Document Image Registration Based on Local Feature Image and Harris Feature Points Detection

Lian-qiang Niu¹, Liu He²,* and Sheng-nan Zhang²

¹School of Software, Shenyang University of Technology, Shenyang 110870, niulq@sut.edu.cn
²,*School of Information Science and Engineering, Shenyang University of Technology, Shenyang 110870, 13898875058@163.com

Keywords: Image registration, Harris feature detection, KD tree nearest neighbor searching, two-way maximum similarity, RANSAC

Abstract. Image registration is the precondition of extracting document image content. A registration method for document image based on local feature image and Harris feature points detection is proposed. Firstly, a small amount of Harris feature points in the reference image are extracted so as to obtain the local feature image, then the Harris feature points of the registration image are also extracted, after that the feature points of the two images are matched by the Euclidean distance similarity measure and KD tree nearest neighbor searching. Secondly, in order to improve the accuracy and the efficiency of matching, the two-way maximum similarity matching is used to filtrate the matching point-pairs for the first time, and then RANSAC and the affine transformation are combined to complete the second filtrating and parameters estimation. Finally, the image is registered by using quadratic linear interpolation. The experimental results show that the proposed method has the high accuracy and efficiency of registration for the document image, and has a good resistance to image rotation, distortion and noise.

Introduction

As an important means of the computer information input, Optical character recognition (OCR) is widely used in the field of bank reconciliation, letters sorting, and questionnaire statistics. Unlike the general image recognition problem, the questionnaire statistic based on template not only involves the characters recognition, but also thinks about matching between the collected images and template so as to intercept the areas related to the answers of the users, and identify them. In the process of image acquisition, the tilt and the distort of documents themselves may reduce the recognition rate of scanned images, and they still need to be further registered with the template images after skew correcting.

A text-based image is called document image, which is usually acquired from the scanner, the degree of its tilt and distort is small, in general it also does not include other deformation and its structure is close to the template of registration image. Since the information of the document image is often more than the reference image, its registration method is different from other images'. At present, the general methods for image registration mainly aim at the questions like image fusion, image retrieval, remote sensing image analysis, target recognition [1], which are used for two or more images in the same scene to align within the same space. The registration includes a series of steps: feature information extraction, matching, estimation of spatial transformation parameters, registration, etc. Every step may adopt a variety of technologies and their combination can produce different registration methods, whose registration effects are different. But the purpose of all the existing methods is to satisfy the diversity of input data and the requirements of different registration applications. Registration methods for document
image mainly include user-defined template method, block template mapping method, hierarchical search method, and the method based on Hausdorff distance and geometric invariance. Among them, the first two methods can register image relatively accurately, but different images need different template definition, the registration progress is cumbersome and difficult to achieve automation. Hierarchical search method divides the image into different layers and hierarchically searches the image so as to improve the accuracy of feature points extraction, in which the typical one is SURF-DAISY [3]. The fourth method calculates the similarity of matching points to match them and it can simplify the calculation and improve efficiency, but its registration accuracy is restricted by the instability of Hausdorff distance.

With identification and statistics of questionnaire based on template image as the background, this paper proposes a fast registration method for document based on local feature image and Harris feature points detection to implement the fast and accurate extraction for the user handwritten regions. Firstly, a small feature image from the reference image is intercepted, and then the feature points of the reference image and registration image are detected by Harris algorithm[4]. Secondly, Euclidean distance similarity measure [5], KD tree nearest neighbor searching [6], and two-way maximum similarity [7] are combined to match feature information. Finally, the affine transformation based on RANSAC [8] and the bilinear interpolation [9] is used to complete the estimates of spatial parameters and image registration. This method not only has the high registration accuracy and speed, but also has the strong resistance to image noise.

Registration Algorithm Construct for Questionnaire Image

In the process of identification and statistics of questionnaire based on template image, the reference image is an unanswered blank image and the registration image is an answered image, shown in Fig.1. This image has the following characteristics:

(1)In the image, words and figures are in the majority and the components are simple, which have no cross and overlap between each other. The inflection points of components constitute the more prominent feature points, which can be used as the basis for the subsequent matching.

(2)The distribution of image information within the same page is uniform, and the local reference image has sufficient feature points for matching, we may utilize local feature image to register image so as to improve efficiency.

(3)Image information is not sensitive to color. In general, we may preprocess the image by using conventional gray or binarization to simplify the image and eliminate information unrelated to registration in order to further increase processing speed.

(4)There is a translation and rotation between images, and no scale variation.

Based on the above analysis, the image registration algorithm is built as follows

Step1: Carry on binary processing for image, and detect the Harris feature points of the reference image to extract its local feature image.

(1) Blank questionnaire (b) Answered questionnaire

Figure 1. The sample images.
Step2: Calculate the feature points of the registration image by Harris algorithm to obtain the effective feature points used to complete the estimation of the spatial transformation parameters.

Step3: According to the similarity of the images’ structure and elements arrangement, match and purify the feature points by two-step method, namely, the feature points are matched and purified quickly for the first time by using Euclidean distance with less calculation, KD tree nearest neighbor searching and the maximum similarity algorithm. Subsequently, the second selection of the feature points and the parameter estimation are carried on by the affine transformation based on RANSAC, which further purify the feature points at the same time of obtaining space parameters so as to make up the accuracy deficiency of the first selection.

Step4: Transform the image in space, then according the estimated spatial transformation parameters, bilinear interpolation is used for image resampling and interpolation operation in order to complete image registration.

Algorithm Implementation

Harris Feature Point Detection

The core of the Harris algorithm is to define a local autocorrelation function of signal, which is used to measure local variation when the image block is translated a small distance along different directions. Let $I$ be image function, $(x, y)$ and $(\Delta x, \Delta y)$ are the coordinates of the specified point and its local translation respectively, $(x_i, y_i)$ are the coordinates in a window which is centered at the point $(x, y)$, then the local autocorrelation function may be defined as:

$$c(x, y) = \sum_w [(x_i, y_i) - I(x + \Delta x, y + \Delta y)]^2$$

(1)

When the translation $(\Delta x, \Delta y)$ is very small, the local translation image can be approximately expressed as first order taylor series:

$$I(x + \Delta x, y + \Delta y) \approx I(x, y) + [I_x(x, y) \Delta x] + [I_y(x, y) \Delta y]$$

(2)

Where, $I_x$ and $I_y$ are the derivatives of the image in the X direction and Y direction respectively. Eq.1 and Eq.2 are combined, we can obtain

$$M(x, y) = \frac{I_x^2 \cdot I_y^2 - (I_x I_y)^2}{I_x^2 + I_y^2}$$

(3)

where, $M(x, y)$ represents the gray structure of the local neighborhood of point $(x, y)$. The eigenvalues of $M(x, y)$, which are $\lambda_1$ and $\lambda_2$, only reflect the lengths of two main axes of the local image, and they are unrelated to the directions of main axes, thus they can form a rotation invariant description. $\lambda_1$ and $\lambda_2$ have three relations:

(1) If $\lambda_1$ and $\lambda_2$ are both very small, then the partial autocorrelation function is smooth, and the grayscale of the local image window is approximately a constant.

(2) If one is larger and the other is smaller, then the partial autocorrelation function appears as ridge shape and the variation of $M(x, y)$ caused by the translation of the partial image along the ridge is small, while the variation of $M(x, y)$ caused by the translation along the orthogonal direction is large, it also means that the feature point is located in the edge of the image.

(3) If $\lambda_1$ and $\lambda_2$ are both very large, then the partial autocorrelation function is a sharp peak. The
translation along any directions will cause large variation of \( M(x, y) \), and indicate that the point is a feature point.

**Local Feature Extraction of Reference Image**

As shown in figure 2, we scan the reference image to obtain Harris feature points, and extract the first \( T \) feature points (\( T \) represents the required number of feature points). The region which contain the feature points is used for registration.

![Figure 2. Local feature extraction of reference image.](image)

**Affine Transform Space Parameter Estimation based on RANSAC**

RANSAC is a typical algorithm of image registration based on feature, whose calculation process is stable and reliable. It not only has the high precision, but also has a good ability to eliminate the mismatched point-pairs. We only need small amount of the necessary initial data. In the process of calculation, we judge which point-pairs should be to be reserved, and which ones are to be removed.

Space transformation of questionnaire image mainly includes translation, rotation and zoom with a certain degree. The steps of space parameter estimation based on affine transform are as follow:

1. Select three point-pairs randomly from \( k \) groups of candidate feature point-pairs to establish equations set, and solve six unknown parameters of matrix \( M \).
2. Carry out \( M \) matrix transformation on the rest of \( k-3 \) feature points, and calculate the distance between them and their candidate matching points.
3. If the distance is less than the set threshold, the feature point is an interior point, otherwise it is an exterior point.
4. Count the number of interior points under the current transformation matrix.
5. Reselect another three candidate matching point-pairs, repeat step (1) to (4) several times, the set that has the largest number of interior points is just the final matching point set.

**Bilinear Interpolation**

We use bilinear interpolation to sample the pixels of the registration image, and insert pixels into the reference image based on the estimated spatial parameters. Bilinear interpolation is a backward mapping algorithm, it outputs image pixel by pixel and line by line. The gray level of each pixel is uniquely determined by the interpolation of four pixels. This method has smooth function and can effectively overcome the lack of the nearest pixel interpolation, which may meet the required interpolation effect of document image, compared with higher order interpolation, it has better efficiency.

**Experiments and Results Analysis**

In order to verify the validation of the proposed algorithm, we used C++ Builder 6.0 to program and tested the real document images that were from the questionnaires of China's National Games in 2013 and some data sets of Google lab. The configuration of the experimental machine was:
Corei5-3210M CPU Intel, 2.5GHz clocked, and 4G memory.

Aim at the possible cases existing in the actual scanning process, such as rotation, distortion and noise pollution, we tested the accuracy and the efficiency of the proposed registration algorithm, and the ability of resistance to rotation, distortion and noise. In the test, the local feature image was extracted at first, as shown in Fig.3, then 100 images with the size of 480×640 were selected as the registration images, the image samples are shown in Fig.4–Fig.6. A variety of test results are shown in Table 1.

![Figure 3. Local feature image of reference image.](image1)

![Figure 4. Rotation images.](image2)

![Figure 5. Distortion images.](image3)

![Figure 6. Images with salt and pepper noise.](image4)

![Table 1. Registration result.](image5)

<table>
<thead>
<tr>
<th>Image category</th>
<th>The number of feature points</th>
<th>Matching Point</th>
<th>Correct matching points</th>
<th>Average matching rate</th>
<th>Average time[s]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rotation</td>
<td>24569</td>
<td>1985</td>
<td>1901</td>
<td>95.77%</td>
<td>2.714</td>
</tr>
<tr>
<td>Distortion</td>
<td>21586</td>
<td>1865</td>
<td>1799</td>
<td>96.46%</td>
<td>3.928</td>
</tr>
<tr>
<td>Pollution</td>
<td>20956</td>
<td>1823</td>
<td>1758</td>
<td>96.43%</td>
<td>3.255</td>
</tr>
</tbody>
</table>

In addition, we also compared the proposed algorithm with the other typical ones in terms of performance. The test results are shown in Fig.7~Fig.8.

![Figure 7. Compared with other methods.](image6)

![Figure 8. Compared with other methods.](image7)

Table 1 shows that the average matching rate of the proposed method can be maintained at more than 95% in three cases, and the average matching time is less than 4s, the accuracy and the efficiency of registration can meet the actual requires.

Through contrast experiments, it shows that this method is more suitable for the document image registration than other methods. It uses a more effective feature extraction and matching process, and by the local feature image and two-step matching and purifying, we obtain the better effect and
efficiency of registration. Although other methods have better resistance to the affine transformation of the image and the light effect, there is still a certain gap with our method in the accuracy and the efficiency due to the image information is processing by image classification or fixed template, and the matching process is not optimized.

Conclusion

In this paper, a fast image registration method based on local feature image and Harris feature point detection is proposed. Aim at the characteristics of document scanning images, the applicable methods are selected and combined, the process of registration is also be improved so as to obtain better result.

(1) Two-step method can not only ensure the matching accuracy, but reduce the computation of the higher order matching algorithm as well.

(2) For the document image with high structural similarity, using local image to extract feature points and register image can meet the requirements of the actual accuracy of registration, furthermore may significantly improve the efficiency.

Acknowledgement

This research was financially supported by Shenyang Science and Technology Project F15-194-5-00.
Liaoning Province Science and Technology Program 2015410010.

References