RESEARCH PAPER | Computer Science | Volume : 3 | Issue : 6 | June 2013 | ISSN - 2249-555X

Featureless Image Mosaicing Of Torn Document

KEYWORDS: cropping, image processing, mosaicing, torn documents, boundary detection

M.Nandhini
Department of Computer Science
Pondicherry University, Puducherry-14

Rahul Kumar
Department of Computer Science
Pondicherry University, Puducherry-14

ABSTRACT *A forensic expert might need to assemble the heap of torn documents to get back the original one. Performing it manually is time consuming and complex. In order to alleviate the manual effort of assembling torn documents that may be ripped up by hand, mosaicing of torn document image is done.*

1. Introduction

The recovery of ripped up documents is often a problem that arises in archival study and investigation sciences such as forensic and crime detection. A forensic expert might need to assemble the heap of torn documents to get back the original one.

A document is a bonded physical or digital representation of information designed with the capacity to communicate. Documents get deteriorated due to insects, moisture, temperature, humidity, constant handling, obliteration and shredding. So reconstruction of torn documents is essential to import information which has wide application in forensic sciences, art conservation, and archaeology. Doing it manually is time consuming job, needs hard work of experienced personals. Digitization makes the job easier.

Algorithmic methodology is applied to reconstruct the torn documents in which the principle technique used is the featureless image mosaicing technique which includes the two scanned fragments of the same documents such that the non-uniform side faces each other. This is for bringing Automation in reconstruction through image processing algorithms which yields effective solution and result says that this approach of mosaicing has proven to be better and can be done with all documents apart of its contents.

2. Related works

Bernardo Esteves Pires and Pedro M.Q.Aguiar [2005] demonstrated the featureless global alignment of multiple images to overcome the propagation errors particularly when dealing with videos that show the same region at disjoint time intervals. Through this approach, the input images are modeled as noisy observations of limited regions of the unknown panorama where Maximum Likelihood estimation is used.

Simon T.Y. Suen, Edmund Y. Lam, and Kenneth K.Y. Wong [2006] had proposed Image Mosaicing with Optimized Matching of Global and Local Contents. The Curvature-domain image stitching approach reduces the artifacts caused by photometric inconsistency and geometric misalignment problems. It would be computationally heavy if the whole mosaic region uses this method. A new concept has been proposed for separating the regions into two sub-regions, global and local, while getting their optimized matching with different energy functions.

Nad'ege Rebi'ere, Marie-Flavie Auclair-Fortier[2008] addressed the image mosaicing using Local Optical Flow Registration in which optical flow is used in order to find a pixel wise registration in the overlapping region, which reduces ghost problems that occurred due to local distortions.

Kantilal P.Rane and S.G.Bhirud [2011] had proposed image mosaicing with Strip Search Algorithm based on a novel similarity measure. It improves non-linearity, accuracy and vertical distortions possibly found in mosaic image.

Ramesh Babu, M.Ravishankar [2012] brought the concept of Automatic Seamless Image Mosaicing through an approach based on Quad-tree Technique. In real life, it might not be possible to capture a large image with a given source of imaging device in a single exposure. It has to be captured as two or more splitted images due to the inherent limitations of the capturing media. A new approach was used to create a single large image by adopting a quad-tree approach, where the image is decomposed into sub images to find the corresponding points in the sub images and thereby reducing the search space.

Image Mosaicing is a fundamental area of work where efforts are being put off to get the mosaiced image more accurate, efficient and distortion free [4]. The ghost problem arriving because of overlapping has been sorted out during pixel wise registrations of the image to make the mosaiced image reflecting the original one [5].

In the existing system, the mosaicing of images is done provided that there will be overlapping regions between the fragments. Based upon contrast, color, the images are finally stitched together. Beside these, loss in the contents of the image also occur.

3. Proposed Methodology

The featureless image mosaicing technique which includes the two scanned fragments of the same torn documents with the hard constraint that non-uniform sides must be facing each other is proposed.

A featureless approach for mosaicing the documents is much easier to apply without losing its contents in which there is no need of looking at the color, contrast and brightness. It is valid for all kinds of images apart of its contents, whether it is torn pieces of map or a document or pieces of anything.

In the existing system, the images are stitched together provided that there will be overlapping region between the two fragments, but in this case the images will be stitched together based upon matching their boundaries (curve matching) of the non-uniform faces of the torn documents [7]. In this algorithm, first important step constitutes building a digital database of the given set of real-world paper fragments.

Volume : 3 | Issue : 6 | June 2013 | ISSN - 2249-555X

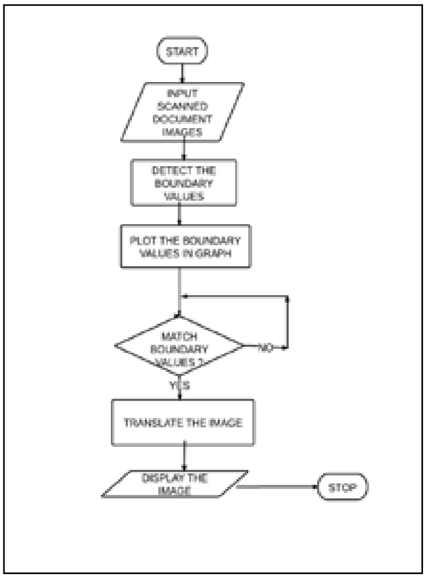

Fig 1 Flow of Proposed Methodology

4. Implementation

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.

A) Inputting of image:

- Select the subset of the fragments and put them on top of a flat bed scanner.
- Put a uniformly colored piece of paper on top of the fragments and a digital image scan is made, which will be input image.

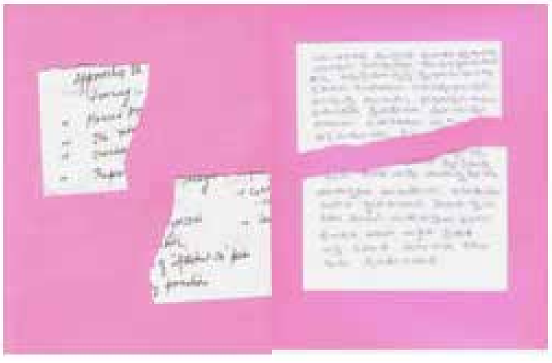

Fig 2 Input Images

B) Gray Image Conversion

The scanned image will be converted into gray image:-

- Get Red, Green and Blue values of pixel
- Gray=(red+green+blue)/3
- Replace the original red, green, and blue values with the gray Value

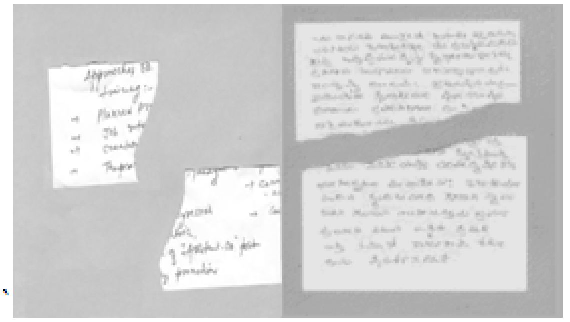

Fig 3 Gray Image Conversion

C) Convert to Binary Image:

The gray image is then converted into binary image as it is easier to work with binary image.

- Compute a global threshold value in the range of 0-255
- If pixels value greater than the threshold value then pixel value is set to 1
- If pixels value less than the threshold value then pixel value is set to 0

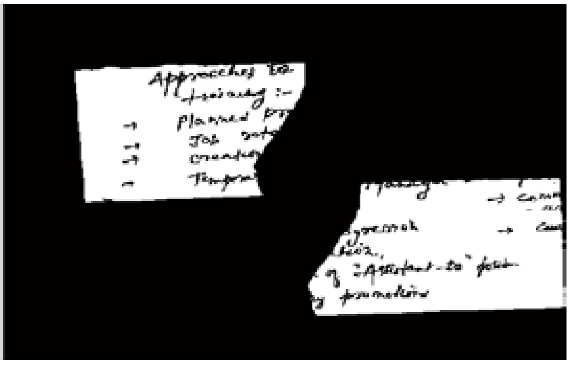

Fig 4 Binary Image

D) Apply Bounding Box

Working area has to be selected from a binary image within the bounding box. So a bounding box is applied using following steps:

- Scan the image width row by row and find the first row and last row pixels values where white pixel is encountered mark them as first row and last row
- Likewise, scan the image height column by column and mark them as first column and last column
- Draw a rectangular box covering the point of intersection of first row, first column, first row, second column, second row, second column, second row, first column

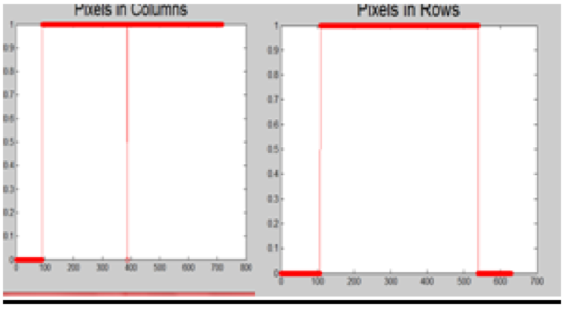

Fig 5 Pixels in columns and rows

Volume : 3 | Issue : 6 | June 2013 | ISSN - 2249-555X

E) Crop the Image:

The working area i.e. bounding box region is cropped using following phases:

- Image width is set to row1 to row2
- Image Height is set to column1 to column2

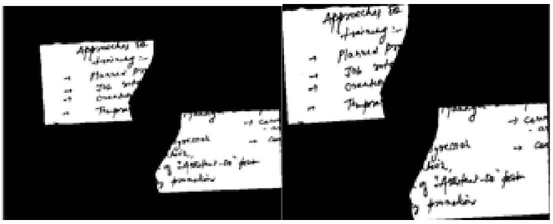

Fig 6 Cropped Image

F) Boundary Detection

Boundary detection of the fragment parts of the cropped image is done using canny edge detection method which runs in 5 steps:

- Blurring of the image to remove noise
- The edges should be marked where the gradients of the image has large magnitudes
- Only local maxima should be marked as edges
- Potential edges are determined by thresholding
- Final edges are determined by suppressing all edges that are not connected to a very certain (strong) edge

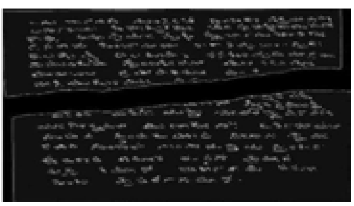

Fig 8 Boundary Detection

G) Plotting Points:

Two distinct points are manually plotted on each fragment with the help of a mouse. Select n points from the current axes and returns the x- and y-coordinates in the column vectors x and y, respectively).

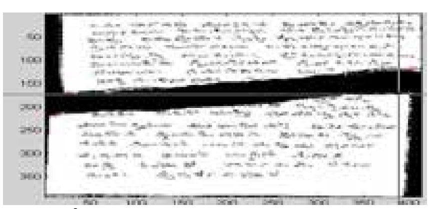

Fig 9 Plot Points

X and Y- Values of the selected points will be

X-Values	252	378	238	292	366
Y-Values	132	196	228	47	221

H) Translation of Fragments:-

- Plotted fragments are translated for reconstruction based on the reference points
- Effectively resize, examine image content and remove "less important" regions
- The pixel R (u, v) will be derived from the input pixel I (u + tx, v + ty)
- Each output pixel can be labeled by a shift (tx, ty). When an output pixel (u, v) should originate from location (x, y) in the input image, the appropriate shift gets zero energy
- Object rearrangement is specified such that one part forces pixels to appear in a new location using. The second part marks these pixels for removal from their original location using to avoid their duplication in the output
- 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

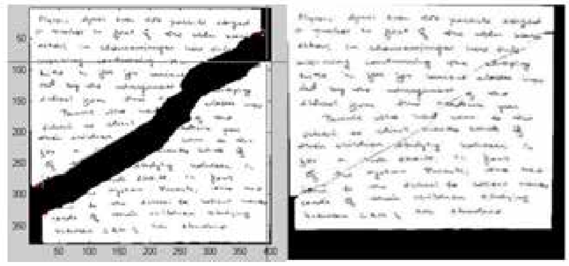

Fig 10 Mosaiced Image

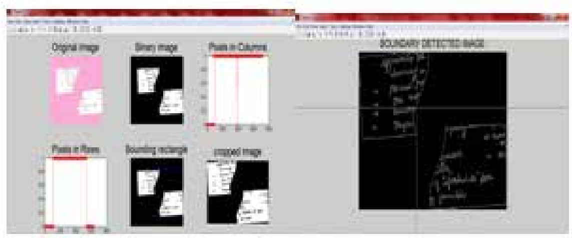

Fig 11 Sample result

While scanning the fragments, even though fragments were placed with utmost care on the scanner bed, skew problem occurred and it was highly impossible to remove this skew manually. Removal of skew from a scanned document is still in research stage. Hence continuing with this procedure to build a mosaicing algorithm was quite tedious. So a new design is taken up where the input image was first converted to a binary image. A bounding box was then applied only to the region of interest. Thus selected region was then cropped, eliminating the unwanted values or pixels from the source image thus identifying the boundaries of the fragments. Further, two points from each fragment were selected along the edges of the non-uniform sides which acted as the reference points for the mosaicing of the two fragments. The final stage was the stitching of the two fragments based on the values of the reference points. The re-constructed document image was displayed for visual verification. The major advantage of this work is that, the algorithm can be applied on all kinds of document irrespective of its content.

5. Conclusion

In the present work, the methodology has been corroborated to reconstruct torn fragments by considering two fragments of the same ripped up document. The working principle used here is the featureless image mosaicing technique. 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.

Volume : 3 | Issue : 6 | June 2013 | ISSN - 2249-555X

REFERENCE [1] Bernardo Esteves Piresa & Pedro M.Q.Aguiar "Featureless Global Alignment of Multiple Images" In ICIP vol.1, pp:57-60, 2005. | [2] Ramesh Babu & M. Ravishankar, Member, "Automatic Seamless Image Mosaicing: An Approach Based on Quad-tree Technique", Proceedings of the World Congress on Engineering, Vol. 1, 2010. | | [3] Kantilal P. Rane & S.G.Bhirud "Image Mosaicking with Strip Search Algorithm based on a Novel Similarity Measure" International Journal of Computer Application,vol.27, no.2, pp.42-47, 2011. | | [4] NadègeRebière, Marie FlavieAuclair-Fortier, François Deschênes: Image mosaicing using local optical flow registration, ICPR, pp.1-5,2008. | | [5] Nakao, A. Kashitani & A. Kaneyoshi "Scanning a Document with a Small Camera Attached to a Mouse" In Proc.WACV'98, pp. 63–68, 1998. | | [6] Nagaraj B. Patil, V. M Viswanatha, & Sanjay Pande M.B "Mosaicing of Torn Documents Image", IJRRAS, Vol.7, Issue 1, April-2011. | | | [7] Simon T.Y. Suen, Edmund Y. Lam, & Kenneth K.Y. Wong, "Digital Photograph Stitching with Optimized Matching of Gradient and Curvature," in Proceedings of the SPIE, vol. 6069, pp.139–154, January 2006. |