The Splicing Scrap Paper and Restoration Based on MATLAB Algorithm

Fengling Wang

Department of Mathematics, Heihe University , Heilongjiang164300, China

wfl45@163.com

Keywords: MATLAB; Grey system theory; Scraps of paper recovery; Mathematical model

Abstract. In this paper, it find out the way of splice the scraps of paper which come from the same page of the printed text documents together by MATLAB software. Firstly, the image is made to the value of two forms of matrix, to establish the algorithm of the scraps of paper splicing restoration model by analyzing all matrix data of the scraps of paper based on the Grey system theory and Pearson correlation coefficient as the test variable. Then the fragments were spliced restoration by MATLAB software. Because of the computer matching has certain systematical error, it must write the time node intervention and intervention with the appropriate human intervention in this paper. Finally, it gains some parts of complete Chinese pictures and tables according to the recovery results.

Introduction

Stitching broken file is important applications in judicial evidence reconstruction, historical documents and military information access to repair and other fields. In the traditional, scraps of paper recovery by artificial completed, it is high accuracy, but the efficiency is very low. Especially when the number of pieces is huge, artificial splicing is difficult to complete the task within a short period of time. With the development of computer technology, people try to find out some ways of automatic splicing technology development of the scraps of paper, in order to improve the recovery efficiency of splicing. In this paper, The establishment of scraps of paper splicing restoration model and algorithm for splice the scraps of paper which come from the same page of the printed text documents were broken by shredder, and it gains some parts of complete Chinese pictures and tables according to the recovery results.

The basic principle of mode

Correlation degree analysis

Fist of all, the establishment of the matching degree to analyze the scraps of the grey based on system model in this paper, the all data of The attachment about Scraps of paper were made the value of the two, that The image pixel gray value is set to 0 or 255, the whole image shows obvious only black and white of visual effect [1].

$f(i, j)$ is the gray value function that the (i, j) in $N \times M$ graph, μ is the level gray. Let the value of $f(i, j)$ is $[0, m-1]$, and its probability value is $p(k)$.

Let

$$p(k) = \frac{1}{MN} \sum_{f(i,j)=k} 1 \tag{1}$$

Assume that the target and background is divided by the gray value t, it is $\{f(i, j) \le t\}$ and $\{f(i, j) > t\}$, then we have the formula of optimal threshold

$$g = Arg \ \underset{0 \le t \le m\text{-}1}{Max} [\omega_0(t) \ (\mu_0(t) - \mu)^2 + \omega_1(t) \ (\mu_1(t) - \mu)^2] \tag{2}$$

Secondly, Let $X_0 = \{x_0(k) | k = 1, 2, ..., n\}$ is the reference sequence, $X_i = \{x_i(k) | k = 1, 2, ..., n\}$ is the comparative sequence, then we have the calculation formula of the correlation coefficient

$$\zeta_{i,0}(k) = \frac{\min_i \min_k |x_0(k) - x_i(k)| + \rho \max_i \max_k |x_0(k) - x_i(k)|}{|x_0(k) - x_i(k)| + \rho \max_i \max_k |x_0(k) - x_i(k)|} \tag{3}$$

There ρ is the resolution coefficient that is between 0 and 1, it usually get 0.5. $\zeta_{i,0}(k)$ is the function which is the relative difference between the comparative sequence x_i and the reference sequence x_0 at the k moment. Then the correlation function is

$$r(x_i, x_0) = \frac{1}{n} \sum_{i=1}^{n} \zeta_{i,0}(k) \tag{4}$$

$r(x_i, x_0) \in (0,1]$, it reflects the similarity comparison sequence and reference sequence on the whole.

The image edge detection

The algorithm is as follows. Fist of all, find the reference sequence and comparative sequence in the graph, Non edge points are those points which are between the pixel value, and the surrounding adjacent pixel values no difference or little difference. Let a pixel and its neighboring pixels forming a reference sequence, which take the same value, there must be non edge points. The values are 1 to 9 as reference sequence, comparative sequence comes from the 8 adjacent pixel position of each pixel and its surrounding points in the image .Then the comparative sequence for in $M \times N$ image is

$$X'_{i,j} = (x_{i-1,j-1}, x_{i-1,j}, x_{i-1,j+1}, x_{i,j-1}, x_{i,j}, x_{i,j+1}, x_{i+1,j-1}, x_{i+1,j}, x_{i+1,j+1})$$

There, $X_0 = \{1,1,1,1,1,1,1,1,1\}$ is a reference sequence, and $i = 2,3,....,M-1, j = 2,3,...,N-1$ .if $i, j = 1$ or $i = M-1, j = N-1$ [2].

Calculate the correlation function $r(X_i, X_0)$ which is comparative sequence and the reference sequence points for each pixel that are formed with the center of the image. To determine the whether pixel edge point or not, when $r(X_i, X_0)$ is some threshold θ, then we known that if the point and the reference sequence has the same characteristics , it is not edge point or it is[3].

Paper splicing by MATLAB

Now, there is some data of Scraps of paper, calculate the correlation degree of each point of paper data according to the above algorithm by MATLAB [4].

It shows that the Correlation matrix which between the left of edge of 014 picture and right of picture is maximum, it is 0.7996.

The following is the input image and extract the edge of matrix by MTLAB[5].

```
I=cell(1,19);
for b=1:19
m1=imread
(['C:\Documents and Settings\Administrator\ appendix \',int2str(b),'
.bmp']);
I{b}=m1;
end
for i=1:19
    Ai=I{1,i}(:,[1]);
    Bi=I{1,i}(:,[72]);
end
```

Thus, it is feasible that Mosaic by comparing with correlation data of the scraps of paper edge .

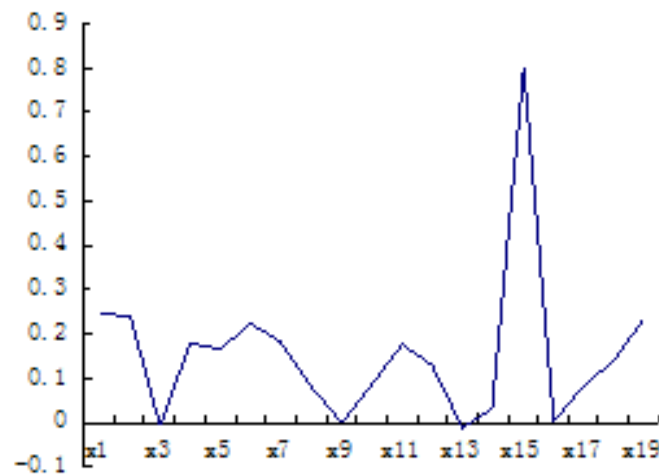Although there may be some errors, but the error is very small.

Fig.1   The correlation comparison of between the number 008 picture right edge and the other pictures left edge

It shows that error is very small, because there is only one high peak period.
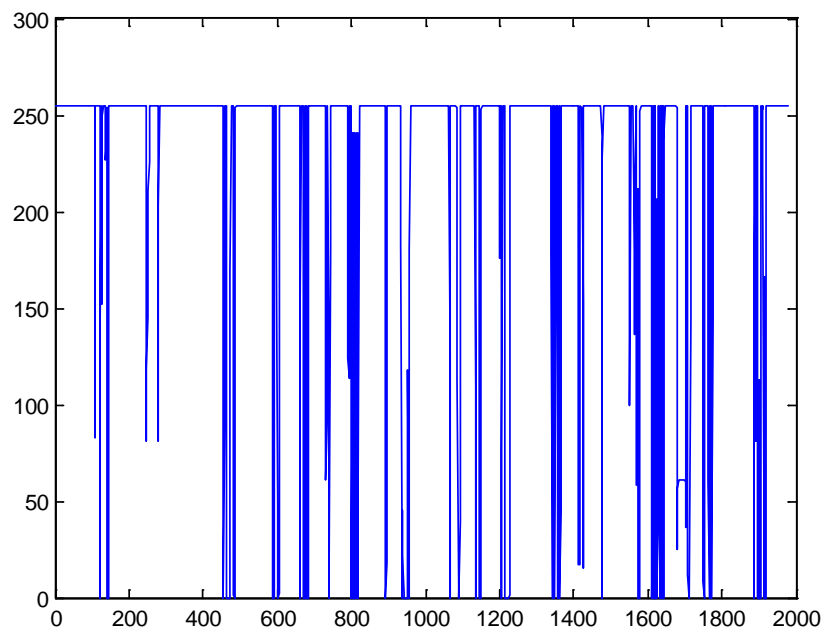

Fig.2   The right edge matrix of 008 picture

We can see that the numerical left edge matrix of 255 represents all white, and The numerical right edge matrix has not only 255 but also 0,I,e there are white points and black points from figure five, figure six.Thus, we have the picture after mosaic graphic.

Summary

It uses the simple theory to analyze the possibility of the problem, and find the correlation between the scraps of paper to establish the mathematical model. Double sided debris on the surface seems to make the problem more complicated, because there are two sided information table, the number data of information is greater, the results get better by this method.

Scraps of paper come from the paper which is cut by transverse and longitudinal, but it is know that the correlation degree of paper independent on cutting method by the grey model ,as long as the

scraps of paper edge correlation degree reaches a certain height ,it can be considered two pieces spliced together. The grey correlation analysis method is to research object and factors influencing factor value of the points as a line. Compared with the object to be identified and the drawn curve about the factors of influence and were quantified, then calculate the close degree of correlation between the factor of the research object and the object to identified，by comparing the size of the correlation to determine how influence the recognition object to the object of study.

But, the above model is implemented base on the regular pattern, although this work has done some research in this paper, due to the special application background of repair technology of automatic the paper shredder, at present, there is almost no reference material. Similar studies are mainly concentrated on the automatic restoration of historic fragment, virtual archaeology, fault analysis, computer aided design and the field of medical analysis etc.

In summary, this algorithm may be use to the irregular graphics. Due to the irregular shape of the fragments, it is not suitable for directly analyzing characteristic points of the curve, we may use the approximation method to find simplified representation, then matching analysis to the approximate contour.

The previously mentioned image segmentation technology is based on grayscale images as processing images. At present, we have the most of images are color, many of the original segmentation method for gray image is not suitable for the color image. To segment a color image, it is not only choose the appropriate color space, but also that take the segmentation methods suitable for the space. Thus, research on these problems will have a broad prospect and application value.

Reference

[1] Z.Z.Lou: Document Scraps   of Text Feature Based   on semi Automatic   Mosaic. Computer engineering and Application, (2012) No 5,P.10 (In Chinese)

[2] Z.Z.Lou: Study on Scraps of Boundary Detection Algorithm Based online Scanning. Chinese Jo urnal of scientific instrument, (2011) No.2 P.20. (In Chinese)

[3] G.L.Zhang: Study on Restoration Model Based on the Recognition of Splicing Chinese Characters Scraps. Technology Square,(2014) No.1,P.51(In Chinese)

[4] J.W.Zhou: *Application of MATLAB in mathematical modeling* (Beijing University of Aeronautics and Astronautics, Beijing 2011) P.54. (In Chinese)

[5] M.S.Yang: *The Optimization Principle, Method and Solving Softwar*e (Science Press, Beijing 2006), P.60. (In Chinese)