

Image Analysis for Historical Japanese Book Archives

Chulapong Panichkriangkrai, Graduate School of Science and Engineering, Ritsumeikan University, Kyoto, Japan

Liang Li, Ritsumeikan Global Innovation Research Organization, Ritsumeikan University, Kyoto, Japan

Ross Walker, College of Information Science and Engineering, Ritsumeikan University, Kyoto, Japan

Kozaburo Hachimura, College of Information Science and Engineering, Ritsumeikan University, Kyoto, Japan

ABSTRACT

*This paper describes methods of image analysis for historical Japanese book archives with a dominant focus on character segmentation. The segmentation methodology includes stain and smear removal, binarization, character line extraction, and character extraction by region labeling with integration and separation techniques. The experimental results show that the proposed method can segment all text lines correctly and can extract more than 79% of the characters from 16 pages of *Chinsetsu Yumiharizuki*, containing 176 text lines and a total of 5181 quite complicated characters.*

Keywords: Character, Extraction, Methodology, Region Labeling, Segmentation

INTRODUCTION

Paper free, easy access and distribution, mobile availability, and interactive multimedia functions make digital books a rapidly growing market. With internet distribution, the digital book market is already a global business. On the other hand, nowadays, there is increasing research on the strategies of developing and introducing national cultural heritages worldwide (Leong, 2010). A large number of historical books were digitized and made available not

only to researchers but to the public as digital media in many digital archives (Art Research Center, Ritsumeikan University, 2010; National Institute of Japanese Literature, 2013). However, only limited items have been transcribed into text based digital books, whereas most of the rest were collected as digital images without text information. It is a challenging work to develop learning/reading support system that analyses document images to discriminate figures, titles, and text area from digital archives. This article focused on these topics. Although our method

DOI: 10.4018/ijabim.2014040101

and results are still in a preliminary stage, the approach will be significant when expanding the target of digital book business so as to include historical books in the future.

The analyses of digital archived historical book images in this paper mainly focuses on page segmentation and identification. For example, text region extraction means identifying the text part of a page image, while text-line extraction means identifying the text-line from the text region. Furthermore, character extraction refers to segmenting each character from the text-line. In this paper we propose techniques for both text-line and character segmentation.

The historical Japanese books that are focused on in this paper are books printed in the Edo period (1603-1867). Figure 1 shows examples of Japanese woodblock printed books from the Edo period. The Edo period was a period of calm that provided an ideal environment for developing commercial art. During that period, while Europeans used moveable type printing processes, the Japanese developed and used a woodblock printing process. This process uses wooden blocks, upon which are engraved reversed images of both the text and

illustrations, as relief printing. For printing, two consecutive pages were carved on one side of the woodblock. During the Edo period, Japan published over 110,000 titles of books with more than 10 million copies in the markets (Hioki, 2009).

Currently, a large number of the books published in the Edo period have been scanned and made available to the public as digital images. However, experts have transcribed only a small number of book titles printed in the Edo period into modern book productions. Furthermore, only a small number of people can recognize and read the characters used in Edo period books. Old style characters and running scripts are different to modern ones. Characters of this type of historical book are difficult to segment because they have ligature-like characters that join two or more characters.

In this paper, we propose a character segmentation system for character shape comparison, character image retrieval, and to make a statistical analysis of the usage of characters in single or multiple books. The proposed concept can be applied for other Asian historical digital archives to offer cultural and social support.

Figure 1. Example of Japanese woodblock printed historical books



9 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product's webpage:

www.igi-global.com/article/image-analysis-for-historical-japanese-book-archives/114694?camid=4v1

This title is available in InfoSci-Journals, InfoSci-Journal Disciplines Business, Administration, and Management.

Recommend this product to your librarian:

www.igi-global.com/e-resources/library-recommendation/?id=2

Related Content

Burnout Among Manufacturing Workers in China: The Effects of Organizational Socialization and Leadership Behavior

Jennifer H. Gao (2013). *International Journal of Asian Business and Information Management* (pp. 35-45).

www.igi-global.com/article/burnout-among-manufacturing-workers-china/77827?camid=4v1a

Effect of Country-of-Origin and Ethnocentrism on Consumer Cognitive Processing in India

Vaibhav Arora, Jyoti Kainth and Lubna Nafees (2014). *International Journal of Asian Business and Information Management* (pp. 17-33).

www.igi-global.com/article/effect-of-country-of-origin-and-ethnocentrism-on-consumer-cognitive-processing-in-india/124299?camid=4v1a

Using Enterprise Architecture to Transform Service Delivery: The U.S. Federal Government's Human Resources Line of Business

Timothy Biggert (2011). *Global Business: Concepts, Methodologies, Tools and Applications* (pp. 311-343).

www.igi-global.com/chapter/using-enterprise-architecture-transform-service/54776?camid=4v1a

Exploring E-Marketing Practises as Tool for Globalisation by Egyptian SBEs

Hatem El-Gohary (2011). *Electronic Globalized Business and Sustainable Development Through IT Management: Strategies and Perspectives* (pp. 181-204).

www.igi-global.com/chapter/exploring-marketing-practises-tool-globalisation/45360?camid=4v1a