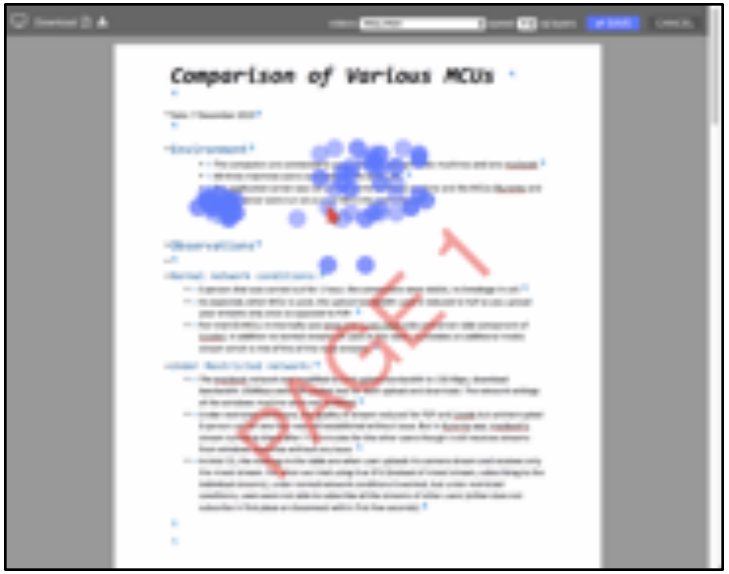
DocuGram: Turning Screen Recordings into Documents



Laurent Denoue, Scott Carter, Matthew Cooper

FXPAL

Sharing documents and teleconferencing

- Modern workflow involves web conferences
  - Google Hangout, Skype, GoToMeeting, etc.
  - People often need to share documents

- People share their screen, because:
  - it's ubiquitous: every conferencing tool has it
  - they can still interact while talking
  - they can share only what is important
  - others might not have the right software or credentials to open the document even if they sent it

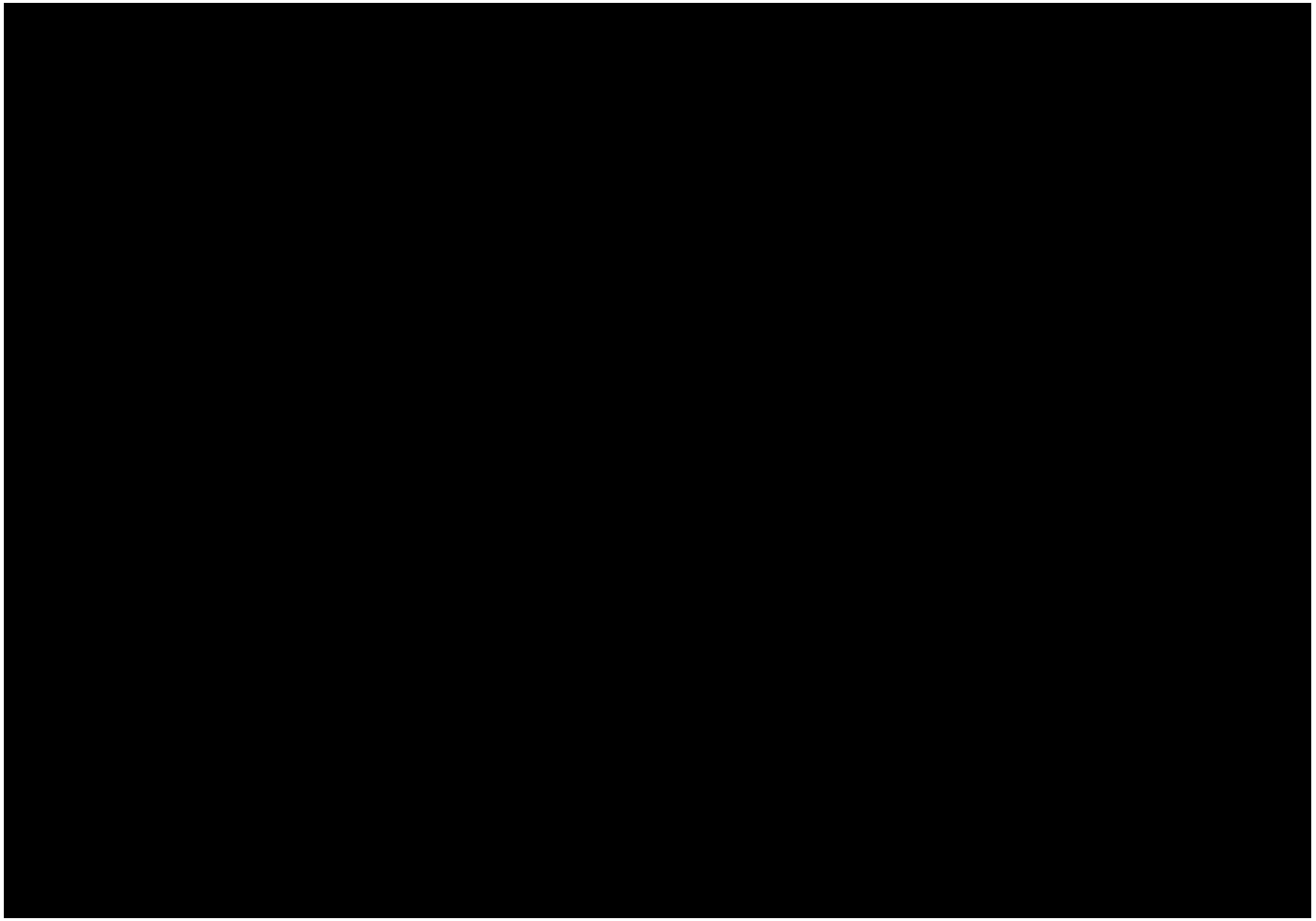
DocuGram solution

- Let users simply capture their screen as they interact with the document: scrolling, mouse actions, voice comments

- Process the recorded frames to reverse engineer the document that was shown
  - Recover document pages
  - Extract user's actions (mouse, text selections)
  - Link user's voice to the actions

Result is a sharable web link that shows the document

Demonstration



Technical Steps

- Region of Interest (ROI) computation
- Based on image stitching
  - Fast feature point extractor
  - Robust and fast to match binary descriptors
  - Matching of detected descriptors between 2 frames
  - Gives a vertical motion deltaY
- When no vertical motion is found
  - Fast frame difference to estimate mouse position
- Stitch all frames into one image and segment tall image into pages based on horizontal breaks

Reversed engineered document from screen recording!

Region of Interest (ROI)

- Goal is to find the region corresponding to the document

- Fast binarization computes horizontal and vertical edges
- Aligned verticals are grouped, aligned horizontals are grouped
- Pairs that roughly join are candidate corners

Pick the biggest of rectangles found by the corners

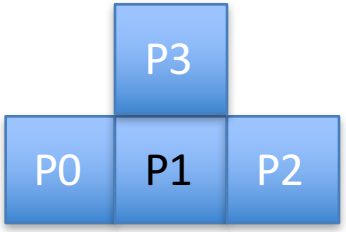
Image stitching (1)

- Used in video processing
  - E.g. creating panoramas in real-time
- Based on 2 insights
  - Extract consistent key-points from each image
  - Compute a robust descriptor of each keypoint

- Consistent = pick corresponding points across frames
- Robust = can tolerate some amount of noise, ~~rotation, translation~~

Image stitching (2)

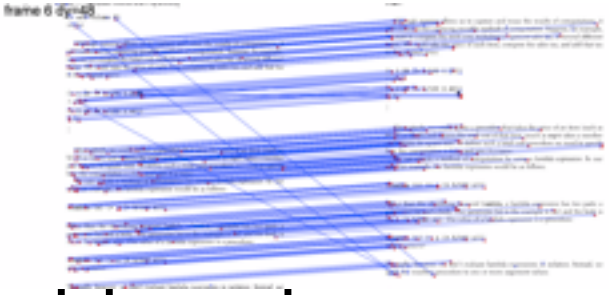
- We use a fast key-point extractor that works well for textual content



$|P0-P1| > 32$ AND $|P0-P2| <= 2$ AND $|P0-P3| > 32$

- We use the BRIEF binary descriptor
  - $BRIEF_{p(x,y)} = \{0,\ldots,511\}$ bits = 64 bytes
  - $Similarity(B_{p1},B_{p2}) = XOR(B_{p1},B_{p2})$ <= VERY FAST!!!

Computing vertical shift



- For every video frame captured at time t, we have key-points and descriptors:

$$(\text{pts, desc})_{\text{frame(t)}}$$

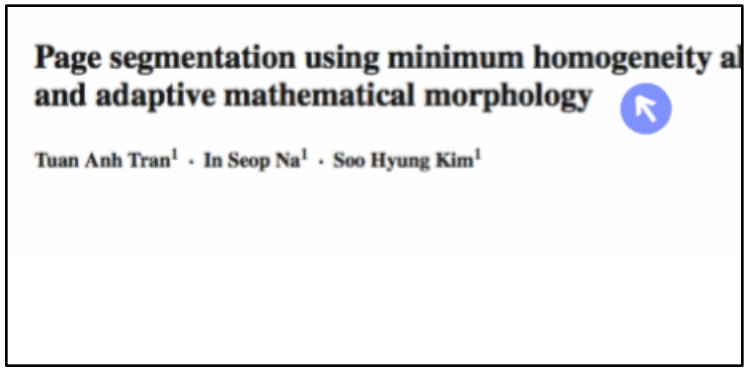
- Now loop through the pairs and find best match:

$$(\text{pts,desc})_{\text{frame(t-1)}} \text{ with } (\text{pts,desc})_{\text{frame(t-1)}}$$

⇨ Gives us the vertical shift between the frames

Detection of mouse position

- If vertical shift is zero, compute frame difference
  - Intuitively, people don't scroll when they move the cursor



- Needs to be fast
  - Reduce images to 64x64 pixels before comparing

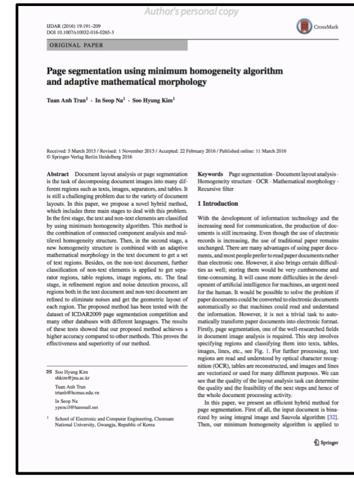
Obtain coarse location of mouse cursor

Final document re-generation

- Allocate a tall image = Sum(deltaY)
- Paint each frame at the corresponding deltaY

- Detect likely page boundaries
  - Binarize tall image using only vertical difference
  - Detect long horizontals and their gaps
  - Cut the long image into individual pages

Obtains one image per page



Author's personal copy

IJDAR (2016) 19:191–209
DOI 10.1007/s10032-016-0265-3

CrossMark

ORIGINAL PAPER

Page segmentation using minimum homogeneity algorithm and adaptive mathematical morphology

Tuan Anh Tran[1] · In Seop Na[1] · Soo Hyung Kim[1]

Received: 5 March 2015 / Revised: 1 November 2015 / Accepted: 22 February 2016 / Published online: 11 March 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract Document layout analysis or page segmentation is the task of decomposing document images into many different regions such as texts, images, separators, and tables. It is still a challenging problem due to the variety of document layouts. In this paper, we propose a novel hybrid method, which includes three main stages to deal with this problem. In the first stage, the text and non-text elements are classified by using minimum homogeneity algorithm. This method is the combination of connected component analysis and multilevel homogeneity structure. Then, in the second stage, a new homogeneity structure is combined with an adaptive mathematical morphology in the text document to get a set of text regions. Besides, on the non-text document, further classification of non-text elements is applied to get separator regions, table regions, image regions, etc. The final stage, in refinement region and noise detection process, all regions both in the text document and non-text document are refined to eliminate noises and get the geometric layout of each region. The proposed method has been tested with the dataset of ICDAR2009 page segmentation competition and many other databases with different languages. The results of these tests showed that our proposed method achieves a higher accuracy compared to other methods. This proves the effectiveness and superiority of our method.

✉ Soo Hyung Kim
shkim@jnu.ac.kr

Tuan Anh Tran
trtanh@hcmus.edu.vn

In Seop Na
ypencil@hanmail.net

1 School of Electronic and Computer Engineering, Chonnam National University, Gwangju, Republic of Korea

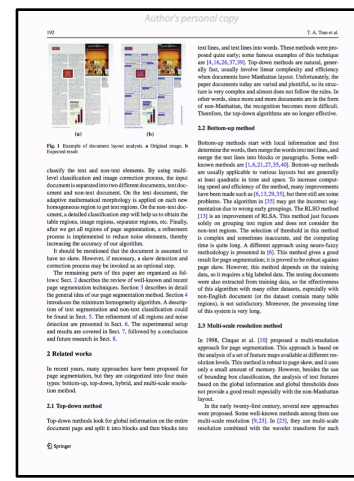
Keywords Page segmentation · Document layout analysis · Homogeneity structure · OCR · Mathematical morphology · Recursive filter

1 Introduction

With the development of information technology and the increasing need for communication, the production of documents is still increasing. Even though the use of electronic records is increasing, the use of traditional paper remains unchanged. There are many advantages of using paper documents, and most people prefer to read paper documents rather than electronic one. However, it also brings certain difficulties as well; storing them would be very cumbersome and time-consuming. It will cause more difficulties in the development of artificial intelligence for machines, an urgent need for the human. It would be possible to solve the problem if paper documents could be converted to electronic documents automatically so that machines could read and understand the information. However, it is not a trivial task to automatically transform paper documents into electronic format. Firstly, page segmentation, one of the well-researched fields in document image analysis is required. This step involves specifying regions and classifying them into texts, tables, images, lines, etc., see Fig. 1. For further processing, text regions are read and understood by optical character recognition (OCR), tables are reconstructed, and images and lines are vectorized or used for many different purposes. We can see that the quality of the layout analysis task can determine the quality and the feasibility of the next steps and hence of the whole document processing activity.

In this paper, we present an efficient hybrid method for page segmentation. First of all, the input document is binarized by using integral image and Sauvola algorithm [32]. Then, our minimum homogeneity algorithm is applied to

Springer



Author's personal copy

(a) (b)

Fig. 1 Example of document layout analysis. a Original image. b Expected result

classify the text and non-text elements. By using multilevel classification and image correction process, the input document is separated into two different documents, text document and non-text document. On the text document, the adaptive mathematical morphology is applied on each new homogeneous region to get text regions. On the non-text document, a detailed classification step will help us to obtain the table regions, image regions, separator regions, etc. Finally, after we get all regions of page segmentation, a refinement process is implemented to reduce noise elements, thereby increasing the accuracy of our algorithm.

It should be mentioned that the document is assumed to have no skew. However, if necessary, a skew detection and correction process may be invoked as an optional step.

The remaining parts of this paper are organized as follows: Sect. 2 describes the review of well-known and recent page segmentation techniques. Section 3 describes in detail the general idea of our page segmentation method. Section 4 introduces the minimum homogeneity algorithm. A description of text segmentation and non-text classification could be found in Sect. 5. The refinement of all regions and noise detection are presented in Sect. 6. The experimental setup and results are covered in Sect. 7, followed by a conclusion and future research in Sect. 8.

2 Related works

In recent years, many approaches have been proposed for page segmentation, but they are categorized into four main types: bottom-up, top-down, hybrid, and multi-scale resolution method.

2.1 Top-down method

Top-down methods look for global information on the entire document page and split it into blocks and then blocks into text lines, and text lines into words. These methods were proposed quite early; some famous examples of this technique are [4,16,26,37,39]. Top-down methods are natural, generally fast, usually involve linear complexity and efficiency when documents have Manhattan layout. Unfortunately, the paper documents today are varied and plentiful, so its structure is very complex and almost does not follow the rules. In other words, since more and more documents are in the form of non-Manhattan, the recognition becomes more difficult. Therefore, the top-down algorithms are no longer effective.

2.2 Bottom-up method

Bottom-up methods start with local information and first determine the words, then merge the words into text lines, and merge the text lines into blocks or paragraphs. Some well-known methods are [1,6,21,27,35,40]. Bottom-up methods are usually applicable to various layouts but are generally at least quadratic in time and space. To increase computing speed and efficiency of the method, many improvements have been made such as [6,13,29,35], but there still are some problems. The algorithm in [35] may get the incorrect segmentation due to wrong early groupings. The RLSO method [13] is an improvement of RLSA. This method just focuses solely on grouping text region and does not consider the non-text regions. The selection of threshold in this method is complex and sometimes inaccurate, and the computing time is quite long. A different approach using neuro-fuzzy methodology is presented in [6]. This method gives a good result for page segmentation; it is proved to be robust against page skew. However, this method depends on the training data, so it requires a big labeled data. The testing documents were also extracted from training data, so the effectiveness of this algorithm with many other datasets, especially with non-English document (or the dataset contain many table regions), is not satisfactory. Moreover, the processing time of this system is very long.

2.3 Multi-scale resolution method

In 1998, Cinque et al. [10] proposed a multi-resolution approach for page segmentation. This approach is based on the analysis of a set of feature maps available at different resolution levels. This method is robust to page skew, and it uses only a small amount of memory. However, besides the use of bounding box classification, the analysis of text features based on the global information and global thresholds does not provide a good result especially with the non-Manhattan layout.

In the early twenty-first century, several new approaches were proposed. Some well-known methods among them use multi-scale resolution [9,23]. In [23], they use multi-scale resolution combined with the wavelet transform for each

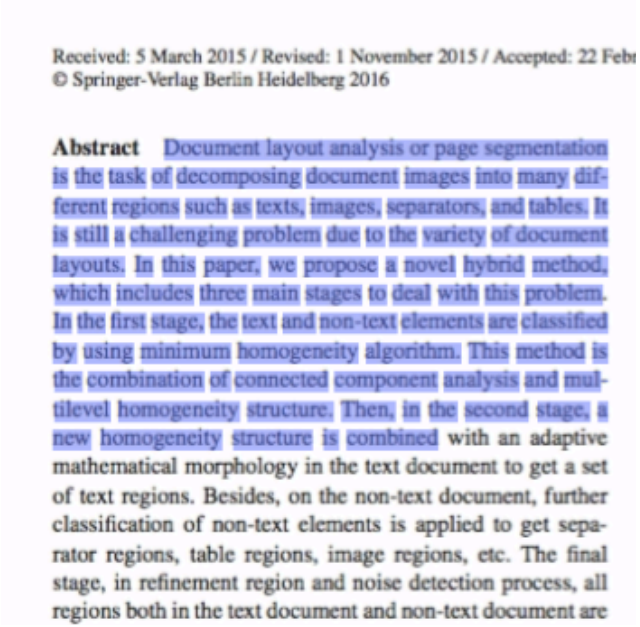
Springer

Text detection

- Users like to select text in documents
  - XY-Cut on binarized document pages
  - Group connected components into text lines
  - Split each text line into word candidates based on vertical projection profiles

- CANVAS detects mouse motion and highlights text rectangles

User feels interacting with the original document


Received: 5 March 2015 / Revised: 1 November 2015 / Accepted: 22 Feb
© Springer-Verlag Berlin Heidelberg 2016

Abstract Document layout analysis or page segmentation is the task of decomposing document images into many different regions such as texts, images, separators, and tables. It is still a challenging problem due to the variety of document layouts. In this paper, we propose a novel hybrid method, which includes three main stages to deal with this problem. In the first stage, the text and non-text elements are classified by using minimum homogeneity algorithm. This method is the combination of connected component analysis and multilevel homogeneity structure. Then, in the second stage, a new homogeneity structure is combined with an adaptive mathematical morphology in the text document to get a set of text regions. Besides, on the non-text document, further classification of non-text elements is applied to get separator regions, table regions, image regions, etc. The final stage, in refinement region and noise detection process, all regions both in the text document and non-text document are

Conclusion

- Easy-to-use tool for users to capture and share any text-based document shown by any application
- Real-time processing inside their browser
- No application to install (only Chrome, Firefox)
- Novel use of video processing techniques to reverse engineer displayed documents
- Capture and replay of user's interactions

Future work

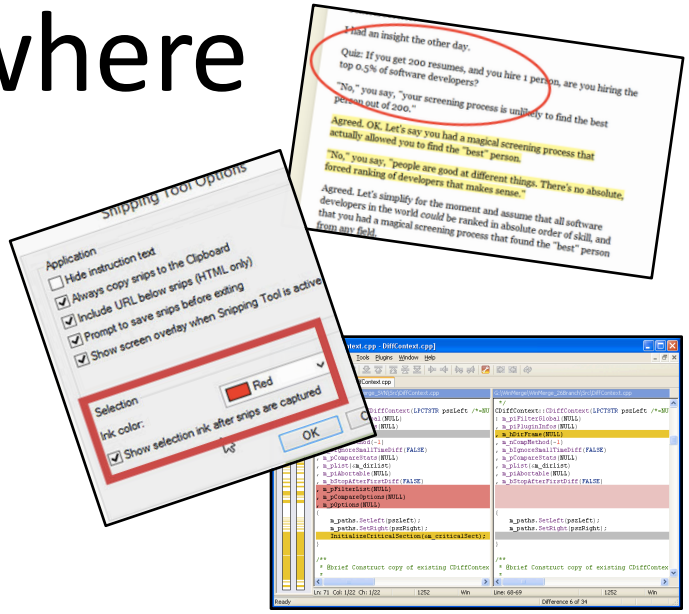
- OCR on captured document page images
  - Problems with low resolution, anti-aliasing
  - Promising method = LSTM networks for OCR
  - They don't assume pre-segmentation
- Beyond cursor motion
  - Users also select text, highlight passages, etc.
- Integration inside live video-conferencing
  - Users simply share their screen
  - DocuGram technology reverse engineers the documents and inserts them into the chat window
    - e.g. Slack, Hangout, HipChat

References

- Calonder, M., Lepetit, V., Strecha, C. and Fua, P., 2010. Brief: Binary robust independent elementary features. Computer Vision–ECCV 2010, pp.778-792.
- Javascript CANVAS image-manipulation and BRIEF implementation https://trackingjs.com/

OLD SLIDES

Screenshots are everywhere



- Everybody takes screenshots
  - Keep receipts, review web-sites

- But how do you screenshot a document?
  - Screenshot first page
  - Scroll down
  - Repeat 1 and 2
  - Stitch inside Photoshop?

  ➡ Not very easy

Problem definition

- Sharing a document requires many hoops
- It’s not easy to share only a part of the document
- It’s next to impossible to comment what we share
  - Export to PDF, highlight?
  - What about voice?
  - What about cursor motion?

We can find a better way

Why document screenshots?

- Digital born documents should be easy to share
  - Just email me the document
    - What format? PDF, DOC, DOCX, DOC 95, PPT, PPTX
    - Will my colleague be able to open this?
    - Can I share just 2 pages?
  - Just send the link of the document
    - Web page, Google Drive, Office 365, GitHub code
    - Will my colleague have the credentials?
    - Can I share just slides 4 and 5?

Application scenarios

- User shares a PDF document shown inside Preview App
- User shares a few slides from a PowerPoint deck
- User shares a few lines of code from his text editor