_____

Identification of Kashmiri Script in a Document Image Using Dwt and Entropy

Rumaan Bashir[1,] Kaiser J. Giri[2,] Javaid Iqbal Bhat[3]*, Zahid Hussain Wani[4]

[1]Department of Computer Science, Islamic University of Science & Technology, Awantipora, J & K (India)
Email:-rumaan.bashir@islamicuniversity.edu.in

[2]Department of Computer Science Islamic University of Science & Technology, Awantipora, J & K (India)
Email:-kaiser.giri@islamicuniversity.edu.in

[3]*Department of Computer Science Islamic University of Science & Technology, Awantipora, J & K (India)
Email:-javaidonnet@gmail.com

[4]Department of Computer Science Islamic University of Science & Technology, Awantipora, J & K (India)
Email:-zahid.uok@gmail.com

*Corresponding Author:- Javaid Iqbal Bhat
*Email:-javaidonnet@gmail.com

Abstract:
Over the past decades, the growth in the development of automated systems and their related technologies for document processing has been exponential to achieve more and more effective & efficient solutions. Since the majority of documents contain text so text processing is an important aspect of the document processing. However for text processing it is important to know in which script a particular document is written. These scripts are used to write languages. The identification of script has been an important area of research and accordingly a substantial amount of research work has been done using different schemes. However, there is very less work reported with respect to the identification of Kashmiri language. The main impetus in this work is to perform the identification of Kashmiri Script alongside Urdu, Devanagari and English. The scheme uses discrete wavelet transformation and entropy evaluation. The results achieved are promising. This work will also pave a way for designing more optimal and better solutions with respect to overall automated solutions to the Kashmiri Language.

Keywords: Automatic Document Processing, Script Identification, Discrete Wavelet Transformation, Entropy.

1 Introduction

Describing & depicting the inner self to the world has always required an arduous effort from man since the olden times. However, humans have discovered numerous ways for this depiction and are even said to have accomplished the ability. Expression of thoughts has been one of the important achievements over the past centuries in the history of human race. In this, spoken expression has taken the first place but written expression is not in any way behind. Writing has become the key and noteworthy manner of the depiction [1]. Writing is a pictorial and perceptible practice of representing language.

Human communication that represents language & thoughts through the inscription or recording of signs & symbols is called 'Writing'. Writing schemes use sets of symbols to depict the sounds of speech. It may also have symbols for punctuation & numerals [2]. As human civilizations progressed, the development of writing was driven by practical necessities such as sharing information, recording keeping, marketing activities, etc. Around the 4th millennium BCE [3], the complexity of government &

trade in Mesopotamia outgrew human recall, and writing became a more reliable system of recording & presenting transactions in a perpetual form. In both Mesoamerica & Ancient Egypt, writing systems may have evolved through calendric events and a political requirement for recording important history.

With the innovation of computer based systems and more recently communication systems, a duplication of the writing systems has transpired [4]. Computer systems are used to write on & read from as we are used to writing on and reading from paper. The electronic documents are the electronic form of items