

T158863

เครื่องช่วยงานส่วนบุคคลแบบดิจิทัล หรือพีดีเอเริ่มมีบทบาทในชีวิตประจำวันและกำลังเป็นที่ได้รับความนิยมอย่างมาก การเขียนตัวอักษรที่เรียกว่ากราฟฟิติเป็นวิธีการบันทึกข้อมูลรูปแบบหนึ่งในการป้อนข้อมูลเข้าสู่เครื่องพีดีเอ ปัญหาอย่างหนึ่งในการใช้ตัวกราฟฟิติภาษาไทยคือปริมาณของตัวอักษรที่ผู้ใช้จะต้องเรียนรู้เพื่อใช้การทำงานมีจำนวนมาก

งานวิจัยชิ้นนี้จะพยายามพัฒนาวิธีการรู้จำตัวอักษรที่มีรูปแบบการเขียนแบบปกติผ่านทางเครื่องพีดีเอ โดยได้นำวิธีการเปรียบเทียบแบบไดนามิกโปรแกรมมิ่งมาใช้ในการรู้จำตัวอักษรที่ผู้ใช้เขียน โดยในงานวิจัยจะแบ่งการทำงานเป็น 2 ขั้นตอนใหญ่ ๆ คือ ขั้นตอนของการจัดทำตัวอักษรต้นแบบ และขั้นตอนของการเปรียบเทียบเพื่อทำการรู้จำตัวอักษร ในขั้นตอนของการจัดทำตัวอักษรต้นแบบจะเริ่มจากนำตัวอักษรที่เขียนไปผ่านประมวลผลขั้นแรกซึ่งประกอบด้วย การกำจัดสัญญาณรบกวนออกจากข้อมูล การปรับขนาดของตัวอักษรให้มีขนาดปกติ การสุ่มจุดบนเส้นใหม่ และการหาลักษณะเด่นของตัวอักษร หลังจากนั้นข้อมูลลักษณะเด่นของตัวอักษรจะถูกนำไปจัดเก็บเพื่อใช้ในการเปรียบเทียบต่อไป สำหรับขั้นตอนของการเปรียบเทียบเพื่อทำการรู้จำตัวอักษรนั้น จะเริ่มต้นจากการรับตัวอักษรเข้ามาด้วยวิธีการเขียนและนำไปประมวลผลขั้นแรกเช่นเดียวกันกับในขั้นตอนของการจัดทำตัวอักษรต้นแบบ หลังจากนั้นจะนำลักษณะเด่นที่ได้ไปจัดกลุ่มเพื่อลดจำนวนครั้งของการเปรียบเทียบลง โดยการนำลักษณะเด่นของตัวอักษรภาษาไทย เช่น จำนวนของการยกปากกาขึ้นในการเขียนตัวอักษรรดับของอักษร มาใช้คัดกรองตัวอักษรที่จะนำมาเปรียบเทียบว่าควรจะนำมาเปรียบเทียบกันหรือไม่ หลังจากนั้นตัวอักษรจะถูกนำไปเปรียบเทียบกับตัวอักษรต้นแบบที่อยู่กลุ่มเดียวกันเพื่อหาตัวอักษรต้นแบบที่มีความแตกต่างของลักษณะเด่นสะสมน้อยที่สุด

หลังจากที่ได้ผลการรู้จำแล้ว ตัวอักษรบางประเภทที่มีลักษณะการเขียนที่คล้ายคลึงกันจะถูกนำมาเปรียบเทียบซ้ำอีกครั้งหนึ่ง อักษรที่จะถูกนำมาเปรียบเทียบซ้ำประกอบด้วย กลุ่มของอักษร ข และ บ ซึ่งจะแยกความแตกต่างได้โดยสัดส่วนของความกว้างต่อความสูงของตัวอักษร กลุ่มของอักษร ช และ บ และ ป, ผ และ ฝ และ พ และ ฟ ซึ่งจะแยกความแตกต่างได้โดยสัดส่วนของความสูงทางด้านซ้ายและขวาของอักษร และกลุ่มของอักษร ถ และ ฏ, ก และ ฏ ซึ่งจะแยกความแตกต่างได้โดยสัดส่วนความสูงทางด้านซ้ายและขวาของอักษรเช่นเดียวกัน

จากการทดลองพบว่า วิธีการที่นำเสนอให้ผลความแม่นยำในการรู้จำตัวอักษรโดดภาษาไทยที่จัดเก็บด้วยเครื่องพีดีเอ 98.19 เปอร์เซ็นต์ โดยมีความเร็วในการรู้จำเฉลี่ยเป็น 4.85 วินาทีต่อ 1 อักษร เวลาโดยเฉลี่ยสำหรับอักษรที่เป็นพยัญชนะคือ 6.37 วินาที และเวลาเฉลี่ยสำหรับอักษรที่ไม่ใช่พยัญชนะคือ 2.17 วินาที และผลการทดลองการรู้จำตัวอักษรโดดโดยใช้ข้อมูลทดสอบที่จัดเตรียมโดยเนคเทคให้ความแม่นยำในการรู้จำอักษร 94.85 เปอร์เซ็นต์

TE 158863

Personal Digital Assistant or PDA becomes more popular and important to daily lifestyle. Graffiti writing is currently a method to enter handwritten text into PDA. However, the major problem on writing Thai Graffiti is its large amount of Graffiti letters which user need to remember.

This research objective is to develop a new method to recognize Thai handwriting character via PDA, utilizing Dynamic Programming Matching. The recognition method can be divided into 2 major phases: Model making process and Recognition process. Model making process begins with preprocessing the input character. This process can be separated into 4 steps: Noise Removal, Size Normalization, Resampling and Feature Extraction. Eventually, the extracted features will be serialized to the feature dictionary for future recognition. Recognition process starts with preprocessing, similar to that of model making process. The character will be classified to reduce the iteration of matching, using Thai character properties, such as the number of Pen-Up-State strokes and position etc to identify the sets of characters that are possibly matched with the input character. Finally, the input character will be compared to each character in the set to find the best matched character.

If the best matched character turns out to be one of similar characters, the input characters need to be reexamined against the similar characters. For instance, group of character ๗ and ๘ can be distinguished by the proportion of its width to its height, ๗ and ๘ and ๙, ๑ and ๒, ๓ and ๔, ๕ and ๖ can be identified by the proportion of its left-part height to its right-part height and ๗ and ๘, ๙ and ๑ can be also distinguished by the proportion of its left-part height to its right-part height.

The experiment shows that the algorithm achieved the character recognition accuracy of 98.19%, for Thai characters input via PDA. The average time for recognition was 4.85 sec/character. The average time for Thai alphabet recognition was 6.37 sec/character while average time for Thai vowel recognition was 2.17 sec/character. In addition, conducted on testing data suite compiled by NECTEC gave the recognition accuracy of 94.85%.