



Journal of Applied Sciences

ISSN 1812-5654

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

Mathematical Programming Model on Joint and Recovery of Paper Scrap

¹Jiang Zhixia, ²Yang Jianjun, ²Luo Haifeng and ¹Wen Zhengqiang

¹Department of Applied Mathematics, School of Science,
Changchun University of Science and Technology,

²School of Computer Science and Technology, Changchun, Jilin, 130022, China

Abstract: The joint and recovery of paper scrap was an optimal matching issue. The automatic restoration technology for the broken document was designed according to different cutting ways of the paper shredder. As the fragment data was one-side print file, a nonlinear programming model with no constraint conditions was established for the straight-cutting broken paper scrap from a given print file of the same page and the nonlinear programming model with restraint programming model was established for the broken scrap with straight cut and cross cut. For the paper scrap data from a one-page print file in English printed on both sides, the joint was accomplished by the ant colony algorithm.

Key words: Joint, least square method, ant colony algorithm

INTRODUCTION

Currently, a lot of researches on the automatic restoration technology for the broken document have been done in most of the developed countries internationally. However, in China, the research on this area started too late and few research accomplishments were reported. The joint of the broken document was widely applied in the judicial physical evidence recovery, historical document restoration and military information acquirement. Traditionally, the joint and recovery work needed to be done manually which would lead to high accuracy but low efficiency. Especially as the fragment amount was huge, the manual joint would cost a lot of human and material resources.

The automatic joint issue belonged to the computer vision and pattern recognition area which was accomplished by the computer processing to obtain the information of the paper scrap, like shape and color. Then the paper scrap was restored automatically or semi-automatically. Currently, most of the paper scrap joint works were accomplished manually. Although, some researchers have been done abroad, little research findings could be found for the application background specificity of the paper scrap automatic restoration technology. The description on the shape matching algorithm of the similar issue was found in some articles (Da Gama Leitao and Stolfi, 2002; Yigitsoy and Navab, 2013; Kong and Kimia, 2001; Zhu *et al.*, 2008; Liu *et al.*, 2011; Richter *et al.*, 2013; Jia *et al.*, 2006). For example, in

literature (Jia *et al.*, 2006), a method to accomplish the paper scrap joint by judging, if two profiles matched by boundary and area principles from extracting the contour line of the paper scrap was given to achieve the paper scrap automatic restoration based on the computer assistance.

BROKEN PAPER SCRAP OF ONE-SIDE PRINT DOCUMENT WITH ONLY STRAIGHT CUT

For a given broken paper scrap from the same print document with only straight cut, the fragment data of the document with 19 items on one page in English was given by the computer processing and then the joint and recovery was done.

The least square method was a mathematic optimization technology which minimized the error square and found the optimal function match. The unknown data could be calculated conveniently and the error square between the calculated and actual data was the minimum. Nonlinear least square method could be seen as a special situation for the unrestraint minimization which was widely applied in the areas like data fitting, parameter estimation and functional approximation (Yuan and Sun, 1997). The optimal matching solution for the paper scrap in the attachment by the least square method and its value was determined.

From the basic knowledge of the image processing, it was known that the pixel value range of each pixel point on the paper scraps to be jointed was 0-255, according to

which, the figure could be analyzed. The figure was firstly broken up into the point collection and the pixel matrix of the point collection of the figure was established for a random figure which was assumed as *j*-th figure. It was known that the matrix of the *j*-th figure was 1980×72 and the pixel matrix of the *i*-th column of the *j*-th figure was:

$$P_{ij} = \begin{bmatrix} P_{1ij} \\ P_{2ij} \\ \vdots \\ P_{hij} \end{bmatrix}, 0 \leq i \leq 72, 0 \leq j \leq 19$$

Firstly, for all the paper scraps to be jointed, the first one on the left was determined if all of the left boundary points were white, namely the pixel was 255, from which the first paper scrap to be jointed was found.

Obviously, for a random figure, the corresponding pixel point color difference of the other figure and it was not the same. The screening function $Q_{p,q}$ was defined and to calculate the square sum of the pixel point color value difference between the right boundary of the *p*-th figure and corresponding left boundary of the *q*-th figure, the expression was as follows:

$$Q_{p,q} = \sum_{h=1}^{1980} (P_{h,72,p} - P_{h,1,q})^2$$

where, P_{hij} meant the *h*-th row and *i*-th column of the pixel point color value matrix of the *j*-th figure.

For figure *p*, as other figures and it matched, each $Q_{p,q}$ was obtained and what needed was the optimal match, namely the nonlinear unrestraint mathematical programming model:

$$\text{Min } Q_{p,q}$$

Then figure *p* and *q* could be jointed. The program was written with the MATLAB software in the Acer computer of 5750G configured with i5 processor, the running time was 0.1 sec. Figure 1 shows the jointed results.

BROKEN PAPER SCRAP OF ONE-SIDE PRINT DOCUMENT WITH STRAIGHT AND CROSS CUT

For a given broken paper scrap from the same print document with straight and cross cut, the fragment data of the document with 209 items on one page in English was given by the computer processing and then the joint and recovery was done.

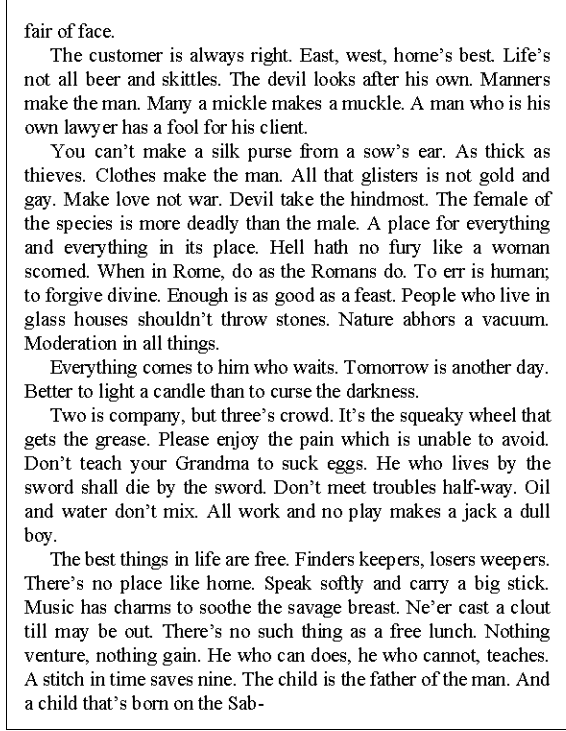


Fig. 1: Straight cut

Establishment of nonlinear programming model: In the situation of the cross cut, the method above was improved. Firstly, each paper scrap in the attachment was analyzed and the minimum vertical distance between the white pixel point on the top of each paper scrap and the first non-white pixel point was utilized and analyzed in the computer to roughly judge the shape match of the font. According to the literatures, the geometrical characteristic of the row where the paper scrap lay could be obtained by existed technology, like the row height of the character and the space between the character lines (Luo, 2012).

As shown in Fig. 2, random paper scraps were picked as the *i*-th and *j*-th page and if the minimum vertical distances between the white pixel point on the top of the *i*-th and *j*-th paper scrap and the first non-white pixel point were d_{mi} and d_{mj} , $\Delta d_{ij} = d_{mi} - d_{mj}$.

The boundary value of Δd was 0 and 5, respectively. If the value of Δd was between 0 and 5, it was considered that the *i*-th and *j*-th paper scrap could be jointed in the same line, or else the two could not be jointed in the same line for the big difference of the character structure which meant that the two paper scrap could not be jointed. Then the distances of the peaks of *i*-th paper scrap on the left and right and the first non-white pixel point horizontally were calculated and analyzed to find the minimum distance x_{si} and x_{si} , the boundary value of which was

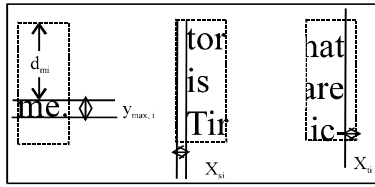


Fig. 2: Characters of the i-th paper scraps

chosen as 12, 16 and 6-28. By the same way, the maximum character space $y_{max,i}$ between the top and bottom point in another restraint condition was determined.

In the process above, the binaryzation method was adopted. By setting the threshold value, the conversion from the gray level image to the binaryzation image was accomplished (Yao and Cheng, 2002). The gray level value of the white pixel point was marked as 255 while the gray values of the other pixel points were marked as 0. In this process, the point, whose pixel value was close to that of the white pixel, was considered as the white point. The form could be expressed as follows:

$$u = \begin{cases} 255, & \text{the point is white} \\ 0, & \text{else} \end{cases}$$

Then, based on the above theory, for all the paper scrap to be jointed, the program was firstly compiled to rank the paper scraps in the descending way according to the minimum principle for the sum of d_{mi} and x_{si} , the paper scrap with the minimum value was chosen as the first element on the left top. Then each of the paper scrap was jointed by the recurrence algorithm based on the principle of from left to right and from top to bottom. In the joint process the least square method in the first model was used to screen the paper scraps to be jointed with the objective paper scrap, the $\text{Min } Q_{p,q}$ was then calculated which meant the difference of paper scrap p and q was the minimum, namely paper scraps p and q were the best match.

The optimized model was a nonlinear restraint optimal problem:

$$\begin{aligned} & \text{Min } Q_{p,q} \\ & \text{s.t.} \begin{cases} 0 \leq d_{mi} - d_{mj} \leq 5 \\ 12 \leq x_{si} \leq 16 \\ 6 \leq x_{ti} \leq 28 \\ 38 \leq y_{max,i} \leq 42 \end{cases} \end{aligned}$$

As the paper scrap p and q were jointed crosswise, the objective function was:

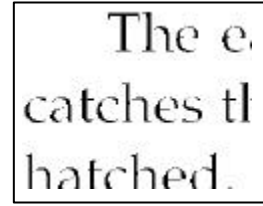


Fig. 3: Jointed correction

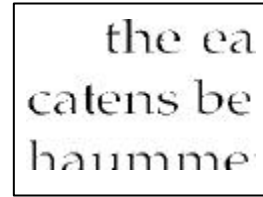


Fig. 4: Jointed error

$$\text{min } Q_{p,q} = \sum_{h=1}^{1980} (P_{h,72,p} - P_{h,1,q})^2$$

As the paper scrap p and q were jointed lengthways, the objective function was:

$$\text{min } \sum_{i=1}^{72} (P_{180,i,p} - P_{i,q})^2$$

According to the model establishment and analysis, the program was compiled with MATLAB software and the initial joint was obtained.

Artificial interference solution: In many situations, as two paper scraps were both considered as the suitable one to be jointed with the objective paper scrap by the computer, the computer would automatically chose one of them which was considered the better to joint without the logical analysis. At this time, it was easy to produce a large-scale and extensive joint error. Hence the artificial interference was of significance as the solution error of the computer happened. As shown in Fig. 3 and 4, obviously the two figures could be totally matched with the objective figure and in this situation, if the wrong figure was picked by the computer (Fig. 1), the other paper scrap would be jointed with other paper scraps wrongly either which would result in the domino effect. Meanwhile, if the artificial interference was used with the logical thinking and analysis of human, it could be realized that the word ‘‘Catens’’ in the figure to be jointed in Fig. 4 didn’t actually exist, hence the joint error would be found and corrected as in Fig. 3, which would lead to a high accuracy and right restoration of the literature.

Hence, the artificial interference needed to be added after the initial joint was obtained and the interference

bath day. No news is good news.
 Procrastination is the thief of time. Genius is an infinite capacity for taking pains. Nothing succeeds like success. If you can't beat 'em, join 'em. After a storm comes a calm. A good beginning makes a good ending.
 One hand washes the other. Take of the Devil and he is bound to appear. Tuesday's child is full of grace. You can't judge a book by its cover. Now drips the saliva, will become tomorrow the tear. All that glitters is not gold. Discretion is the better part of valour. Little things please little minds. Time films. Practice what you preach. Cheats never prosper.
 The early bird catches the worm. It's the early bird that catches the worm. Don't count your chickens before they are hatched. One swallow does not make a summer. Every picture tells a story. Softly, softly catchee monkey. Though is already is late, exactly is the earliest time. Less is more.
 A picture paints a thousand words. There's a time and a place for everything. History repeats itself. The more the merrier. Fair exchange is no robbery. A woman's work is never done. Time is money.
 Nobody can casually succeed, it comes from the thorough self-control and he will. Not matter of the today will drag tomorrow. They that sow the wind, shall reap the whirlwind. Rob Peter to pay Paul. Every little helps. In for a penny, in for a pound. Never put off until tomorrow what you can do today. There's many a slip twixt cup and lip. The law is an ass. If you can't stand the heat get out of the kitchen. The boy is father to the men. A nod's as good as a wink to a blind horse. Practice makes perfect. Hard work never did anyone any harm. Only has compared to the others early, diligently

Fig. 5: With straight and cross cut

way was: The joint couple of the paper scraps in the running result of the program was firstly found and then the serial number and the joint order were recorded. Another section of the program was then added in the initial joint program to achieve the descending order of the sum of d_{mi} and x_{si} for each paper scrap, the second sequencing value obtained was chosen as the first element on the left top to find the paper scrap to be jointed again, the serial number and joint order were recorded, then the third, the forth etc., were found until no single paper scrap was left or after the program was run for three times or more, no new paper scraps were found to be jointed with other paper scraps. Then the artificial interference was finished and the final right result was achieved by the artificial joint with the logical analysis on the character and language.

In addition, for English documents, the variable value range of the restraint condition could be changed appropriately because of the difference between English and Chinese characters, the following was used:

$$\begin{cases} 0 \leq d_{mi} - d_{mj} \leq 16 \\ 12 \leq x_s \leq 16 \\ 6 \leq x_t \leq 28 \\ 23 \leq y_{max} \leq 42 \end{cases}$$

The Fig. 5 shows the joint results:

CONCLUSION BROKEN PAPER SCRAP OF BOTH SIDES PRINT DOCUMENT

For a given both sides print document of English version from the same page with both straight and cross cut by the shredder, the joint and restoration of the paper scrap were accomplished with the help of the both sides fragment data of 209 documents given by the computer.

The paper scrap collection of the first line and column and the last line and column after being jointed was firstly judged. For which, a new variable d_{mi} was introduced to indicate the minimum distance between the white pixel point of all the peaks at the bottom of the i -th paper and the first non-white pixel point vertically. From the computer program, the d_{mi} , d_{mj} , x_{si} and x_{sj} could be obtained and the four matrixes were established for the paper scrap. According to the principle that d_{mi} of the first line elements was bigger than other elements and the difference between d_{mi} and the elements in each line should not be too big and by the artificial interference to observe the value of each element in the matrix, the 19 paper scraps in the first line, the last line, the first column and the last column could be found soon. Then according to the principle that only one common element existed in a random line or column, the first element in the first line could be determined soon and the optimal match for each paper scrap could be found according to the principle of from left to right and from top to bottom. In the same way, the last element of the first line, the first element of the last line and the last element of the last line were determined and each match result could be obtained by taking the three elements as the initial objective paper scrap to be matched.

For the restriction of the both sides print, the algorithm accuracy was lowered. Hence, in the matching process of the objective paper scrap and other paper scraps, the ant colony algorithm was adopted.

The ant colony algorithm was a probabilistic and overall situation search method and the uncertainty property in this algorithm would lead to more opportunities to find the optimal solution in the overall situation. The ant colony algorithm could increase the match accuracy for the paper scrap, its optimization process did not rely on the rigorous mathematic property of itself and it had the potential parallelism which increased the efficiency and instant reaction ability of the algorithm. In the matching process, the paper scrap number became larger and the pixel of the figures was smaller, hence the ant colony algorithm was used in this paper to process optimally and increase the match accuracy.

The four matching results from the above process were observed and the artificial interference was added at this time which was to record the right joint results, adjust the wrong jointed paper scraps which was artificially

What can't be cured must be endured. Bad money drives out good. Hard cases make bad law. Talk is cheap. See a pin and pick it up, all the day you'll have good luck; see a pin and let it lie, bad luck you'll have all day. If you pay peanuts, you get monkeys. If you can't be good, be careful. Share and share alike. All's well that ends well. Better late than never. Fish always stink from the head down. A new broom sweeps clean. April showers bring forth May flowers. It never rains but it pours. Never let the sun go down on your anger.

Pearls of wisdom. The proof of the pudding is in the eating. Parsley seed goes nine times to the Devil. Judge not that ye be not judged. The longest journey starts with a single step. Big fish eat little fish. Great minds think alike. The end justifies the means. Cowards may die many times before their death. You can't win them all. Do as I say, not as I do. Don't upset the apple-cart. Behind every great man there's a great woman. Pride goes before a fall.

You can lead a horse to water, but you can't make it drink. Two heads are better than one. March winds and April showers bring forth May flowers. A swarm in May is worth a load of hay; a swarm in June is worth a silver spoon; but a swarm in July is not worth a fly. Might is right. Let bygones be bygones. It takes all sorts to make a world. A change is as good as a rest. Into every life a little rain must fall. A chain is only as strong as its weakest link.

Don't look a gift horse in the mouth. Old soldiers never die, they just fade away. Seeing is believing. The opera ain't over till the fat lady sings. Silence is golden. Variety is the spice of life. Tomorrow never comes. It is ain't broke, don't fix it. Look before you leap. The road to hell is paved with good

However, for the paper scrap with big information, a great amount of artificial interferences was needed in this model.

REFERENCES

- Da Gama Leitao, H.C. and J. Stolfi, 2002. A multiscale method for the reassembly of Two-dimensional fragmented objects. *IEEE Trans. Pattern Anal. Machine Intell.*, 24: 1239-1251.
- Jia, H. Y., L.J. Zhu, Z.T. Zhou and D.W. Hu, 2006. A shape matching method for automatic reassembly of paper fragments. *Comput. Simul.*, 23: 180-183.
- Kong, W. and B.B. Kimia, 2001. On solving 2D and 3D puzzles using curve matching. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Volume 2, December 8-14, 2001, Toronto, Canada*, pp: II-583-II-590.
- Liu, H., S. Cao and S. Yan, 2011. Automated assembly of shredded pieces from multiple photos. *IEEE Trans. Multimedia*, 13: 1154-1162.
- Luo, Z., 2012. Semi-auto stitching of scrapped paper based on character characteristic. *Comput. Eng. Appl.*, 48: 207-210.
- Richter, F., C.X. Ries, N. Cebren and R. Lienhart, 2013. Learning to reassemble shredded documents. *IEEE Trans. Multimedia*, 15: 582-593.
- Yao, Y.H. and Y. Cheng, 2002. A binarization method to process scanned images of id card. *Mach. Build. Autom.*, 5: 65-68.
- Yigitsoy, M. and N. Navab, 2013. Structure propagation for image registration. *IEEE Trans. Med. Imag.*, 32: 1657-1670.
- Yuan, Y.X. and W.Y. Sun, 1997. *Optimization Theory and Method*. Science Press, Beijing, China.
- Zhu, L., Z. Zhou and D. Hu, 2008. Globally consistent reconstruction of Ripped-up documents. *IEEE STrans. Pattern Anal. Machine Intell.*, 30: 1-13.

Fig. 6: Joint results

recognized and re-matched the wrong jointed paper scraps which could not be artificially recognized from the paper edge. The process should be repeated until the right matching results, Fig. 6 was obtained.

CONCLUSION

In the joint process for the one-side paper scrap with only straight cut, the artificial interference was not needed for either English or Chinese version while in the joint process for the one-side paper scrap with straight and cross cut, the artificial interference was needed. For the both sides paper scrap with straight and cross cut, the joint was correct with the help of the ant colony algorithm.