2013

Camera-Based Document Analysis and Recognition

Program Booklet

of the

Fifth International Workshop on Camera-Based Document Analysis and Recognition

August 23, 2013

Omni Shoreham Hotel, Washington D. C., USA

Edited by

Masakazu Iwamura
Osaka Prefecture University, Japan

Faisal Shafait
University of Western Australia, Australia

Endorsed by IAPR

Timetable

8:00-	Registration Desk Open
8:45- 9:00	Opening
9:00- 9:50	Keynote Real-life Activity Recognition - Recognizing Reading Activities Dr. Kai Kunze (Osaka Prefecture University)
10:00-10:30	Morning Break
10:30-12:00	Oral 1: Scene Text
12:00-13:30	Lunch
13:30-15:00	Poster & Demo
15:00-15:30	Afternoon Break
15:30-16:30	Oral 2: Camera-Based Capture
16:30-17:00	Discussion
17:00-17:15	Closing

Program

Oral Session 1 - Scene Text (10:30-12:00)

O1-1 **10:30-10:50**	Spatially Prioritized and Persistent Text Detection and Decoding *Hsueh-Cheng Wang, Yafim Landa, Maurice Fallon and Seth Teller*
O1-2 **10:50-11:10**	Scene Text Detection via Integrated Discrimination of Component Appearance and Consensus *Qixiang Ye and David Doermann*
O1-3 **11:10-11:30**	Saliency inside Saliency - A Hierarchical Usage of Visual Saliency for Scene Character Detection *Renwu Gao, Faisal Shafait, Seiichi Uchida and Yaokai Feng*
O1-4 **11:30-11:50**	Font Distribution Analysis by Network *Chihiro Nakamoto, Rong Huang, Sota Koizumi, Ryosuke Ishida, Yaokai Feng and Seiichi Uchida*

Oral Session 2 - Camera-Based Capture (15:30-16:30)

O2-1 **15:30-15:50**	Hyperspectral Document Imaging: Challenges and Perspectives *Zohaib Khan, Faisal Shafait and Ajmal Mian*
O2-2 **15:50-16:10**	A Dataset for Quality Assessment of Camera Captured Document Images *Jayant Kumar, Peng Ye and David Doermann*
O2-3 **16:10-16:30**	Dewarping Book Page Spreads Captured with a Mobile Phone Camera *Chelhwon Kim, Patrick Chiu and Surendar Chandra*

Poster Session (13:00-15:00)

P1	A Robust Approach to Extraction of Texts from Camera Captured Images *Sudipto Banerjee, Koustav Mullick and Ujjwal Bhattacharya*
P2	Accuracy Improvement of Viewpoint-Free Scene Character Recognition by Rotation Angle Estimation *Kanta Kuramoto, Wataru Ohyama, Tetsushi Wakabayashi and Fumitaka Kimura*
P3	A Morphology-Based Border Noise Removal Method for Camera-Captured Label Images *Mengyang Liu, Chongshou Li, Wenbin Zhu and Andrew Lim*
P4	Robust Binarization of Stereo and Monocular Document Images Using Percentile Filter *Muhammad Zeshan Afzal, Martin Kramer, Syed Saqib Bukhari, Mohammad Reza Yousefi, Faisal Shafait and Thomas Breuel*
P5	Mobile Phone Camera-Based Video Scanning of Paper Documents *Muhammad Muzzamil Luqman, Petra Gomez-Krämer and Jean-Marc Ogier*
P6	Sign Detection Based Text Localization in Mobile Device Captured Scene Images *Jing Zhang and Rangachar Kasturi*

Demonstration of a Human Perception Inspired System for Text Extraction from Natural Scenes

Lluís Gómez and Dimosthenis Karatzas
Computer Vision Center
Universitat Autònoma de Barcelona
Email: {lgomez,dimos}@cvc.uab.es

This demonstration presents a prototype implementation of the method described in [1], able to detect textual content in scenes in real time. The text detection algorithm makes use of a high level representation of text as a perceptually significant group of atomic regions. This representation is independent from the particular script or language of the text, allowing the system to detect textual content in multi-script scenarios without any extra training.

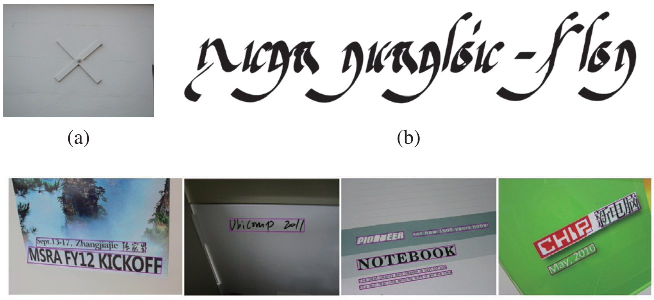
Mainstream state of the art methodologies are usually based on the classification of individual regions (connected components) or image patches. The approach followed here is distinctly different as text detection is not performed based on classifying individual components, but through searching for groups of components that share certain characteristics.

The inspiration comes from human perception. Humans make strong use of perceptual organisation to detect textual content, through which text emerges as a perceptually significant group of atomic objects (disjoint text parts such as characters or ideograms). Therefore humans are able to detect text even in languages and scripts never seen before (see Figure 1b), as long as such gestalts emerge by the way the text parts are arranged. As a result, the text extraction problem can be posed as the detection of such gestalts, as we show in [1].

From an implementation point of view (see Figure 2), the system creates group hypotheses, considering a number of different modalities in which atomic regions might be similar to each other (size, colour, background colour, stroke width, etc) and subsequently tests these hypotheses in terms of their meaningfulness. Meaningfulness is defined in a probabilistic way according to [2]. Given a group hypothesis comprising a set of regions which have a feature in common (they are similar in terms of that particular feature), meaningfulness measures the extent to which this common feature is happening by chance or not (thus it is a significant property of the group).

Similarity is explored in a number of modalities separately, but always in association with the spatial proximity of the atomic objects. The collaboration of the different similarity laws is taken into account at the end of the process through evidence accumulation [3], which provides a flexible way to identify maximal perceptual groups without any strict definition of the exact similarity laws invoked.

The system runs on a laptop computer using a Web camera to sense the environment. The current implementation offers real time performance at VGA resolution. The processing time of a frame is mainly dependent on the complexity of the scene and lesser on the frame resolution.



(a) (b) (c)

Fig. 1: (a) Should a single character be considered "text"? (b) An example of automatically created non-language text[1]. (c) Example result images from the MSRA-TD500 dataset.

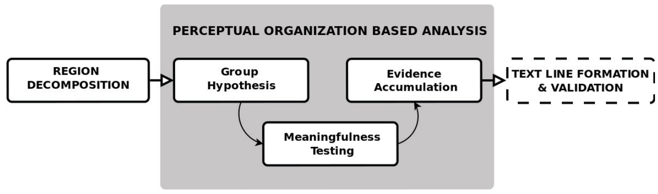


Fig. 2: Text Extraction algorithm pipeline.

Acknowledgements

The creation of this demo was supported by the Spanish projects TIN2011-24631 and 2010-CONE3-00029, the fellowship RYC-2009-05031, and the Catalan government scholarship 2013FI1126.

References

[1] L. Gomez and D. Karatzas, "Multi-script text extraction from natural scenes," in *Proceedings of the 12th Int. Conf. on Document Analysis and Recognition (ICDAR 2013)*, 2013. 1

[2] A. Desolneux, L. Moisan, and J.-M. Morel, "A grouping principle and four applications," *IEEE Trans. PAMI*, 2003. 1

[3] A. Fred and A. Jain, "Combining multiple clusterings using evidence accumulation," *IEEE Trans. PAMI*, 2005. 1

[1]Daniel Uzquiano's random stroke generator: http://danieluzquiano.com/491

Fast and Layout-Free Camera-Based Character Recognition on Complex Backgrounds

Masakazu Iwamura, Takuya Kobayashi, Takahiro Matsuda and Koichi Kise
Graduate School of Engineering, Osaka Prefecture University
{masa, kise}@cs.osakafu-u.ac.jp, {kobayashi, matsuda}@m.cs.osakafu-u.ac.jp

I. Overview of Demonstration

Recognizing characters in a scene is a challenging and unsolved problem. In this demonstration, we show an effective approach to cope with the problems: recognizing Japanese characters including complex characters such as Kanji (Chinese characters), which may not be aligned on a straight line and may be printed on a complex background.

In the demo, we address the problems above. Our demo is based on our recognition method [1], [2] and the method is applied to image sequences captured with a web camera. Figure 1 shows an overview of the recognition method. The recognition method is based on local features and their alignment. The idea is that if the local features locate in the query image in the same alignment as ones in a reference image, the character of the reference image should exist in the region of the query image. Using a tracking method, recognition results and extracted features are accumulated so as to increase recognition accuracy as time goes on. The demo runs about 1 fps on a standard laptop computer.

A result is shown in Fig. 2. The original query image is shown on the left and the corresponding recognition result is shown on the right. The recognition result was obtained by a capturing printed query image with a web camera and the recognition method is applied to the image. Red rectangles represent bounding boxes of recognized characters and recognition results were superimposed on them.

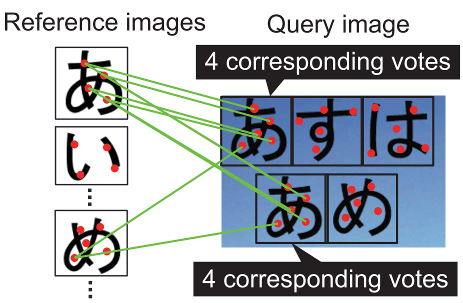
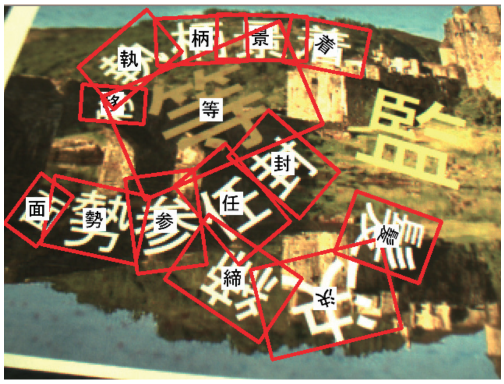


Fig. 1: An overview of the proposed method. Red points represent local features extracted and green lines do correspondences of features. Recognition results (characters and their bounding boxes) are determined at once based on correspondences of local features and their alignment.

(a) Query image.



(b) Recognition result.

Fig. 2: An example that was recognized by the demo system. The recognition result was obtained by capturing a part of the printed query image.

Acknowledgment

This work was supported in part by JST CREST project and JSPS KAKENHI Grant Number 25240028.

References

[1] M. Iwamura, T. Kobayashi, and K. Kise, "Recognition of multiple characters in a scene image using arrangement of local features," *Proc. ICDAR2011*, pp. 1409–1413, 2011.

[2] T. Kobayashi, M. Iwamura, and K. Kise, "An anytime algorithm for faster camera-based character recognition," *Proc. ICDAR2013*, 2013.