



RECONSTRUCTION OF TORN DOCUMENTS USING SUPPORT VECTOR MACHINE (SVM) METHOD

Jemika Gabani¹, Divyang Patel²

¹Electronics and Communication Department, SVBIT

²Electronics and Communication Department, SVBIT

Abstract — The main task of my thesis is to reconstruct the torn pieces of any paper. This paper can be a document or an image. Documents could be shredded for any reason such as, one may do it intentionally and it could happen by mistake also but the main disadvantage is that, the information will be lost which are for certain cases extremely important. Reconstruction of a torn document can be very beneficial for people from various fields, such as in defence, investigative science, archaeology and forensics fields.

Torn documents can be hand shredded or manually shredded. One can shred it vertically or horizontally. The task will be solved by using Support Vector Machine (SVM) method. Torn pieces are combined by comparing edge length, angles, page margin and by matching the colors of images or text documents. Therefore, my task is divided in three stages: read the image, find edges and fragmentation of the edges. In addition to that, SVM is a newly invented method which is also suitable for the documents which have high resolution.

Keywords- Torn pieces or documents; Image processing; Support Vector Machine (SVM) method; Edge detection; Using in forensic science; Do core recovery algorithm; Documents reconstruction; Reconstruction techniques; Document image processing

I. INTRODUCTION

Reconstruction of torn documents is the process of creating a full position of scan documents. Many times in short intervals too large to capture all the way through a single scan or in a single frame, by attaching stitches assemble many small fragments. The quality of being able to perform do away with hard (an alternative papers) copies of documents has become particularly relevant in the past a couple of years, with become greater in identity fraud, and large scandals (private lives of other people) such as the accountancy (financial reports for a business) case. Paper shredders can be easily obtained from any stationary store, and many people use them to destroy their important documents, as do most corporations.

A state of difficulty that needs to be resolved of shredded document recovery can be majorly time overwhelming or even impossible, fundamentally piecing together an extended to edge detection, with the added trouble of all pieces being of identical shape and size and possibly double-sided. There is thus self-evident motivation to automate the process as much as possible, with the potential to increase both accuracy and speed of reconstruction. In spite of the fact there is currently very little evidence in the literature of academic study into the process.

There are mainly two types of paper shredded documents or images: (1) Machine cut documents (2) Hand or manually shredded documents. Show the figures of torn documents. Figure 1 is demonstrated machine cut documents and the surface is sharp cut. On the other hand, figure 2 is display the hand shredded or manually cut documents and the documents surface is not sharp. There is main difference between shredded documents technique.



Figure 1. Machine shredded image



Figure 2. Manually or hand shredded document (unshredded)

There are creating so many problems while reconstructing the documents or any images. For instance compare the page margin, distance between paragraphing, edge detection, blur output of image or document, fix camera position, limits of torn pieces, so on. On the contrary, overcome this whole problem by using support vector machine method (SVM) with the help of algorithms. Also use the PYTHON language for the shredded and reconstruction operations.

II. GENERAL FRAMEWORK OF RECONSTRUCTION OF TORN DOCUMENTS

A general framework for reconstruction of torn documents is proposed in this section. The block diagram for the proposed framework is shown in figure 3. The algorithm for the framework is given as follows:

Firstly, choose the images of torn or ripped-up document. Secondly, arrange them in correct order and orientation by using the Features extracted, SVM (support vector machine) etc. Thirdly, build the reconstruction image of document. Finally, the high resolution reconstruct document image is received.

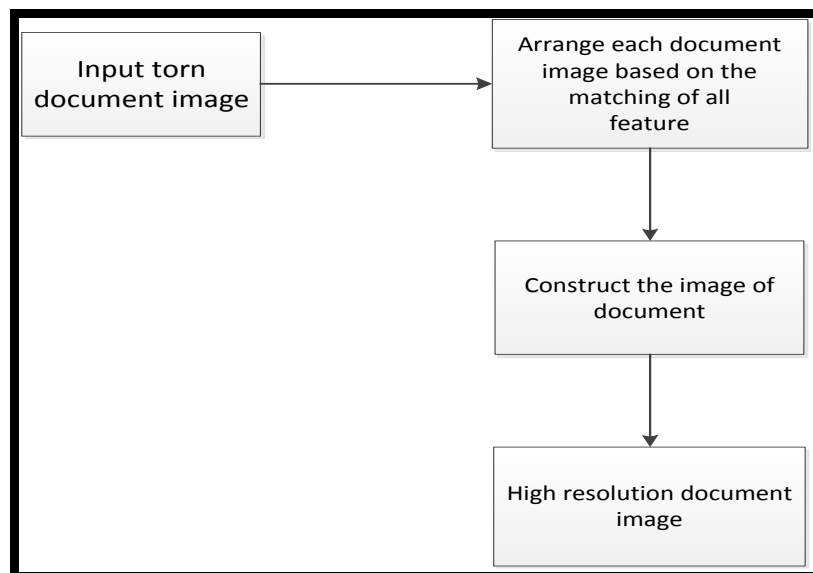


Figure 3. General framework of document image reconstruction

Show the steps for reconstruction image or documents for machine cut or shredded documents. The basic steps are taken the image, read the image, open in normal window, write the image in gray level, shred the image for various shredded values and generate the image in three steps: step 1: gray level transformation; step 2: blurring for gray area searching; step 3: blank area searching, up to the end of this all operation reconstruction image or document generate. Now show the figures for (shredded) image reconstruction.



Figure 4. Take the image



Figure 5. Read the image



Figure 6. Read the image in normal window



Figure 7. Shred image



Figure 8. Reconstruction image

III. EXPERIMENTAL RESULTS

As per the flow graph, the put up the value of torn pieces in the operation and observe the torn pieces size as well as value of pixels after image reconstruct. Shown in table 1, using the value of shredded documents are 5, 10, 15, 20, 25, and 30. The original image pixels are width with height.

Sr. No.	Torn Pieces	Original image size (Pixels)	Shredded pieces (For each piece of pixels)	Reconstruction image size (Pixels)
1.	5	1920*1080	384	No Change
2.	10	1920*1080	192	No Change
3.	15	1920*1080	128	No Change
4.	20	1920*1080	96	No Change
5.	25	1920*1080	76	No Change
6.	30	1920*1080	64	No Change

Table 1. Experimental results

IV. CONCLUSION

Document or image reconstruction is the process of rebuilding torn or ripped-up document images in order to produce one bright and sharp image. Reconstruction of shredded document image is very difficult task. The majority of proposed approach is fulfilled it takes an image as input, process it, and return a complete reconstruction image.

While the approach has demonstrated some good results on several image. It can clearly be seen that, when image quality is in high pixels, the image fragmentation is shown in colored, and the torn pieces are re-built in fast time, image edge is correct in operation. Up to the end of the operation the torn pieces pixels value is same as the original image. There is no change in the reconstruction image or documents.

V. FUTURE WORK

Possible area to examine in the future might include further consideration of the effect of hand shredded documents or images. The various torn pieces will join to torn pieces for key point matching and character recognition are main task for next performance. Using more torn pieces for re-built (shredded or unshredded). Improve the image or documents quality for show and read the reconstruction outputs. Also calculate the image processing parameters for find the reconstruction ratio and so on.

REFERENCES

- [1] DHARUVEN PRAJAPATI, KRUTI J. DANGARWALA, "Various Document Image Mosaicing Method In Image Processing : A Survey", *IEEE conference on signal processing and communication engineering systems (SPACES) 2015* (12 March 2015) 281-285
- [2] BHATTARABHORN WATTANACHEEP, "Plane Alignment Algorithm For Torn Document Reconstruction", *IEEE conference on electrical engineering/electronics ,computer, telecommunications and information technology (ECTI-CON) 2015* (20 August 2015)
- [3] YEHONG LIU, JIAQI LU, YONG FANG, HONGHAO QIU, "Shredded Document Reconstruction Based on Intelligent Algorithms", *IEEE conference on Computational Science and Computational Intelligence (CSCI) 2014* (29 May 2014) 108-113
- [4] V.C KIEU, NICHOLAS JOURNET, MURIEL VISANI, RE'MY MULLOT AND JEAN PHILLIPE DOMENGER, "Semi-synthetic Document Image Generation Using Texture Mapping on Scanned 3D Document Shapes", *IEEE Conference on Document Analysis and Recognition (ICDAR) 2013* (15 October 2013) 489-493
- [5] ANKUSH ROY, UTPAL GARAIN, "A Probabilistic Model for Reconstruction of Torn Forensic Documents", *IEEE Conference on Document Analysis and Recognition (ICDAR) 2013* (15 October 2013) 494-498
- [6] POOJA SHARMA, SHANU SHARMA, "Image Processing based Degraded Camera Captured Document Enhancement for Improved OCR Accuracy", *IEEE conference on cloud system and big data engineering (confluence) 2016* (11 July 2016) 441-444