Licensing and Technology Transfer Opportunity: Manipal University

Title of Technology Available:
A SYSTEM AND METHOD FOR ONLINE RECOGNITION OF KANNADA HANDWRITTEN CHARACTERS

Brief Description of Invention:
The primary object of the present invention is to provide a system and method for online script recognition of Kannada handwritten characters using directional feature-based segmentation approach. The system uses a structural decomposition approach to obtain the primitive stroke classes to recognize Kannada handwritten characters. The present invention provides a system and method that helps in representing a vast data set using less number of symbol sets as compared to existing systems and methods, thereby resulting in reduced complexity.

The present system provides the segmentation of the top stroke from the connected component of the Kannada handwritten characters with the validation of the segmentation process in recognizing Kannada handwritten characters. By this novel segmentation method, the present invention provides a system and method for reducing the total number of classes required for recognizing Kannada handwritten characters.

Brief Background of Invention:
The present invention focuses on Online Handwriting Recognition (OHR) of Kannada characters which is the state language of Karnataka. In the case of OHR, data is written over a digitizer with the help of a digital pen which provides easy and effective human-computer interaction as compared to keyboard, especially for languages having a large vocabulary.

For Indian scripts, the main challenge in designing an OHR is to handle complex shapes of compound characters which are formed by combining two or more basic characters. The Kannada language has 16 vowels (V), 34 consonants (C). Consonants, when combined with vowels, modify the base consonant giving rise to distinct symbols called Consonant-Vowel (CV) modifiers. Further, these CV combinations can be combined with one or two consonants and lead to CCV or CCCV. The total possible number of character combinations becomes 6,47,921, including 10 numeral symbols. To recognize such a massive dataset, segmentation methods are necessary.

A segmentation process is used to decompose a character into primitive sub-strokes. But an erroneous segmentation affects the recognition efficiency. Earliest means of segmentation involves a system that operates on an intermediate representation of characters called Stroke Level Representation. Some techniques include segmenting the Kannada character into the main unit, Right Auxiliary and Bottom auxiliary to create 282 stroke class. Hence there is a need for an adequate alternative method to reduce the number of stroke class for online script recognition of Kannada handwritten characters.

Describe the final product:
The present invention provides a system and method for online script recognition of Kannada handwritten characters. The handwritten data are smoothened using a low pass filter to remove noise due to jitter. The smoothened data is further checked for the existence of a segmentation

point, indicating a point of the sudden change in the writing direction of the first stroke in character. The existence of the segmentation point indicates the presence of the top and middle portions in the first stroke. Segmentation point is validated to confirm the presence of the top portion in the first stroke using a validation algorithm. The system makes use of directional features for segmentation of characters into four primitive strokes, namely Middle, Top, Auxiliary and Bottom strokes. Then the primitive strokes are recognized to identify the character. Based on this concept, 14 top classes, 45 middle classes, 15 auxiliary classes and 37 bottom classes with the total of 111 stroke classes are identified in the present invention which is much less than 282 classes found by using earlier techniques. Thus, a large data set is represented by a less number of symbol set. The condensed symbol set reduces the complexity of the recognition system.

Technological Domain (Keywords): Online Kannada Handwriting Recognition, Directional Features, Segmentation, Pattern Recognition.

Proof of Concept:
1) Collected Kannada Online Handwritten Dataset, by visiting nearby schools and colleges.
2) Novel segmentation method based on directional features is tested on the dataset.
3) A Fuzzy rule-based stroke segregation method is implemented to cluster into the primitive strokes namely Middle, Top, Auxiliary and Bottom strokes
4) Finally, Character is recognized using the deep autoencoder, KNN classifier and combining module

Stage of Development: Advanced Prototype

Provide Information on Competitors who manufacture and/or sell similar products: NA

What are the unique advantages your innovation has compared to the competition:
The present invention provides a novel method of segmentation method using directional features. Thus massive data set is represented by less number of symbol set. Based on this concept, the total of 111 classes is identified, which are much less than 282 classes found by using earlier techniques. The condensed stroke classes reduce the complexity of the recognition system. As the complexity is reduced, the memory requirement is also reduced. The present invention is also extended for another language with some prior knowledge of lexicon. Since Telugu and Kannada characters are derived from Kadamba language, they share the similarity. Hence the present method is extended for the Telugu language also.

A few potential companies who might be interested in this technology:
Stylus enabled Smartphone companies, E-Educational Systems, Conducting Digitizer based survey, Rural Health Services/Data collection (like ASHA Health worker) using local languages, Government to collect census data

Intellectual Property Status: Application Published
Indian Patent application with number : 201741021107   filed in (mention year) 16/06/2017