

วิทยานิพนธ์ฉบับนี้มีวัตถุประสงค์เพื่อนำเสนอวิธีการรู้จำตัวอักษรภาษาไทยแบบใหม่ที่เรียกว่าการวิเคราะห์องค์ประกอบสำคัญแบบหลายประเภทสำหรับการรู้จำตัวอักษรภาษาไทย โดยวิธีการนี้มีแนวคิดพื้นฐานมาจากการวิเคราะห์องค์ประกอบสำคัญดั้งเดิม ซึ่งการวิเคราะห์องค์ประกอบสำคัญดั้งเดิมมีข้อดีที่สามารถลดปริมาณข้อมูลทำให้ข้อมูลที่ได้มีขนาดกะทัดรัดโดยใช้กระบวนการแปลงเชิงเส้นและการตัดลดคุณลักษณะที่ไม่สำคัญออก อย่างไรก็ตามการวิเคราะห์องค์ประกอบสำคัญดั้งเดิมนี้อย่างขาดประสิทธิภาพในการแบ่งแยกข้อมูลที่มีประเภทที่มีจำนวนมาก ดังเช่นตัวอักษรภาษาไทย ส่วนการวิเคราะห์องค์ประกอบสำคัญแบบหลายประเภทที่นำเสนอนี้มีประสิทธิภาพในการแบ่งแยกข้อมูลโดยการสร้างเซตขององค์ประกอบสำคัญโดยที่แต่ละเซตสร้างจากข้อมูลในแต่ละประเภท แต่การสร้างเซตขององค์ประกอบสำคัญหลายเซตนั้นมีจุดอ่อนตรงที่ต้องใช้ทรัพยากรและเวลาในการคำนวณมากเกินไปซึ่งเป็นการสิ้นเปลืองและเสียเวลาจึงต้องมีวิธีการในการลดจำนวนองค์ประกอบสำคัญให้น้อยลง ดังนั้นในวิทยานิพนธ์ฉบับนี้จึงนำเสนอวิธีการสำหรับการกำหนดจำนวนองค์ประกอบสำคัญที่เหมาะสมสำหรับข้อมูลแต่ละประเภทที่แตกต่างกัน 4 วิธีด้วยกัน โดยผลการทดลองแสดงให้เห็นว่าวิธีการที่นำเสนอนี้ให้ความถูกต้องในการรู้จำตัวอักษรที่สูงขึ้นกว่าวิธีการวิเคราะห์องค์ประกอบสำคัญดั้งเดิม

The objective of this thesis is to present a novel method, called multi-class principal components analysis (MCPCA), for Thai optical character recognition(Thai OCR). The method is based on the original principal components analysis (PCA) for Thai OCR. The original PCA reduces the original data to smaller size data. It has the advantage of compact representation of data by linear transformation and reduction of unimportant features. However, PCA lacks of discriminative power. The proposed MCPCA possesses strong discriminative power by constructing several sets of principal components, each for one class of data. Each set is the representative for the corresponding class of data. However, constructing several sets of principal components consumes lot of resources and computational time. We propose four methods for determining the appropriate number of principal components for each class. The experimental results show that our proposed methods provide higher accuracy than the original PCA for Thai OCR.