

**THAI HANDWRITTEN CHARACTER RECOGNITION SYSTEM  
(THW-CR)  
IMPROVING FEATURE EXTRACTION PROCESS BY  
THE ANALYSIS OF CONTOUR CHARACTERISTICS**

**URAIRAT LIMKONGLAP**

**A THESIS SUBMITTED IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR  
THE DEGREE OF MASTER OF SCIENCE  
(COMPUTER SCIENCE)  
FACULTY OF GRADUATE STUDIES  
MAHIDOL UNIVERSITY  
2006**

**COPYRIGHT OF MAHIDOL UNIVERSITY**

ระบบการรู้จำลายมือเขียนภาษาไทยโดยการปรับปรุงกระบวนการคัดแยกลักษณะบ่งชี้ด้วยการวิเคราะห์ลักษณะเฉพาะของโครงร่าง (THAI HANDWRITTEN CHARACTER RECOGNITION SYSTEM (THW-CR) IMPROVING FEATURE EXTRACTION PROCESS BY THE ANALYSIS OF CONTOUR CHARACTERISTICS)

อุไรรัตน์ ลิ้มกองกลาง

4437177

SCCS/M

วท.ม. (วิทยาการคอมพิวเตอร์)

คณะกรรมการควบคุมวิทยานิพนธ์ : เจริญศรี มีตรภานนท์, Ph.D., สุกัญญา พงศ์สุภาพ, Ph.D.

**บทคัดย่อ**

งานวิจัยนี้เป็นงานวิจัยต่อเนื่องในการพัฒนาระบบการรู้จำลายมือเขียนภาษาไทยจากการวิเคราะห์ลักษณะบ่งชี้ของตัวอักษร ซึ่งได้นำเสนอการปรับปรุงกระบวนการคัดแยกลักษณะบ่งชี้ (Feature Extraction Process) ด้วยการวิเคราะห์ลักษณะเฉพาะของโครงร่าง (The Contour Characteristics Analysis) โดยมุ่งเน้นที่จะเพิ่มความถูกต้องในการคัดแยกลักษณะบ่งชี้และการรู้จำตัวอักษร โครงสร้างของงานวิจัยแบ่งออกเป็น 4 กระบวนการ ประกอบด้วย การปรับปรุงเบื้องต้น การคัดแยกลักษณะบ่งชี้ การรู้จำ และการปรับปรุงขั้นตอนสุดท้าย อย่างไรก็ตามงานวิจัยนี้เน้นที่การปรับปรุงกระบวนการคัดแยกลักษณะบ่งชี้ประกอบด้วย 3 ส่วนหลักคือ Feature Extraction Algorithms, Feature Representation และ Feature Conflict Resolution Rules โดยการหาจุดบนโครงร่าง ซึ่งเป็นตัวแทนของจุดทั้งหมดในลักษณะบ่งชี้นั้น เพื่อนำมาเทียบกับฟังก์ชันต้นแบบที่เหมาะสมกับลักษณะบ่งชี้นั้นๆ และสามารถอ้างอิงถึงลักษณะบ่งชี้นั้นได้ในที่สุด ซึ่งจะช่วยให้การคัดแยกลักษณะบ่งชี้และการรู้จำมีความถูกต้องมากยิ่งขึ้น

ดังนั้นระบบต้นแบบ THW-CR จึงถูกสร้างขึ้นมา เพื่อวัดความถูกต้องของการคัดแยกลักษณะบ่งชี้และการรู้จำตามวิธีการที่นำเสนอข้างต้น ระบบต้นแบบได้ถูกนำไปทดสอบกับเอกสารที่ scan เข้ามา จากผู้เขียน 32 คน ซึ่งมีจำนวนตัวอักษรทั้งหมดประมาณ 103,000 ตัวอักษร การทดลองแบ่งเป็น 3 ส่วน คือทดสอบกับข้อมูลที่ได้มาจากงานวิจัยก่อนหน้านี้ ทดสอบกับลายมือเขียนที่มี feature ก่อนข้างสมบูรณ์ และทดสอบกับลายมือเขียนที่มี feature ไม่สมบูรณ์ จากผลการทดลองพบว่า อัตราความถูกต้องของการคัดแยกโดยเฉลี่ยประมาณ 95.97% และสูงถึง 97.73% สำหรับอัตราความถูกต้องของการรู้จำโดยเฉลี่ยประมาณ 90.76% และสูงถึง 95.35% นอกจากนี้อัตราความถูกต้องของการคัดแยกและการรู้จำเมื่อเทียบกับระบบต้นแบบของงานวิจัยขึ้นก่อนเพิ่มขึ้นเป็น 3.62% และ 8.33% ตามลำดับ

**THAI HANDWRITTEN CHARACTER RECOGNITION SYSTEM (THW-CR)  
IMPROVING FEATURE EXTRACTION PROCESS BY THE ANALYSIS OF  
CONTOUR CHARACTERISTICS**

URAIRAT LIMKONGLAP 4437177      SCCS/M

M.Sc. (COMPUTER SCIENCE)

THESIS ADVISORS: JARERNSRI L. MITRPANONT, Ph.D., SUKANYA  
PHONGSUPHAP, Ph.D.**ABSTRACT**

This research proposes an approach to the analysis of contour characteristics for the feature extraction process of an off-line Thai handwritten character recognition program (THW-CR). The program aims at increasing the accuracy of both feature extraction rate and character recognition rate. We introduce the contour to capture the movement of features. The start contour point is initiated from the location of the region. Then, it is chained along the feature shape until it meets the termination of the region. Subsequently, the contour points along the contour are retrieved and fit by the fit function which is compatible with the feature shape. The characteristic of the fit function is finally interpreted to feature type. The research framework consists of four major components: preprocessing, feature extraction process, recognition and post-processing. However, this work focuses on the part of feature extraction, which consists of three parts: feature extraction algorithms, feature representation and feature conflict resolution rules. Specifically, the feature representation is the main focus for the improvement in which the contour is established to represent a more detectable structure and accurate feature of the character. In addition, the feature conflict resolution rules are proposed to adjust the conflict features of either position or feature type compared with the possible structure.

For the implementation part, the prototype of the THW-CR was developed to validate our proposed solution. All proposed techniques were integrated into the HWR system, which was the prototype of the original work extended by this research. The 230 documents written by 32 people and containing 103,000 characters, approximately, were tested. The results were evaluated by two evaluation criteria: Feature Extraction Rate (FER) and Character Recognition Rate (CRR). The FER was tested to verify the accuracy of detected features of each character while the CRR was tested to measure the accuracy of recognized characters of the whole paragraphs or sentences. From the experiment consisting of three test scenarios of the classification of writing types, the average FER was 95.96% and the maximum FER reached 97.73%. Meanwhile, the average CRR was 90.76% and the highest CRR reached 95.35%. Additionally, the result of the overall evaluation showed that the THW-CR system generated reliable results for improving the accuracy of the feature extraction rate and the character recognition rate compared with the previous research by 3.62% and 8.33%, respectively.

**KEY WORDS: THAI HANDWRITTEN CHARACTER RECOGNITION /  
CONTOUR ANALYSIS / CONTOUR FITTING / CURVE FITTING**