Video Object Representation for Content-based Video Retrieval (CBVR)

A THESIS

submitted by

A. DYANA

for the award of the degree

of

DOCTOR OF PHILOSOPHY

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY, MADRAS.

February 2010

ABSTRACT

Due to rapid advances in information highways, large amount of digital video data is now publicly available for use in various applications. Users often pose a query based on content, instead of viewing/displaying the raw video data in a random or sequential order. Content-based search and retrieval of video is thus a vital and challenging task for processing digital video database. Traditional ways of processing videos using text annotations or image-based features are inefficient to solve the problem of retrieving similar videos based on the content of user's query. Existing approaches primarily use text data from closed captions or speech transcription, image retrieval using key-frames, and specialized detectors for faces or vehicles. Color, shape and texture features are commonly used for indexing key-frames, but temporal features have often been ignored. Algorithms used for Content-based video retrieval (CBVR) lack in efficiently computing a compact, unified (joint) representation of spatio-temporal information and then use the same effectively in measuring similarity between a pair of video shots.

In this thesis, an efficient system for Content-based video retrieval based on shape and motion trajectory features of a video object is proposed. The key contributions of this work are : (i) A novel motion trajectory representation using multichannel spectral features derived using Gaborfilters; (ii) Combining the similarity costs of Curvature Scale Space (CSS) representation for shapes and Multi-spectral representation for motion trajectory for efficient CBVR. Then, a joint representation of shape and motion features, termed as 3D Curvature Scale Space (3D-CSS) is also proposed for a compact description of video objects. (iii) A novel Multi-Spectro-Temporal Curvature Scale Space (MST-CSS) representation has been proposed as a unified spatio-temporal descriptor of video objects, to overcome the limitations of the proposed 3D-CSS.

In the first part of the work, we propose a global representation of the motion trajectory,

using multi-channel spectral features along with its semantic meaning. A set of salient points are detected on the trajectory using a Gabor filter bank, which represent the temporal locations and degree of the deviations in the direction of the paths. Edit distance matching algorithm for string matching is adapted to design an efficient trajectory similarity measure. The system produced good performance when compared with the state-of-the-art methods, using large benchmark databases and real-world video shots.

In the second part of the work, two approaches are proposed to represent shape and motion features for the purpose of retrieving similar videos. In the first approach, we use CSS (standardized in MPEG-7) for representing shapes and the proposed multi-spectral representation for motion trajectory. The cost of matching two video objects is the weighted sum of the two match-costs obtained from shape and trajectory matching, to retrieve similar video shots. In the second approach, a first attempt of a joint representation of shape and motion of video objects (termed as 3D Curvature Scale Space -3D CSS) is proposed. Curvature scale space theory proposed by Mokhtarian is extended (in 3-D) to represent shape as well as motion trajectory of video objects. The video object (VOB) is convolved using a 2-D Gaussian kernel and the zero-crossing curvature contours (ZCCs) of the evolving VOB surface are stacked in the form of layers to form a hilly surface for representation of VOB.

For retrieving similar videos, the proposed 3D-CSS representation was found to be only sensitive to variations in shape and not for trajectories. To overcome this limitation, a novel Multi-Spectro-Temporal Curvature-Scale Space (MST-CSS) descriptor has been developed for a unified representation of a video object, in the last part of the work. The unified descriptor (MST-CSS) incorporates the advantages of both the multi-scale curvature based features (CSS) for shape and the proposed multi-spectral representation for the motion trajectory of a video object. The surface of the video object volume is iteratively processed using a pair of orthogonal 1-D filters to obtain a gradually evolving (smoother) surface. Zero-crossing contours of the mean curvature computed from this evolving VOB surface yields

(stacked in layers) a hilly surface for a unique representation of a video object.

Experiments are performed on a synthetic (benchmarked) database consisting of 20,000 virtual video objects, using Precision-Recall metric for our performance study. The system is also tested using real-world video shots. The proposed methods based on MST-CSS and multi-spectral motion trajectory representation have shown better performance, when compared with MPEG-7 standardized representation, as well as a few recent state-of-the art methods available in literature.

KEYWORDS: Content-based video retrieval, Video Object Representation, Spatio-temporal descriptor, Curvature Scale Space, Gabor filter, 3D-CSS, MST-CSS, Precision-Recall