Story Segmentation and Retrieval of News Videos in a Multi-modal Framework

Abstract

Shot segmentation, categorization, indexing, and news story formation are the most important and primary steps in building an efficient and well-sorted video storage and retrieval system. News channels have evolved as one of the primary sources of information. However, in recent times, with the increase in the number of news channels, a plethora of news content is available on air, and it has become difficult to store and retrieve the news videos effectively. Commercials are also included in a news video, containing considerably less information. These commercials are to be filtered out, and the remaining news video will be segmented meaningfully. Segmentation of news videos is a crucial process for efficient storage and categorizing of the videos. The segmented stories also facilitate the easy retrieval and finding of the desired news. In this work, we developed different algorithms for shot segmentation, categorization, indexing, and retrieval of news videos. Our methods are independent of different temporal and spatial structures of various news channels and require a minimal manual input.

The spatial and temporal features of a news channel are captioned texts, overlaid texts, logo position, anchorperson of the actual screen, and the position of multiple screens in case of a split-screen configuration. The content of each overlaid text has some unique characteristics. We employed the number, position, and contents within text boxes features for shot segmentation categorization, indexing, and story segmentation. However, the earlier methods do not consider the contents within text boxes, which could have been employed for shot segmentation and categorization.

In Chapter 2, we proposed an adaptive threshold-based multistep approach to detect shot boundaries based on a YIQ colour model. The shot segmentation algorithm compares the histograms between two consecutive frames based on an adaptive threshold. A shot boundary is detected if the similarity measure is less than the threshold. For the measurement of similarity and dissimilarity, cosine comparison is used. Proposed shot segmentation algorithms based on the YIQ model can successfully detect the presence of all shot boundaries through segmentation at fade-in/out, dissolve and cut transition.

Commercials constitute a significant chunk of the videos aired on a news channel. These videos are substantially different from the ones containing news, and the inclusion of commercials in between the news videos makes it challenging to segment the stories. In Chapter 2, we also proposed two shot categorization algorithms, where we utilized the information on the presence of text blocks in the spatial arrangement of the news frame, and the color content present in the news frame is selected as a feature for classification.

In Chapter 3, we proposed a novel indexing method based on the text appearing in a news frame. The indexed shot corresponding to similar news content are combined to form individual videos depicting a single story.

We proposed a novel technique for story segmentation and indexing of broadcast news video in a multi-modal framework in Chapter 4. The visual similarity, silence in the audio, and the text in the text boxes of a news video are parameters to define the story boundaries. Each of these parameters is used to create an index, and these three indices are fed to a probabilistic multi-modal algorithm, which then predicts the story breaks.

In Chapter 5, we proposed novel image and video searching and retrieval algorithms based on edge-based and compressed domain features. In image-based searching, the query image is converted into a gray-scale image, and its pixel values are mapped and matched with those in the broadcast news database. In video-based searching, an edge detection algorithm is applied, and the similarity between the detected edges is determined for the data extracted from the video shot and the broadcast news video. For rectification, we checked the similarity between the detected edges of a particular video shot and the broadcast news. To refine our algorithm of edge similarity, we also performed a Structural Similarity Index (SSIM) measure to achieve the most desired results.

Finally, we proposed a video searching and retrieval algorithm using scene classification using compressed domain features. The performance of the said feature in classifying the scenes is evaluated by considering different feature lengths. Subsequently, a novel searching algorithm is proposed based on the entropy of the image. This method incorporates temporal information from the video during video retrieval. Additionally, the proposed system has the facility to query both by video and by image.

The efficacy of the proposed methods is evaluated by performing a large no of experiments on news videos.

**********end**********