CBDAR 2011

Camera-Based Document Analysis and Recognition

Program Booklet

of the

Fourth International Workshop on Camera-Based Document Analysis and Recognition

September 22, 2011

Beijing Friendship Hotel, Beijing, China

Edited by

Masakazu Iwamura
Osaka Prefecture University, Japan

Faisal Shafait
DFKI GmbH, Germany

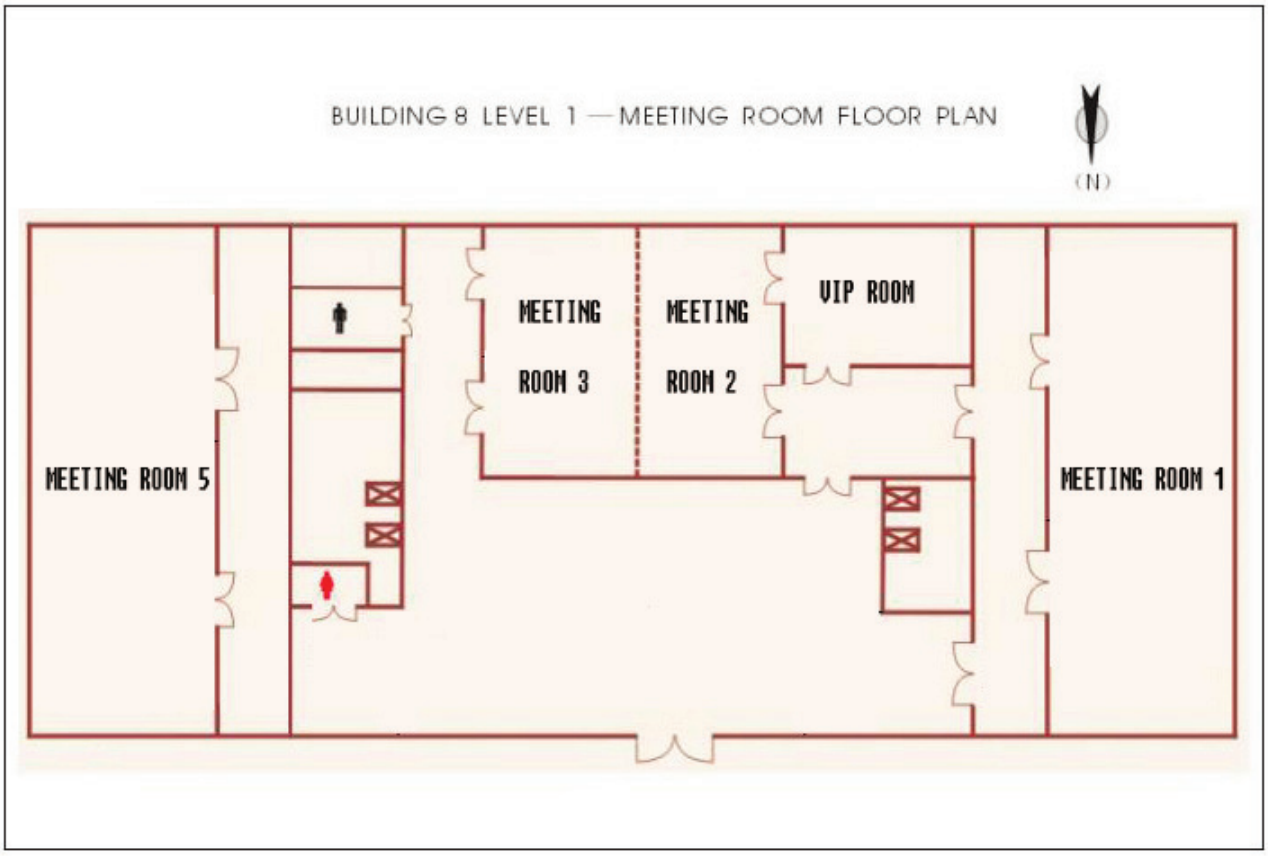
FX PAL PALO ALTO LABORATORY

Google

DFKI

Venue

The main site of CBDAR 2011 is Room #2 of Building 8. Poster and Demo Sessions will be held in the neighboring VIP room.



Time Table

8:00-	Registration Desk Open
8:45- 9:00	Opening
9:00- 9:50	Keynote 1 PaperUI Dr. Qiong Liu (FXPAL)
9:50-10:20	Coffee Break
10:20-11:40	Oral 1
11:40-12:40	Lunch
12:40-14:10	Poster & Demo
14:10-15:10	Oral 2
15:10-15:40	Coffee Break
15:40-16:30	Keynote 2 Reading text in Google Goggles and Streetview images Dr. Alessandro Bissacco (Google Inc.)
16:30-17:00	Discussion
17:00-17:15	Closing

Program

Oral Session 1 (10:20-11:40)

O1-1 10:20-10:40	Multiscript and Multioriented Text Localization from Scene Images *Thotreingam Kasar and A G Ramakrishnan*
O1-2 10:40-11:00	Assistive Text Reading from Complex Background for Blind Persons *Chucai Yi and Yingli Tian*
O1-3 11:00-11:20	A Head-Mounted Device for Recognizing Text in Natural Scenes *Carlos Merino-Gracia, Karel Lenc and Majid Mirmehdi*
O1-4 11:20-11:40	Recognizing Natural Scene Characters by Convolutional Neural Network and Bimodal Image Enhancement *Yuanping Zhu, Jun Sun and Satoshi Naoi*

Oral Session 2 (14:10-15:10)

O2-1 14:10-14:25	Decapod: A Flexible, Low Cost Digitization Solution for Small and Medium Archives *Faisal Shafait, Michael Patrick Cutter, Joost van Beusekom, Syed Saqib Bukhari, Thomas M. Breuel*
O2-2 14:25-14:40	An Image Based Performance Evaluation Method for Page Dewarping Algorithms Using SIFT Features *Syed Saqib Bukhari, Faisal Shafait and Thomas M. Breuel*
O2-3 14:40-14:55	NEOCR: A Configurable Dataset for Natural Image Text Recognition *Robert Nagy, Anders Dicker and Klaus Meyer-Wegener*
O2-4 14:55-15:10	Memory Reduction for Real-Time Document Image Retrieval with a 20 Million Pages Database *Kazutaka Takeda, Koichi Kise and Masakazu Iwamura*

Poster Session (12:40-14:10)

P1	An Algorithm for Colour-based Natural Scene Text Segmentation *Chao Zeng, Wenjing Jia and Xiangjian He*
P2	Text Detection of Two Major Indian Scripts in Natural Scene Images *Aruni Roy Chowdhury, Ujjwal Bhattacharya and Swapan K. Parui*
P3	QUAD: Quality Assessment of Documents *Deepak Kumar and A G Ramakrishnan*
P4	An Experimental Comparison of Image Retrieval Methods Using Character String Images Distorted by Perspective Transformation *Tatsuo Akiyama and Daisuke Nishiwaki*
P5	A Camera-Based Interactive Whiteboard Reading System *Szilard Vajda, Leonard Rothacker and Gernot A. Fink*
P6	Fast Text Binarization for Scenery Images Under Rough Color Specifications *Keiichiro Shirai, Guofeng Ma and Masayuki Okamoto*
P7	Border Noise Removal of Camera-Captured Document Images Using Page-Frame Detection *Syed Saqib Bukhari, Faisal Shafait and Thomas M. Breuel*
P8	The IUPR Dataset of Camera-Captured Document Images *Syed Saqib Bukhari, Faisal Shafait and Thomas M. Breuel*

Demo Session (12:40-14:10)

D1	smartFIX - The Multichannel Document Analysis Product by Insiders Technologies *Florian Deckert and Michael Gillmann*
D2	Hypothesis Preservation Approach to Scene Text Recognition with Weighted Finite-State Transducer *Takafumi Yamazoe, Minoru Etoh, Takeshi Yoshimura and KousukeTsujino*
D3	Real-Time Document Image Retrieval with a 1 Million Pages Database Running on a Laptop *Kazutaka Takeda, Koichi Kise and Masakazu Iwamura*
D4	MAST: Multi-Script Annotation Toolkit for Scenic Text *T Kasar, D Kumar, M N Anil Prasad, D Girish and A G Ramakrishnan*

smartFIX - The Multichannel Document Analysis Product by Insiders Technologies

Florian Deckert, Michael Gillmann
Insiders Technologies GmbH
Brüsseler Straße 1, 67657 Kaiserslautern, Germany
{*f.deckert, m.gillmann*}*@insiders-technologies.de*

smartFIX [3] is a document analysis product for knowledge-based extraction of data from any document format. Paper documents as well as any type of electronic document format (e.g. faxes, e-mails, MS Office, PDF, HTML, XML, etc.) can be processed. Regardless of document format and structure, smartFIX recognizes the document type and any other important information during processing.

smartFIX imports the documents to be processed from various sources. Scanned paper documents, incoming fax documents, e-mails, and other electronic documents are processed. Basic image processing, like binarization, despeckling, rotation and skew correction is performed on each image page. If desired, smartFIX automatically merges individual pages into documents and creates processes from individual documents. For each document, its document class is determined, defining the business process to be triggered in the company. smartFIX subsequently identifies all relevant information on the document belonging to the respective business process. In this step, smartFIX can use customer relation and enterprise resource planning data (ERP data) provided by a matching database and other sophisticated knowledge-based methods to increase the detection rate. The quality of extracted data is enhanced by automatic mathematical and logical checks over all recognized values. To do this e.g. Constraint Solving [2] and Transfer Learning methods [4] are used. Values that are accurately and unambiguously recognized are released for direct export; uncertainly [5] recognized values are forwarded to a verification workplace for manual checking and verification. The quality-controlled data is then exported to the desired downstream systems e.g., an enterprise resource planning system like SAP for further processing.

smartFIX provides self-learning mechanisms as a highly successful method for increasing recognition rates. The self-learning mechanisms use the post-verification quality-controlled data as ground truth (GT) in order to find rules for the analysis step. Data that has been validated automatically is used to evaluate the reliability of the learned rules as well. Typical learned rules are the position of fields relative to stationary layouts or keywords, regular expressions, relative positions of extracted information (e.g. net amount and total amount) and many more.

smartFIX uses a strategy that searches for all entries contained in the customers database on the document. The procedure is independent of location, layout and completeness of the data on the document. Within smartFIX this strategy is called “Top Down Search”. It supplies results normally within less than one second even on large databases.

Many of the documents processed in smartFIX contain tables of vital information. Examples are position tables on invoices or tables containing requested items on orders. Our experience is that there is no clear layout that helps to identify columns and the BP only needs a subset of the provided information. Therefore smartFIX does not only rely on a physical structure. Table extraction in smartFIX is based on expectations about the presence and semantics of certain data entities in order to understand a table’s content [1].

ACKNOWLEDGMENT

The work presented in this paper was performed in the context of the Software-Cluster project EMERGENT (www.software-cluster.org). It was partially funded by the German Federal Ministry of Education and Research (BMBF) under grant no. 01IC10S01. The authors assume responsibility for the content.

REFERENCES

[1] F. Deckert, B. Seidler, M. Ebbecke, and M. Gillmann, *Table Content Understanding in smartFIX*, 11th Int. Conf. on Document Analysis and Recognition (ICDAR), Beijing, China, 2011.

[2] A. Fordan, *Constraint Solving over OCR Graphs*, 14th Int. Conf. on Applications of Prolog (INAP), Tokyo, Japan, 2001.

[3] B. Klein, A. Dengel, and A. Fordan, *smartFIX: An Adaptive System for Document Analysis and Understanding*, in: A. Dengel, M. Junker, A. Weissbecker (Eds.), *Reading and Learning - Adaptive Content Recognition*, LNCS 2956, Springer, 2004.

[4] F. Schulz, M. Ebbecke, M. Gillmann, B. Adrian, S. Agne, and A. Dengel, *Seizing the Treasure: Transferring Layout Knowledge in Invoice Analysis*, 10th Int. Conf. on Document Analysis and Recognition (ICDAR), Barcelona, Spain, 2009.

[5] B. Seidler, M. Ebbecke, and M. Gillmann, *smartFIX Statistics – Towards Systematic Document Analysis Performance Evaluation and Optimization*, 9th IAPR Int. Workshop on Document Analysis Systems (DAS), Boston, MA, USA, 2010.

Hypothesis Preservation Approach to Scene Text Recognition with Weighted Finite-State Transducer

Takafumi Yamazoe, Minoru Etoh, Takeshi Yoshimura, and Kousuke Tsujino
Service & Solution Development Department and Research Laboratories, NTT DOCOMO 3-6, Hikarino-oka, 239-8536 Japan
yamazoet at nttdocomo.com, {etoh, yoshimura.takeshi, tsujino} at nttdocomo.co.jp

The demonstration contains two mobile applications that shows scene text recognition. The applications, automatic image repository tagging and real-time multilingual menu recognition, use a language model based method which the authors newly propose at ICDAR 2011. The proposed method uses Weighted Finite-State Transducer (WFST) that greatly suppresses large-scale ambiguity in scene text recognition, especially for Japanese Kanji characters. The details appear as the same title in ICDAR 2011 proceedings.

DEMO1: IMAGE TAGGING

This shows a cloud-based prototype service that recognizes photo galleries on a mobile through the recognition of words in the scene images. The system is commercially available on the web for Japanese general public. <http://tangochu.jp/en/>

Over-extracted text regions

Elimination of Incorrect hypothesis

Recognition results

Recognition results (Japanese lexicon)

DEMO2: REAL-TIME MULTILINGUAL MENU RECOGNITION

The developed portable system locally extracts and recognizes multilingual scene texts in real-time. The target language and text types are generic though, the demonstration is tuned to local menu translation for which the system combines multiple language processing based on WFST and an existing OCR engine for traditional Chinese, simplified Chinese, Korean, English, and Japanese.

Application for the food menu translation

CBDAR2011 Demo Abstract D2

Real-Time Document Image Retrieval with a 1 Million Pages Database Running on a Laptop

Kazutaka Takeda, Koichi Kise and Masakazu Iwamura
Dept. of CSIS, Graduate School of Engineering
Osaka Prefecture University
1-1 Gakuen-cho, Naka, Sakai, Osaka, 599-8531 Japan
takeda@m.cs.osakafu-u.ac.jp, {kise, masa}@cs.osakafu-u.ac.jp

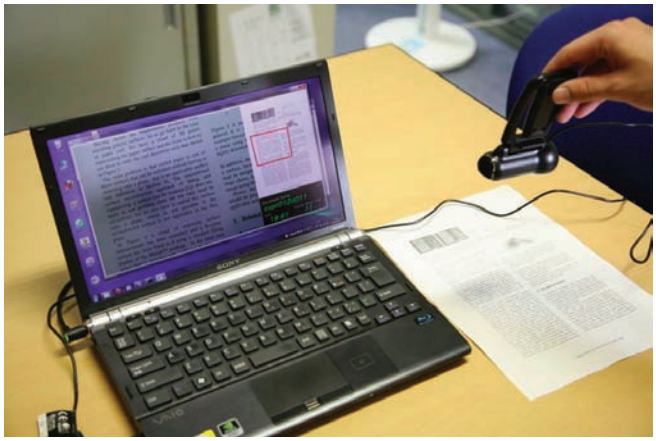


Figure 1. A scene of real-time demo system.

We propose a real-time document image retrieval method for a large-scale database in CBDAR2011 oral session [1]. In this paper, we introduce the real-time demo system using a web camera. As shown in Fig. 1, this system runs on a laptop and a user can retrieve document images by capturing a printed page with a web camera.

Figure 2 shows retrieval processing. As shown in Fig. 2, a retrieval result is calculated based on correspondences of feature points.

This demo system has several characteristics as follow.

- This system runs with a large-scale database.
  - 1 million pages on laptop with 8GB memory
- This system runs efficiently.
  - About 15 frame per second
- This system has robustness against following various types of disturbances.
  - Rotation
  - Scaling
  - Perspective distortion
  - Occlusion
  - Curvature
- Augmented Reality to the printed documents is realized.
  - An example is shown in Figure 3; the top left figures have been altered.
- Without character recognition, a user enable to obtain words and text in the captured region.

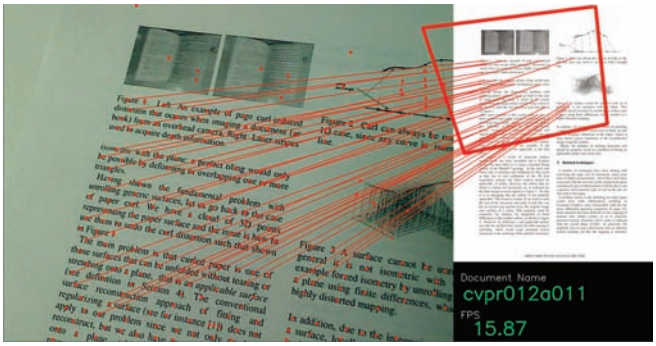


Figure 2. Correspondence of feature points.

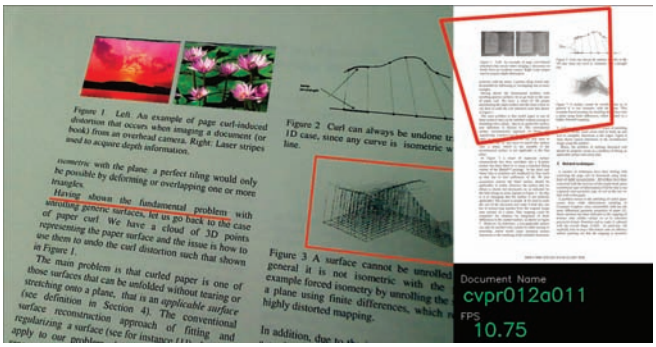


Figure 3. Augmented Reality to a printed document.

REFERENCES

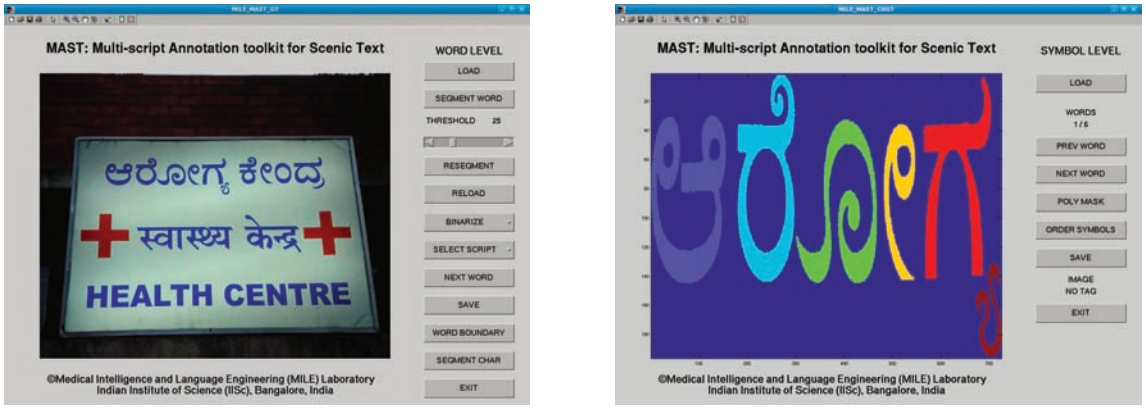
[1] K. Takeda, K. Kise, and M. Iwamura, "Memory reduction for real-time document image retrieval with a 20 million pages database," *Proceedings of the Fourth International Workshop on Camera-Based Document Analysis and Recognition (CBDAR2011)[to appear]*, 2011.

MAST: Multi-Script Annotation Toolkit for Scenic Text

T Kasar, D Kumar, M N Anil Prasad, D Girish and A G Ramakrishnan
Medical Intelligence and Language Engineering Laboratory
Indian Institute of Science, Bangalore, INDIA - 560 012
{tkasar, deepak, anilprasadmn, dasarigirish, ramkiag}@ee.iisc.ernet.in

Abstract

There is a significant need for methods to extract and recognize text in scenes. Unlike the case of processing conventional document images, natural scene text understanding usually involves a pre-processing step of text region location and extraction before subjecting the acquired image for character recognition task. There are no standard, pixel-level annotated databases containing camera-captured multi-script, multi-oriented text. The availability of annotated datasets for scenic images will aid in testing and quantifying the performances of various document analysis and recognition algorithms. We have developed a semi-automatic tool to aid the creation of such annotated databases for research in camera-based document analysis. The procedure involves manual seed selection followed by a region growing process to segment each word present in the image. The threshold for region growing can be varied by the user so as to ensure pixel-accurate character segmentation. The text present in the image is tagged word-by-word. A virtual keyboard interface has also been designed for entering the ground truth in ten Indic scripts, besides English. The keyboard interface can easily be generated for any script, thereby expanding the scope of the toolkit. Optionally, each segmented word can further be labeled into its constituent characters/symbols. Polygonal masks are used to split or merge the segmented words into valid characters/symbols. The ground truth is represented by a pixel-level segmented image and a '.txt' file that contains information about the number of words in the image, word bounding boxes, script and ground truth Unicode. The toolkit, which we call MAST, can be used to generate ground truth and annotation for generic document images. The software, developed on Matlab, is available online[1] *along with a detailed description of the functionalities of each of the menu items. We hope that researchers worldwide will find it useful in creating ground truth database for any generic document image and performance evaluation.*



(a) (b)

Figure 1. Screenshot of the user-interface for (a) word-level and (b) symbol level annotation.

[1] http://mile.ee.iisc.ernet.in/mast