

Research Article

Reconstruction of Shredded Paper Documents by Feature Matching

Peng Li,¹ Xi Fang,² Lianglu Pan,² Yi Piao,² and Mengjun Jiao²

¹ School of Science, Department of Mechanics, Wuhan University of Technology, Wuhan 430000, China

² School of Science, Department of Mathematic, Wuhan University of Technology, Wuhan 430000, China

Correspondence should be addressed to Xi Fang; fxwhut@163.com

Received 27 October 2013; Accepted 23 January 2014; Published 6 March 2014

Academic Editor: Manuel Ruiz-Galan

Copyright © 2014 Peng Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Splicing the shredded paper means the technology that, according the paper, which has been shredded to design a particular algorithm to splice and recover the original paper. This paper introduced the algorithm of splicing the shredded paper based on the matching to texture feature. By means of this algorithm, we modeled for the problem of splicing the shredded paper and solved it. And, we used the algorithm to splice both pieces of English shredded paper and Chinese shredded paper. The recovered paper had proved the accuracy of the algorithm.

1. Introduction

Repairing of fragments is important in that it can be applied to many areas such as restoration of archeological findings, the obtainment of modern war information. A direct and valid way is splicing the fragments manually. But when there was a huge quantity of the fragments, splicing manually would consume vast human resources [1]. Even during the process of splicing some valuable fragments might be damaged by workers unintentionally. Along with the development of technology, more and more people tried to resolve this type of problem which could be attributed to the problem of how to repair the 2D fragments. As the typical problem of repairing of 2D fragments, the technology of splicing the shredded paper had received much more attention than before. In many cases, people need to recover the shredded paper which may carry some important information. Those pieces of paper might be shredded by shredder or some other ways. For now, to achieve the solution of this problem, researchers had proposed various algorithms. However, there were some limitation existed in many main trend of the splicing technologies [2]. This paper introduced the algorithm based on the matching to texture feature. This algorithm is modeled for the paper which had been shredded only in lengthwise or been shredded horizontally and vertically. In the process of solving the problem, we also designed the

different details for both kinds of shredded paper whose text was printed in English and Chinese.

This paper described the main problems of splicing the shredded paper and proposed the core algorithm of the matching to texture feature. Afterwards, this paper discussed the solutions for the all those problems in two conditions in which the text was printed in English and Chinese. Results showed that the proposed algorithm is efficient. And finally, there would be a summary of the recovered result.

2. Main Problem and Assumption

Based on the paper which had been shredded in different ways, there were the following problems to be solved.

- (1) For shredded paper which was cut in lengthways from the same page of file, it was necessary to build the corresponding mathematical model and algorithm.
- (2) Given the condition that the paper had been shredded either horizontally or vertically, we should model the corresponding mathematical model and algorithm which is certainly different from (1).
- (3) Under the condition in Problem 2, we took the further consideration that the document had been printed double-sided which had made it more difficult to

complete the reconstruction of shredded paper. Based on the model from the above problem, we should propose an advanced algorithm to deal with this problem.

- (4) There was a conclusive difference that height of the word was not consistent between the English text and Chinese text.

In this paper, the model of those problems was presented based on the following necessary and reasonable assumptions.

- (1) The text form from exactly the same paper was consistent basically.
- (2) There was no piece of shredded paper lost.
- (3) The size of each piece of the shredded paper was identical.
- (4) The paper was shredded paper regularly.
- (5) The direction of the shredded paper was consistent.

All the conclusions which yield from this paper were based on the above assumptions. And it was worth noting that at some point the shredded document might need to be preprocessed to make sure that all pieces of paper satisfied the assumptions. For instance, when the text of paper was oblique or distorted for reason of wet, we should have to do further preprocess to correct it before the pieces of paper were scanned into computer.

3. Modeling and Solving for Problem: Repairing of Shredded English Paper

3.1. Analysis for Problem 1. For shredded paper which was cut in lengthways from the same page of file, we take the shredded paper in Figure 1 for an example while modeling the mathematical model and algorithm. From Figure 1 we could see that the paper had been shredded to 19 pieces.

The scanning copies shown in Figure 1 had never been handled with some preprocessing software. And in fact, nearly more than half of the pieces need to be preprocessed for various reasons such as wrinkled surface of paper. The scanning copies mentioned in this paper were the results of this kind of preprocessing.

To solve this problem, firstly, we would quantify the pixel from the left and right edge of the shredded paper, which could yield matching degree-matrix for each piece of shredded paper. In this case, the match degree expressed the similarity for splicing [3]. Judgment of the left edge of each piece could help us determine the right piece which was from the left edge of the original file as the first piece. Then, we would select another piece of shredded paper and splice both pieces with each other. This selected piece should have the highest match degree with the right edge of the first piece. By this analogy, we could deduce that when processing the i th splicing, we would also select the adjacent piece by judging the match degree. Finally, we can get the correct splicing result of the original file.

st, it, we, e, r of ys, st, fai, fac, rig, wa, er, i, e's, be, o, m, on, Ea, st, l, s al
il, k, The, ook, use, Th, tle, m, e, c, s, T, ski, er, a, his, o, w, b, et, ter, dev, s, al, nd
e, m, mi, iaku, all, e's, fan, A, Lif, not, y, a, M, h, n, u, c, k, es, a, han
'or, a, fc, his, m, i, m, n, v, er, l, M, ers, has, w, y, own, n, is, n, t, o, l, clie, a, la
so, v, or, v's, -is, l, n, v, pur, as, ma, who, se, il, k, ike, s, th, ick, -ma, A, n, a, ear, a, s
at, s, Al, glis, a, n, Yo, e, m, l, d, u, c, a, n, the, s, m, not, go, the, s, i, th, t, ers, ake
ir, a, l, It, nd, Clot, eve, sw, od, thi, s, C, ord, ian, a, t, ie, nd, go, n, g, i, gl, se, s, fa, wi, r, th
ke, i, evi, the, is, n, e, p, var, he, Th, en, : D, o, t, v, lo, v, o, s, t, T, ike, ldr, I, ta, hir, e, n
han, adl, the, Ma, d, g, o, re, ice, a, n, a, y, A, de, s, m, pec, A, ple, z, sp, ale, y, tl, m, es, i
ace, i, its, H, i, th, n, a, h, i, n, t, y, fer, e, of, g, i, r, e, y, and, n, o, n, f, u, g, a, h, t, s, pl, ell, l, eve
ne, in, I, do, thin, 'ev, A, h, o, ns, for, e, y, en, : l, V, cor, R, co, m, a, n, s, the, Rom, as, nec
e, l, di, Enc, i, m, a, e, a, f, o, r, o, d, l, ike, wc, gi, ve, to, i, m, a, as, go, s, h, h, is, vi, n, u, g, u, n,
use, l, a, s, es, s, r, i, . T, r, e, i, w, do, o, e, n, g, l, i, v, i, l, e, r, i, t, t, thr, e, o, p, al, d, i, s, h, h, o, u, wh
er, a, t, M, i, o, n, . P, a, fe, acu, as, e, a, t, um, a, v, i, b, b, i, h, ir, i, g, s, re, a, all, l, o, d, i, o, n, o, r, s
. To, o, w, m, c, atu, nes, i, m, i, n, e, r, sto, : N, wh, o, h, om, sar, o, t, h, i, g, c, w, i, a, t, s, o, r, o, e, s
t, s, e, t, t, o, he, t, h, i, n, E, v, le, t, l, e, r, e, y, n, a, n, d, g, h, t, ss, . S, o, l, i, g, k, ne, cur, d, a, c
wd, s, a, . I, er, t, y, B, t, h, r, ky, da, ette, ee, 's, ou, pan, sq, aea, om, q, the, cro, t, s, y, l
joy, ase, the, s, c, Tw, e, is, 'o, i, Ple, eas, t, w, i, h, i, c, h, gets, sin, en, e, p, i, g
idn, ir, C, i, a, t, at, g, i, eel, i, ch, He, w, t, t, h, y, t, e, D, o, e, g, j, s, l, o, d, i, c, k, br, a, r, o, s, t, n
re, s, e, b, two, a, v, a, b, l, e, hal, e, t, un, e, t, o, I, d, i, d, s, e, s, e, n, t, m, u, t, D, o, y, t, r, d, w
m, i, i, d, c, x, A, s, b, y, o, l, l, I, w, n, o, w, i, v, e, a, t, e, a, n, c, i, y, C, k, a, n, d, -w, a, w, o, r, n, t, l, l
half, ub, l, y, . t, r, o, e, s, l, b, o, i, a, c, l, ack, l, ack, l, ull
ind, ree, e, r, s, J, y, n, e, a, r, s, p, l, a, n, a, k, e, f, i, l, i, f, i, n, g, r, s, l, o, s, e, t, h, i, s, p, e, F, kee, i
Spe, o, m, a, k, s, est, Th, e, lib, r, y, e, b, a, h, l, a, c, 's, n, i, n, d, car, e, r, e, ly, e, e, t, soft, o, p
the, t, o, s, Th, e, p, e, r, a, r, i, s, t, w, e, r, s, n, s, c, d, i, s, i, c, y, e, b, t, r, e, a, Mt, v, a, g, s, o, o, : s, a, ha
r, e, t, I, no, k, b, i, g, 'be, s, a, a, t, s, t, i, c, ou, d, a, y, at, t, h, i, n, g, a, c, l, o, t, h, t, 'he, i, su, ill, M
ain, hin, . H, e, t, a, 'e, r, i, r, e, es, Ne, cas, not, n, t, u, n, i, n, g, can, do, l, o, t, h, g, g, z, w, ; ve
e, s, a, i, n, t, v, e, s, i, n, e, l, u, s, t, i, l, d, f, r, e, i, n, t, c, h, s, A, t, e, a, t, h, e, 'ch, o, t, 'e, e, 'i, m, n, r, i, t
h, e, chil, 's, b, a, n, n, a, wh, A, n, t, b, -h, e, o, c, a, n, . t, he, t, h, e, t, h, e, S, e, o, f, t, o, n, l, o, t, o, m, m
n, r, i, t, f, a, t

FIGURE 1: Scanning copies of the 19 pieces of shredded paper.

For the shredded paper in Figure 1, we can do the binarization of pixel from the edge of each piece of the shredded paper by programming. There were n pixels which composed the left and right edge of the piece of shredded paper [4]. Suppose that vector \mathbf{x}_i is the binarization of the pixel from the left edge of the i th piece and vector \mathbf{y}_i is the binarization of the pixel from right edge. In addition,

$$\begin{aligned} \mathbf{x}_i &= (x_{i1}, x_{i2}, \dots, x_{in})^T, \\ \mathbf{y}_i &= (y_{i1}, y_{i2}, \dots, y_{in})^T. \end{aligned} \quad (1)$$

And then, we could define the number of the pixel matching of the right edge of the i th piece and the left edge of the j th piece as

$$\begin{aligned} a &= \sum_{i=1}^n x_{ik} y_{jk}, \\ b &= \sum_{i=1}^n y_{jk} (1 - x_{ik}), \\ c &= \sum_{i=1}^n x_{ik} (1 - y_{jk}), \\ d &= \sum_{i=1}^n (1 - x_{ik}) (1 - y_{jk}). \end{aligned} \quad (2)$$

There was a necessary introduction that in the result of the binarization the value "1" represented the black pixel and "0" represented the white [5]. Thereinto, the symbol a is the (1-1) matching number of vector \mathbf{x}_j and vector \mathbf{y}_j . Analogously, the symbol b is the (0-1) matching number, the symbol c is the 1-0 matching number, the symbol d is the (0-0) matching number.

The matching measure of binary vector is obtained by method of simple match coefficient:

$$\delta_{ij} = \frac{a+d}{a+b+c+d}. \quad (3)$$

In the above equation, the numerator is the sum of the matching number of (1-1) and (0-0). The denominator is the sum of all the numbers.

By that analogy, we can get the matching degree of left and right edge of any pieces within m pieces of shredded paper:

$$\delta = \begin{pmatrix} \delta_{11} & \delta_{12} & \cdots & \delta_{1m} \\ \delta_{21} & \delta_{22} & \cdots & \delta_{2m} \\ \vdots & \vdots & & \vdots \\ \delta_{m1} & \delta_{m2} & \cdots & \delta_{mm} \end{pmatrix}_{m \times m}. \quad (4)$$

The symbol δ_{ii} represented the matching degree of both the left and right edges of the i th piece. It meant nothing.

Through running the program we could regard the certain piece ("008" piece) as the part of the left edge of the original paper for the reason that the "008" piece of shredded paper had the regular blank on its left edge. So we chose 008 as the beginning of our splicing work.

We would get vectors of the "008" and other pieces of shredded paper.

$$\delta_8 = (\delta_{81}, \delta_{82}, \dots, \delta_{8m}). \quad (5)$$

From the matching vector we can know that the "014" piece has the maximum similarity with the 008 piece of shredded paper. So we spliced the 014 piece and the 008 piece as the first step.

Then, we would get vectors of the "014" and other pieces of shredded paper:

$$\delta_{14} = (\delta_{141}, \delta_{142}, \dots, \delta_{14m}). \quad (6)$$

Following the above steps until all pieces of the shredded paper had been spliced we could recover the original paper. The correct order was 3-6-2-7-15-18-11-0-5-1-9-13-10-8-12-14-17-16-4.

The result of the repairing of English shredded paper was showed as Figure 2.

3.2. Analysis for Problem 2. For this problem we should design for the model that the paper had been shredded either horizontally or vertically. Here we took the following shredded paper for example in Figure 3. The paper has been shredded into 11(lines) * 19(rows) pieces.

3.2.1. Solving Idea and Method. Supposed that all pieces of the shredded paper had been splicing regularly so we could get a 11*19 grid. Each piece of the shredded paper could be located by its (x, y) coordinate. To simplify the whole problem, we might as well ignore the row coordinate y at first. At this point we should only determine the line coordinate x of each piece of shredded paper.

We supposed further that the pieces of shredded paper in each line had already been spliced in correct sequence [6].

fair of face.

The customer is always right. East, west, home's best. Life's not all beer and skittles. The devil looks after his own. Manners maketh man. Many a mickle makes a muckle. A man who is his own lawyer has a fool for his client.

You can't make a silk purse from a sow's ear. As thick as thieves. Clothes make the man. All that glitters is not gold. The pen is mightier than sword. Is fair and wise and good and gay. Make love not war. Devil take the hindmost. The female of the species is more deadly than the male. A place for everything and everything in its place. Hell hath no fury like a woman scorned. When in Rome, do as the Romans do. To err is human; to forgive divine. Enough is as good as a feast. People who live in glass houses shouldn't throw stones. Nature abhors a vacuum. Moderation in all things.

Everything comes to him who waits. Tomorrow is another day. Better to light a candle than to curse the darkness.

Two is company, but three's a crowd. It's the squeaky wheel that gets the grease. Please enjoy the pain which is unable to avoid. Don't teach your Grandma to suck eggs. He who lives by the sword shall die by the sword. Don't meet troubles half-way. Oil and water don't mix. All work and no play makes Jack a dull boy.

The best things in life are free. Finders keepers, losers weepers. There's no place like home. Speak softly and carry a big stick. Music has charms to soothe the savage breast. Ne'er cast a clout till May be out. There's no such thing as a free lunch. Nothing venture, nothing gain. He who can does, he who cannot, teaches. A stitch in time saves nine. The child is the father of the man. And a child that's born on the Sab-

FIGURE 2: Result of the reconstruction of English shredded paper.

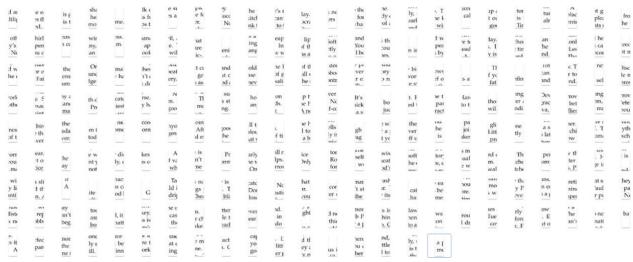


FIGURE 3: Scanning copies of the 209 pieces of shredded paper.

And then, we only needed to get the sequence of all the 11 lines of shredded paper. We defined symbol d as the distance of pixels between the bottom of the text in the first line and the top of the piece of shredded paper. By using inequality and enumeration method we could get the d values of all the 11 lines. And by comparing the d values of lines with the d value of each piece of shredded paper we can get the line coordinate x of the latter (based on the actual situation, this paper considers that the d values of the shredded paper from the same line in original paper were identical).

Up to this point, essentially, the Problem 2 was simplified into Problem 1. We could easily recover the shredded paper by using the model and algorithm that were proposed in Problem 1.

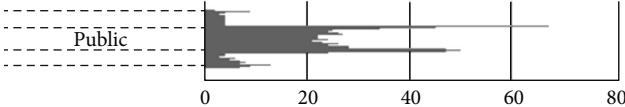


FIGURE 4: Frequency distribution of black pixel.

3.2.2. Preprocessing of Data. To get the d values of all the 11 lines, it was necessary to do the preprocessing of data of shredded paper. We defined the height of row h as the pixel distance between the bottom of the text in the n th line and the bottom of the text in the $(n - 1)$ th line. For the English shredded paper, there was the significant difference in values of h between each other [7]. On the basis of this characteristic, we could get the area which had the high concentration of pixels by progressive scanning technology of pixel. Through this method, we could get the height of row h .

As we all know, some letters occupied additional space like “*d*,” “*b*,” “*p*,” “*q*,” and so on. When there was a sentence which was constituted by letters, the black pixel proportion of the middle part of the sentence on the paper was not uniform [8]. Take the word “public,” for example, the frequency distribution of black pixel was showed as Figure 4.

It was obvious from Figure 4 that the black pixels concentrated on the middle part of the whole word. In view of the above, this paper selected the way of scanning the shredded paper from top to bottom to get the line which has the maximum sudden-change of black pixel. From this we might judge the location of the line from the original paper. In this case, we defined the bottom of the text as the third dotted line. Then we could get the height of the row $h = 64$ pixel.

3.2.3. Modeling and Solving. As previously stated, we would ignore the sequence of the line of the shredded paper. At this point, we supposed that all pieces of the shredded paper had been splicing correctly as before. We modeled 9 pieces of shredded paper for analysis.

As Figure 5 showed, the rectangular boxes which were circled by thick lines meant a piece of shredded paper. While the rectangular boxes with the dotted lines represented the lines of text.

The symbol H was the pixel height of the line of the shredded paper. From the pieces of shredded paper we could get that the H of each piece of shredded paper was 180 pixels and the line coordinate $x = 1, 2, \dots, 11$. The symbol h was the height of the word. Thus, we might get the pixel distance d by the following method.

Suppose that the lines of the text were m ; we could get the minimum positive integer which meet the inequality that $mh - xH > 0$. Do plug m into the following formula:

$$d_{x+1} = mH - xH. \quad (7)$$

Then we could obtain the value of d of $(x + 1)$ th.

For example, assuming the d_1 value of first line of the shredded paper, we used the following process to get the value of d_2 and d_3 .

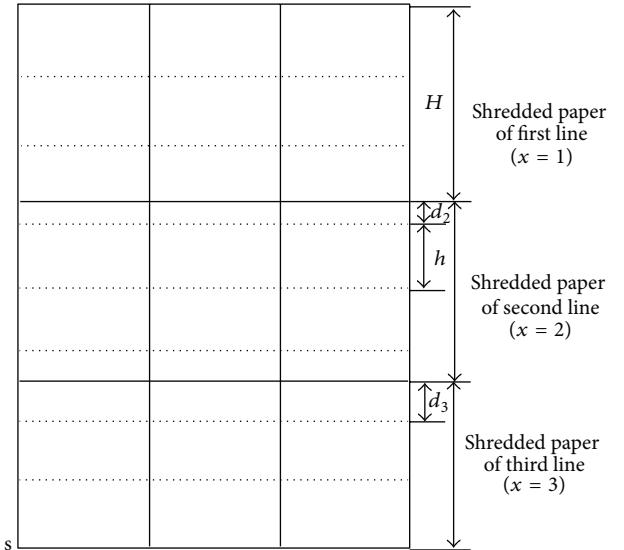


FIGURE 5: 9 pieces of shredded paper model.

When $x = 1$, we could know the $d_2 = 68m - 1 * 180$. From this equation we could easily work out that $m = 3$. So d_2 is 24 pixels.

And when $x = 2$, with the same method, we can know that d_3 is 48 pixels.

By that analogy, we could obtain all the d values of each line of shredded paper, and then we would get the right order of each line of shredded paper.

By using of the model in Problem 1, we could get the row coordinate y . When we got all the coordinates (x, y) we could locate all pieces of the shredded paper. And the final result is shown as Figure 6.

3.3. Analysis for Problem 3. Based on the Problem 2, the purpose of the Problem 3 was to repair the shredded paper which was printed double-sided. It posed a new problem that before repairing the shredded paper we should identify whether the shredded paper was the front or the back side of the original paper. Through the relation between the d value and line coordinate of shredded paper we could obtain the row coordinate of shredded paper. And then, the problem came down to an easier problem: sorting and splicing of shredded paper which was printed double-sided in the same line of the original paper.

3.3.1. Solving Idea and Method. As previously stated, we should identify the front and back of the shredded paper while sorting and splicing them. Given that the text of both sides has the same line height, firstly, we can still use the model in Problem 2 to get the correspondence of the row coordinate of the shredded paper and the value of d . By calculating the value of d of each piece of shredded paper, we could get the row coordinate of each of them. Then, the problem was simplified into an easier problem of how to sort and splice both sides of the shredded paper which were on

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. Talk 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 flies. 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. Thought 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 the 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 man. 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

FIGURE 6: The result of the reconstruction of English shredded paper.

the same row. This paper introduced a new concept: sum-similarity [9]. While we were splicing two pieces of shredded paper, both sides of the shredded paper would generate the similarity of splicing on the edge of themselves [10]. And, the sum-similarity is the summation of both two values of similarity. For the shredded paper which was printed on both sides, there would have been two conditions which would generate two sets of sum-similarity. Consequently, the identification of the front and back of the shredded paper was turned into the comparison between the two sets of the sum-similarity. Afterwards, we would always take out the piece of shredded paper which had the maximum sum-similarity splicing with the special piece of shredded paper at every turn. Finally, we could get the recovered paper.

3.3.2. Modeling and Solving. By the conclusion based on Problem 2, we could obtain the d value of each piece of the shredded paper which would yield the line coordinate of each piece. Suppose that the two sides of each piece of shredded paper had been named A and B side as well as the original paper. And we supposed further that the text in A side of the shredded paper ought to belong to the A side of the original paper. So, there were two cases to be considered when the t th piece was splicing with $(t+1)$ th piece of shredded paper under

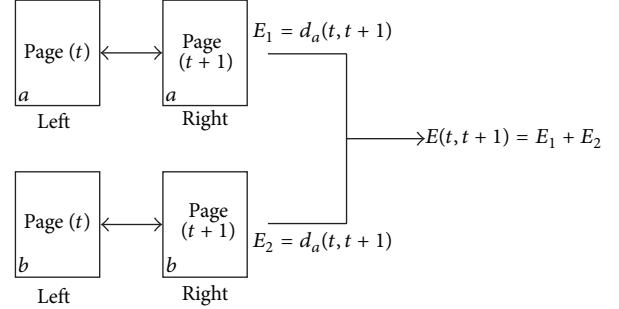


FIGURE 7: Flowchart for the process in Case 1.

the condition when both pieces came from the same line of the original paper.

Case 1. When we were splicing t th piece of shredded paper with $(t+1)$ th piece, the A side of t th piece should be spliced with the A side of $(t+1)$ th piece as well as their B sides. And at this point there would lead to two similarities of splicing. Then we could get the overall similarity by summing the two similarity of splicing. The exact process was showed as Figure 7.

Case 2. When we were splicing t th piece of shredded paper with $(t+1)$ th piece, the A side of t th piece should be spliced with the B side of $(t+1)$ th piece while the B side of t th piece was spliced with the A side of $(t+1)$ th piece, there would also lead to two similarities of splicing refer to Case 1. Certainly, we would get the overall similarity by the same way.

After that, we could choose the most special piece of shredded paper through the program such as the piece which was from the left edge of the original paper. We defined this piece of shredded paper as the first piece. This way, when we try splicing the rest of 18 pieces of shredded paper with the first piece, there would be 36 values of overall similarity. By the algorithm in Problem 1, we chose the piece which had the maximum similarity to splice with the first piece.

By that analogy, we would foresee that when we sort the i th piece, there would be $2 * (19 - i)$ values of overall similarity. We would always take out the piece of shredded paper which had the maximum sum-similarity splicing with the special piece of shredded paper at every turn until all pieces of shredded paper had been spliced. By that point, we might obtain the result of reconstruction of shredded paper as Figures 8 and 9 showed.

Finally, a summary of the time performance of this algorithm could be described by Figure 10.

Figure 10 had shown the relation between the number of fragments in a certain line and the time consumption when the algorithm was used to splice the double-sided of shredded paper. For an 11(lines) * 19(rows) double-sided shredded paper the algorithms would consume no longer than 10 milliseconds. From Figure 10 we could learn that there was a nonlinear relationship between the time consumption

He who laughs last laughs longest. Red sky at night shepherd's delight; red sky in the morning, shepherd's warning. Don't burn your bridges behind you. Don't cross the bridge till you come to it. Hindsight is always twenty-twenty.

Never go to bed on an argument. The course of true love never did run smooth. When the oak is before the ash, then you will only get a splash; when the ash is before the oak, then you may expect a soak. What you lose on the swings you gain on the roundabouts.

Love thy neighbour as thyself. Worrying never did anyone any good. There's nowt so queer as folk. Don't try to walk before you can crawl. Tell the truth and shame the Devil. From the sublime to the ridiculous is only one step. Don't wash your dirty linen in public. Beware of Greeks bearing gifts. Horses for courses. Saturday's child works hard for its living.

Life begins at forty. An apple a day keeps the doctor away. Thursday's child has far to go. Take care of the pence and the pounds will take care of themselves. The husband is always the last to know. It's all grist to the mill. Let the dead bury the dead. Count your blessings. Revenge is a dish best served cold. All's for the best in the best of all possible worlds. It's the empty can that makes the most noise. Never tell tales out of school. Little pitchers have big ears. Love is blind. The price of liberty is eternal vigilance. Let the punishment fit the crime.

The more things change, the more they stay the same. The bread always falls buttered side down. Blood is thicker than water. He who fights and runs away, may live to fight another day. Eat, drink and be merry, for tomorrow we die.

FIGURE 8: The result of the reconstruction of text from A side in English shredded paper printed double-sided.

and the number of fragments. The time complexity of the algorithm could be expressed as $o(n^2)$.

4. The Application of Improved Algorithm Used in Chinese Shredded Paper

The algorithm above had been used in the repairing English shredded paper. But obviously, the algorithm was not absolutely appropriate for the Chinese shredded paper for the reason that there was a crucial difference between the two kinds of text—the height of the row.

More specifically, to the Chinese characters which were from the text, the height of the characters was mainly identical. Based on the actual situation we could get the height of the row h by the program easily rather than calculate the frequency distribution of black pixel. And once we got the h , we might model and solve the problem with the same way used in English shredded paper. Totally speaking, the algorithms for Chinese and for English shredded paper were similar in essence. Both kinds of the shredded document should be preprocessed to satisfy the five assumptions mentioned in chapter 2. Both executions of the two algorithms were also similar with each other. The main difference was the course of getting the height of the row h . Here we gave

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. If it ain't broke, don't fix it. Look before you leap. The road to hell is paved with good

FIGURE 9: The result of the reconstruction of text from B side in English shredded paper printed double-sided.

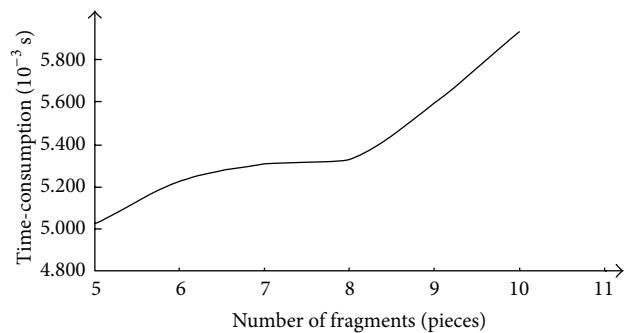


FIGURE 10: Time performance of the algorithm for shredded document.

the concrete steps which could be described as the following Figures 11, 12, and 13.

And, Figure 12 showed the scanning copy of shredded paper and Figure 13 showed the result of recovery of Chinese paper which was shredded either horizontally or vertically.

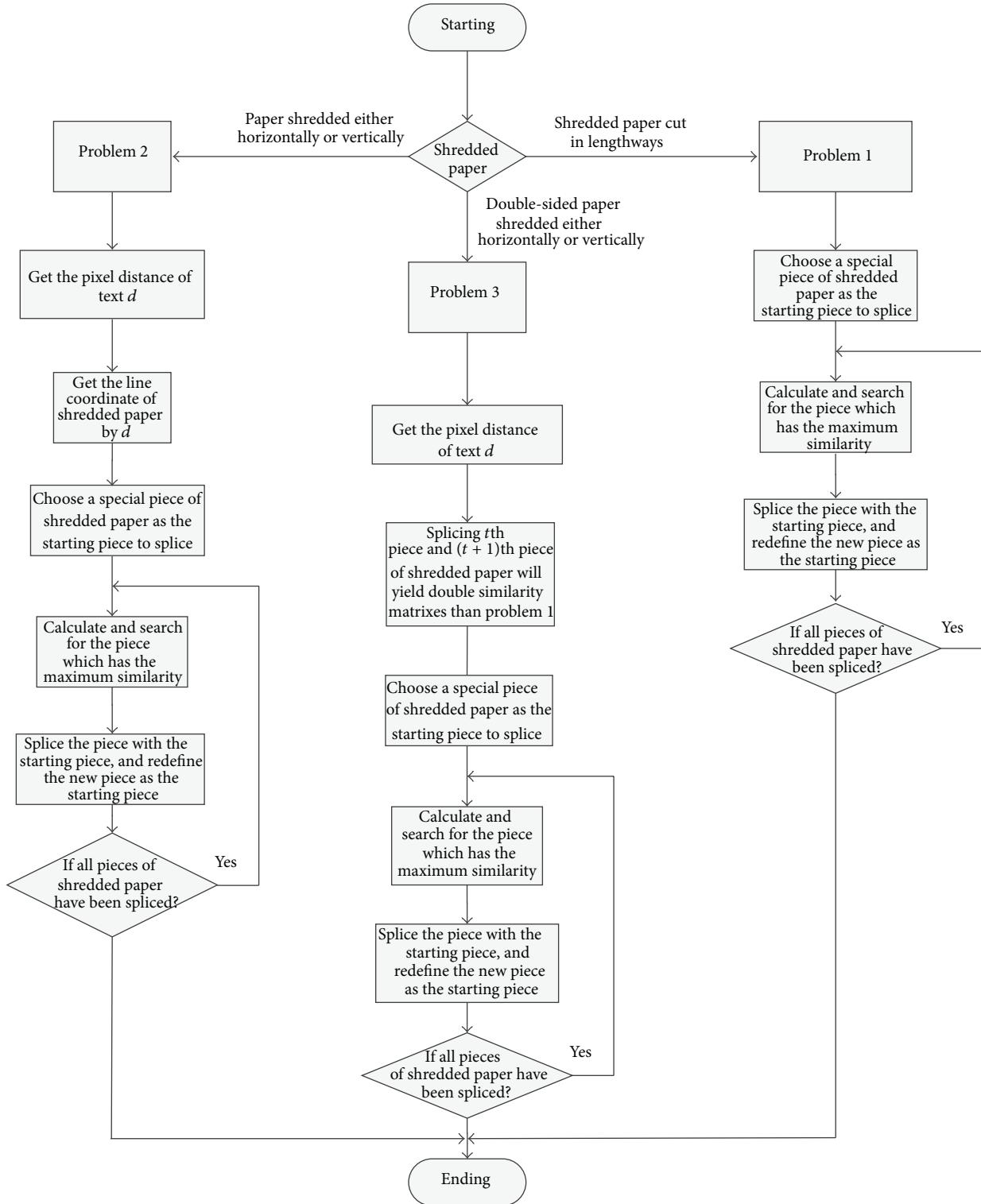


FIGURE 11: The concrete steps of recovery of Chinese shredded paper.

5. Conclusion

In this paper the model and algorithm for reconstruction of shredded paper had been discussed. For the condition in Problem 1, this paper compared the similarity degree

between the rightmost pixels of a certain shredded paper and the leftmost pixels of other pieces of shredded paper. The time complexity of this algorithm was low. Based on this model and algorithm the problem was solved without any manual intervention. For the condition in Problem 2, this



FIGURE 12: Scanning copy of Chinese shredded paper.

便邮。温香熟美。醉慢云鬟垂两耳。多谢春工。不是花红是玉红。一颗樱桃樊素口。不爱黄金，只爱人长久。学画鵝儿犹未就。眉尖已作伤春皱。清泪斑斑，挥断柔肠寸。嗔人问。背灯偷搵拭尽残妆粉。春事阑珊芳草歇。客里风光，又过清明节。小院黄昏人忆别。落红处处闻啼鵙。岁云暮，须早计，要褐裘。故乡归去千里，佳处辄迟留。我醉歌时君和，醉倒须君扶我，惟酒可忘忧。一任刘玄德，相对卧高楼。记取西湖西畔，正暮山好处，空翠烟霏。算诗人相得，如我与君稀。约他年、东还海道，愿谢公、雅志莫相违。西州路，不应回首，为我沾衣。料峭春风吹酒醒。微冷。山头斜照却相迎。回首向来潇洒处。归去。也无风雨也无晴。紫陌寻春去，红尘拂面来。无人不道看花回。惟见石榴新蕊、一枝开。

九十日春都过了，贪忙何处追游。三分春色一分愁。雨翻榆荚阵，风转柳花球。白雪清词出坐间。爱看才器俱健全。异乡风景却依然。团扇只堪题往事，新丝那解系行人。酒阑滋味似残春。

缺月向人舒窈窕，三星当户照绸缪。香生雾縠见纤柔。搔首赋归欤。自觉功名懒更疏。若问使君才与术，何如。占得人间一味愚。海东头，山尽处。自古空槎来去。槎有信，赴秋期。使君行不归。别酒劝君君一醉。清润潘郎，又是何郎婿。记取钗头新利市。莫将分付东邻子。西塞山边白鹭飞。散花洲外片帆微。桃花流水鳜鱼肥。主人瞋小。欲向东风先醉倒。已属君家。且更从容等待他。愿我已无当世望，似君须向古人求。岁寒松柏肯惊秋。

水涵空，山照市。西汉二疏乡里。新白发，旧黄金。故人恩义深。谁道东阳都瘦损，凝然点漆精神。瑤林终自隔风尘。试看披鹤氅，仍是谪仙人。三过平山堂下，半生弹指声中。十年不见老仙翁。壁上龙蛇飞动。暖风不解留花住，片片著人无数。楼上望春归去。芳草迷归路。犀钱玉果。利市平分沾四坐。多谢无功。此事如何到得侬。元宵似是欢游好。何况公庭民讼少。万家游赏上春台，十里神仙迷海岛。

虽抱文章，开口谁亲。且陶陶、乐尽天真。几时归去，作个闲人。对一张琴，一壶酒，一溪云。相如未老。梁苑犹能陪俊少。莫惹闲愁。且折

FIGURE 13: The result of the reconstruction of Chinese shredded paper.

paper sorted the shredded paper according to the location of each piece of shredded paper and satisfactory result was obtained. For the condition in Problem 3, this paper sorted the shredded paper by the optimum matching for the double-sided printed of shredded paper. This model and algorithm could recover the shredded paper with only a little manual intervention. And, given the difference between English text and Chinese text, this paper had proposed the improved algorithm to complete the reconstruction of Chines shredded paper.

At last, it must also be pointed out that the proposed algorithms had their own limitation and bound. Through extensive tests, we ought to admit that in the practical problem, not every piece from various shredders was exactly

identical. And at the same time, the text color, the texture of the paper, the cutting direction (such as documents were shredded askew and promiscuously), or the large stains on the paper, and so on, all of the conditions above would have an impact more or less on repairmen of the shredded document. In some cases, the proper preprocessing would reduce this kind of impact. But in many other much more serious cases, the algorithms introduced in this paper were not performing precisely enough. To improve the general applicability of the algorithm, we had been trying some other methods. As we explore more kinds of shredded paper, we believe that the technology would become more and more mature.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

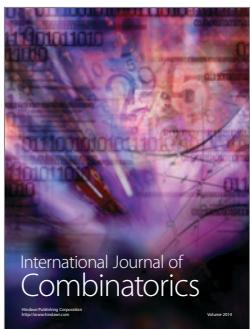
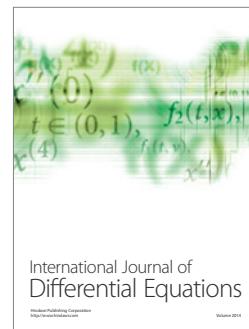
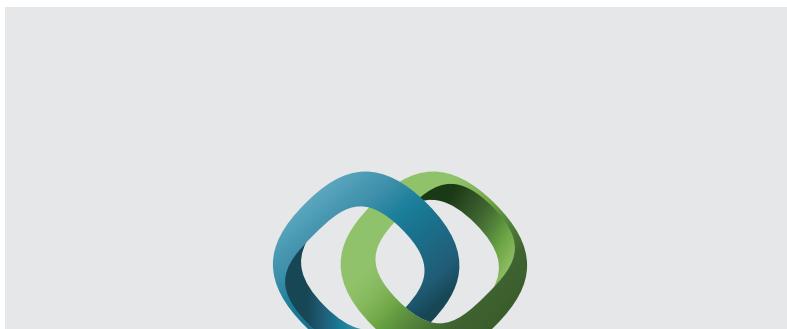
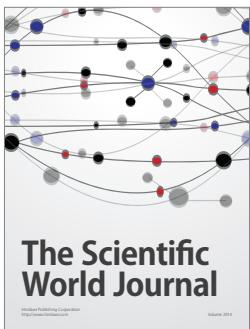
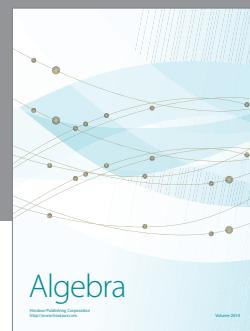
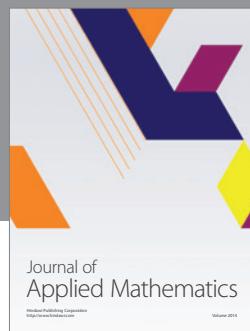
Acknowledgments

The authors gratefully acknowledge the service of WUT Mathematical Computation and Modeling Simulation Center for supplying the relating materials. This work was inspired by the activities of IMP Action of “the Fundamental Research Funds for the Central Universities (2012-IV-057)” and was also supported by the National Natural Science Foundation of China (nos. 10672128, 50878169).

References

- [1] B. T. Ávila and R. D. Lins, “A fast orientation and skew detection algorithm for monochromatic document images,” in *Proceedings of the ACM Symposium on Document Engineering (DocEng ’05)*, pp. 118–126, November 2005.
- [2] C. Schauer, *Reconstructing cross-cut shredded documents by means of evolutionary algorithms [M.S. thesis]*, Institute of Computer Graphics and Algorithms, Vienna University of Technology, 2010.
- [3] P. de Smet, J. de Bock, and E. Corley, “Computer vision techniques for semi-automatic reconstruction of ripped-up documents,” in *Visual Information Processing XII*, vol. 5108 of *Proceedings of SPIE*, pp. 189–197, April 2003.
- [4] M. Werman and D. Weinshall, “Similarity and affine invariant distances between 2D point sets,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 8, pp. 810–814, 1995.
- [5] J. Brassil, “Tracing the source of a shredded document,” in *Proceedings of the 5th International Workshop on Information Hiding (IH ’02)*, pp. 387–399, 2003.
- [6] L. Zhu, Z. Zhou, and H. Dewen, “Globally consistent reconstruction of ripped-up documents,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 1, pp. 1–13, 2008.
- [7] D. Goldberg, C. Malon, and M. Bern, “A global approach to automatic solution of jigsaw puzzles,” *Computational Geometry*, vol. 28, no. 2-3, pp. 165–174, 2004.
- [8] A. Ukovich and G. Ramponi, “Features for the reconstruction of shredded notebook paper,” in *Proceedings of the IEEE International Conference on Image Processing (ICIP ’05)*, pp. 93–96, September 2005.

- [9] P. Faber, "A theoretical framework for relaxation processes in pattern recognition: application to robust nonparametric contour generalization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, pp. 1021–1027, 2003.
- [10] G. Papaioannou, E.-A. Karabassi, and T. Theoharis, "Reconstruction of three-dimensional objects through matching of their parts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 114–124, 2002.



Submit your manuscripts at
<http://www.hindawi.com>

