

อักขระตัวเขียนภาษาไทยเป็นตัวอักขระที่ยากในการรู้จำเนื่องจากมีขนาดความสูงและความยาวไม่แน่นอน นอกจากนี้ยังมีความซับซ้อนและยืดหยุ่นตามลักษณะการเขียนของผู้เขียนแต่ละคน ในการรู้จำจึงต้องอาศัยการสกัดลักษณะเฉพาะจากลักษณะเด่นหลาย ๆ อย่าง ทั้งนี้เพื่อให้สามารถอธิบายเอกลักษณ์ที่สำคัญของอักขระแต่ละตัวได้อย่างครบถ้วนมากที่สุด งานวิจัยนี้มุ่งเน้นศึกษาการรู้จำอักขระตัวเขียนภาษาไทยโดยอาศัยวิธีการสกัดลักษณะเฉพาะแบบออฟไลน์ โดยในการสกัดลักษณะเฉพาะที่สำคัญได้ใช้วิธีการปรับการกระจายของพิกเซลของตัวอักขระแบบไม่เชิงเส้นและคุณลักษณะเด่นของอักขระหลายวิธีร่วมกันโดยใช้ตัวแบบฮิดเดนมาร์คอฟชนิดต่าง ๆ เพื่อศึกษาเปรียบเทียบประสิทธิภาพ โดยมีการกำหนดจำนวนสถานะของตัวแบบในลักษณะที่แน่นอนและไม่แน่นอน ผลที่ได้จากการวิจัยพบว่าจำนวนสถานะที่เหมาะสมและน้อยที่สุดที่ให้ผลอัตราการรู้จำเฉลี่ยสูงสุดสำหรับตัวแบบชนิด Left-Right มีจำนวนสถานะแน่นอน 35 สถานะ โดยมีอัตราการรู้จำอักขระในกลุ่มพยัญชนะเท่ากับ 90.07% และอักขระในกลุ่มสระ วรรณยุกต์ และอักขระพิเศษเท่ากับ 94.09% ในขณะที่ผลการทดสอบกับตัวแบบ 5 ชนิด ที่มีจำนวนสถานะไม่แน่นอน ให้ผลลัพธ์ในการรู้จำตัวอักขระกลุ่มพยัญชนะเฉลี่ยที่ดีที่สุดคือ 91.67% โดยใช้ตัวแบบ Left-Right Allow Skip State แบบที่ 2 ขณะที่ในกลุ่มตัวอักขระ สระ วรรณยุกต์ และอักขระพิเศษ ให้ผลลัพธ์โดยเฉลี่ยดีที่สุด คือ 96 % โดยใช้ตัวแบบชนิด Left-Right-Left Allow Skip State

Thai handwritten characters exhibit many different characteristics and complexity ranging from size, height, width to variations on individual's dependent handwritten style. It is a very difficult task to build up an acceptable algorithm for recognizing them. In this regard, promising handwritten features play an important role for recognition capability. A vector of such features can be feed to the recognition algorithm, resulting in high recognition performance. In this research, we are specific to investigate Thai handwritten character recognition by using offline feature extraction algorithm. The research takes a character pixel distribution in nonlinear in corporation with a number of distinctive patterns of handwritten characters to determine features for the recognition algorithm which utilizes Hidden Markov Models (HMM). For comparison of recognition performance purpose, a number of HMM topologies is compared. The research shows that, when test on equal state policy, HMM with left-right topology with 35 states gives the best recognition rate. It results on average 90.07% for consonant, and 94.09% for tone, consonant, and vowel. When test on unequal state policy among five different topologies, HMM with type-2 left-right allow skip state topology gives the best recognition on average at 91.67% for consonant, while HMM with left-right-left allow skip state topology provides 96% for tone, consonant and vowel.