ADVANCED ENCODING TECHNIQUE FOR SCANNED MULTIFACETED MANUSCRIPTS USING CLUSTERING TECHNIQUES

NISHA JOSEPH & DIVYA MOHAN

Assistant Professors, CSE, SAINTGITS College of Engineering, Kottayam, Kerala, India

ABSTRACT

This paper deals mainly with the compression of scanned multifaceted manuscripts. Here we propose an enhanced encoder for scanned composite manuscripts. Effective compound document compression algorithms require that scanned document images be first separated into regions such as text, pictures, and background. The proposed algorithm first classifies the document as image and text regions. These regions are then compressed using different encoders, suited for encoding either type of regions. The adaptive use of different types of encoders using clustering techniques resulted in performance gains. An efficient algorithm can be used to separate the text from the image in a complicated document where the text overlays the picture. Then text regions are compressed using optimized grid based clustering algorithm and image regions are compressed using optimized density based clustering algorithm. This procedure will reduce the pixel classification errors in compound manuscript and thereby increase the overall efficiency of the compression scheme. Experiments using different databases show that the proposed method can improve the document compression to a great extend.

KEYWORDS: Multifaceted Manuscript, Multidimensional Multiscale Parser (MMP), Grid Based Clustering, Varied Density Based Spatial Clustering of Applications with Noise (VDBSCAN)