd. plans to continue the development of this seamless image stitching method to contribute to the preservation and effective use of historical documents in China.

References

1) L. G. B. Mirisola et al.: Exploiting inertial sensing in mosaicing and visual navigation. Intelligent Autonomous Vehicles, Vol. 6, Part 1, pp. 306–311 (2007).
2) V. Kwatra et al.: Graphcut textures: image and video synthesis using graph cuts. ACM Transactions on Graphics, Vol. 22, No. 3, pp. 277–286 (2003).
3) C. V. Veena: Minimizing seam artifacts in image stitching. Asian Journal of Information Technology, Vol. 6, No. 2, pp. 209–214 (2007).
4) J. Jia et al.: Image stitching using structure deformation. Pattern Analysis and Machine Intelligence, IEEE Transactions on, Vol. 30, No. 4, pp. 617–631 (2008).
5) H. Fang et al.: Textureshop: texture synthesis as a photograph editing tool. ACM Transactions on Graphics, Vol. 23, No. 3, pp. 354–359 (2004).
6) W. Fan et al.: Paper stitching using maximum tolerant seam under local distortions. ACM Symposium on Document Engineering, pp. 35–44 (2014).
7) D. P. Bertsekas: Dynamic programming and optimal control, volume 1. Athena Scientific, Belmont, MA, 1995.
8) D. Levin: The approximation power of moving least-squares. Mathematics of Computation of the American Mathematical Society, Vol. 67, No. 224, pp. 1517–1531 (1998).
9) S. Scott et al.: Image deformation using moving least squares. ACM Transactions on Graphics, Vol. 25, No. 3, pp. 533–540 (2006).
10) G. D. Forney Jr.: The Viterbi algorithm. Proceedings of the IEEE, Vol. 61, No. 3, pp. 268–278 (1973).

Wei Liu
Fujitsu Research and Development Center Co., Ltd.
Ms. Liu is currently engaged in research related to image processing and recognition.

Jun Sun
Fujitsu Research and Development Center Co., Ltd.
Mr. Sun is currently engaged in research related to image processing and deep learning.

Wei Fan
Fujitsu Research and Development Center Co., Ltd.
Mr. Fan is currently engaged in research related to image processing and recognition.

Satoshi Naoi
Fellow, Fujitsu Laboratories Ltd.
President, Fujitsu Research and Development Center Co., Ltd.

Li Chen
Fujitsu Research and Development Center Co., Ltd.
Mr. Chen is currently engaged in research related to image processing and recognition.