Exploiting Multimedia Content: A Machine Learning Based Approach

Ehtesham Hassan

Department of Electrical Engineering, Indian Institute of Technology Delhi, India.
Prof. M Gopal and Prof. Santanu Chaudhury
Date and location of PhD thesis defense: 10th September 2013, Indian Institute of Technology Delhi

1 Abstract

This thesis explores use of machine learning for multimedia content management involving single/multiple features, modalities and concepts. We introduce shape based feature for binary patterns and apply it for recognition and retrieval application in single and multiple feature based architecture. The multiple feature based recognition and retrieval frameworks are based on the theory of multiple kernel learning (MKL). A binary pattern recognition framework is presented by combining the binary MKL classifiers using a decision directed acyclic graph. The evaluation is shown for Indian script character recognition, and MPEG7 shape symbol recognition. A word image based document indexing framework is presented using the distance based hashing (DBH) defined on learned pivot centres. We use a new multi-kernel learning scheme using a Genetic Algorithm for developing a kernel DBH based document image retrieval system. The experimental evaluation is presented on document collections of Devanagari, Bengali and English scripts.

Next, methods for document retrieval using multi-modal information fusion are presented. Text/Graphics segmentation framework is presented for documents having a complex layout. We present a novel multi-modal document retrieval framework using the segmented regions. The approach is evaluated on English magazine pages. A document script identification framework is presented using decision level aggregation of page, paragraph and word level prediction. Latent Dirichlet Allocation based topic modelling with modified edit distance is introduced for the retrieval of documents having recognition inaccuracies. A multi-modal indexing framework for such documents is presented by a learning based combination of text and image based properties. Experimental results are shown on Devanagari script documents.

Finally, we have investigated concept based approaches for multimedia analysis. A multi-modal document retrieval framework is presented by combining the generative and discriminative modelling for exploiting the cross-modal correlation between modalities. The combination is also explored for semantic concept recognition using multi-modal components of the same document, and different documents over a collection. An experimental evaluation of the framework is shown for semantic event detection in sport videos, and semantic labelling of components of multi-modal document images.

Correspondence to: <hassan.ehtesham@gmail.com>