Novel Curve Signatures and a Combination Method for Thai On-Line Handwriting Character Recognition

Ekawat Chaowicharat[1], Nick Cercone[2], Kanlaya Naruedomkul[3]

ekawatchaow@gmail.com, ncercone@yorku.ca, scknr@mahidol.ac.th

[1,3] Department of Mathematics, Faculty of Science, Mahidol University, Bangkok, Thailand

[2] Department of Computer Science and Engineering, Faculty of Science and Engineering, York University, Toronto, Canada

Abstract

There is no commercial character recognition software that supports Thai handwriting. Thai handwritten character recognition is needed to convert handwritten text written on mobile and tablet devices into computer encoded text. We propose a novel method that joins three curve signatures. The first signature is the normalized tangent angle function (TAF), which provides rough classification. The other two novel curve signatures are the relative position matrix (RPM), which is used to compare global curve features, and the straightened tangent angle function (STAF), which is used to compare the tangent angle along the cumulative unsigned curvature domain. In the recognition process, an input curve is extracted for these three signatures and the similarity against each character in the handwriting templates is measured. Then, the similarity scores are weighted and summed for ranking. Our experiment is done on 48 handwriting sample sets (44 Thai consonants appear in each set, and there are 4 sets per handwriting). Our methods yield an accuracy of 94.08% for personal handwriting, and 92.23% for general handwriting.

1. Introduction

The use of tablet PCs is widespread and operating one is understood easily. However, the language input method for most tablet PCs is a virtual keyboard; a set of virtual on-screen buttons that allow typing by touching. Operating with virtual keyboard is not easy because there is no actual button.

There is a solution to make the input method more convenient. On-line optical character recognition (OCR) is applied to acquire machine-understandable text from our handwriting on the tablet screen. For example, WritePad (2012) and PhatPad (2012) are on-line OCR software systems for the iPad and android tablets. Unfortunately, this software still does not support Thai.

Character recognition also extends the ability of the visually impaired to be able to read any document. We can construct a reading machine by connecting a camera/scanner to an OCR and text-to-speech (TTS) software. A visually impaired person who wishes to read a book without braille can simply scan the book, and then OCR and TTS will provide the voice reading the scanned text. OCR for a reading machine is "off-line" and processes written text without pen movement data. An off-line OCR can be built from an on-line OCR by adding a writing trace extraction process.

In this research, we propose two novel curve signatures, namely, the straightened tangent angle function (STAF) and the relative position matrix (RPM) and their similarity measurement. These signatures are amalgamated with the classic curve signature, namely, the tangent angle function (TAF) to create a novel online handwritten Thai character recognizer. Our experiment is to find the best configuration of each method and to set the most efficient collaboration of all these methods in order to construct a reliable Thai handwritten OCR.

The relevant issues about Thai character recognition and curve signatures are summarized in section 2. In section 3, our novel curve signatures, i.e., STAF and RPM are proposed. Implementation and evaluation can be found in section 4 and 5 respectively.

2. Relevant Issues

2.1 Difficulty of Thai Character Recognition

The Thai character set consists of 44 consonants, 21 vowel characters that can be used directly or can be combined into 11 additional compound vowels (i.e., 32 vowels in total) and 4 tone marks. Table 2.1 shows Thai consonants, vowel and tone marks. In this research, we are interested in recognition of 44 consonants only to avoid the identical curve shape between some consonant and vowel.

One of the main difficulties in handwritten Thai character recognition is character similarity. Different from other languages such as English, in which character identities are distinct, Thai characters look similar because the distinctive features are relatively small compared to the character size, e.g., ก ภ ถ ฎ ฏ share beak feature, ข ฃ ข ช ซ share head and body features.

In handwriting characters, small differences among characters in the same group can be made unclear. A common challenging handwriting style is the lack of a complete head, or even lack of head. When the head feature is flawed, the detection of head existence, orientation and style cannot discriminate from similar characters in some groups shown above. Similarly, the notch on the bottom and broken head can be too shallow and cause confusion with the regular bottom and head.

The most difficult situation in handwritten character recognition is the case when a character is distorted and degenerates into another character. For example, the intended character "ข" becomes "บ" when it was written with a regular head and the bottom is too wide. In this research, the term handwriting does not include the degenerate case. The distinctive features of handwritten characters must be visible, at least partially.

Consonants	ก	ข	ฃ	ค	ฅ	ฆ	ง	จ	ฉ
	ช	ซ	ฌ	ญ	ฎ	ฏ	ฐ	ฑ	ฒ
	ณ	ด	ต	ถ	ท	ธ	น	บ	ป
	ผ	ฝ	พ	ฟ	ภ	ม	ย	ร	ล
	ว	ศ	ษ	ส	ห	ฬ	อ	ฮ	
Vowels	-ะ	อั	อ็	-า	อิ	อี	อึ	อื	อุ
	อู	เ-	แ-	ไ-	โ-	-อ	-ย	-ว	ฤ
	ฤๅ	ฦ	ฦๅ	Tone		อ่	อ้	อ๊	อ๋

Table 2.1: Thai consonants and vowels set

2.2 Tangent Angle and Curvature

Most Thai characters are written in a single stroke. These characters differ in stroke patterns. A writing stroke can be represented in the form $C(t) = (x(t), y(t))$ where x(t) and y(t) are horizontal and vertical coordinate of at t. This parametric representation is derived into a 1-dimensional function for similarity measurement convenience.

A signature S applied to a curve C is invariant under a transformation T if the signature function still remains the same no matter the transformation is applied on the curve C or not, i.e., $S(TC) = S(C)$. In pattern recognition, the important invariant properties are translation, rotation and scaling invariance.

The tangent angle function $(\theta(s))$ represents the direction of the tangent vector of a curve as a function of arc length. It is used in a number of similarity measurements such as the work of Li and Jiang (2007) and Yu and Guo (2008). The tangent angle function of a curve $C(t)$ can be derived mathematically by $\theta(s) = \theta(t) \circ t(s)$ where,

$$\theta(t) = \begin{cases} \frac{\pi}{2}, & x'(t) = 0 \\ \arctan\left(\frac{y'(t)}{x'(t)}\right), & otherwise \end{cases}$$

$$s(t) = \int_0^t \sqrt{\left(x'(t)\right)^2 + \left(y'(t)\right)^2}\, dt$$

$$t(s) = s^{-1}(t)$$

Curvature was mentioned frequently in pattern recognition as a signature that represents the "bending" of a geometric curve. Mathematically, the curvature $\kappa(s)$ is the derivative of unit tangent vector of a curve (or the second derivative of the given space curve) (Shih, 2010).

Given a parametric curve $C(t) = (x(t), y(t))$, the curvature as a function of arc-length ($\kappa(s)$) is derived by $\kappa(s) = \frac{d\theta}{ds} = \frac{d\theta}{dt}\frac{dt}{ds}$ where $\frac{ds}{dt} = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}$ and $\arctan\theta = \frac{y'}{x'}$.

A benefit of curvature over the tangent angle function is that curvature is rotation invariant. However, tangent angle function is more sensitive to the change in stroke pattern than the curvature, especially when comparing two curves that differ in just finite points of curvature.

2.3 Thai Online OCR in the Past and Relevant Curve Signatures

Thongkamwitoon et al (2002) proposed a bilingual online OCR based on distinctive feature classification. There are three levels of classification starting from rough to fine. Most of the features are based on the number of loops, the number of junction points, ripple, width, height, and ratio of these features. The authors claim 86.34% and 95.42% accuracy for Thai and English respectively.

Budsayaplakorn et al (2003) proposed a fuzzy feature representation to emphasize the existence of distinctive character features before applying the Hidden Markov Model (HMM). This method achieve 91.2% accuracy.

Bounnady et al introduced (2008) a multiple representation of writing stroke using a binary tree where binary branching represents the uncertainty of each unclear inflection point. The recognition is performed by choosing the smallest Euclidean distance between each path in the binary tree of the unknown and templates. The authors claim the accuracy of 97.5%, but the experiment was done on only 37 single stroke characters.

Karnchanapusakij et al. (2009) applied a simple curve signature to handwriting recognition. The signature consists of a rotation angle and direction of the 10 sample points of a curve. The error metric is measured by weighted sum of the mean absolute error of rotation angle and rotation direction. The accuracy of this method was 90.88%. Jamjuntr and Dejdumrong (2011) improved the curve signature to be more effective by converting a curve into a matrix storing the relative position (up, down, left, right or identical) of each sample point pair. Then the number of identical relative positions is counted and returned as the similarity. This improved method gave 87.89% and 94.93% accuracy for general and personal handwriting respectively.

M. Cui et al (2009) proposed a novel curve matching metric by using a scale invariant signature constructed by curvature (κ) as a function of integral of unsigned curvature (K). This curve signature is a composition of regular curvature as a function of arc length and the inverse of cumulative absolute value of curvature. The matching is performed by sliding the unknown curve along the domain of K and finds the maximum correlation.

Our novel curve signature RPM and STAF are inspired by Bounnady's, Jamjuntr's and Cui's works. The detail of our curve signature can be found in the next section.

3. Novel Curve Signatures

We introduce two novel curve signatures and analyze their strength and weakness for Thai OCR use. The first signature is called the Standard Tangent Angle Function (STAF). The second signature, called the Relative Position Matrix (RPM), is used for handling noisy curves.

3.1 Straightened Tangent Angle Function (STAF)

The difference in local character feature size is an obstacle when comparing an arc-length-based function like TAF and curvature because the analogous feature might not be at the same position on the arc length domain. By changing the TAF domain from arc length into the cumulative unsigned curvature presented by Cui et al. (2009), we get a novel signature called the straightened tangent angle function (STAF) that could solve for local feature size.

Mathematically, STAF is the composition of TAF and inverse of cumulative unsigned curvature function. It is similar to curvature segment multiple representations proposed by Khampheth et al. (2008) except that STAF is uniquely represented as a function instead of a tree. Each line segment in a STAF represents a sub-stroke in the original curve segmented by inflection points. The derivation of STAF is as shown below.

Given a parametric curve $C(t) = (x(t), y(t))$ where t is a parametric variable. Let $\theta(s)$ be the tangent angle function of C as a function of arc length (s). The cumulative unsigned curvature $K(s)$ is defined as follows.

$$K(s) = \int_0^s \left|\frac{d\theta}{ds}\right| ds$$

STAF $\theta(K)$ can be found by composition of $\theta(s)$ and $s(K)$, where $s(K) = K^{-1}(s)$.

Each segment in a STAF graph represent the orientation and total bending of an arc from the original curve, so STAF is totally independent to scale. This property is useful for Thai character recognition because handwritten Thai characters can vary in local feature sizes.

Although STAF has some distinctive benefit, it also has some limitations. First, the STAF of a line segment is shrunk to a point. Since the domain of STAF is the integral of unsigned curvature, the curvature of a straight line is 0 so the integral does not change. Any line segments embedded into the original curve cannot be seen from the STAF graph. Another weakness of STAF is that it is sensitive to noise. Adding noise into the original curve will result in the presence of a saw tooth in the STAF graph. These weaknesses can be overcome by joining STAF with other curve signatures. This reasoning leads to another novel curve signature, the relative position matrix.

3.2 Relative Position Matrix (RPM)

Two curves with identical STAF can look different because one STAF represents a group of curves possessing the same $\theta(K)$. In order to discriminate each curve pattern in a group, we consider the "relative position" of each point pair in a curve. The idea of the relative position matrix (RPM) comes from map memorizing. Suppose we are standing at a point in a city. By looking around and memorizing the position of all buildings around us, and doing this repeatedly along the walking trail, we can sketch the trail and all buildings in the entire map.

RPM can be constructed by resampling a given curve along the arc length into N points, say $p_1, p_2, \ldots, p_N$. Then we measure the angle between each point pair with respect to x-axis, i.e., from p_1 to p_2, p_1 to p_3, …, p_1 to p_N until reaching $p_{N-1}$to p_N and put all of these angles into a matrix. Each row in an RPM describes the direction to go to the other sample points in the curve.

The benefit of including just the angle (excluding distance) is that the RPM becomes scale invariant. An advantage of RPM over the STAF is that the relative position information can distinguish the difference in straight line and other global structures of the curve because each row of the RPM contains information derived from the entire curve. Moreover, noise does not affect the value of the RPM significantly because the small curve ripple caused by noise can change only small angle values.

Both STAF and RPM functions are angle-valued functions. The similarity measurement between two curves can be done by counting the number of sample points that contain the identical angle value divided by the total number of sample points. The identical angle obtains a one score and the 180 degree difference obtains a zero score. For RPM, the angle $\Delta\theta$ between 0 and 180 degree is mapped by $\cos(\frac{\Delta\theta}{2})$ into a score between 0 and 1. The RPM similarity measurement can be defined as follows.

$$Similarity_{RPM}(C_1, C_2) = \frac{\sum_{i=1\ldots N} \cos\left(\frac{\Delta\theta_i}{2}\right)}{N}$$

For STAF, the sample points of those two curves can be unequal. The number of identical angle sample points will be divided by the total sample point number of each curve and then these proportions are averaged by using the harmonic mean. From our experiment, the best way to map $\Delta\theta$ to similarity score is to use binary threshold, i.e. $sim(\Delta\theta) = 1$ if $\Delta\theta < 30$, otherwise $sim(\Delta\theta) = 0$.

$Similarity_{STAF}(C_1, C_2) = \frac{2Sim_{12} \times Sim_{21}}{Sim_{12} + Sim_{21}}$, where
$Sim_{12} = \frac{\sum_{i=1 \ldots N} sim(\Delta\theta_i)}{N_1}$ and $Sim_{21} = \frac{\sum_{i=1 \ldots N} sim(\Delta\theta_i)}{N_2}$

Sim_{12} and Sim_{21}describe how similar in angle is the curve C_1 to C_2 when restricted to $C_1's$ domain. The harmonic mean of Sim_{12} and $Sim_{21}$is exploited to ensure the symmetry property of the metric. The best alignment of STAF is found by *shifting* the STAF of a curve along that of the other. STAF similarity is measured at every shifting step. The max argument of STAF similarity is returned.

In the next section, we will present the implementation of the whole recognition system.

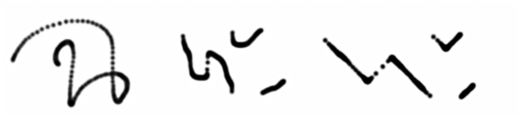

Figure 3.1: Example of writing trace (left), TAF (center) and STAF (right) of character ฉ.

4. Implementations

Three selected curve signatures have different recognition capability. The TAF representation is for point-wise comparison along the arc length domain, as is the STAF along the cumulative curvature domain, while the RPM is good at overall structure comparison. The set of correctly recognized characters by using each signature is different. Our intent is to find the most appropriate mixture of all three signatures to maximize the recognition accuracy.

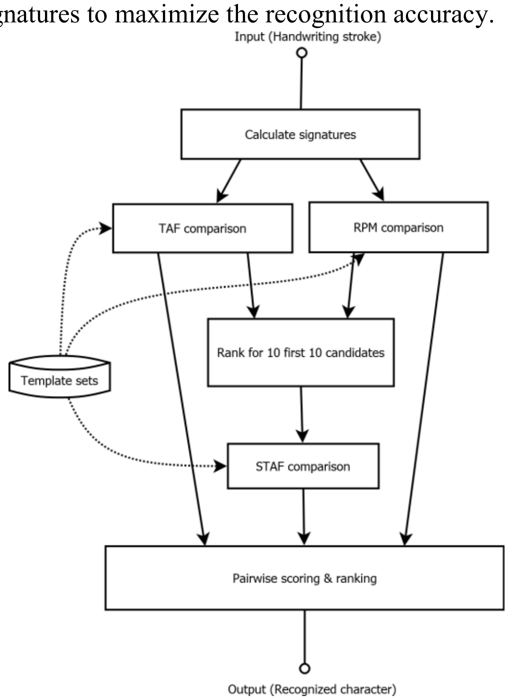

Figure 4.1: Overall recognition process

4.1 Joining of three curve signatures

Tangent angle function is chosen as one of three curve signatures in our research because of its simplicity and efficiency. The TAF can match two normalized curves accurately if no deformation occurs. The benefit of TAF over STAF and RPM is that TAF is the least computationally intensive signature. The TAF similarity can be calculated within N loops of angle comparison for N sample points (STAF takes MN loops where M is the number of shifting steps, and RPM takes $\frac{N^2-N}{2}$ loops).

The overall processes are as shown in the figure 4.1. The recognition begins by reading all template sets into memory. An input curve is acquired from a pen tablet and is smoothed. Then all curve signatures of the input curve are calculated. We set the TAF to work as a rough classification. An input curve will get the TAF similarity measured to all characters in every template set. Candidates corresponding to the first few TAF score ranking will be selected.

The same set of templates will get RPM similarity measured against the input. Then some candidates will be selected from the ranking of the summed score of the TAF and the RPM. From our experiment methods, TAF+RPM correct candidates never go beyond 10^{th} rank, so the number of selected candidate is set to 10.

The STAF comparison is performed on one these 10 candidates to reduce computation. At the end of similarity measurement, we have 10 candidates with three similarity scores ready for overall ranking.

4.2 Overall ranking

The selected candidates that pass the second round will have all three similarity scores from three signatures. The last process is to sum these scores and rank for the fittest candidate. Our ranking scheme follows the "round-robin tournament" rule. Each candidate pair must compete and the score summed as a match is done. The detail of ranking scheme is described below.

A competition match between two candidates calculates the weighted sum of three similarity scores against each other. Since each character can be recognized well with different score attributes, the weight must be specific to each character. The weight vector for each candidate pair is trained independently by increasing the weight of the attribute that can be used for specifically discriminating these character pairs. The product of similarity score vector of each candidate and the pair-specific weight vector will be recorded as the score of the match. Each candidate computes the sum of competition score from every match before the overall ranking is computed. The most likely character is the character corresponding to the maximum total weighted sum score.

For demonstration, suppose the correct character is ถ and the top 4 character candidates are ก ถ ภ and ค. The pairwise weight vector provides the "bias rule" for each character pair. For example, let a weight vector $W_{\text{กถ}}$ = (w_{taf}, w_{rpm}, w_{staf}) is trained for discriminating ก from ถ and $W_{\text{ถก}}$

for reversion. $S_ก$ and $S_ถ$ are score vector of the input for being similar to ก and ถ respectively. If the input is actually ถ (or similar to ถ rather than ก), the product $W_{ถก} \circ S_ถ$ is greater than $W_{กถ} \circ S_ก$.

Every combination of characters will compete, i.e., ก-ถ, ก-ค, ก-ภ, ค-ถ, ค-ภ and ถ-ภ are all six rounds in the tournament. The correct character (ถ) is expected to gain the highest score among all candidates. Then ถ is returned as the recognized output.

5. Training and evaluation

To obtain an optimal set of weight vectors, we require training with a corpus of labeled handwritten characters. In this section, we will describe the dataset collection, training and evaluation method.

We are interested in two applications of OCR; personal usage on a mobile device, and general handwriting recognition for document reader development. The former must be trained by a few sets of the owner's handwriting, but the handwriting style is quite stable. The latter should be trained by utilizing more handwriting sets from a number of writers. The second method is expected to handle higher handwriting variation than the individual usage.

5.1 Dataset collection

Our training and testing sets are collected from 48 Thai participants by using the Wacom Bamboo Pen tablet and a handwriting canvas program. The handwriting canvas records each Thai character as a sequence of coordinates. Two strokes character is concatenated and recorded as a single stroke. We collected 4 handwriting sets from each participant. Each character is stored in a text file. The example of collected dataset is as shown in figure 5.1.

Personal training is performed by using three sets as templates and the remaining set is used for accuracy measurements. For general handwriting training, a handwriting set of each participant is randomly selected, so we have 48 template sets in total for training.

Figure 5.1: Example of collected dataset

5.2 Self-training

Three score attributes are calculated from three signature similarity measurements. These scores must be weight summed into a final score used for ranking. We want to find the most appropriate weight vector for every character. The weight vector must be trained so that it can amplify the score difference between each character pair. The self-training method starts by selecting one set as a testing set. All remaining sets are used as templates. After running recognition for each testing character, the weight vector will be modified when misrecognition occurs by increasing the value of the attribute so that the incorrect character is defeated. Each hand-writing set will be chosen as a testing set and trained repeatedly until the weight vector is stable.

For example, if we are to train the weight vector of character ภ against ก, the weight attribute to update depends on the comparison of the similarity score attributes. Suppose that the TAF, RPM and STAF of ภ and ก are as shown below.

	TAF	RPM	STAF
ภ	0.78	0.87	0.92
ก	0.83	0.69	0.75

Table 5.1 Example of TAF, RPM and STAF score of character ภ and ก during the weight training

From table 5.1, RPM and STAF of ภ is greater than ก, so we promote this difference by increasing the weight attribute of RPM and STAF of ภ against ก. The weight updating rate of those attributes is increased by the score difference.

5.3 Evaluation

From our experiment, the signature combination method gains 94.08% of first accuracy for personal handwriting and 92.23% for general handwriting. The average running time is 68 milliseconds and 365 milliseconds for personal and general handwriting respectively on Lenovo ThinkPad X220i with Intel Core i3 2.30 GHz and 4 GB RAM. The recognition accuracy is as shown in table 5.2.

6. Experimental Results

From table 5.2, there are nine characters that have the 1st accuracy below 90%: ซ บ ฎ ร ข ษ ธ ป and ฝ. Each of these characters is similar to another character by default. In some handwriting, the discrimination is more difficult because these characters are written unclearly.

For the most part, the general handwriting accuracy is significantly lower than the personal testing, because a character in a handwriting set can be similar to another character in another handwriting set rather than the correct character in the same set. For some characters such as ข, ซ and พ (personal accuracy < 80%), each of them has its similar twin, i.e., ข-บ, ซ-ช, พ-ฟ so the chance that it is misrecognized as its twin is higher than other characters.

For some character such that the general accuracy is higher than personal accuracy, i.e. ฃ, ฒ, ง, ฉ, ฎ, ฏ, ฑ, น, ป, ผ,

อ, ฮ, there are two possible reasons. The first reason is that most of these characters are used not so often (ฃ ฒ ฉ ฎ ฏ ฑ ฮ) so the participants wrote them “more carefully” in the way that expresses the character features clearer than the more often characters. Another reason is that these characters have not so much variation in curve signatures as the others characters, hence the more templates being used, the more accuracy it gains.

7. Conclusion

7.1 Summary

We proposed a novel online Thai OCR comprising three curve signatures, i.e., TAF, RPM and STAF. An input character stroke is recognized by calculating these three signature similarity and summing up by using character pair specific weight. The weight must be trained by running the recognition on the labeled datasets and use the incorrect result to adjust weights iteratively.

Two recognition schemes were conducted. The first one called personal scheme, uses three template sets from each writer to train, gives 94.08% accuracy. Another scheme called general scheme, uses 48 template sets from different writers to train, gives 92.23% accuracy.

7.2 Further research

In this research, the recognition is performed individually on each input character. We can improve the accuracy by applying context information (such as N-gram analysis) to correct the misrecognition of similar characters. Our current research on OCR error correction using N-gram and Hidden Markov Model looks promising.

Ch	Psnl	Gnrl	Ch	Psnl	Gnrl	Ch	Psnl	Gnrl
ก	97.92	97.92	ฐ	100	95.83	ฟ	93.75	83.33
ข	87.5	70.83	ฑ	93.75	97.92	ภ	93.75	93.75
ฃ	97.92	100	ฒ	100	93.75	ม	95.83	93.75
ค	95.83	95.83	ณ	97.92	91.67	ย	97.92	97.92
ฅ	97.92	95.83	ด	97.92	91.67	ร	89.58	83.33
ฆ	89.58	93.75	ต	95.83	93.75	ล	93.75	87.5
ง	91.67	93.75	ถ	93.75	93.75	ว	95.83	91.67
จ	89.58	83.33	ท	95.83	93.75	ศ	97.92	97.92
ฉ	97.92	100	ธ	93.75	93.75	ษ	95.83	95.83
ช	87.5	83.33	น	93.75	97.92	ส	97.92	97.92
ซ	81.25	77.08	บ	85.42	83.33	ห	97.92	95.83
ฌ	100	95.83	ป	89.58	95.83	ฬ	97.92	93.75
ญ	95.83	93.75	ผ	89.58	91.67	อ	91.67	93.75
ฎ	91.67	95.83	ฝ	93.75	91.67	ฮ	95.83	97.92
ฏ	89.58	95.83	พ	91.67	75	Avg	94.08	92.23

Table 5.2 Recognition accuracy for personal and general handwriting recognition.

We can create an offline OCR from an online system by adding writing trace extraction to provide coordinate sequences of character curves from a scanned document to our system. Our goal is to assemble an offline OCR and text-to-speech system to create an automatic Thai book reading machine for visually impaired people.

Since the three curve signature in this system do not depend on languages, we can apply this method for characters in any language. Our demonstration on 26 English characters using the same algorithm with the author’s handwriting shows that 25 out of 26 characters are recognized correctly except the character x which is misrecognized as r.

Acknowledgement

The authors would like to thank the Development and Promotion of Science and technology Talents project (DPST) for providing the scholarship. We also would like to thank Miss Pennapa Tungjiratthitikan for her kindness to collect the handwriting datasets we used.

References

WritePad software, Retrieved Oct 6, 2012, from http://www.phatware.com/index.php?q=product/details/writepad

PhatPad software, Retrieved Oct 6, 2012, from http://www.phatware.com/index.php?q=product/details/phatpad

Bounnady, K., Kruatrachue, B., Matsura, T. (2008). Online unconstrained handwritten Thai character recognition using multiple representations, *Int. Symp. on Communications and Information Technologies (ISCIT)*, 2008, 135-140.

Budsayaplakorn, R., Asdornwised, W., Jitapunkul, S. (2003, September). On-line Thai handwritten character recognition using hidden Markov model and fuzzy logic, *Proceeding of the IEEE 13th Workshop on Neural Networks for Signal Processing, 2003. (NNSP'03)*, 537- 546.

Cui. M., Femiani J., Hu J., Wonka P., Razdon A.(2009, January). Curve matching for open 2D curves, *Pattern Recognition Letters*, 30(1), 1-10.

Jamjuntr, P., Dejdumrong, N. (2011) An Improved Online Handwriting Thai Character Recognition, 8th *Int. Conf. Computer Graphics, Imaging and Visualization*, 2011. 87-91.

Karnchanapusakij, C. et al. (2009). Online Handwriting Thai Character Recognition, *6th Int.Conf. on Computer Graphics, Imaging and Visualization*, 2009, 323-328.

Li, G., Jiang, X. (2007, August). Feature abstraction of time-sequence curve and its similarity calculation, *Proceedings of the Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007)*, volume 4, 374-378.

Shih, F. Y. (2010). *Image processing and pattern recognition*, John Wiley & Sons, Inc., Hoboken, New Jersey.

Thongkamwitoon, T., Asdornwised, W., Aramvith, S., Jitapunkul, S. (2002). Online Thai-English and written character recognition using distinctive features, *Asia-Pacific Conference on Circuits and Systems (APCCAS '02)*, volume 2, 259- 264.

Yu, X., Guo, L. (2008, May). Image registration by contour matching using tangent angle histogram, *Congress on image and signal processing (CISP'08)*, volume 4, 746-749.