Whole of Word Recognition Methods for Cursive Script

C. Weliwitage
S2114330@student.rmit.edu.au

A. Harvey
harvey@rmit.edu.au

A. Jennings
ajennings@rmit.edu.au

School of Electrical & Computer Systems Engineering, RMIT
GPO Box 2476V, Melbourne, Victoria 3001, Australia

Abstract

Most cursive script recognisers, segment the words into characters, either prior to recognition or during recognition. Whole of word recognition removes the needs for segmentation of the word into characters, eliminating problems associated with poor placement or missing segmentation points. The clear problem with this is that instead of a finite alphanumeric A-Z, 0-9 vocabulary, an unbounded word vocabulary is needed for the unrestrained case. However, in many cases, the application context means that there will be a strictly finite number of words in the application vocabulary. Therefore word recognition becomes feasible. Some examples are signature recognition and the words indicating the dollar amount on a cheque.
This paper examines the current methods of whole word or holistic word recognition methods giving special attention to useful features in recognition algorithms. Some useful features mentioned in the literature are ascenders, descenders, holes, loops, near loops, number and direction of strokes, the direction and orientation of the outer contour of the word, endpoints, cross points, and word length.

An attempt is made to introduce a new feature, 1D word profile, based on horizontal average of the word image. Method is tested with a sample of 72 machine printed cursive words and the results are compared with existing holistic features. Applying the technique for handwriting is under evaluation.

Current Word Recognition Methods

According to a survey on off-line cursive word recognition by **Steinherz, Rivlin and Intrator [1]**, features useful in recognition of off-line segment-free recognition of cursive word recognition can be classified into three categories based on representation level, ie:

1. low level
2. medium level
3. high level

Low level features include, smoothed traces of the word contour, pieces of strokes between anchor points, edges of the polygonal approximation etc.
Medium level category is an aggregation of low level features to serve as primitives. Medium level features are continuous in nature in contrast to low level features.

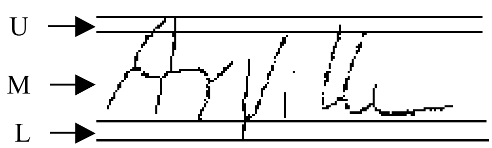
High level features are holistic or global features such as ascenders, descenders, loops, i dots, t strokes etc.

The paper summarises the different algorithms proposed for off-line cursive word recognition. They include minimum edit distance calculations based on dynamic programming, Hidden Markov Models or other specialised methods.

Govindaraju and Krishnamurthy [2] presents an algorithm which uses temporal information derived from off-line word images in the form of uptrends and downtrends of each stroke for the holistic recognition of off-line cursive words. This method basically applicable, only to cursive words and small lexicons. Method represents the off-line cursive word as a set of strokes. Each stroke is traversed to extract global contour features relating to the upstroke and downstroke movements of the pen. First the binarised word image is skeletonised using a standard thinning algorithm. Then the image is subdivided into three zones: Upper, Middle and Lower [Fig. 1]. For the recognition process, an array of inception and terminal points of each stroke is generated. A feature vector [Fig 2] is created using the peak and valley points identified along the word contour. The attributes of the feature vector subset of each stroke piece between two feature points are as follows:

1. Orientation (Up or Down)
2. Slope of Stroke
3. Length of Stroke
4. Stroke piece start zone (Upper, Middle and Lower)
5. Stroke piece end zone (Upper, Middle and Lower)

To minimise the Influence of ligatures, components of the vectors, which have the same start and end zones and the slope that is almost zero are deleted.



U – Upper zone
M – Middle zone
L – Lower zone

Fig. 1 Reference lining on word image (Ref 2)

$$\bar{F} = \begin{Bmatrix} 1 & 47 & 5525 & 2 & 1 \\ 0 & 71 & 1145 & 1 & 2 \\ 0 & 3 & 257 & 2 & 2 \\ 1 & -87 & 485 & 2 & 2 \\ 0 & 26 & 5 & 2 & 2 \\ 0 & 78 & 2197 & 2 & 3 \\ 1 & -85 & 2417 & 3 & 2 \\ 1 & -55 & 3826 & 2 & 1 \\ 0 & -74 & 2696 & 1 & 2 \\ 1 & -74 & 2696 & 2 & 1 \\ 0 & 10 & 457 & 1 & 1 \\ 0 & 82 & 1885 & 1 & 2 \\ 1 & -80 & 148 & 2 & 2 \\ 0 & -50 & 485 & 2 & 2 \\ 1 & 37 & 269 & 2 & 2 \end{Bmatrix}$$

where:
Column 1: Orientation (Up: 1; Down: 0)
Column 2: Slope of stroke piece
Column 3: Length of stroke piece
Column 4 & 5: Stroke_piece_Start and End_zones
(Upper: 1; Middle:2; Lower: 3)

Fig.2 The feature vector matrix of the image in Fig. 1 (Ref 2)

Parisse [3] presents a method of use of simplified profiles of word shapes for the global recognition of off-line handwriting.

This method, first extracts the complete contour of a digitised word. By eliminating internal contours, the upper and lower parts of the contour of the image are obtained [Fig 3], and transformed into a series of vectors. The upper and the lower profiles extracted correspond to two series of vectors representing the top and the bottom of the word. Vectorization, ie. series of points obtained through contour extraction is transformed into a series of vectors, is the next process before attempting comparison. Dynamic time warping technique is used for comparison of two vector series, ie. upper profile of the word to be recognised with the upper profile of the known word.

This global method is limited to a smaller lexicon as training of each individual word is required. To generalise it to a larger lexicon, the use of sub profiles [Fig 4], that are the profiles of strings of two or three letters or n-grams are extracted and the extraction procedure is totally automatic. In contrast to global comparison of word profiles, recognition will be based in seeking of all the profiles of known n-grams in the shape of the unknown word using dynamic time warping algorithm.

Guillevic and Suen [4] propose a method for recognizing unconstrained, writer independent, handwritten cursive words belonging to a small static lexicon, ie amounts written in bank cheques. After pre-processing, slant correction mainly, amount segmentation into words and extraction of global features for the recognition module are performed.

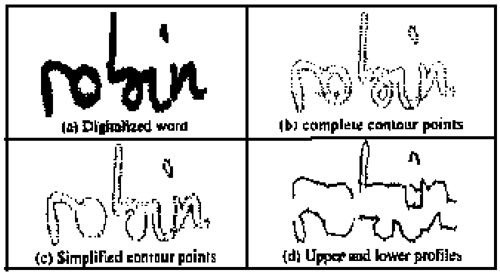


Fig 3 Processing of the word 'robin' to extract upper and lower profile vectors (Ref 3)

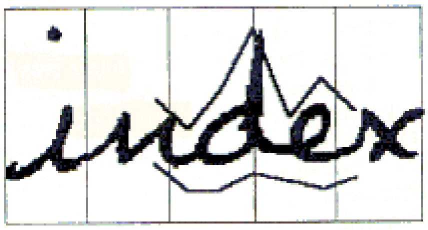


Fig 4 Extracted sub-profile for the 'de' in the word 'index' (Ref 3)

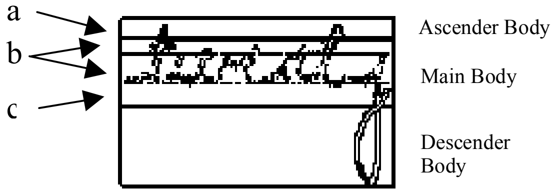
Seven types of global features are extracted from the word image, [Fig 5] and [Fig 6].
They are:

1. Ascenders
2. Descenders
3. Loops
4. Estimate of the word length
5. Vertical Strokes
6. Horizontal Strokes
7. Diagonal Strokes

Thresholds for ascenders and descenders are determined empirically and are expressed as a percentage of the main body height. Ascenders and descenders are detected by following the upper and the lower contour of the word respectively. Word length is estimated as the number of central threshold crossings. Strokes are extracted using mathematical morphology operations.

Input feature vector is different to class feature vector. Input or word feature vector consists of eleven features, relative position of ascenders, descenders, loops, strokes, number of ascenders, descenders, loops and the word length. When classifying, this input feature vector is converted to the class feature vector, which is build up of eleven sub vectors. Class feature vector is compared to the vectors obtained from the training sample.

Nearest neighbour classifier is the classifying technique used. Minimum shift distance is defined to get the distance between two feature vectors



a. Ascender threshold
b. Reference lines
c. Descender threshold

Fig 5 Ascender, Descender and Loop features (Ref 4)

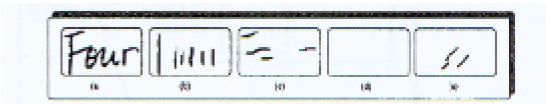


Fig 6 Stroke features: (a) Original Image (b) Vertical (c) Horizontal (d-e) Diagonal Features (Ref 4)

Madhvanath and Krpasundar [5], present a technique for pruning of large lexicons for recognition of cursive script words. This technique involves extraction and representation of down-ward strokes from the cursive word to obtain a generalised descriptor, which is matched with ideal descriptors.

A new approach to the method presented in [11] is used in obtaining shape descriptors (M-medium, A-Ascender, D-Descender, F-f-stroke and U-unknown), from the down ward strokes of off/on line strokes. In this approach, each downward stroke has been represented by an ordered pair (u, l) where u and l are in the range [-1, +1], and a word is represented as a sequence of such (u, l) pairs. Limiting contour extrema has been used to approximate the end points as shown in Fig 7.

To compute the distance between the descriptor extracted from the image and the 'ideal' descriptor corresponding to a given word (ASCII string derived), elastic matching technique is used and implemented using a trie-representation, ie. organising the lexicon entries and their ideal descriptors as a trie of stroke classes.

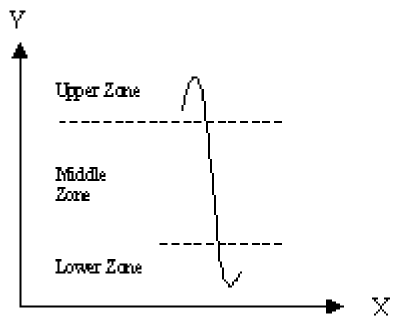


Fig 7 Upper and lower extensions of a stroke (Ref 5)

Guillevic and Suen [6], describe a fast reader system in which the recognition is done at the sentence level. Information from the graphical input is supplemented with the knowledge of context, orthography, syntax and semantics as the skilled human reader does in text reading. Method is tested on bank check processing. Theoretically this method is more robust against spelling mistakes, missing letters, unreadable letters etc.

The primary features extracting should be invariant enough and at the same time discriminative enough. So the features selected here are

1. Loops
2. Near loops
3. Ascenders
4. Descenders
5. Horizontal and vertical strokes

When these primary features are not sufficient, secondary features such as the characters that are adjacent to blank spaces and the estimation of the number of characters in an input word are useful.

Cai and Liu [7] present a new method to automatically determine the parameters of Gabor filters to extract structural features from word images. Features used in this system are the parameters of word image line segments, ie orientation, length and line centroid. Extraction of these line segments would be based on the output of the Gabor filter. Since it is difficult to order 2D word image features in 1D domain, all line segments in a word are divided into eight groups according to their orientation and each group is further divided into three sub-groups based on four base lines. Dynamic programming is used to calculate the distance between two words.

Madhvanath and Govindaraju [8], discuss the use of holistic features in their address classifier implemented at CEDAR. Features used by the system are the word length, number and positions of ascenders, descenders, loops and points of return. Macro features, or composite features such as 'ff' and 'ty' are also extracted and used in the classifier to enhance the scores. Feature equivalence rules provide means of normalization among different writing styles.

Madhvanath, Kim and Govindaraju [9], present a method of chain code based representation and manipulation of hand written images. Techniques that are applicable to word level recognition as well as image level and character level , are described.

For chain code representation, binary image is first scanned and the contour is traced and expressed as an array of contour elements which contain x,y coordinates of the pixel, slope/direction of the contour into the pixel and auxiliary information such as curvature. Slope convention is as shown in Fig 8.

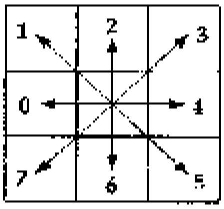


Fig 8 Slope convention *(Ref 9)*

Determination of following word recognition techniques are described:

- Upper and lower contour
- Local contour extremas
- Reference lines
- Word length

Useful features such as word length, number and the location of ascenders and descenders are extracted from the extremas of the upper and lower contours of the word image.

Madhvanath and Govindaraju [10], discuss the method of applying holistic recognition techniques to a large, dynamic lexicon of handwritten words. In the past, these methods are mainly applied to small, static lexicons. For large, dynamic lexicons, this approach can be used for lexicon reduction which will eventually improve computational efficiency as well as combining this technique serially with analytical classifiers will improve the performances of word recognizers.

In this paper, two different lexicon reduction methods are described.

1. Use of constrained, bipartite graph matching scheme to match perceptual features such as ascenders, descenders, for unconstrained handwritten words.
2. A system that operates on pure cursive script, which captures relative heights of downstrokes in the word and form a string descriptor and matched with lexicon entries using a syntactic matching scheme [12].

In first scheme, perceptual features, either scalar (eg. word length) or positional (eg. ascenders) are collected. A confidence and a weight is associated with every feature type. Features used in the system are:

- Natural word length
- Ascenders
- Descenders
- Facts, ie. Description of the existence of certain features in specific regions of the word

The images and the lexicons are represented by wordgraphs, or sets of feature nodes, and a match between nodes of the image and of a given lexicon entry is obtained using constrained bipartite matching.

In second method, the word contour is extracted and represented as chain code. Downstrokes are then extracted from the contour and a shape descriptor is created using the relative heights of downstrokes. This descriptor is matched against predicted descriptors.

New Feature for holistic HWR based on centroid running average of word profile

A novel feature is introduced in this paper, which will be useful in holistic word recognition. Preliminary testing has been done using printed script characters, and applying this method to handwriting is under investigation.

As the first stage of the method, the average vertical distance of the pixels in each column of the word image is calculated and converted to a 1D profile of the word image [Fig 10]. Running average of this profile is determined and used as the new feature. For the calculations of this average, centroid or the base line is selected as the peak line of row-wise histogram of each word image.

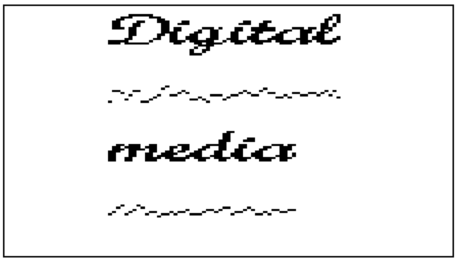


Fig. 10 Abstraction of 1D word profile

Results are analysed and compared with the word length feature, which is a widely used **basic** feature in holistic word recognition. Table 2 summarises the preliminary statistics of the results obtained for a sample of 72 words. Results are analysed for a lexicon of 10 words selected randomly. Table 3 shows few examples of word classes having the same word length and combination of word length and the running average of 1D centroid profile improve the discrimination factor of these word classes.

Conclusion

Existing methods of offline whole-word recognition have been summarised in this paper. These are listed in references [2] to [10]. According to the survey, most useful features for the holistic word recognition are, word length, number and positions of ascenders/descenders, holes, loops, near loops, number and direction of strokes, information on upper and lower contour of the word profile, endpoints and cross points.

A new feature method based on the word horizontal average is presented here. Preliminary studies with machine generated cursive text show promising results. According to the results shown in Table 2 and Graph 1, it can be proved that this new feature is as powerful as word length feature in HWR. Most importantly, as shown in Table 3, combination of these two features will enhance the discrimination factor among word classes.

An evaluation of this method using handwritten cursive script is now in progress.

Table 1: A summary of methods presented in the literature under survey

Ref.	Pre-Processing Method(s)	Recognition Method(s)	Test Data	Result
2	Thinning, zoning	Not Given	552 test words, 10 word lexicon	80% -92%
3	Upper/lower contours, n-grams	Dynamic time warping, dynamic programming	16,200 words	50% - 96%
4	Slant correction	Nearest neighbour classifier, genetic algorithms	5,322 training words, 2, 515 test words	72% - 98.5%
5	Down-ward stroke descriptor	Elastic matching, Trie implementation	21,000 words lexicon, reduced to 1000 words	> 95%
6	Line removal, slant correction	Mathematical Morphology	Not Given	Not Given
7	Slant/tilt correction, baseline finding, image normalisation, line width calculations	Dynamic programming, fuzzy inference	105 training words, 113 test words	64.6% – 94.7%
8	Baseline skew correction, character slant correction, finding reference lines	Euclidean distances between the feature vector and the description vector	103 images of street names	88% - top choice, 97% within top three
9	Noise removal, Slant correction, smoothing contours	Chain code processing	768 images of city names, lexicon of 1000 random city names	> 98% for the reduction of half of the lexicon
10	Chain coding, skew and slant correction	Constrained bipartite matching, syntactic matching	1. 768 images of city names, lexicon of 1000 words 2. 825 cursive words	1. > 98% for the reduction of 50% of lexicon 2. 75% for the reduction of 99%

Table 2: Preliminary results showing a comparison of word length feature with the suggested new feature for a page of 72 machine generated cursive script with a lexicon of 10 words (15 words belong to the lexicon)

			Feature 1 (Word Length)			Feature 2 (Running average)		
	Word Class	No. of Occurrences	Word Length in pixels	No. of detected words with similar features	% Discrimination[1]	Running average of 1D profile	No. of detected words with similar features	% Discrimination[1]
1	image	2	48	3	67%	-6	3	67%
2	media	1	48	3	33%	11	7	14%
3	extract	1	56	1	100%	14	1	100%
4	Digital	1	59	1	100%	8	3	33%
5	specific	1	60	3	33%	-5	1	100%
6	features	3	63	1	100%	-4	7	43%
7	difficult	1	67	2	50%	13	1	100%
8	different	1	70	1	100%	7	1	100%
9	database	3	72	2	50%	15	2	50%
10	corresponding	1	112	1	100%	-21	1	100%

Graph 1: Graphical representation of the results shown in Table 2

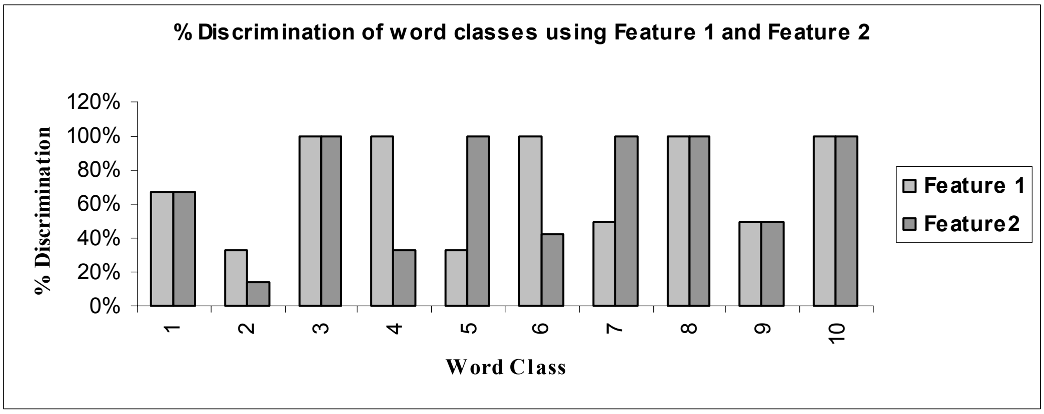


Table 3: Examples of few word classes having same word length

Word Class	Feature 1	Feature 2	% Discrimination using both features
are	26	0	100%
for	26	-2	100%
from	40	3	100%
these	40	6	100%
image	48	-6	100%
media	48	11	100%
specific	60	-5	100%
domain	60	8	100%
format.	60	1	100%

References:

[1]. T Steinherz, E Rivlin, N Intrator, Off-line cursive word recognition – A survey, International Journal on Document Analysis and Recognition, Volume 2, Issue 2-3, 1999, Pages 90-110

[2]. V. Govindaraju, R. K. Krishnamurthy, "Holistic handwritten word recognition using temporal features derived from off-line images", Pattern Recognition Letters, 1996, Vol 17, No. 5, pg 537 – 540.

[3]. C. Parisse ,"Global word shape processing in off-line recognition of handwriting", IEEE transactions on pattern analysis and machine intelligence, 1996, Vol 18 , No 4, pg 460 – 464.

[4]. D Guillevic , C Y Suen, "Cursive script recognition applied to the processing of bank cheques", Proceedings, International Conference on Document Analysis and Recognition, 1995, pg 11-14.

[5]. S Madhvanath, V Krpasundar, "Pruning large lexicons using generalised word shape descriptors", Proceedings, International Conference on Document Analysis and Recognition, 1997, pg 552-555.

[6]. D Guillevic , C Y Suen, "Cursive script recognition : A fast reader scheme ", Proceedings, International Conference on Document Analysis and Recognition, 1993, pg 311-314 .

[7]. J Cai, Z Q Liu , "Off-Line Unconstrained Handwritten Word Recognition", Proc. 1996 Australian New Zealand conf. On Intelligent Information Systems, 1996, pg 199-202.

[8]. S Madhvanath, V Govindaraju, "Using Holistic Features in Handwritten Word Recognition", Proceedings of U.S. Postal Service 5th advanced Technology Conference, 1992, pg 183-198.

[9]. S Madhvanath, G Kim, V Govindaraju, "Chaincode Contour Processing for Handwritten Word Recognition", IEEE transactions on pattern analysis and machine intelligence, 1996, Vol 21 , No 9, pg 928 – 932.

[10]. S Madhvanath, V Govindaraju, "Holistic Lexicon Reduction for Handwritten Word Recognition", SPIE, Vol 2660, pg 224- 234.

[11]. S Madhvanath and S N Srihari, "Effective reduction of large lexicons for recognition of off-line cursive script, Proceedings of the Fifth International Workshop of Frontiers in Handwritten Recognition, 1996.

[12]. G. Seni, N. Nasrabadi and R. Srihari, "An online cursive word recognition system", Proceedings of the IEEE CVPR-94, Seattle, Wa, Jun 17-23, 1994.

[1] ***% Discrimination =(no. of Occurrences of word class/ no. of detected words with similar features)*100***