Content-Based Document Image Indexing and Retrieval using Hybrid of Quadtree and Sparse Matrix Structure

Javaid Iqbal Bhat[1],, S.M.K. Quadri[2]*

[1]Ph.D. Scholar, Department of Computer Science, University of Kashmir, Srinagar,
Jammu and Kashmir, India
[2]Professor, Department of Computer Science, Jamia Millia Islamia, Jamia Nagar, New Delhi, India

Abstract

The fundamental consideration in the design and development of a content-based image retrieval system is to extract the image features that best represent the image contents in a database. Currently, several image indexing approaches based on Quadtree do exist. In this paper, the proposed approach is primarily concentrating on database classification for efficient image representation and effective retrieval using Quadtree decomposition technique augmented by sparse based index structure. In this proposed framework, the efficient storage of images has been carried out, in the beginning, using Quadtree decomposition by routing the images in their respective subsets. This consequently enables us to directly retrieve the closely matching images from the appropriate subset(s) while discarding the other disparate. The fine retrieval process is eventually achieved with the help of already maintained sparse based information vector, prepared at the time of storage of images. Thus, the proposed approach has the competence to significantly increase the accuracy of the image retrieval while simultaneously minimizing the computational complexity.

Keywords: Content-based image retrieval, image retrieval, index table, quadtree representation, sparse matrix structure

***Author for Correspondence** E-mail: javaidonnet@gmail.com*

INTRODUCTION

Information and Communication Technology (ICT) has influenced the people by facilitating large storage, instantaneous transmission, efficient processing and enhanced retrieval of data. With its economic feasibility along with the extraordinary development of the technological tools, including the availability of high-end image acquisition and capturing devices such as digital cameras & scanners, the possibility of maintaining large databases of digital images has realized and is growing constantly. Maintaining the large repositories of document images has however created an incredible need to access and manipulate document images using effective permissible ways. Proficient image searching and retrieval tools are demanded by users from various domains that include medicine, remote sensing, entertainment, architecture and many more in order to bring efficiency in their respective extents. In this direction, many general purpose image retrieval systems have been developed particularly in the form of Text and Content-based frameworks. In the Text-based systems, the document images are manually annotated by text descriptors which are then used by a conventional database management system to perform the identification and retrieval. This approach, however, has two significant disadvantages: first, a reasonable level of human labor & skill is required for manual annotation, second is the annotation inaccuracy due to the subjectivity of human perception. In Content-based Image Retrieval (CBIR) systems, images are indexed by their visual content and the fundamental idea requires the user to provide a description of some of the prominent visual features of the image such as color, texture, and shapes. The designed system can search the archive subsequently and return the images that best match the description. Many researchers focused on content-based image indexing and retrieval using the only color as a visual feature. There are several projects in this field such as QBIC (Query by Image Content), VIR (Visual Information Retrieval)