Journal of Applied Sciences

ISSN 1812-5654

Journal of Applied Sciences 13 (9): 1537-1543, 2013
ISSN 1812-5654 / DOI: 10.3923/jas.2013.1537.1543

Vector Constraint and Ncc Based Chinese Document Image Mosaic

[1]Lijing Tong, [1]Quanyao Peng, [2]Sam Li, [1]Huiqun Zhao and [3]Guoliang Zhan
[1]College of Information Engineering, North China University of Technology, Beijing, China
[2]Juniper Networks Incorporation, Sunnyvale, California, USA
[3]Zhuhai Digital Power Technology Company, Zhuhai, China

Abstract: Document images which captured by camera-based systems are often suffered from warping distortion that deteriorates the performance of current Optical Character Recognition (OCR) approaches. To overcome this difficulty and get better Chinese document image OCR recognition accuracy, a document image mosaic method based on vector constraint and Normalized Cross Correlation (NCC) is proposed. In this method, two images, one is reference image and the other is auxiliary image, are captured for the same warped document from left side and right side firstly. Then, for two reference points in the inflection point position in reference image, NCC registration method are used to find two matching points in the auxiliary image. During registration, the area limiting method and vector constraining method are proposed to improve the registration performance. At last, the two flatter parts of each image are mosaiced together. Experimental results show that the document image mosaic method based on vector constraining and NCC is more robust than the classical cross correlation registration method. The OCR recognition rate of the new mosaiced image is markedly higher than the two original distorted document images and the time cost is reduced.

Key words: Image processing, image registration, OCR, image mosaic, normalized cross correlation

INTRODUCTION

Recently, with the advancement of processing speed and internal memory of hand-held mobile devices such as high-end cell-phones, Personal Digital Assistants (PDA), smart phones, iPhones, iPods, etc. having built-in digital cameras, the Optical Character Recognition (OCR) of document image is more and more becoming a research focus (Stamatopoulos *et al*., 2011; Mollah *et al*., 2011). However, the OCR recognition rate of the warped document is very low (Stamatopoulos *et al*., 2012). To solve the problem of the warped document image recognition, two images captured from left side and the right side of the same warped document, one is the primary image and the other is auxiliary image, can be mosaiced together with each flatter part.

Image mosaiking requires estimating the displacement between spatially close images that are not necessarily temporally close (Rosa *et al*., 2013). Currently, some image mosaic methods have been proposed.

Koo *et al*. (2009) have proposed an algorithm to compose a geometrically dewarped and visually enhanced image from two document images. However, their method was based on the assumption that the document surfaces conformed to the cylindrical surface model.

Miao and Yue (2011) proposed an image mosaicing method for camera-captured document images and it can be used to stitch multiple overlapping document images into a large high resolution image through the calculation of a document's horizontal vanishing point and the vertical vanishing point. However, their method was based on those flat document images, especially for the perspective distortion.

Dong *et al*. (2012) proposed an image mosaic algorithm which is called immune memory clonal selection algorithm. However, the Susan algorithm just uses the brightness difference threshold to find the matching points and is too simple to be fit for the Chinese document image registration.

Xu *et al*. (2012) proposed an image mosaic algorithm based on the Harris corner detection. However, this algorithm is mainly applied to the medicine image. For Chinese document image, there are too many similar corner points need to be considered.

Jiang *et al*. (2012) have put forward an image stitching method based on SIFT feature matching for images with distortions. However, a distortion correction of the acquired images was needed before image stitching and the distortion correction need some known internal parameters.

Tang *et al*. (2012) proposed an intuitive technique to find a stitching line based on energy map which is essentially a combination of gradient map which indicates the presence of structures and prominence map which

Corresponding Author: Lijing Tong, College of Information Engineering, North China University of Technology, Beijing, China

determines the attractiveness of a region. However, this method is mainly for the scene image and is not suitable for document image.

Qureshi *et al*. (2012) approach is based on the use of structural similarity index between the high-frequency information of both geometrically corrected unstitched images and the stitched panoramic image for geometric quality assessment. The complementary low-frequency information from the same pair of images is used for assessing the photometric quality of the stitched image using the spectral angle mapper and intensity magnitude ratio measures. However, this method is just a quality metrics.

Cho *et al*. (2013) proposed a method that made well-aligned stereo panoramas with impressive stereo perception. However, in our method, we only use the image from the left view point and the image from the right view point, need not the third image, i.e., the reference panorama in their method.

In our previous research works, we have also proposed some warped document image mosaic methods, which are mainly based on cross correlation image registration (Tong *et al*., 2012; Tong *et al*., 2013a; Tong *et al*., 2013b). However, the accuracy of the registration need be improved further.

Unlike other previous work and our previous work, to realize an accurate recognition of the matching points, this paper takes the luminance differences of two document images into account and uses the Normalized Cross Correlation (NCC) registration method to mosaic two Chinese document images. And the regional limit method and the vector constraint method are also introduced to improve the speed of the registration.

NCC BASED IMAGE REGISTRATION

For a document, two images can be captured from the left side and the right side as shown in Fig. 1(a-b). After binarization, morphological dilation and text line exaction, the inflection point horizontal coordinate can be identified at the curved position of the text line (Tong *et al*., 2013c; Fan and Wang, 2011; Kumar *et al*., 2008). The vertical coordinates of two reference points, y_1 and y_2, can be calculated by (1):

$$\begin{cases} y_1 = k \times y_{Top} + (1-k) \times y_{Bottom} \\ y_2 = (1-k) \times y_{Top} + k \times y_{Bottom} \end{cases} \tag{1}$$

Here, the y_{Top} and y_{Bottom} is the vertical coordinates of the first and the last text line at the inflection point horizontal coordinate. k is a position control parameter, which can be taken as 0.1.

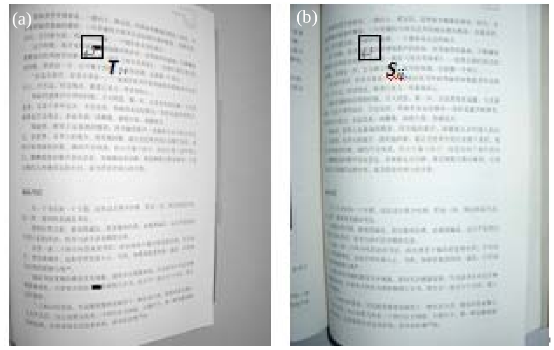

Fig. 1(a-b): Image captured from the left side (a), Image captured from the right side (b)

According to 1, two reference points can be obtained as shown as the centers of the two small black rectangle blocks in Fig. 1(a).

In order to f ind a matching point for each reference point in the auxiliary image, a template block T can be defined in Fig. 1(a) around each reference point respectively. The size of T is X×Y. Then, a same size template S_{ij} is defined and moved in Fig. 1(b) to find a matching point. Here, i and j are the center coordinates of the template S_{ij}. The two template similarity R(i,j) is calculated by the cross correlation coefficient as (2).

$$R(i,j) = \frac{\sum_{m=1}^{X}\sum_{n=1}^{Y}\left[s_{ij}(m,n) * T(m,n)\right]}{\sqrt{\sum_{m=1}^{X}\sum_{n=1}^{Y}[s_{ij}(m,n)]^2}\sqrt{\sum_{m=1}^{X}\sum_{n=1}^{Y}[T(m,n)]^2}} \tag{2}$$

where, S_{ij} (m,n) and T (m,n) are the gray value of pixel with the coordinates (m,n) in template S_{ij} and T.

However, the most of Chinese characters are composed of few basic strokes. This will greatly impact the registration accuracy when the two document images have different luminance. Considering this situation, a Normalized Cross Correlation (NCC) method is taken in our warped document image registration. In the NCC registration method, the similarity is considered according to the relative gray value of each pixel in the template. The NCC registration is performed as (3):

$$R(i,j) = \frac{\sum_{m=1}^{X}\sum_{n=1}^{Y}\left[s_{ij}(m,n) - E(S_{ij})\right] * \left[T(m,n) - E(T)\right]}{\sqrt{\sum_{m=1}^{X}\sum_{n=1}^{Y}[s_{ij}(m,n) - E(S_{ij})]^2}\sqrt{\sum_{m=1}^{X}\sum_{n=1}^{Y}[T(m,n) - E(T)]^2}} \tag{3}$$

where, the E (S_{ij}) and E (T) are the average gray of the template S_{ij} and template T.

VECTOR CONSTRAINT METHOD AND REGIONAL LIMIT

The document image registration method based on NCC has greatly improved the registration robustness. However, sometimes, it still cannot get a correct registration result. The reason is that most of the Chinese characters are composed of few basic strokes, especially, the "", "", "", "". So the image registration of Chinese document images is more complicated and difficult than the image registration of English document images. For example, if we directly take the NCC registration algorithm for the 2 reference points in the Fig. 1(a), two wrong matched points are c alculated in the auxiliary image Fig. 2(a).

To improve the document image registration method based on NCC, we proposed a novel document image registration method, called NRL-VC registration method. The NRL-VC registration method means NCC and Regional Limit-Vector Constraint based registration method.

Based on the NCC method, the regional limit is taken into account in the document image registration firstly. The coordinates of the matched points in the auxiliary image should not far from the coordinates of the reference points in the reference image. According to this reason, the matching search region can be limited in a small area near the coordinates of the reference points as Fig. 3.

The limited area can be defined by x_{left}, x_{right}, y_{botton} and y_{top}. They can be calculated by the follow equations.

$$\begin{cases} x_{left} = x - k * width \\ x_{right} = x + k * width \\ y_{bottom} = y - k * height \\ y_{top} = y + k * height \end{cases} \tag{4}$$

where, the variable x is the horizontal coordinate value of the reference point. The variable y is the vertical coordinate of the reference point. k is a scale factor to define the rectangle size. Normally, k can be taken about 0.15. The parameter width and height is the width and height of the auxiliary image.

The second improvement of NRL-VC from the NCC is the Vector Constraint. The vector composed by two reference points should be approximately equal to the vector composed by the two matched points.

The vector constraint algorithm is described as follows:

- Calculating 2n matched points which have largest cross correlation values as (3)
- Composing the set S and T to save those 2n matched points

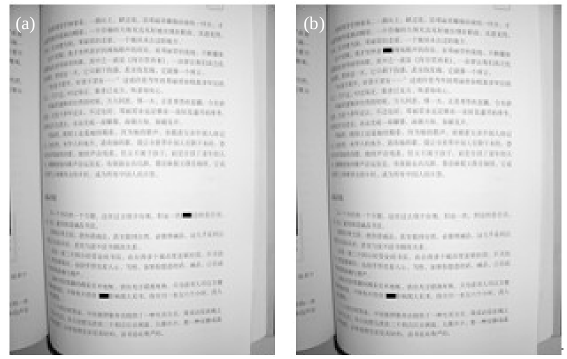

Fig. 2(a-b): Two wrong matched points by NCC method (a), Two correct matched points in the auxiliary image (b)

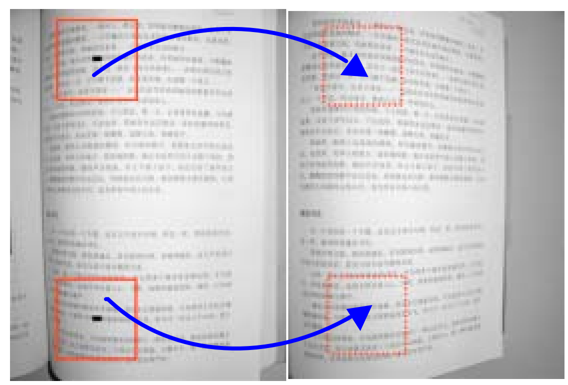

Fig. 3: Limited area in the auxiliary image

$$\begin{cases} S = \{P_1, P_2, P_3 ... P_i ..., P_n\} \\ T = \{Q_1, Q_2, Q_3 ... Q_j ..., Q_n\} \end{cases} \tag{5}$$

- Calculating the reference vector V_0 of the reference point pair

$$\begin{cases} V_{0x} = P_{0x} - Q_{0x} \\ V_{0y} = P_{0y} - Q_{0y} \end{cases} \tag{6}$$

where, (P_{0x}, P_{0y}) and (Q_{0x}, Q_{0y}) are reference point at the upside and downside.

- Calculating n×n matched vectors of the matched point pair V_{ij} from set S and T

$$\begin{cases} V_{ijx} = P_{ix} - Q_{jx} \\ V_{ijy} = P_{iy} - Q_{jy} \end{cases} \quad (i = 1, 2, 3, ...n, j = 1, 2, 3, ...n) \tag{7}$$

where, (P_{ix}, P_{iy}) and (Q_{jx}, Q_{iy}) are the ith matched point at the upside and the jth matched point at the downside:

- Calculating n×n vector differences between V_{ij} and V_0, i.e., M_{ij}

$$M_{ij} = \sqrt{(V_{ijx} - V_{0x})^2 + (V_{ijy} - V_{0y})^2} \quad (i = 1, 2, 3, \ldots\ldots n, j = 1, 2, 3, \ldots\ldots n) \tag{8}$$

- Taking the matched point pair (P_i, Q_j) which has the minimum M_{ij} as the best matched point pair P_{best}, Q_{best}:

$$(P_{best}, Q_{best}) = \arg\min_{(P_i, Q_j)} (M_{11}, M_{12}, \ldots M_{ij} \ldots M_{nn}) \quad (i = 1, 2, 3, \ldots\ldots n, j = 1, 2, 3, \ldots\ldots n) \tag{9}$$

After the area limiting and vector constraining, the robustness is improved. For example, in Fig. 2(a), a wrong registration result is calculated by NCC algorithm. However, a correct result is calculated by NRL-VC registration method in Fig. 2(b).

DOCUMENT IMAGE MOSAIC

For image mosaic, 3 atom transforms of the affined transform are used: scaling, rotation and translation. The transform can be written in (10):

$$\begin{Bmatrix} x' \\ y' \end{Bmatrix} = K \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{Bmatrix} x \\ y \end{Bmatrix} + \begin{Bmatrix} t_x \\ t_y \end{Bmatrix} \tag{10}$$

where, in (10), the K is the scaling coefficient. The θ is rotation angle. The (t_x, t_y) are the translation coefficients. The (x, y) are the original coordinates in the reference image. The (x', y') are the destination coordinates in the auxiliary image. Let:

$$\begin{cases} a = K\cos\theta \\ b = K\sin\theta \end{cases} \tag{11}$$

Then, (10) can be further reduced to (12):

$$\begin{Bmatrix} x' \\ y' \end{Bmatrix} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{Bmatrix} x \\ y \end{Bmatrix} + \begin{Bmatrix} t_x \\ t_y \end{Bmatrix} \tag{12}$$

To solve the 4 parameters, a, b, t_x and t_y, the 2 reference point coordinates and the 2 matched point coordinates can be taken into (12).

According to (12), the destination coordinates (x', y') can be calculated by the original coordinates(x, y). However, the (x', y') maybe are not integer. Hence, the bilinear interpolation is used to get an appropriate gray value for the point (x, y). Let:

$$\begin{cases} u = \lfloor x' \rfloor \\ v = \lfloor y' \rfloor \end{cases} \tag{13}$$

as shown in Fig. 4.

Then the gray value of the point (x', y') in the auxiliary image can be interpolated as (14):

$$g(x', y') = (1-\alpha)\cdot(1-\beta)\cdot f(u, v) + \alpha\cdot(1-\beta)\cdot f(u+1, v) + (1-\alpha)\cdot\beta\cdot f(u, v+1) + \alpha\cdot\beta\cdot f(u+1, v+1) \tag{14}$$

where,

$$\begin{cases} \alpha = x' - u \\ \beta = y' - v \end{cases} \tag{15}$$

After image mosaicing, two flatter parts of each warped document image can be mosaiced together to

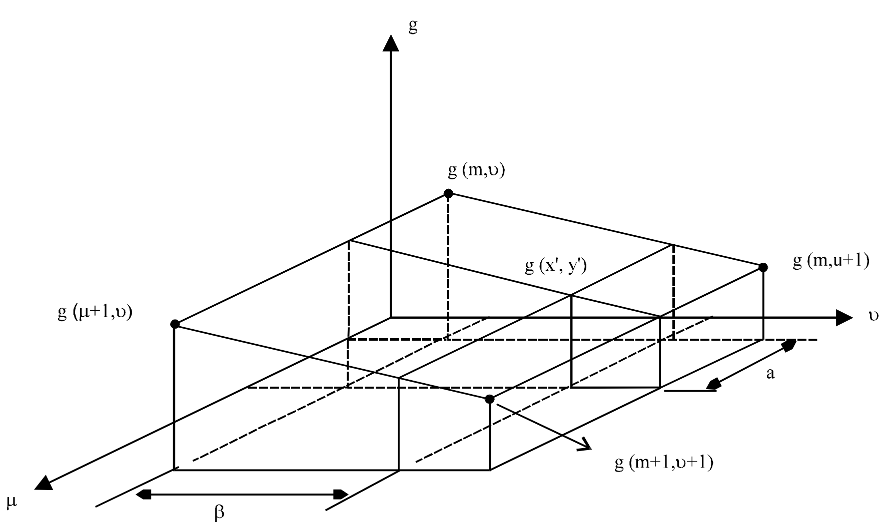

Fig. 4: Bilinear interpolation for the gray value of (x', y')

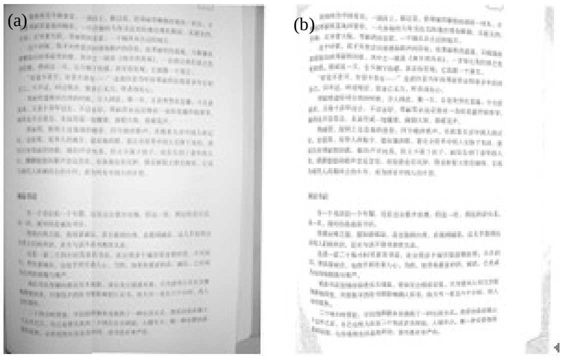

Fig. 5(a-b): Mosaiced gray image based on NRL-VC (a), Mosaiced binarization image based on NRL-VC (b)

one image. For example, Fig. 1(a-b) can be mosaiced to Fig. 5(a). The Fig. 5(b) is the mosaiced result for two binarized images of Fig. 1(a-b).

EXPERIMENTAL RESULTS AND ANALYSES

We made the program for the proposed method in this paper on Windows operating system in our personal computer. The programming environment is the Visual C^{++} 6.0. A lot of warped document images' testing demonstrated that this method can effectively solve the distorted document recognition problem. Some comparing experiments are also performed to compare the NRL-VC method with the classical cross correlation method used in (Tong *et al*., 2012). Three pairs of experiment results are summarized in Table 1.

In fact, Fig. 1(a-b) are the reference and auxiliary images of the comparing experiment A. Fig. 2(a) is the wrong result from classical cross correlation method. Fig. 2(b) is the correct result from NRL-VC method.

As to the experiment B, the reference image, the auxiliary image with two matched points by the classical cross correlation method, the auxiliary image with two matched points by the NRL-VC method and the mosaiced binarization image are shown in Fig. 6.

In experiment B, both the classical cross correlation registration method and the NRL-VC registration method can get the same correct matched result.

As to the comparing experiment C, the reference image, the auxiliary image with two matched points by the classical cross correlation method, the auxiliary image with two matched points by the NRL-VC method and the mosaiced binarization image according to NRL-VC result are shown in Fig. 7.

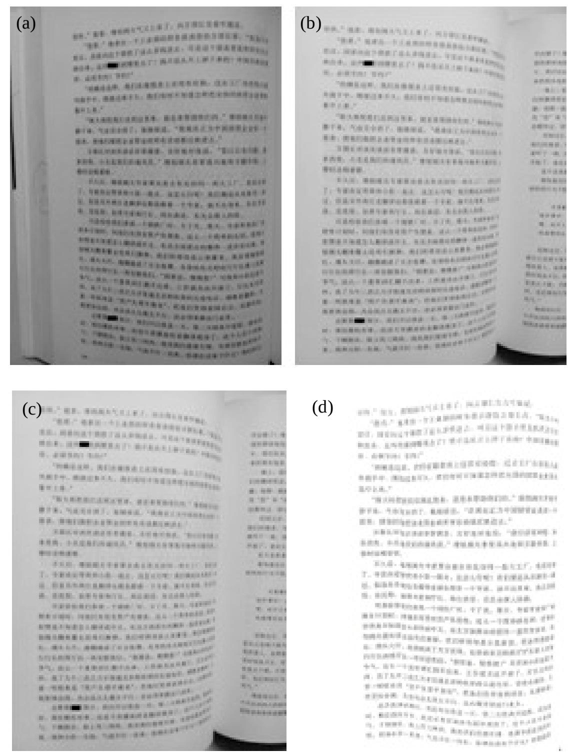

Fig. 6(a-d): Reference image B (a), Auxiliary image B with two matched points by the classical cross correlation method (b), Auxiliary image B with two matched points by the NRL-VC method (c), Mosaiced binarization image (d)

Table 1: Comparing experiment results

Comparing experiment	Classical cross correlation method	NRL-VC method
A	Wrong matching	Correct matching
B	Correct matching	Correct matching
C	Wrong matching	Correct matching

This time, the classical cross correlation registration method got a wrong matched result. But the NRL-VC registration method got a correct matched result.

We also tested the OCR rate for experiment A, B and C. These results are listed in Table 2, 3 and 4.

From Table 2, 3 and 4, we can see that the OCR rate of the reference and the auxiliary images are relatively low. However, after mosaic, the OCR rates are all higher than the original images' OCR rates.

We also tested the registration time consuming comparison for experiment A, B and C. The results are listed in Table 5, 6 and 7.

To assess different template size's influence to the registration time, the sizes were taken 20, 16 and 30 in the

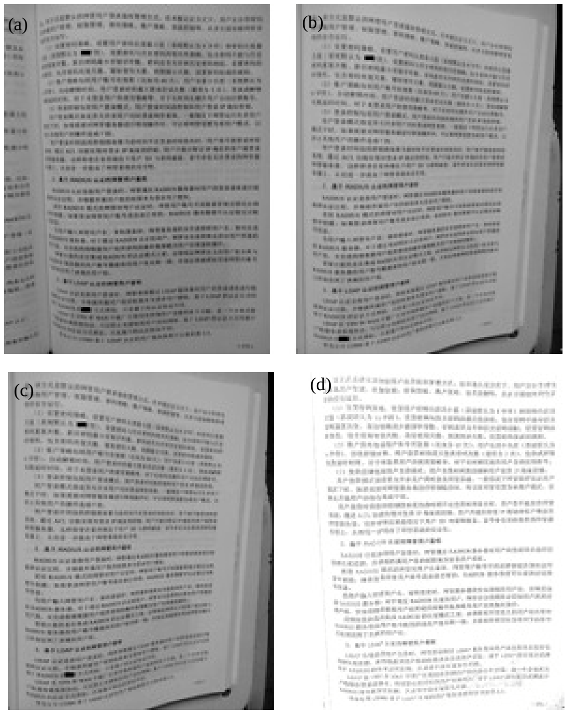

Fig. 7(a-d): Reference image C (a), Auxiliary image C with two matched points by the classical cross correlation method (b), Auxiliary image C with two matched points by the NRL-VC method (c), Mosaiced binarization image according to the NRL-VC result (d)

Table 2: OCR comparison results of experiment A

Experiment A	Total characters	Recognized characters	Recognition rate (%)
Reference image Fig. 1(a)	830	566	68.10
Auxiliary image Fig. 1(b)	830	459	55.30
Mosaiced image Fig. 5(b)	830	787	94.80

Table 3: OCR Comparison Results of Experiment B

Experiment A	Total characters	Recognized characters	Recognition rate (%)
Reference image Fig. 6(a)	847	689	81.30
Auxiliary image Fig. 6(b)	847	781	92.20
Mosaiced image Fig. 6(d)	847	801	94.50

Table 4: OCR comparison results of experiment C

Experiment A	Total characters	Recognized characters	Recognition rate (%)
Reference image Fig. 7(a)	1223	880	71.90
Auxiliary image Fig. 7(b)	1223	887	72.50
Mosaiced image Fig. 7(d)	1223	1061	86.70

A, B and C separately. From Table 5, 6 and 7, we can see that, the time consumed by the NRL-VC method is greatly shorter than the time consumed by the classical cross correlation method.

Table 5: Time consuming comparison for experiment A

Experiment pair A	Registration time (s)
The classical cross correlation registration method	536
NRL-VC registration method	47

Table 6: Time Consuming Comparison for Experiment B

Experiment pair B	Registration time (s)
The classical cross correlation registration method	366
NRL-VC registration method	32

Table 7: Time consuming comparison for experiment C

Experiment pair C	Registration time (s)
The classical cross correlation registration method	1036
NRL-VC registration method	90

CONCLUSION

A document image mosaic method based on vector constraint and normalized cross correlation, especially for Chinese document image, was present in this paper. This method can eliminate the luminance difference for two document images, obtain higher registration accuracy and reduce time cost than the classical cross correlation method. Experimental results showed that the proposed method can effectively solve the mosaic issue of two document images.

ACKNOWLEDGEMENT

This research is funded by Funding Project for Academic Human Resources Development in Institutions of Higher Learning under the Jurisdiction of Beijing Municipality (PHR201107107, PHR20110865) and the National Natural Science Foundation of China (61371142).

REFERENCES

Cho, J., J.H. Cha, Y.M. Tai, Y. Moon and S. Lee, 2013. Stereo panoramic image stitching with a single camera. Proceedings of the IEEE International Conference on Consumer Electronics, January 11-14, 2013, Las Vegas, NV., USA., pp: 256-257.

Dong, L., D. Fu, X. Yu and T. Yang, 2012. The study on infrared image mosaic application using immune memory clonal selection algorithm. Proceedings of the 10th World Congress on Intelligent Control and Automation, July 6-8, 2012, Beijing, China, pp: 4831-4836.

Fan, S. and H. Wang, 2011. Multi-direction fuzzy morphology algorithm for image edge detection. J. Networks, 6: 895-898.

Jiang, Z., J. Wu, D. Cui, T. Liu and X. Tong, 2012. Stitching method for distorted image based on SIFT feature matching. Proceedings of the 8th International Conference on Computing and Networking Technology, August 27-29, 2012, Gueongju, Korea, pp: 107-110.

Koo, H.I., J.H. Kim and N.I. Cho, 2009. Composition of a dewarped and enhanced document image from two view images. IEEE Trans. Image Process., 18: 1551-1562.

Kumar, V.V., B.E. Reddy, A. Nagaraja Rao and U.S.N. Raju, 2008. Texture segmentation methods based on combinatorial of morphological and statistical operations. J. Multimedia, 3: 36-40.

Miao, L. and Y. Yue, 2011. Automatic document image mosaicing algorithm with hand-held camera. Proceedings of the 2nd International Conference on Intelligent Control and Information Processing, July 25-28, 2011, Harbin, China, pp: 1094-1097.

Mollah, A.F., N. Majumder, S. Basu and M. Nasipuri, 2011. Design of an optical character recognition system for camerabased handheld devices. Int. J. Comput. Sci. Issues, 8: 283-289.

Qureshi, H.S., M.M. Khan, R. Hafiz, Y. Cho and J. Cha, 2012. Quantitative quality assessment of stitched panoramic images. IET Trans. Image Process., 6: 1348-1358.

Rosa, B., M.S. Erden, T. Vercauteren, B. Herman, J. Szewczyk and G. Morel, 2013. Building large mosaics of confocal edomicroscopic images using visual servoing. IEEE Trans. Biomed. Eng., 60: 1041-1049.

Stamatopoulos, N., B. Gatos and I. Pratikakis, 2012. Performance evaluation methodology for document image dewarping techniques. IET Image Process., 6: 738-745.

Stamatopoulos, N., B. Gatos, I. Pratikakis and S.J. Perantonis, 2011. Goal-oriented rectification of camera-based document images. IEEE Trans. Image Process., 20: 910-920.

Tang, Y., J. Shin and H.C. Liao, 2012. De-ghosting method for image stitching. Int. J. Digital Content Technol. Appl., 6: 17-24.

Tong, L., G. Zhan, Q. Peng, Y. Li and Y. Li, 2012. Warped document image mosaicing method based on inflection point detection and registration. Proceedings of the 4th International Conference on Multimedia Information Networking and Security, November 2-4, 2012, Nanjing, China, pp: 306-310.

Tong, L., G. Zhan, Q. Peng, Y. Li and Y. Li, 2013b. Warped document image correction method based on heterogeneous registration strategies. Proceedings of the 5th International Conference on Machine Vision, March 13, 2013, Wuhan, China.

Tong, L., J. Chen, Q. Peng and Y. Li, 2013c. Normalized SAD method for Chinese document image registration. J. Multimedia, 8: 121-128.

Tong, L., Q. Peng, G. Zhan and Y. Li, 2013a. Distorted document image rectification based on inflection detection and heterogeneous registration. Proceedings of the International Conference on Information Technology and Software Engineering, December 8-10, 2012, Beijing, China, pp: 557-567.

Xu, J., X.H. Yang, X.X. Shao and X.Y. Meng, 2012. An image mosaic algorithm based on characteristic point matching. Proceedings of the International Conference on Audio, Language and Image Processing, July 16-18, 2012, Shanghai, China, pp: 194-197.