International Journal of Advanced Technology in Engineering and Science
Vol. No.3, Special Issue No. 01, September 2015
www.ijates.com
ijates
ISSN 2348 - 7550

Mosaicing of Text Contents in Real Time for Microcontroller Based Text Reading System

Vimuktha Evangeleen Jathanna[1], Nagabhushan P[2]

[1] Research Scholar, [2]Professor, Department of Studies in Computer Science, University of Mysore, (India)

ABSTRACT

Text Reading Systems are often used as an assistive reading tool for the visually impaired. Such tools utilize image processing algorithms to segment extract and recognize text from images and videos before reading. Due to which there is a scope for light and efficient algorithms that can easily execute text reading with the hardware and microcontroller. The available microcontroller based text reading system is mechanized to videograph text from the documents and utilizes text segmentation, extraction and recognition on each individual frames of the video. In this paper for the same available hardware of text reading system, we are proposing another approach that can mosaic all the video frames into one image composite and then apply the text processing algorithms. The paper also describes the effectiveness and efficiency of both the algorithms by comparing the execution time and memory.

Keywords: Text Reading System, Block Scan, Mosaicing, Text Segmentation, Text extraction, OCR.

I. INTRODUCTION

Text processing from scanned documents, natural scene images and videos are extensively popular among vision community due to the information availability in them for content understanding and retrieval. Also, due to the availability of handy, low cost, portable cameras as in case of mobile phones, tablets and IPAD the text localization, segmentation, extraction and recognition in images has become prevalent rather than the scanned documents. In this paper we are mosaicing the text contents which is videographed using a mechanized microcontroller based text reading system.

Nagabhushan et.al. in [1] described and developed a microcontroller based mechanized text reading system for real time which auto generates voice text after recognizing the text from printed documents. We are utilizing the same hardware of the text reading system. The only difference between the approaches, which we are proposing in this paper is in the software design in which we mosaic text contents from the videographed frames rather than applying text extraction, localization and OCRing algorithm for each frame.

The mosaicing algorithm is used based on the vertical strip based method proposed in [2] using SIFT – match algorithm and then the text processing algorithms are applied on the mosaiced image A comparative study between the available method and proposed method is done in this paper.

The remainder of this paper is structured as follows: In section 2 we have done a survey on the available literature and devices. In section 3 we present the motivation for developing the text reading system. Section 4

presents the proposed design, development of the text reading system using mosicing. Section 5 describes the results and discussion of the system design and in section 6 the conclusion obtained are discussed.

II. LITERATURE SURVEY

We have done a literature review on the research work that describes different text reading systems for visually challenged with the text processing approaches.. Also, we have done a brief appraisal on the image and document mosaicing algorithms present in the literature. The following are the few literatures that we have reviewed.

Nagabhushan P et.al [1] developed a microcontroller based mechanised videographing of text and auto-generation of voice text in real time. This is a dedicated text reading system that is able to videograph text and read the text. This utilized text processing algorithms for each individual frames and the recognized text were stored in notepad file. Appending of text from each block scan was done by file handling functions. We could see few hitches while appending text due to pointers manipulation.

Rajkumar N et.al [2] proposed a camera-based text labels and product packaging reading system for hand-held objects. The method used region of interest (ROI) by a mixture-of-Gaussians-based background subtraction technique. In the extracted ROI, text localization and recognition are conducted to acquire text details.

Nanayakkara, S et.al [3] developed a text reading device that can be worn on finger. It contained a microcontroller and a button camera. The device mainly assisted the visually challenged by reading paper-printed text. It is a novel and real time application giving auditory feedback.

Majid Mirmehdi et.al [4] developed a mobile head-mounted device for detecting and tracking text. A flat cap to which a web camera was connected as acquisition device to the laptop. A microcontroller based remote control was connected through wireless to identify text regions. The text processing included Maximal Stable Extremal Regions (MSERs) for image segmentation, text detection and extraction

Lowe et al's Scale Invariant Feature Transform was used in [5] to form panorama of images. The SIFT features were extracted from video frames and were matched using k - nearest neighbourhood. RANSAC was used to estimate homographies between the matched pairs and was verified by probabilistic model. Each of the connected components derived was bundle adjusted based on graph search method with joint camera parameters and was subjected to multi band blending to provide panoramic view,. the method is invariant to scaling, rotation and geometric distortions

Nagabhushan P et al [6] proposed a vertical strip based mosaicing technique based on SIFT for left to right videographed video frames. The reference frame was matched initially with the other adjacent frames and then the vertical strips were created. The false matches from SIFT were fitted using RANSAC and an affine transformation was solved for blending frames.

Hemanth Kumar et.al [7] and [8] proposed two novel approaches for mosaicing split images based on simple pixel correspondence and Euclidian distance

Anil K et.al [9] surveyed various ongoing research based on text segmentation, localization, extraction and recognition. The survey included the use of different approaches for text processing based on region, edge, textures.

Nirmala Shivananda and Nagabhushan [10] proposed a hybrid method for separating text from color document images which combines connected component analysis and an unsupervised thresholding for separation of text from the complex background. The proposed approach identifies the candidate text regions based on edge detection followed by a connected component analysis.

III. BACKGROUND

The main intention of dedicated text reading system is to read the text present in the document and to develop an assistive text reading tool for visually challenged. The software design for microcontroller based text reading system should be light and accommodative for real time applications so that the execution becomes faster, easier and effective. But it is obvious that the document cannot be read directly from a video. Thus, it is needed to mosaic frame contents and then extract segment and recognize text from the mosaiced image and at last read the text contents. The detailed description of video acquisition, mosaicing and text segmentation is described in the sections below.

IV. PROPOSED SYSTEM

The design of text reading system imitates the reading pattern of human beings. The hardware of the text reading system developed in [1] is designed in such a way that the camera moves from left to right videographing the text present in the document. Certain number of lines that can be captured in the camera's field of view is considered as one block. Once a block is completed the camera shifts vertically and starts acquiring the next set of lines as another block. Each block contains pretty number of frames since it is a video. Hence mosaicing algorithm is applied for the consecutive frames and then the text is extracted and recognized. Figure 1 shows the text reading system developed in [1].

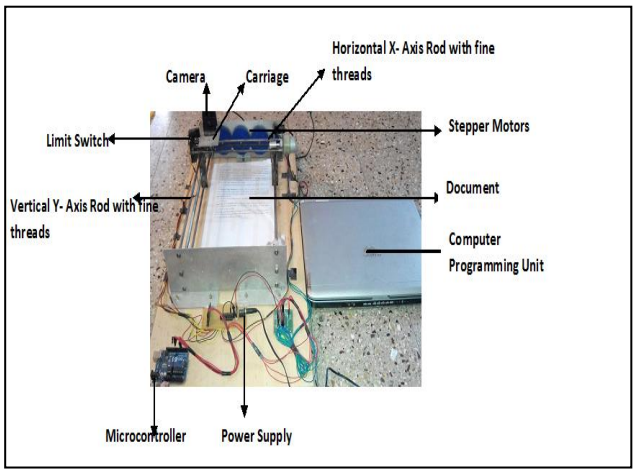


Fig 1: Text Reading SYSTEM as DEVELOPED in [1]

4.1Video Acquisition

As discussed in the above section, with the aid of the hardware unit developed in [1] the text present in the document is videographed using a web camera. Block scan refers to the number of frames captured during the

left to right parsing of camera along horizontal X-axis before a vertical shift of the camera along Y-axis is taken place.

Due to the camera's field of view, multiple lines of the document is captured and each frame contains atleast four to five lines which is either fully (short sentences) or partially (longer sentences) captured in horizontal direction. Hence it is useful to mosaic the consecutive frames contents of each block to acquire complete sentences of the four lines and then extract and recognize text captured in the respective block. The video acquisition procedure is described in figure 2 given below.

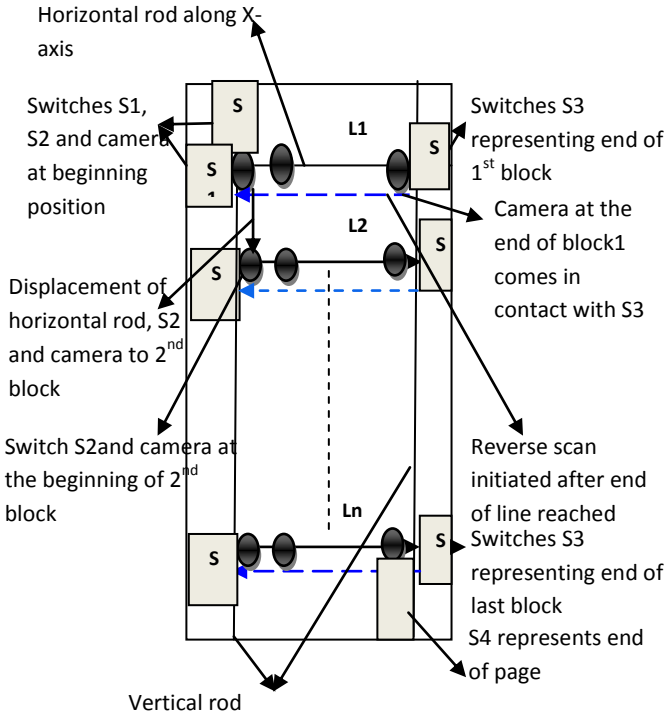


Fig 2 Video Acquisition as Proposed in [1]

Fig 3 Frame 1 of Block 1From Acquisition Device

Fig 4 Frame 10 of Block 1 from the Acquisition Device

4.2 Acquisition Procedure

The hardware design of the text reading system consists of a movable horizontal rod, a vertical rod, a movable carriage and a web camera mounted on the carriage. The camera is connected to the computer through USB. There are four limit switches S1, S2, S3 and S4 switches that acts as sensors for starting and stopping the carriage movement.. For moving the camera along X- axis and Y-axis we have motors M1 and M2 The initial

scan position is at the left top most corner where switches S1 and S2 are closed to initiate the scan along horizontal direction. The camera starts videographing the document and stores the frames in the temporary buffer. The frame trigger is set to 30 frames per minute between the switch S1 and S3. The carriage reaches and closes switch S3 which is present at right top most corner declaring the end of block is reached. The camera is parsed in reverse direction indicating the reverse scan i.e. from right to left and stops video acquisition. At this stage, the frames in the buffer are subjected frame selection, registration, mosaicing and text recognition algorithms.

After switch S3 is closed the camera rolls back to the initial position at S1 and S2 and empties the buffer, moves the hotizontal X-axis rod along with the camera and carriage vertically and initiates the next block scan. The frame acquisition, frame selection, frame registration, text processing and text reading modules are executed independently for each block of horizontal scan containing single / multiple lines of text till the end of page is reached.

4.3 Frame Selection

As the text is captured using a moving web camera there will be sufficient amount of noise , blur and translational changes as stated in [1]. We are utilizing the same blur metric algorithm for each block scan that selects the best frames using no reference perceptual blur metric that is described in [1][11] and [12].

// Algorithm Frame selection
Step 1: Start
Step2 : for I = 1 < number of frames
Step 3: IM = read(frames(I))
Step4: blur = blur_metric(IM)
Step 5 if blur == T
Step 6 Store IM in new_frames
Step 7 end if

End for

4.5 Frame Registration

Image registration involves aligning the images geometrically that are acquired from different sources. But here we register the frame by setting the reference frame from the same source that is present in the buffer for each block scan. After the best frame is selected, the first frame in the buffer of frames is set as the reference frame. The rest of the frames are decomposed into horizontal strips, are matched and stitched to this reference frame.

4.6 Decomposition of frames into horizontal strips and its matching

The effectiveness of vertical strip decomposition method is stated in [5] by Nagabhushan P et.al. We have used the same algorithm to perform vertical strip decomposition and matching for each individual block scan. Strip decomposition is used because of its ability to align the image correctly, easy to determine image manifolds [5]. The strip decomposition according to [5] is computationally inexpensive as all the features of two images is not matched but match only the part of the image strip where there is spatial transition in the content.In the frame

decomposition stage we are dividing the reference frame and its consecutive frames into three vertical strips creating three sub images from that of the original and the third sub image is stored in a buffer memory. The third sub image is matched using SIFT- match as stated in [6]. The outliers are mapped by estimating the homograhies using RANSAC. Homography maps the points in two images with one to one correspondence. Mathematically homography refers to projective linear transformation. Once the falsely detected outliers are removed the blending function is called to mosaic the images. The algorithms for strip decomposition, strip matching, homography estimation, blending stated in [5] is as given below.

//Algorithm for strip decomposition for each block

Step 1: Start

Step 2: [row, col] = size(Image)

Step 3: for i = 1 to row

Step 4 : for j = 1 to col/3

Step 5 VS1(i, j) = Image(i, j)

End

End

Step 6: for i =1 to row

Step 7: for j = (col /3)+ 1 to (2 x col)/ 3

Step 8: VS2(i, j) = Image(i, j)

End End

Step 9 for i = 1 to row

Step 10: for j = (2 xcol)+ 1 / 3 to col

Step 11: VS 3(i, j) = Image(i, j)

End End

Step 12: Return (HS3)

End

// Algorithm for strip matching for each block scan

Step 1: Start

Step 2: Set reference frame as fr = 1

Step 3: for I = 1 to strip _HS3

Step 4: u = image_strip(fr)

Step 5: v = image_strip(I)

Step 6: Match_Score = SIFT_match(u, v)

Step 7 : if (Match_Score > theshold)

Step 8 : IH =SIFT_Match(fr, v)

Step 9: Store outliers in table

Step 10 : Compute homography

Step 11: Call Blending Function

Step 10: End

// Algorithm for RANSAC Estimation

1. Choose number of samples N

2. Choose 4 random potential matches

3. Compute H using normalized DLT

4. Project points from x to x' for each potentially matching pair:

5. Count points with projected distance < t – E.g t = 3 pixels

6. Repeat steps 2-5 N times – Choose H with most inlier

7. Call affine blending function

The mosaiced image after wrapping is as shown below. The mosaiced image from each the block scan is stored in memory. The text segmentation, extraction and recognition algorithm is applied for the mosaiced image and the result is stored in the notepad file. The text segmentation, extraction and recognition algorithms are discussed below.

Fig 5 Mosaiced Image of block scan 1

4.7 Text Segmentation

The mosaiced image from the block scan is subjected to segmentation. The documents used for our experiments contain uniform white background with black text. Also, the block scan is free from skew therefore we can consider horizontal projection profile for line segmentation. Text regions are classified based on the connected component in each line. As discussed in [1] character cannot be too small in the video therefore we consider the region as noise if the height and width of a connected component are lesser than that of the smallest character in the video. The segmented text contents are extracted using the bounding box around the connected components. The result of segmentation is given in figure *6.*

Fig 6 Text Segmented and Extracted in the Frame Using Bounding Box

4.8 Text Recognition

The extracted text is recognized using Tesseract OCR engine to recognize the text. The recognized text is stored in the notepad file. The new mosaiced images after every block scan are recognized and the text contents are appended at the end of the notepad file. Any repetition in the notepad can be found using sentence duplicate finder algorithm.

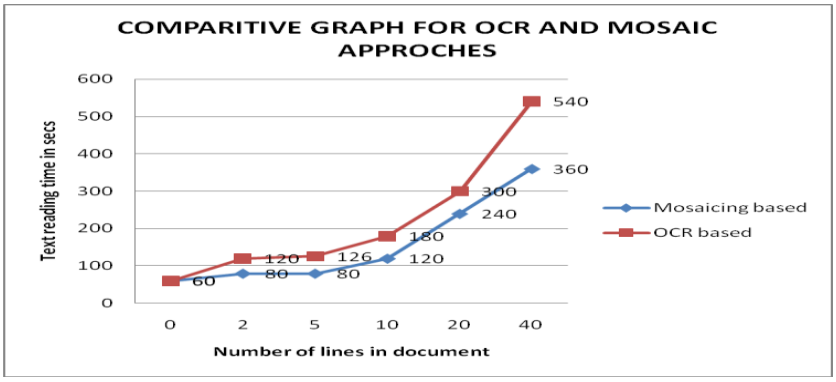
4.9 Text Reading

As in [1] the text reading system is used to read the contents of the document aloud. We have utilized the espeak software to read the notepad contents at the end of notepad creation. The process of frame selection, registration, frame decomposition, mosaicing, text segmentation, text recognition are repeated for each new block scan till the end of the page is reached.

V. RESULTS AND DISCUSSION

The software module was implemented using Matalb 2014(a). We tested reading English printed text containing single, multiple lines with different font styles size and spacing of text which we had created for the experimentation purpose. The proposed approach was compared with the existing approach. The findings derived are as listed below

1. The text extraction and recognition was much efficient in case of the existing system than that of the proposed system. The Tessaract OCR achieved 98 % accuracy in case of the existing system due to text extraction and recognition was from noise free individual video frames. Where as in the proposed system the efficiency reduced to 85% due to seam lines and brightness issues due to blending.
2. Memory for notepad file creation multiple times and pointer manipulations for text appending were almost nullified due to single composite image creation at the end of each block scan and single notpad creation from the recognized text.
3. Execution time for reading the complete document was more efficient in case of mosaicing based proposed system. The comparative graph that depicts the execution time of each approach is as given in Graph 1.



Graph 1. Comparitive Graph for OCR and Mosaiced Based Approach

VI. CONCLUSION

The text reading system developed in [1] is flexible in its software design. Hence in this paper we have proposed mosaicing based approach that forms single image composite for every block scan and then utilizes the text processing algorithms for reading text. The approach proposed executes faster for the same hardware. Also in this paper we have derived the comparative finding between the existing system and the proposed OCR based system for each individual frames in terms of memory, time and reading efficiency.

REFERENCES

[1] Nagabhushan. P and Vimuktha Evangeleen Jathanna, Microcontroller Based Mechanised Videographing of Text and Auto-Generation of Voice Text in Real Time, IJCSIT, International Journal of Computer Science and Information Technologies, Vol. 6 (3) , 2015, 2419-2425

[2] Rajkumar N, Anand M.G, Barathiraja N, "Portable Camera-Based Product Label Reading For Blind People", International Journal of Engineering Trends and Technology (IJETT) – Volume 10 ,Issue No : 1 - Apr 2014.

[3] Nanayakkara, S., Shilkrot, R., Yeo, K. P., and Maes, P." EyeRing: a finger-worn input device for seamless interactions with our surroundings". In Augmented Human ,2013.

[4] Carlos Merino-Gracia,, Karel Lenc,Majid Mirmehdi, " A Head-Mounted Device for Recognizing Text in Natural Scenes "Lecture Notes in Computer Science Volume 7139, pp 29-41,2012.

[5] Nagabhushan P, Vimuktha Evangeleen Jathanna, Mosaicing of Text Contents From Adjacent Video Frames. International Journal of Machine Intelligence. Vol 3, Issue 4, 2011

[6] Brown, M. Lowe, David G. (2003). Recognizing Panoramas. Proceedings of the ninth IEEE International Conference on Computer Vision. 2.pp. 121825.doi:10.1109/ICCV.2003.1238630

[7] Hemanth Kumar, P.Shivkumar, D.S.Guru, P.Nagabushan (2004) Document Image Mosaicing: A novel approach. Sadhana Vol 29, Part 3, June 2004, pp-329=341, India

[8] Hemanth Kumar, P.Shivkumar, D.S.Guru, P.Nagabushan (2004) Sliding window based approach for document image mosaicing. Elsevier Image and Vision Computing Volume 24, Issue 1, 1 January 2006, Pages 94-100

[9] K. Jung, K.I. Kim, A.K. Jain, Text information extraction in images and video: a survey, Pattern Recognition, 37 977-997S, 2004

[10] Nirmala Shivananda and P. Nagabhushan, Separation Foreground Text from Complex Background in Color Images, IEEE Transactions Processing vol.10, 306-0

[11] R. Ferzli and L. J. Karam, A No-Reference Objective Image Sharpness Metric Based on Just-Noticeable Blur and Probability Summation, ICIP'07, pp.445-448, 2007

[12] H. R. Sheikh, A. C. Bovik, L. Cormack. No-reference quality assessment using natural scene statistics: JPEG2000, IEEE Trans. on Image Processing, 14(11), pp.1918–1927,2005.