International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:1, No:5, 2007

On-line Lao Handwritten Recognition with Proportional Invariant Feature

Khampheth Bounnady, Boontee Kruatrachue, and Somkiat Wangsiripitak

***Abstract*—**This paper proposed high level feature for online Lao handwritten recognition. This feature must be high level enough so that the feature is not change when characters are written by different persons at different speed and different proportion (shorter or longer stroke, head, tail, loop, curve). In this high level feature, a character is divided in to sequence of curve segments where a segment start where curve reverse rotation (counter clockwise and clockwise). In each segment, following features are gathered cumulative change in direction of curve (- for clockwise), cumulative curve length, cumulative length of left to right, right to left, top to bottom and bottom to top (cumulative change in X and Y axis of segment). This feature is simple yet robust for high accuracy recognition. The feature can be gather from parsing the original time sampling sequence X, Y point of the pen location without re-sampling. We also experiment on other segmentation point such as the maximum curvature point which was widely used by other researcher. Experiments results show that the recognition rates are at 94.62% in comparing to using maximum curvature point 75.07%. This is due to a lot of variations of turning points in handwritten.

***Keywords*—**Handwritten feature, Chain code, Lao handwritten recognition

I. Introduction

THE on-line handwriting recognition has become an area of active research since 1960 [4], And have many researchers in the field of handwritten alphanumeric character recognition to use difference method for recognition such as geometrical and topological feature [5,6], statistic feature [7], and other algorithms [8,9] to perform recognition based on character shape. But for a good handwritten recognition system depends on main two attributes, first selected feature gathering from a handwritten character, second the recognizers that trained to remember feature of each character in order to cluster and recognize each input character. There are many papers on handwritten system proposed various features such as chain code sequence, chain code histogram, height and width ratio, number of transition from black and white pixel [2-3]. Most of these features are gather in zone position in order to make each character feature more distinct.

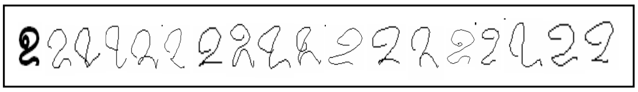


Fig. 1. Variation of handwritten character with and without head, short and long tail.

A lot of these features have been used successfully and reported good recognition rate. But all these good results depend on how recognition characters were written. For example, If a character is written very fast, its proportion changes and all the zone feature and height and width are not working. Fig. 1, show Lao character when written faster it tends to have larger head or some time no headed. It also has longer tail. All these variation destroy zone features. This paper concentrate on the feature construction with proportion and rotation invariant. A character should have the same feature even if it is out of proportion (bigger or smaller head).

The handwritten character set tested in this paper is Lao character set shown in Fig 2, Lao handwritten characters are usually written isolated with few connected cursively. All most all character can be written cursively without raised up pen (only one stroke). Although the tested character in this paper is Lao, the proposed feature can be used in English and other character also.

The high-level character feature proposed in this paper is viewing a handwritten character as a sequence of curve segments. Each curve segment is characterized by its degree of curvature which is measure by the cumulative angle difference from all sampling points with in the segments. If the cumulative is minus, it is a clockwise curve. Since some characters may consist of the same number of segments with the same curve (ex. one segment with clockwise curve) other features of segments are gathering to distinguish these characters.

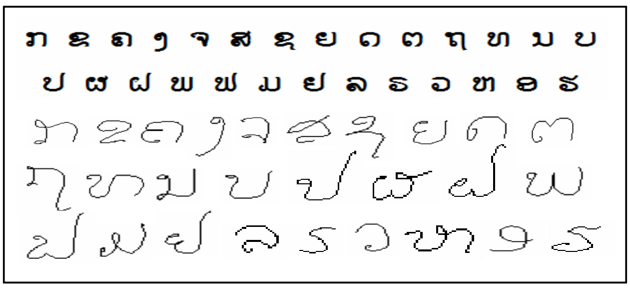


Fig. 2. Example of Lao isolated handwritten and printed characters

Khampheth Bounnady was with National University of Laos, Vientiane, Lao P.D.R. He is now graduate student in Computer Engineering Department, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok Thailand 10520. (e-mail: s6060215@kmitl.ac.th).

Boontee Kruatrachue is with the Computer Engineering Department, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok Thailand 10520. (e-mail: kkboonte@kmitl.ac.th).

Somkiat Wangsiripitak is with the Computer Engineering Department, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok Thailand 10520. (e-mail: kwsomkia@kmitl.ac.th).

This additional segment feature cumulative curve length, cumulative length of left to right, right to left, top to bottom and bottom to top (cumulative change in X and Y axis of segment in both direction increasing and decreasing).

Since the main top level feature is curvature, the character can even be written by 180 degree rotation and still have the same curvature. Unfortunately, some characters when written rotately will become other characters, hence additional feature that are less robust to rotation are added to distinguish them.

II. Feature Extraction

The handwritten character input are the sequence of XY points of pen location captured by an electronic writing tablet sampling with the same timing interval (equi-time sampling). Since the degree of curvature is the feature of interested, there is no need to re-sampling the pen location to make equi-distance sampling. The angle between two sampling points is measure in 360 degree and mapping into chain code of 1 to 8.99 [1] as shown in Fig. 3, Since the length of the points between two sampling point are not adjacent to each other the direction number is real number from 1 to 8.99 instead of integer chain code (1.5 direction number for 22.5 degree). If a character is written with a loop as letter 0 with one whole rotation counter clockwise, this character will have the same curvature feature of plus 6 independent from starting location or number of points in the segments. As long as the loop is closed back to the starting point, the curvature feature is 6. The curvature feature of a segment is a summation of direction difference between adjacent points from the starting point to the last point in the segment as shown in “Fig 3”.

Before extract segmentation of a character, the noises that usually occur in the start and ending of pen trace are removed. This step is perform near the start and ending of the sequence of direction number where the point that has high direction difference to the previous point is a new start and ending point of a character as shown in Fig. 4.

A character is model as consecutive segments of clockwise and counter-clockwise direction. Consecutive segments of the same rotation (ex. both clockwise) is treated as one segment with one cumulative direction difference. Therefore, the points

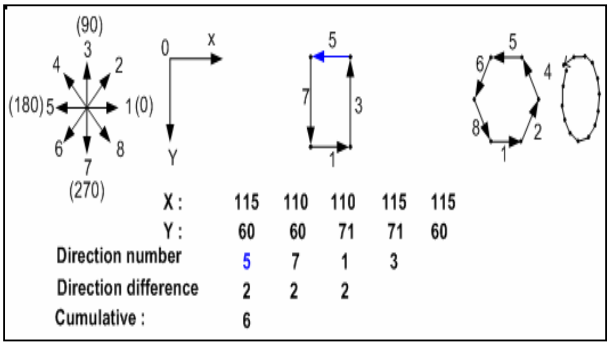
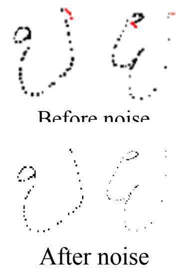


Fig. 3. Mapping of 360 degree to 1- 8.99 chain code and Calculation of Curvature feature.



Before noise reduction			
Index	Direction	Length	Difference direction
0	4	1	0
1	4	1	-2.5
2	1.5	1.414214	0.5
3	2	1.414214	0
4	2	1	
After noise reduction			
0	1.5	1.414214	0.5
1	2	1.414214	0
2	2	1	1
3	3	2	

Fig. 4. Noise at starting and ending of characters.

that separate a handwritten character into segments is a point where pen trace start reverse rotation from clockwise to counter clockwise and vice versa. This point can be identified by monitor the change of sign of the difference at each point. Some researchers [1] used the dominant point or point where the pen traces turn direction dramatically (commonly located at local maximum of direction difference).

We find that number of dominant points varied due to writing style, while the reverse rotation points are more stable (generate less and consistent number of segment from the same shaped character). In order to extract consistent number of segments, the reverse rotation point calculation from direction difference has to cancel noised as follow: If direction difference around point i is diff[i], the reverse rotation point is identified by the sign of diff[i] as follow table 1.

Table I
The Reverse Rotation Point

Sign of diff[i]			Comment
i-1	i	i+1	
+	-	-	The reverse rotation point is at i.
+	-	+	• Reverse rotation point is at i if abs(diff[i])>=1 • In case of diff[i]==diff[i-1] or diff[i]==diff[i+1] point i is not a reverse rotation point.

The change from - to + is treated the same as + to -. In some case the character consists of many small turn segments, all the segment with cumulative direction difference less than 1 is add to the adjacent segments to reduce variation due to very small turn, Jitter generated from writing tablet is canceled by delete segment with very small segment length (cumulative length with in the segment, example is segment number 4 in the fig. 5), The calculation of reverse rotation point and cumulative direction difference of segment is show in Fig. 5.

Some different characters may have the same curvature feature (with the same number of segments and cumulative direction difference, angle). In order to distinguish them, we use the differences in the XY coordinate between adjacent points in both X and Y axis.

International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:1, No:5, 2007

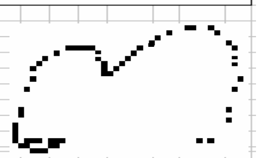
Index	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Direction	1	1	1	1	1	6	5	6	5	5	4	5	4	3	3	3	2.4
Length	1	1	2	1	1	1.4	1	1.4	1	1	1.41	1	1.41	1	1	1	2.23
Direct_Diff	0	0	0	0	-3	-1	1	-1	0	-1	1	-1	-1	0	0	-0.6	0.59
Sign	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

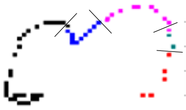
Index	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
Direction	3	2.68	2.4	3	2	2	1.59	1.59	1	1	1	1	8	7	8	7
Length	1	4.12	2.23	2	1.41	1.41	2.23	2.23	1	1	1	1	1.41	1	1.41	1
Direct_Diff	-0.3	-0.3	0.59	-1	0	-0.4	0	-0.6	0	0	0	-1	-1	1	-1	0
Sign	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Segment	1
Cum_direct_diff	-9.98
Cum_Length %	48.48

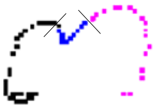
Index	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57
Direction	7	1	2	2	2	2	2	1.6	2	2	1.4	1	1	8	8	8	7	7	7.4	6.4	7	7	6.18	5	5
Length	1	1	1	1	1	1.4	1.4	2.2	1	2	3.16	1	3	1	3	1.41	2	1	3.16	2.23	4	2	5	1	1
Direct_Diff	2	1	0	0	0	0	-0	0.4	-0	-0	-0.4	0	-1	0	0	-1	0	0.4	-1	0.27	0	-0.8	-1.2	0	0
Sign	+	+	+	+	+	+	+	+		-	-	-	-	-	-	-	-	-	-	+	+		-	-	-
Segment	2																			4					
Cum_direct_diff	3								-3.58											0.58		-1.99			
Cum_Length %	11.81								23.65											6.65		9.42			

Direct_Diff = Direction difference, Cum_direct_diff = Cumulative direction difference, Cum_Length = Cumulative length





Segment	1	2	3	4	5
Cum_direct_diff	-9.98	3	-3.58	0.58	-1.99
Cum_Length %	48.48	11.81	23.65	6.65	9.42



Segment	1	2	3
Cum_direct_diff	-9.98	3	-4.99
Cum_Length %	48.48	11.81	39.72

Fig. 5. Curvature feature calculation of reverse rotation point and cumulative direction difference with small turn segment reduction.

Index[i]	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
X	350	350	351	352	354	355	356	355	354	353	352	351	350	349	348	348	348
Y	361	361	361	361	361	361	361	362	362	363	363	363	362	362	361	360	359
X[i+1]-X[i]	0	1	1	2	1	1	-1	-1	-1	-1	-1	-1	-1	-1	0	0	0
Y[i+1]-Y[i]	0	0	0	0	0	0	1	0	1	0	0	-1	0	-1	-1	-1	-1

Index[i]	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
X	348	349	349	350	351	351	352	353	355	357	358	359	360	361	362	362
Y	358	356	355	351	349	347	346	345	344	343	343	343	343	343	344	345
X[i+1]-X[i]	1	0	1	1	0	1	1	2	2	1	1	1	1	1	0	1
Y[i+1]-Y[i]	-2	-1	-4	-2	-2	-1	-1	-1	-1	0	0	0	0	1	1	1

Segment	1			
	X +	X -	Y +	Y-
	21	-8	5	-20
	X + (%)	X - (%)	Y + (%)	Y- (%)
	72.4137931	-27.5862069	20	80

Index[i]	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58
X	363	363	363	364	365	366	367	368	369	371	372	374	377	378	381	382	384	385	385	385	386	385	384	384	381	380
Y	346	347	348	348	347	346	345	344	343	342	341	340	339	339	339	340	342	343	345	346	349	351	355	357	361	361
X[i+1]-X[i]	0	0	1	1	1	1	1	1	2	1	2	3	1	3	1	2	1	0	0	1	-1	-1	0	-3	-1	
Y[i+1]-Y[i]	1	1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0	1	2	1	2	1	3	2	4	2	4	0	

Segment	2				3			
	X +	X -	Y +	Y-	X +	X -	Y +	Y-
	6	0	2	-5	17	-6	22	-4
	X + (%)	X - (%)	Y + (%)	Y- (%)	X + (%)	X - (%)	Y + (%)	Y- (%)
	100	0	28.57	71.42	73.91	26.08	84.61	15.38

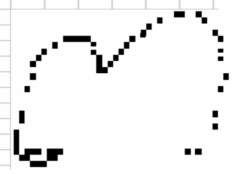


Fig. 6. Calculation of the summation of axis differences.

The summation of difference in X and Y axis in both direction (increasing and decreasing), X+,X-,Y+,Y- are used to characterize the curve. For example if the segment has only X+ and Y+ the segment is written from left to right and top to bottom is show in Fig. 6. The lengths of each segment are another features added to help differentiate character with same curvature and same number of segment. These percentages of length in segment and summation of axis differences reduce the rotation robustness of curvature feature, but they are necessary to distinguish characters with the same curvature.

The character and its summation axis differences is shown in "Fig. 6", "Fig. 7", show the value after normalize of the differences in the X and Y coordinate to distinguish characters with the same curvature, "Fig. 8" shows different writing style can cause the same character to have many different number of segments.

	Character	Segment 1				Segment 2			
		Cum_direct_diff	%Length	% X+	% Y+	Cum_direct_diff	%Length	% X+	% Y+
1	ລ	-0.5842	0.3904	0.7500	0.4167	0.5842	0.6096	0.1026	0.2188
2	ລ	-0.3686	0.2783	1.0000	0.3750	0.5174	0.7217	0.1795	0.1071
3	ລ	-0.5529	0.3411	0.6316	0.5000	0.5529	0.6306	0.1633	0.2286
4	ລ	-0.5013	0.2539	0.5833	0.2000	0.5013	0.7275	0.2250	0.1905
5	ລ	-0.5842	0.3500	0.6471	0.4286	0.5661	0.6500	0.1739	0.1818
6	ລ	-0.5013	0.3570	0.6875	0.4286	0.5718	0.6081	0.1750	0.1200
1	ສ	-0.2921	0.1597	1.0000	0.0000	0.6122	0.8403	0.5500	0.0417
2	ສ	-0.5529	0.1554	1.0000	0.1111	0.6274	0.8446	0.5362	0.1875
3	ສ	-0.3399	0.1529	1.0000	0.0000	0.6274	0.8471	0.5373	0.1282
4	ສ	-0.3686	0.1507	1.0000	0.0000	0.6375	0.8493	0.5000	0.1429
5	ສ	-0.3686	0.1572	1.0000	0.0000	0.6314	0.8247	0.4630	0.2000
6	ສ	-0.3686	0.1464	1.0000	0.0000	0.6696	0.8412	0.6000	0.1111

Cum_direct_difff =Cumulative direction differnce

Fig. 7. The use of summation axis differences to distinguish similar segments character.

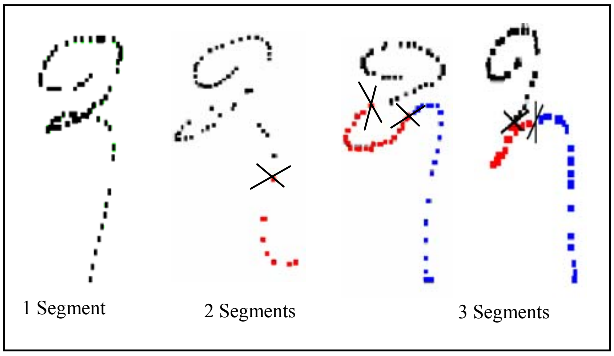


Fig. 8. Show the segments of character that write reference style.

III. Recognition Results

Each character of the same letter is divided into groups that have the same number of segments. In order to recognition, an unknown character number of segments must be determined first, then a template matching is performed for all the character in the training set that has the same number with the unknown character. The closet match will be the recognition character. The number of prototype for each number of segment on average is 268 (max 864, min 5). The recognition time is not quite fast due to the simple feature extraction and comparison time to small number of prototype.

The trained character set consists of 27 letters with the total of 3755 characters. The test characters are 14183 characters. The number of correct characters recognition is 14183/13420 (94.62%), when compare with the dominant point the recognition is 14183/10647 (75.07%) as show in table2 and "Fig. 9", show recognition comparison of segmentation point between reverse rotation point and dominant point.

TABLE II
COMPARISONS OF RECOGNITION

Method	Characters set	Recognition rate			
		Number characters	Right	wrong	% Recognition
Reverse rotation point	Train	3755	3755	0	100
	Test	14183	13420	763	94.63
Dominant point	Train	3755	3755	0	100
	Test	14183	10647	3536	75.07

IV. Conclusion

The curvature and summation axis differences features have been proposed for isolated on-line handwritten recognition. The features are simple and robust for rotation and proportion invariant with high recognition rate.

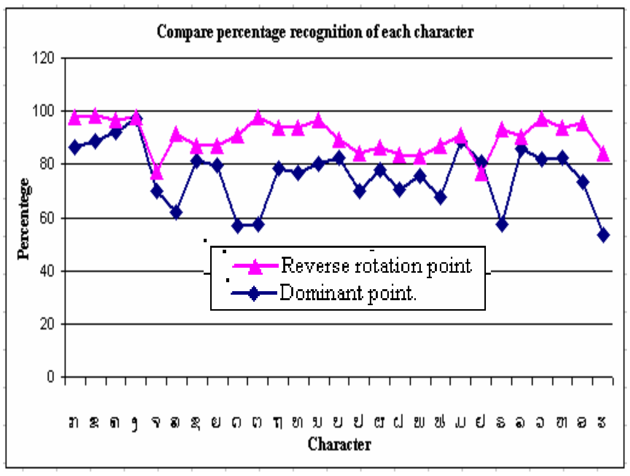


Fig. 9. Recognition rate comparison of two kinds of segmentation point, reverse rotation point and dominant point.

References

[1] H. Yuen, *A chain coding approach for real-time recognition of on-line handwritten characters*, ICASSP'96, Atlanta, USA, 1996.

[2] S. R. Veltman and R. Prasad, *Hidden Markov models applied to on-line isolated character recognition*, IEEE trans. Image Processing, Vol. 3, no. 3, May 1994, pp. 314-318.

[3] Y. H. Tay, P.M. Lallican, M. Khalid, C. Viard-Gaudin, S. Kneer, *An offline cursive handwritten word recognition system*, Proceedings of IEEE Region 10 International Conference on Electrical and Electronic Technology, TENCON 2001, Vol. 2, 2001, pp. 519 -524

[4] C. C. Tappert, C. Y. Suen ans T. Wakahara, *The state of art in on-line handwriting recognition, IEEE Transactions on Pattern Analysis and Machin Intelligence*, Vol. 12, no. 8, 1990, pp. 787-808.

[5] D. D. Kerrick and A. C. Bovik, *Microprocessor-base recognition of handpirnted characters from a talet input*, Pattern recognition, Vol. 21, no. 5, 1988, pp. 525-537.

[6] K. F. Chan and D. Y. Yeung, *Recognition on-line handwritten alphanumberic characters through flexible structural matching*, Pattern recognition, Vol. 32, no. 7, 1999, pp. 1099-1114.

[7] T. Text, J. B. Olafsdottir and M. Daehlen, *Recognition of handwritten symbol*, Pattern recognition, Vol. 23, no. 11, 1990, pp.1155-1166.

[8] D. J. Burr, *Designing a handwriting reader*, IEEE Transaction on Pattern Analysis ans Macihine Intelligence, Vol. 5, no. 5, pp. 554-559.

[9] J. S. Huang and K. Chuang, *Heuristic approach to handwrittern numberal recognition*, Pattern Recognition, Vol. 19, no. 1, 1986, pp. 15-19.