Abstract

This paper aims at applying primitive elements composing Persian/Arabic letters to recognize offline handwritten letters. To do so, eight primitive elements are used that with them all Persian/Arabic letters are shaped. Having undergone these modifications, the letters get thinned, and then enter the primitive extraction phase. At this stage through using strokes, the letters’ primitive elements would be extracted and by making Stroke Identification Vector (SIV) and then comparing with Character Identification Vectors (CIV) the recognition is gained. Then the number and location of dots on the letters and also the location of the letters towards the baseline was extracted. According to this method, unlike common recognition methods of handwritten methods, no training is required and two processes of separation and recognition are accomplished simultaneously. This algorithm is rule-based and according to several rule recognition is done. The accuracy of this algorithm in digit recognition is 98.8%. Also recognition on isolated Persian/Arabic words that approximately constitute 50% of the sub words in Persian/Arabic texts is 88.7%, for two letters sub word that make up about 32% of text is 81.4%, for three letters sub word that approximately constitute 12% of the texts is 73.6% and four
letters that make up 5% of text the accuracy is 69.7%.

References


Index Terms

Computer Science Pattern Recognition
Keywords

Persian/Arabic OCR, Handwritten Recognition, Primitive, joint manuscripts.