



**PRINTED THAI CHARACTER RECOGNITION  
USING TOPOGRAPHIC STRUCTURE**

**JITRLADA ROJRATANAVIJITR**  
Z

อภินันท์นาการ  
จาก  
บัณฑิตวิทยาลัย มหาวิทยาลัยมหิดล

**A THESIS SUBMITTED IN PARTIAL  
FULFILLMENT OF THE REQUIREMENTS FOR  
THE DEGREE OF MASTER OF SCIENCE  
(COMPUTER SCIENCE)  
FACULTY OF GRADUATE STUDIES  
MAHIDOL UNIVERSITY  
2002**

TH  
J61 ๒  
2002  
C.2

**ISBN 974-04-2082-6  
COPYRIGHT OF MAHIDOL UNIVERSITY**

4137572 SCCS/M : MAJOR : COMPUTER SCIENCE ; M.Sc.(COMPUTER SCIENCE)

KEY WORDS : THAI CHARACTER / TOPOGRAPHIC STRUCTURE  
JITRLADA ROJRATANAVIJITR: PRINTED THAI CHARACTER  
RECOGNITION USING TOPOGRAPHIC STRUCTURE. THESIS ADVISORS:  
SUKANYA PHONGSUPHAP, Ph.D., CHOMTIP PORNANOMCHAI, Ph.D.,  
115 p. ISBN 974-04-2082-6

Character recognition is an interesting topic for a number of researchers. Researches have been consistently conducted both in Thailand and foreign countries since the research in this field can be applied to various kinds of work. for instance, optical character recognition, license plate recognition and so on. Currently, car license plate recognition is increasingly becoming an interesting topic because it can be applied to such work as the systems of automatic parking as well as the entry and exit of a place, and the expressway's toll collection, etc.

In Thailand, it has been found that, although there have been researches on character recognition for a long time, most of them are on optical character recognition concentrating on solving the problem of a wide variety of character fonts appearing in documents. Research on car license plate recognition is rarely found. It can be stated that it is just the beginning of a study in this field. A significant problem found in the research on car license plate recognition is a wide variety of environments affecting characters on license plates, for example, an illumination condition which is too dark or bright, a shadow appearing on characters, etc.

This thesis proposed an approach to printed Thai character recognition using the topographic structure. It was aimed at analyzing printed Thai characters under various illumination conditions. In so doing, we employed two global features, clumpiness and miscibility, to select ten candidate characters. Then, the principle of a histogram projection was used for calculating local features. With these local features, the output character category was finally selected.

The proposed approach was tested with images of printed Thai single characters derived from an image scanner, including images of characters on car license plates taken under different environments by a digital camera. The testing consequences revealed that the proposed approach could be effectively used to recognize printed Thai characters under several illumination conditions. It was found that an average rate of accurate recognition of printed Thai single character images derived from a scanner was 99.52% while an average rate of correct recognition of images of characters on car license plates, taken under different environments, was 97.73%.

4137572 SCCS/M : สาขาวิชา : วิทยาการคอมพิวเตอร์; วท.ม. (วิทยาการคอมพิวเตอร์)

จิตรลดา โรจนรัตน์จิตร : การรู้จำตัวพิมพ์ภาษาไทยโดยใช้โครงสร้างเชิงภูมิลักษณะ (PRINTED THAI CHARACTER RECOGNITION USING TOPOGRAPHIC STRUCTURE). คณะกรรมการควบคุมวิทยานิพนธ์ : สุกัญญา พงษ์สุภาพ, Ph.D., ชมทิพ พรพนมชัย, Ph.D., 115 หน้า. ISBN 974-04-2082-6

การรู้จำตัวอักษรเป็นหัวข้อวิจัยที่ได้รับความสนใจ และมีการวิเคราะห์วิจัยกันมาอย่างต่อเนื่องทั้งในประเทศไทยและต่างประเทศ เนื่องจากสามารถนำงานวิจัยด้านนี้ไปประยุกต์ใช้ในด้านต่างๆ ได้มากมาย ยกตัวอย่างเช่น การจัดทำโปรแกรมแปลงเอกสารตัวอักษรเป็นแฟ้มข้อความ (Optical Character Recognition) และการนำมาใช้ในเรื่องการรู้จำแผ่นป้ายทะเบียนรถยนต์ (License Plate Recognition) เป็นต้น โดยเฉพาะอย่างยิ่งในปัจจุบันการรู้จำตัวแผ่นป้ายทะเบียนรถยนต์ ถือว่าเป็นหัวข้อที่เป็นที่นิยมอย่างมาก เพราะสามารถนำมาประยุกต์ใช้ได้กับงานหลายอย่าง เช่น การจัดทำระบบการจราจรอัตโนมัติ ระบบการผ่านเข้า-ออก สถานที่ต่างๆ และการจัดเก็บค่าทางด่วน เป็นต้น

ในประเทศไทยนั้น ถึงแม้ว่าจะมีการศึกษาและวิจัยเกี่ยวกับการรู้จำตัวอักษรมานานแล้วก็ตาม แต่งานวิจัยส่วนมากจะทำศึกษาการแปลงเอกสารตัวอักษรเป็นแฟ้มข้อความ โดยมุ่งเน้นที่จะแก้ไขปัญหาในเรื่องความหลากหลายของตัวอักษรที่ปรากฏอยู่บนเอกสาร ส่วนงานวิจัยทางด้านการรู้จำแผ่นป้ายทะเบียนรถยนต์นั้น ถือว่ายังมีอยู่น้อยมาก อาจกล่าวได้ว่าเป็นช่วงเริ่มแรกของการศึกษาในด้านนี้ โดยประเด็นปัญหาที่สำคัญของงานวิจัยด้านนี้คือ ความหลากหลายของสภาพแวดล้อมที่มีผลต่อตัวอักษรบนแผ่นป้ายทะเบียนรถยนต์ เช่น สภาวะแสงที่มีมืดหรือสว่างมากเกินไป และการเกิดแสงเงาบนตัวอักษร ฯลฯ

วิทยานิพนธ์นี้ได้เสนอวิธีการรู้จำตัวพิมพ์ภาษาไทยโดยใช้โครงสร้างเชิงภูมิลักษณะ เพื่อวิเคราะห์ตัวอักษรพิมพ์ไทยที่อยู่ภายใต้สภาวะแสงต่างๆ โดยใช้ Clumpiness และ Miscibility ซึ่งมีคุณสมบัติในการบ่งบอกรูปแบบของภาพ วิเคราะห์ตัวอักษรที่น่าจะเป็นตัวอักษรที่ต้องการจำนวน 10 ตัวอักษร จากนั้นจึงใช้หลักการสะสมความถี่ในทิศทางต่างๆ ของภาพ วิเคราะห์หาตัวอักษรที่ต้องการอีกครั้งหนึ่ง

การทดสอบกับภาพอักษรพิมพ์ไทยเดี่ยวที่ได้มาจากการใช้เครื่องสแกนเนอร์ และภาพตัวอักษรบนแผ่นป้ายทะเบียนรถยนต์ในสภาวะแวดล้อมต่างๆ ที่ได้มาจากการถ่ายภาพด้วยกล้องดิจิตอล ผลการทดสอบสามารถสรุปได้ว่า วิธีการที่นำเสนอสามารถรู้จำตัวอักษรพิมพ์ไทยในสภาวะแสงต่างๆ ได้เป็นที่น่าพอใจ โดยมีอัตราการรู้จำที่ถูกต้องเฉลี่ยของตัวอักษรพิมพ์ไทยเดี่ยวที่ได้มาจากการใช้เครื่องสแกนเนอร์ 99.52% และมีอัตราการรู้จำที่ถูกต้องเฉลี่ยของภาพตัวอักษร ที่ปรากฏบนแผ่นป้ายทะเบียนรถยนต์ในสภาวะแวดล้อมต่างๆ 97.73%