Abstract

Extraction and recognition of text from image is an important step in building efficient indexing and retrieval systems for multimedia databases. Our primary objective is to make an unconstrained image indexing and retrieval system using neural network. We adopt HSV based approaches for color reduction. This approach show impressive results. We extract a set of features from each ROI for that specific color plane and use them further in a feature-based classifier to determine if the ROI contains text or non-text blocks. The blocks identified as text are next given as input to an OCR. The OCR output in the form of ASCII characters forming words is stored in a database as keywords with reference for future retrieval.

References


**Index Terms**

Computer Science  
Pattern Recognition

**Keywords**

Text extraction, recognition, multimedia, indexing, image retrieval