

Content-based Retrieval of Compressed Images

Gerald Schaefer

Department of Computer Science
Loughborough University
Loughborough, U.K.
`gerald.schaefer@ieee.org`

Abstract. Content-based image retrieval allows search for pictures in large image databases without keyword or text annotations. Much progress has been made in deriving useful image features with most of these features being extracted from (uncompressed) pixel data. However, the vast majority of images today are stored in compressed form due to limitations in terms of storage and bandwidth resources. In this paper, we therefore investigate a different approach, namely that of compressed-domain image retrieval, and present some compressed-domain image retrieval techniques that we have developed over the past years. In particular, a method for retrieving images compressed by vector quantisation, that uses codebook information as image features, is presented. Retrieval of losslessly compressed images obtained using lossless JPEG, can be retrieved using information derived from the Huffman coding tables of the compressed files. Finally, CVPIC, a 4-th criterion image compression technique is introduced and it is demonstrated that compressed-domain image retrieval based on CVPIC is not only able to match the performance of common retrieval techniques on uncompressed images, but even clearly outperforms these.

Keywords: content-based image retrieval (CBIR), image compression, compressed-domain image retrieval, vector quantisation, lossless JPEG, CVPIC