Reconstruction of Shredded Document using image Mosaicing Technique

Bhavisha Patel
Student of Master Engineering In Computer Department
L.J.Institute of Technology
Ahmadabad 392-001, Gujarat, India

Jay Amin
Asst.Prof. Of Computer Engineering In PG Department
L.J.Institute of Technology
Ahmadabad 392-001, Gujarat, India

Abstract

Reconstruction of sharedded pieces is requisite in archaeology, art conservation, and forensics for Recovery of important information from shredded and damage documents. Documents can be machine-shredded or hand-shredded. Reconstructing the document manually with the help of glue, sellotape etc is a tedious and time consuming process as well as the output is not satisfactory. Hence, here we are illustrating a process to recover the original document from its shredded pieces by using the concept of Image Mosaicing. Image Mosaicing is a process where two or more shredded images are merged together to form a single, continuous image. Here, in this paper we have to reconstruct the shredded fragments and then find Accuracy of two types of torn fragments one is strip shredded and another one is hand shredded fragments. The input of shredded pieces is scanned or capture by camera and various processing is applied to match the two or more shredded piece and in the last stage the shredded pieces are finally stitched and the original document is recovered.

Keywords: - *Hand Torn Documents, Image Mosaicing, Image Processing, Strip Shredded Document*

I. Introduction

Reconstruction of shredded document is similar to solving 2-D card board puzzles game [1]. However, the shape of the hand and machine shredded document is different from that of the 2-D puzzle since its shape of hand shredded is irregular when shape of machine shredded is fixed shape and the content that are on shredded peaces may contain text lines with different directions. Even though shredded document reconstruction process is like to 2D puzzles, Documents can be shredded by two types it can be machine-shredded or hand-shredded documents that is shown in fig.1.

(a) **(b)**

Figure 1.Hand shredded and Machine shredded documents

Documents get damage due to insects, moisture, temperature, humidity, constant handling, and Shredding [2]. Reconstruction of Shredded documents is very important to extract important information which has wide application in forensic sciences, art conservation, archaeology, and corporation [2].

Reconstruction of shredded document is a very tedious and laborious task that should be performed by forensic document examiners. The amount of time that take to reconstruct a document depends on the size and the number of fragments, and it can be measured in days or even weeks. In order to reduce the manual effort of the forensic examiner Automatic software based solution is proposed for reconstruction shredded documents that are major requirement in the field of document reconstruction.

To overcome these tedious and laborious task technique in image processing that are Image Mosaicing is used. Image mosaicing is process of reconstructing continuous seamless image from number of images. But here we take hand shredded image to reconstruct seamless image. Image mosaicing is one of the important steps for automatically reconstructing shredded documents. The main benefit of image mosaicing is that there is no involvement of physical reconstructions of a document (i.e. using glue, adhesives, etc). Mosaicing is free from destructive analysis. Image mosaicing is a methodology which is easy to implement.The main focus of our project is to implement recovery of shredded documents with accurate result.Example of image mosaicing is explain belowed fig.2. Hence, this application plays a major role in archaeological study.

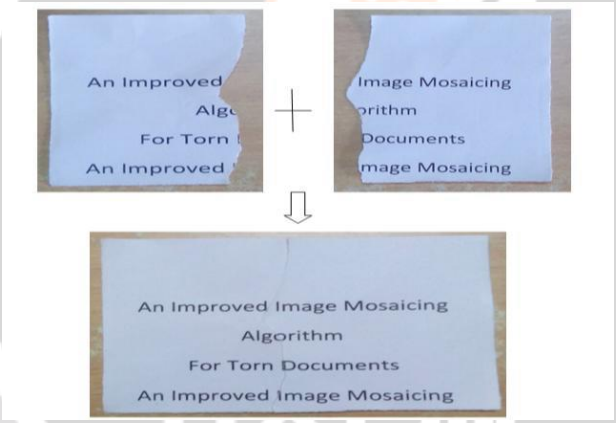


Figure 2. Reconstruction of Original Document using shredded pieces

II. Literature Survey

Mosaicing is the process of reconstructing two or more images or fragments to form a single continuous image. Image mosaicing does not consist of a single specific algorithm. For mosaicing of shredded documents various combination of algorithm or method are used for implementing image reconstructing.

There are various approaches described to implement image mosaicing technique that are used to reconstruct shredded documents. This reconstruct is trial and error base. Carlos Solana In [4] has proposed methodology for reconstruction of hand shredded document based on feature matching. In this paper they apply two steps first one is a polygonal approximation that reduce complexity of the boundaries and second one is extracts features of the polygon to apply the local reconstruction. For using this method overall complexity can be dramatically reduced because fewer features of documents are used to perform the matching. The uncertainty resulting from the local reconstruction is solved and the hand shredded pieces are joining together as we search for a global solution. In this paper time complexity is reduce but the accuracy should not increase.

Andre Pimenta In [8] proposed dynamic programming for reconstruction of hand shredded documents. They used polygonal approximation to compress the complexity of the boundaries and then apply feature extraction. Thereafter, these aspects are used to feed the LCS dynamic programming algorithm. Then feature are used to feed the longest common subsequence (LCS) dynamic programming algorithm. The score get by LCS algorithm is then used to modify prim's algorithm to find the best match among all the pieces. Here accuracy is good as compare to above paper but it take more time in processing.

Wu Youguang In [12] a three step method is proposed, First texture/color fragment information. Second one is contour guided discovery of adjacent image fragments detect the inflection points on the contour and divide the contour in to several segments. Finally match the segment of all adjacent fragments and reassembled the image. The proposed method improved reliability and efficiency of finding adjacent image fragments and in Reconstructed image between fragments still have a gap. In this paper we have to improved accuracy using color of the pixel also.

M.nandhini In [10] proposed Featureless image mosaicing technique. In this technique she proposed Algorithmic methodology in that first ripped up document can scan first and then detect the boundary edge and then plot the boundary value to reconstruct the image. The algorithm can apply to all kind of document irrespective of its content but one disadvantage is of more number of fragments this kind of fragment can't apply.

In exsiting system Yehong Liu In [13] makes an attempt to strengthen a computerized reconstruction system by dividing the matching algorithm into three steps. First, Blank area searching algorithm used to pick up the pieces of shredded part for a given grey level matrix.. Second algorithm is Rightward Education algorithm search for the right hand aspect adjacent document pieces from the primary column pieces until each and every row is reconstructed thoroughly. Revised Education algorithm is used searching in special directions to reconstruct the original document in correct order. Also false searching disorders or sample recognition approach is used for scanning system which is potent but time consuming and the accuracy of join fragment also low.

In this paper we focus on the reconstruction of documents shredded by machine or hand, which is similar to the *automatic assembly of jigsaw puzzle*. A document that contains several important information for further study and analysis purposes. Loss of this information will stop number of important investigation and further studies. Our algorithm effectively matches the two torn fragments and gives accurate results.

The main steps in digital image processing are our proposed work (i) preprocessing, which is a data preparation step for contrast enhancement, noise reduction or filtering, (ii) feature extraction, for retrieving non-redundant and significant information from an image. This operation is targeted at achieving time efficiency at the cost of data reduction followed by object detection, localization and recognition, which determine the position, location and orientation of objects.

In this work we propose a method that can calculate the accuracy and speed of two types of shredded documents (a) strip shredded (Rectangle shape) (b) hand torn (Irregular shape) documents. For strip shredded document use three algorithm and for hand torn we use two methods for reconstruct shredded documents.

The remaining of this work is organized as follows: Section2.1 presents an overview of the proposed methodology; Section 2.2 describes the feature set we have used to carry out the local matching. Section 2.3 shows how we compute the similarity between the polygons as well as the global search algorithm. Finally, Section 3 reports the experimental results and Section 4 present some perspectives of future works and concludes this work.

III. Implementation

In the existing system [13] they implement on strip shredded (machine shredded) documents that are fixed on size and shape. We have to calculate the accuracy and speed of execution of process that how much time takes to reconstruct the shredded piece. Here, we take two different types of shredded documents that are Machine shredded [13] and we proposed hand shredded documents and calculate both type of accuracy and speed of processing time. Here, we explain machine shredded first and then going to proposed system.

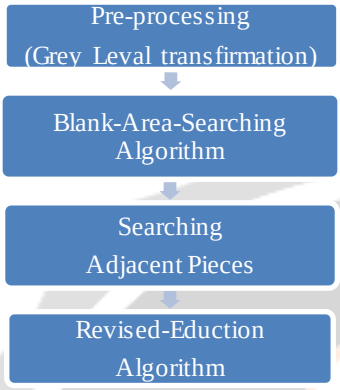


Figure 3. Flow Diagram of base paper

Step 1: Take an input as an image that are shredded into fix size and rectangle shape and perform preprocessing on that. In preprocessing grey level transformation is applied. Each document piece is transformed into a grey level matrix, of which each element's value varies from 0 to 255.

Step 2: This step attempts to pick out pieces in the first column using blank-area-searching algorithm by recognizing the left-hand side blank area of these pieces.

Step 3: The next step is completing each row separately using the rightward-eduction algorithm, which searches the adjacent pieces by selecting the minimum distance between them. Euclidean distance formula is use for minimizing the distance.

Step 3.1: During the rightward searching process, two special situations are taken into considerations. The first one is "endless loop" (when a piece is blank in its right-hand side area, and distance is zero and hence it is consider as adjacent by eduction algorithm and actually they are not adjacent) and Second one is "false searching"(when two shredded pieces both contains minor information on the edge of its grey level matrix at left -hand and right-hand side respectively, perhaps being cut at the edge of a word, the eduction algorithm may consider them adjacent since lacking information makes the distance between them considerably close even when they are actually not).

Step 3.2: Overcome two special cases we perform Fixed-Distance-Based Clustering Method for endless loop and Pattern Recognition Method.

Step 4: Revised-Eduction Algorithm is use for final reconstruction of shredded documents.

III. Proposed system

In the proposed system we have to reconstruct hand shredded document in such a way that it is seen like an original image. We have to reconstruct shredded piece accurately and it take less time to reconstruct. Proposed flow diagram is shown bellowed in fig.4.

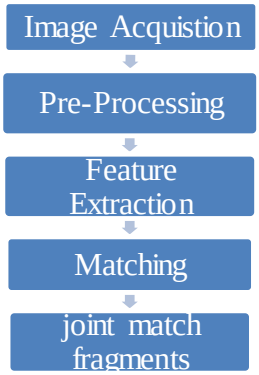


Figure 4. Flow of Proposed Diagram

A. Image Acquisition

This is the first step that includes simple process of taking an image as an input. The input to the system is the shredded fragment that is to be joined or reconstruct to fetching important information. The input images consists of both sides i.e. uniform as well as nonuniform images. The images that are to be taken as input can be captured using camera or can be scanned. Here we have to take input image that are capture by camera that shown in fig.5. There are various formats of images that can be used. They include .jpg, .tif, .bmp, etc but we have to use .png format and then process is performing on that .png image. When it comes to size of images, images having size of images within range of 100kb-400kb are accepted as best. It is not necessary that the images taken satisfy this condition. Hence, the size of images has to be reduced or resized before giving it as input to the system. We are implementing our system in ubntu software. We are implementing in python language using opencv library in ubntu. It is necessary that the images are stored in the current working folder. If not so, correct path to the folder where the images are stored is to be mentioned in the program.

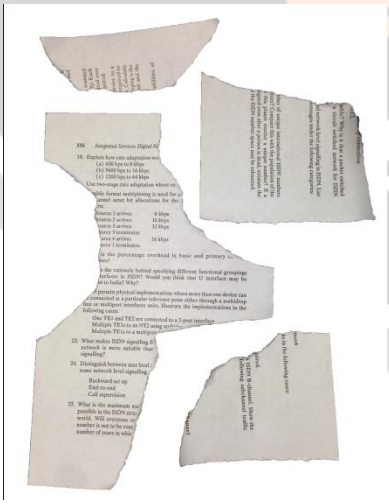


Figure 5.Input image

B. Pre-Processing

The images are given input to the system as specified in the first step and shown in fig.5. These images are not clean because while giving input to the system the images are blurred and noise gets added to it. Hence it becomes necessary to improve the quality of the image and to remove the noise. For remove the noise and improve quality of shredded image we have to use image smoothing and pixel averaging method. Using this we can also improve

contrast of image. In pre-processing step we can use edge detecting the non-uniform edges of the torn fragments which are necessary in matching of the edges for stitching purpose. We are also giving the fragment number to the each piece so we can easily recognize which fragment is match. So, we are use fragmentation for giving number to the system. In preprocessing we are use dilation method to extract the background that is shown in fig.6. For the text extraction we use the dilation process of the image. Text can be extracted by subtracting the dilated image from grayscale image. Dilated images may be like erased image written with pencil on paper.

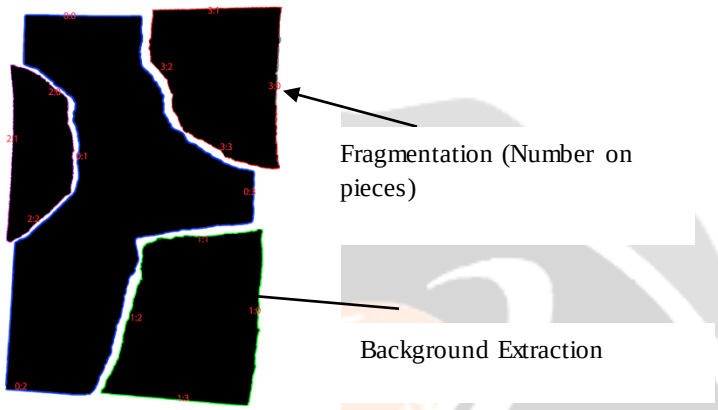


Figure 6. Dilated and fragmentation image

C. Feature Extracting

This stage includes edge detection, boundary extraction and corner detection for further processing of characteristics. This is very important stage in this algorithm which extracts the feature. And these features are used in merging stage to join the fragments. Edge detection is accomplished by using edge detector on each piece separately, which give edge map of individual torn pieces. From this edge map boundary of the individual pieces are extracted and stored in the variables accordingly for further use. Boundary may be oriented clockwise or counter clockwise. In implementations the boundary is traced clockwise. Boundary extraction is nothing but saving co-ordinates of boundary pixels. This extracted boundary is use for corner detection. For corner detection we use angle by three point method that is used for detecting the corner point in each separated points. Here, also use interpolation method describe feature for each shredded fragments and find the connected pixel in the image. With the help of this we can find matching piece of shredded image.

Connected pixel value is found using this formula = $$y = y_0 + (y_1 - y_0)\frac{x - x_0}{x_1 - x_0}$$

Angle and distance is calculated using this formula = $$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Using this two formula we can find the feature point to the shredded piece and three angles among the piece that are describe in fig.7.

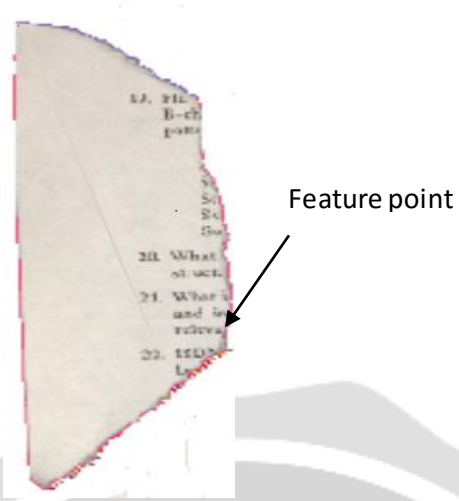
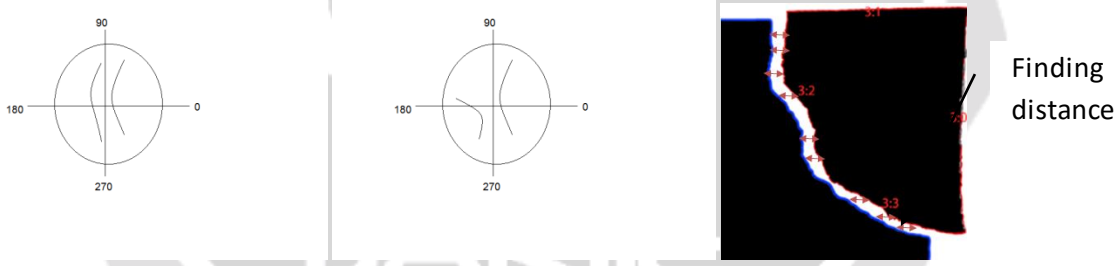


Figure 7. Feature point Extraction

D. Matching

In matching step we have to match the shredded pieces that are finding that two pieces are match with each other. For that we have two types (1) match and (2) finding Best Match. We try to match using rotation we have to rotate the fragment is match than it is consider in match part and it is lies between 0 to 90 degree. Otherwise it goes in best match loop and it may be rotate at 270 degree. With rotation it interpolation is used for distance measurement if distance is minimize than it is match otherwise it search best match. Are describe this using this fig.8.



(a) (b) (c)

Figure 8. (a) Match (b) Best Match (c) distance finding

E. Joint match Fragments

Image Match or mosaic is the most crucial step in recovery of the torn document. After performing all the above steps the matched sides of the torn fragments are joint together to recover the original document. There are cases where in the recovered original document have some gaped because of irregular boundaries. So in our experiment smaller gape is remaining. For reducing or minimizing this gape the researcher can research on that. Here we have to experiment this on three page that are break into three pieces.

F. Experiments Results

Machine Shredded

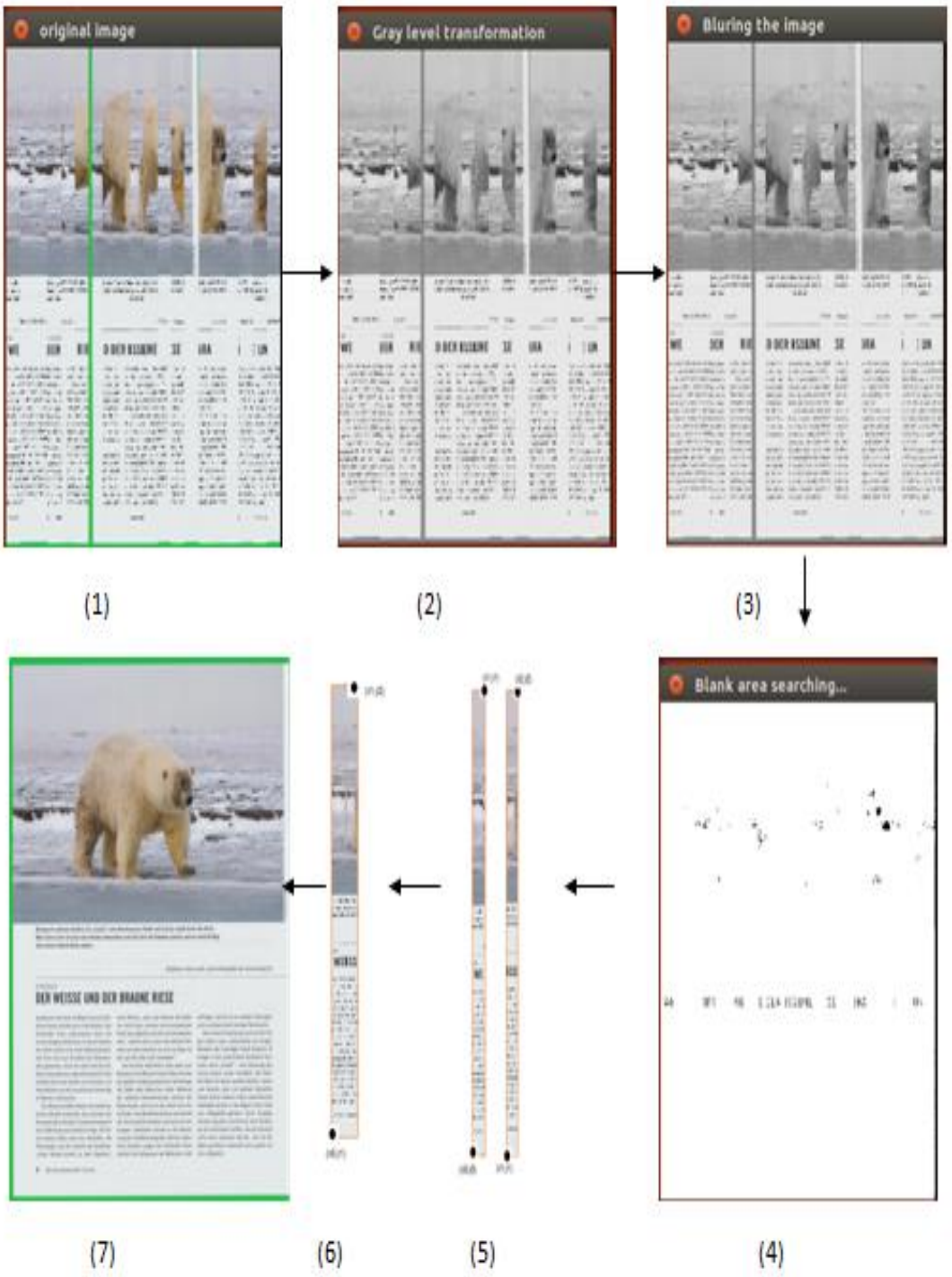


Figure 9. (1) Original image that are shredded as 20 strip (2) Grey level transformation (3) Blurring the image (4) Blank area searching (5) finding adjacent pieces (6) joint the fragment (7) Reconstruct the image.

Hand Shredded

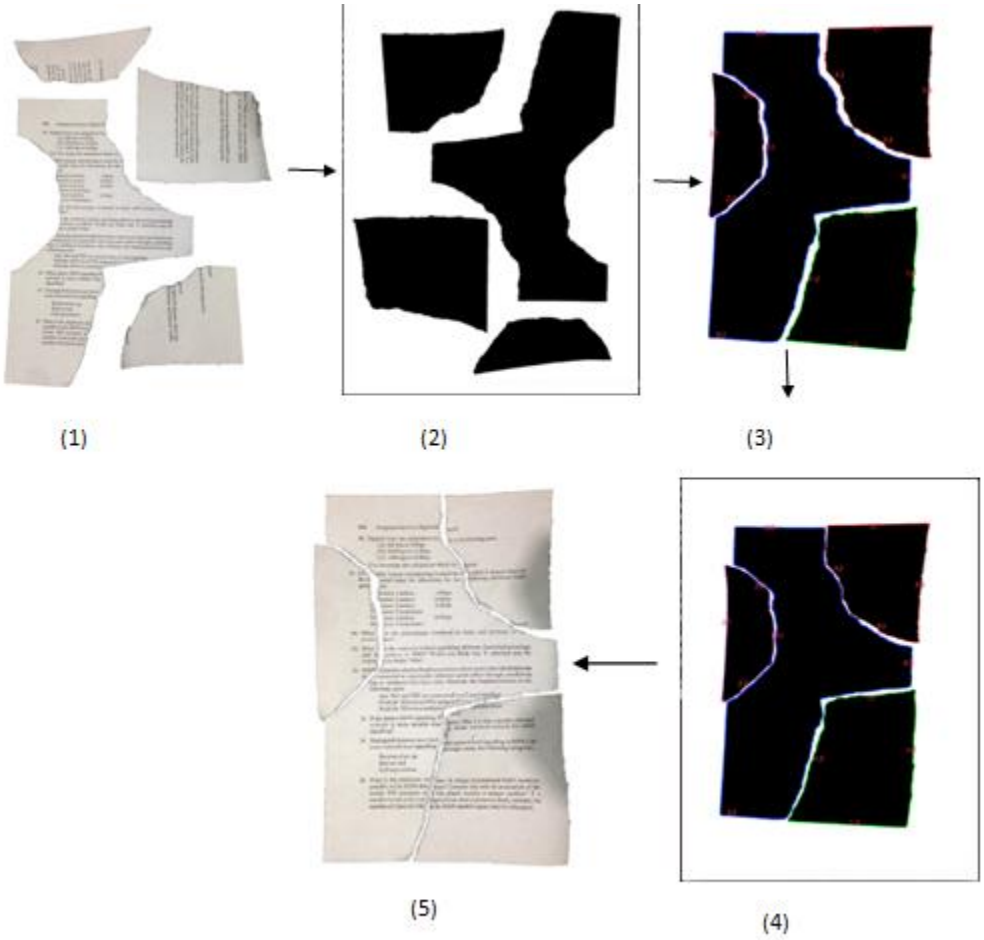


Figure 10. (1) Hand Shredded piece (2) Background Extraction (3) Feature Extraction and find Matching (4) Joint Match Fragments (5) Reconstructed Image

Results

Table 1: Result Analysis Table

Fragment Type	**Accuracy**	**Speed**
Rectangle shape/Fixed Size	35%	15 sec
Hand Shredded	45%	13 sec

IV. Conclusion and Future work

Shredded Document Reconstruction Using image mosaicing Technique is the process of rebuilding torn or ripped-up document images in order to produce one large view of image. Reconstruction of hand shredded document image is very difficult task and also manual reconstruction of shredded document is a time consuming job, digitization makes the job easier. The proposed approach is fulfilled it takes an image as input, process it, and return a complete

reconstruction image, which is give accurate results. Our proposed work gives accurate result to reconstruct shredded documents.

In future work we can reconstruct hand shredded document using color, texture, content for Reconstruct documents etc. We have to find other algorithm to preprocessing and feature extraction that take less time and efficient. So we can accurately and efficiently reconstruct shredded documents.

References

[1]. Florian Kleber, Markus Diem and Robert Sablatnig, "Torn Document Analysis as a Prerequisite for Reconstruction" 15th International Conference on Virtual Systems and Multimedia, IEEE, 2009

[2]. Rintu Abraham and Philomina Simon, "Review on Mosaicing Techniques in Image Processing" International Journal of Software Engineering Research & Practices Vol.3, Issue 1, presented at ACCT2013,April, 2013

[3]. S. Battiato, G. Di Blasi, G. M. Farinella and G.Gallo, "A Survey of Digital Mosaic Techniques", Eurographics Italian Chapter Conference, pp. 129-135, 2006.

[4] Carlos Solana, Edson Justino, Luiz S. Oliveira,and Flavio Bortolozzi, "Document Reconstruction Based on Feature Matching" XVIII Brazilian Symposium on Computer Graphics and Image Processing, IEEE,2005.

[5] T Kasar, A G Ramakrishnan, "Block-Based Feature Detection and Matching for Mosaicing of Camera-Captured Document Images", TENCON 2007 - 2007 IEEE Region 10 Conference, pp: 1 - 4,Nov. 2007.

[6] Tong Lijing, Zhang Yan, Zha Huiqun, "A Warped Document Image Mosaicing Method Based on Registration and TRS Transform", IEEE/ACIS International Conference on Computer and Information Science, pp: 179- 183, May. 2011.

[7] Lijing Tong, Guoliang Zhan, Quanyao peng, Yang Li, Yifan Li "Warped Document Image Mosaicing Method Based on Inflection Point Detection and Registration", IEEE International Conference on Multimedia Information Networking and Security,pp: 179- 183, May. 2012.

[8] Andre Pimenta, Edson Justino, Luiz S. Oliveira,And Robert Sabourin, "Document Reconstruction Using Dynamic Programming", The National Council For Scientific And Technological Development , IEEE 2009.

[9] Aaron Deever, Andrew Gallagher, "Semi-Automatic Assembly of Real Cross-Cut Shredded Documents", IEEE, International Conference on Image Processing (ICIP), pp: 233-236, 2012.

[10] Patrick Butler, Prithwish Chakraborty, Naren Ramakrishan, "The Deshredder: a Visual Analytic Approach to Reconstructing Shredded Document",IEEE Symposium on Visual Analytics Science and Technology (VAST), pp: 113-122, 2012.

[11] M.Nadhini, Rahul Kumar, "Featureless Image Mosaicing Of Torn Document" ,Indian journal of applied research, Volume : 3 , Issue : 6,ISSN - 2249-555X , June 2013.

[12] Wu Youguang, Xin Li, Maoqing Li, "Color and Contour Based Reconstruction of Fragmented Image", IEEE 8th International Conference on Computer Science and Education (ICCSE), pp: 999-1003, 2013.

[13] Yehong Liu, Jiaqi Lu, Honghao Qiu, Yong Fang, "Shredded Document Reconstruction Based on Intelligent Algorithms", IEEE International Conference on Computational Science and Computational Intelligence,pp: 108 - 113, 2014.