International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2014): 5.611

Reconstruction of Shredded Document Using Image Mosaicing Technique-A Survey

Bhavisha Patel[1], Jay Amin[2]

[1]Student of Master Engineering, Computer Engineering, PG Department, L. J. Institute of Technology, Ahmadabad 392-001, Gujarat, India

[2]Assistant Professor, Computer Engineering, PG Department, L. J. Institute of Technology, Ahmadabad 392-001, Gujarat, India

Abstract: ***In Image processing, image mosaicing is one of technique that are tilling or reconstruct shredded images.Reconstruction of hand shredded. Document is challenging task in forensic, investigation science, in corporation, historical artifact reconstruction. Creating of document mosaic images from some shredded piece views is provide valuable information from getting a larger view of a document image than available within a single view and it has been used in many applications. A general framework for shredded document image mosaicing is proposed in this paper. This paper also discusses a review on different applications of image mosaicing and techniques of shredded document image mosaicing. These techniques are helpful to reconstruct shredded documents effectively.***

Keywords: Document image mosaicing, Image processing, Shredded Document

1. Introduction

Reconstruction of shredded or torn document is like a solving 2-D card board puzzles game [1]. However, the shredded shape of the hand torn document is different from that of the 2-D puzzle since its shape is irregular and the content may contain text lines with different directions. Even though torn document reconstruction process is similar to 2D puzzles, it is possible that hand torn or shredded document have different types of layers, it may be outer boundary or inner boundary that is shown in fig.1.

Documents get worse due to insects, moisture, temperature, humidity, constant handling, and shredding [2]. Reconstruction of Shredded or hand torn documents is very important to extract important information which has wide application in forensic sciences, art conservation, archaeology, and corporation [2]. Documents may be ripped up by hand or shredded by a machine. In both cases, the automatic or semiautomatic reconstruction of the original document would alleviate the manual effort, which is difficult and time-consuming. The amount of time necessary to reconstruct a document depends on the size and the number of fragments.



Figure 1: Hand shredded document with inner and outer boundary [1]

One problem faced when reconstructing documents by hand lies in its manipulation. Digitization makes the job easier. Automation of reconstruction through image processing algorithms yields effective solution. Mosaic images can be classified into four types, crystallization mosaic [3], ancient mosaic[3], photo mosaic[3], and puzzle image mosaic[3]. The first two types of mosaics decompose a source image into tiles (with different color, size, content based and rotation); rebuild the image by properly painting the tiles. So they can be grouped together under the denomination of tile mosaic. The last two kinds are obtained by fitting images from a database to cover an assigned source image. Hence they may be grouped together under the denomination of multi-picture mosaics. This classification system should not be intended as a rigid one.

Paper shredders come in many different shapes, sizes and price ranges. There are 3 main categories of shredders and these categories are strip-cut, cross-cut and others (hand shredded or torn).In strip-cut the paper is cut in long vertical strip. In cross-cut the paper is cut in both horizontally and vertically into small rectangles and the other the shape may be torn or grinder type.

Automatic mosaic construction has been applied in many fields such as photogrammetric, computer vision, image processing and computer graphics [2]. We categorize various techniques of document reconstruction as shape based document reconstruction (Apictorial Reconstruction) and document reconstruction using pictorial information(Pictorial Reconstruction) such as texture, color etc. The images can be captured by scanner, camera, mobile and many other sources.

The paper is organized as follows: Section 2 deals with Reconstruction of shredded document framework, Section 3 deals with image mosaicing technique Section 4 deals with various techniques of Shredded or hand torn document reconstruction and concludes the paper.

2. Reconstruction of Shredded Document Framework

A general framework for shredded or hand torn document reconstruction is proposed in this section. The block diagram

Licensed Under Creative Commons Attribution CC BY

for the proposed framework is shown in Fig.2. The algorithm for the framework is given as follows:

1) We have shredded or hand torn document
2) Then performing preprocessing on hand torn document
3) Applying Feature Extraction to extract the feature
Points for matching torn documents
4) Applying methods and match the torn fragments
5) Torn fragments are match it may be partially or totally reconstruct torn or shredded documents.

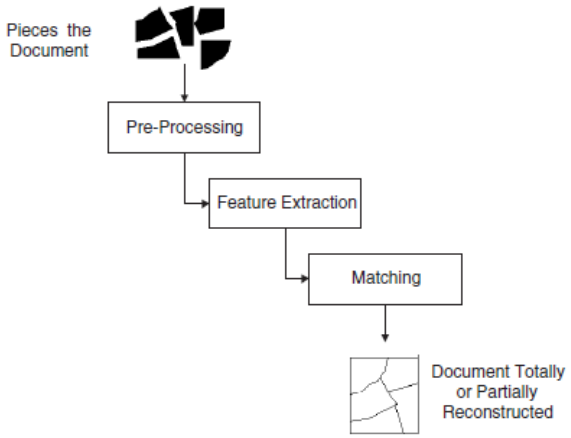


Figure 2: Framework for reconstruction of shredded document [4]

3. Document Image Mosaicing Techniques

Document image analysis requires mosaicing when shredded document are there or such document that are not caputer in large image document. Such piece is scanned and again rebuilt same as original piece of document or in large view. This motivates the use of "mosaicing", whereby piece of shredded document or small size pieces of document are stitched together.

In [5] a novel block-based layout is employed to make sure that corners can be reliably detected over a wide range of images. 2-D separate cosine change is calculated for image blocks defined around each of the detected corners and a small subset of the coefficients is used as a feature vector. A 2-pass feature matching is performed to establish point back and forth writings from which the homography relating the input images could be calculated. The Harris corners have been extremely carefully selected so that we have the features fairly evenly spread throughout the image.

In[6] presented a warped document image mosaicing method based on registration and a composing change of Translation, Rotation and Scaling, called TRS change. This method mosaics two sick images of the same document from different viewpoints through feature extraction, image registration, etc. After mosaicing, the distortions are mostly removed and the OCR results are improved very much.

In[7] document image mosaicing method based on the inflection point detection and registration, Firstly, document image is gray, binaries and expanded. Than document images text line extracted, curve fitting and inflection point are determined, registration is performed. Finally image registration based on the template matching, TRS transform, gray level interpolation and image mosaicing are done. Using this method OCR recognition rate of the mosaiced images is higher than original image.

4. Technique For Shredded Document Reconstruction

Florian Kleber In [1] has proposed an automated assembly of torn documents. A precalculation of document snippets is described for clustering of the provided data. The main orientation is determined according to the printed or handwritten text information. Thus a combined shape and pictorial approach is used. To minimize the search space for matching the following calculations are performed on each snippet: (a) a rotational analysis to determine the alignment (b) The color of the ink/paper is distinguished. One advantage is that the characteristics of snippets are calculated using combined shape and pictorial approach such that the tearing paper problem is solved and disadvantage is that the algorithm does not work if the torn pieces are from multiple documents.

Carlos Solana In [4] has proposed methodology for reconstruction of methodology based on feature matching. In this it take two steps first step applies a polygonal approximation in order to reduce the complexity of the boundaries and then extracts relevant features of the polygon to carry out the local reconstruction. In this way the overall complexity can be dramatically reduced because few features are used to perform the matching. The ambiguities resulting from the local reconstruction are resolved and the pieces are merged together as we search for a global solution.

Andre Pimenta In [7] has proposed the use of dynamic programming for document reconstruction. Firstly, polygonal approximation is used to reduce the complexity of the boundaries and extract features from them. Thereafter, these features are used to feed the LCS dynamic programming algorithm. The scores yielded by the LCS algorithm are then used into a modified Prim's algorithm to find the best match among all pieces. Comprehensive experiments on a database composed of 100 shredded documents support the efficiency of the proposed methodology. When compared to global search algorithms, this approach brings an improvement of 18% in the number of fragments reconstructed. Advantage is that Polygonal approximation to reduce the complexity of the boundaries and overcome specific problems faced in document reconstruction and disadvantage is the other kind of features such as texture and color can be added.

Aaron Deever In [8] a unique three step approach is follows, First preprocessing steps to extract and orient each fragment of shred image in which connected components algorithm is used to extract connected groups of shred pixels. Second perform feature extraction and likely matching fragment of each shredded document main idea is to capture location of markings and lines on the fragments. Finally human interface needed to evaluate the correct matches. Advantage of this approach is that fragment contains only information is considered to reassemble that directly effect to efficiency

Licensed Under Creative Commons Attribution CC BY

and Correct matches are proposed with correct offset than accuracy of match is increased. Disadvantage of this approach is that Shredded document have not been completely rebuild and Efficiency of human-computer step is dependent on the accuracy of automatic ranking of potential match for each piece based on the computed features.

Patrick Butler In [9] the Deshredder represents shredded piece as time series and use nearest neighbor matching techniques that enable matching both the outline of shredded piece as well as the content of shreds themselves. Each shred into single image and apply, image straightening algorithm to guarantee that each shred has a known orientation. Next texture matching technique is applied to find the best match shreds. Than after human interface to identify the best matches for a given shredded document and seeing similarity to see overall progress and to help weed out unhelpful shreds. Next it takes the verified matches and orientation information to eliminate potential matches. This step is repeated until the original document is rebuilt. This approach is focus on the vertical edges only, for more efficient matching rotationally invariant algorithm is developed.

M.nandhini In[10] proposed Featureless image mosaicing technique, Initially the proposed algorithm starts with the inputting of image in which the fragments are scanned with the help of a scanner and taken as input to the algorithm. The algorithm detects the boundary of each fragment, marks them explicitly. After that boundary values are plotted with the help of a graph. The fragment contains both uniform and non-uniform boundaries. As only the non-uniform boundary values are considered, it must be separated from the uniform boundary values. Next the non-uniform boundary values have to be recursively matched. Once the boundary values are matched, the fragments are translated and the reconstructed document is displayed. If the boundary values are not matched, then the matching process is continued. The major advantage of our work is that, the algorithm can be applied on all kinds of document irrespective of its content. The present work has been carried out only on two fragments. It can be extended further on more than two fragments so that it can mosaic heap of documents one by one.

Fabian Richter In [11] a novel two step framework is created, first use a SVM classifier to identify location on fragment pairs, shape and content based local features. In order to restore the layout of the document, create a document graph in which nodes represent pieces and edges go along with alignments and assign weight to the edges. Second align iteratively fragment groups that choose the edge

having highest weight during iteration until document is rebuilt. Main advantage of this approach is to Reduction of the number of potential fragment matches while preserving a sufficient number of correct matches and In case of some fragment is missing than also reconstruct the document. Also in Rebuilt image between the fragments still have a space.

Wu Youguang In [12] a three step method is developed, First texture/color fragment information. Second 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. Also result shows that the proposed method improved reliability and efficiency of finding adjacent image fragments and In Reconstructed image between fragments still have a gap.

Yehong Liu In[13] a three algorithm is used first, blank-area-searching algorithm is used to pick up the shredded fragment for a gray level matrix this algorithm check column by column, only pieces which have two columns in blank area in left of gray level matrix is reserved while others are discarded. Second, rightward-eduction algorithm is used to identify adjacent document fragment by selecting the minimum distance between them. Finally revising the eduction algorithm into upward and downward directions to frame original document in right order. Also False searching issue, the pattern recognition method is effective but time consuming.

Bhattarabhorn Wattanacheep In [14] Here, plane alignment algorithm is used for torn document reconstruction. The proposed technique analyzes the contents inside the snippet such as the direction of the character alignment based on the histogram of the accumulated radius of the fitted ellipses. The direction result is then used to revert the snippet to its original position. Hough transform based local descriptor is extracted as shape feature. These parameters are helpful for accurate reconstruction. This technique can also achieve approximately 5.07 decrease in relative orientation error thus increase 24.11 percent in reverting precision. This can demonstrate the significant performance improvement of the proposed algorithm. Estimating geometric shape representation in term of Hough transform descriptor is analyzed and used later in the reconstruction process. Performing plane alignment prior to reconstruction can help to accelerate the reconstruction with higher success.

5. Conclusion

Image mosaicing is a powerful tool for generating larger view of a scene. Various image mosaicing techniques related to document images are discussed. A general framework for image mosaicing in the document images is proposed in the paper. A review on document image mosaicing in the field of image processing is also discussed. Document image mosaicing helps to capture a large document at a high resolution in single exposure.

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

Licensed Under Creative Commons Attribution CC BY

[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] Fabian Richter, Christian X. Ries, Nicolas Cebron, Rainer Lienhart,"Learning to Reassemble Shredded Documents", IEEE Transactions on multimedia, Vol. 15, No. 3, pp: 582-593, 2013.
[13] 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.
[14] 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.
[15] Bhattarabhorn Wattanacheep, Assoc. Prof. Dr. Orachat Chitsobhuk, "Plane alignment algorithm for torn document reconstruction",IEEE, 978-1-4799-7961,2015

Licensed Under Creative Commons Attribution CC BY