Abstract

Digital multimedia archives have seen an enormous growth over the last few years. However, since many multimedia collections are still poorly indexed or annotated, there is a need for developing automated, content-based image and video retrieval methods that would help users to navigate and retrieve necessary information from these archives. In order to improve the performance of retrieval that is traditionally based only on keywords, recent research is focused on providing efficient access to large archives using features like color, texture, shape and embedded text. Although multimedia information retrieval systems deal with image, video, graphics, audio and text, we concentrate only on image and video. Several image and video retrieval systems have recently been developed in the academia as well as in the industry. In most of these retrieval systems, color based features have a predominant role. In order to improve the recall and precision of retrieval, other low-level features like texture and shape are also considered. These features can be used either alone or in combination with each other.

We have analyzed and utilized the properties of the HSV color space to generate a color histogram for content-based retrieval of images and videos. We propose a robust histogram construction technique named as Human Color Perception Histogram (HCPH) that consists of two sets of bins, one for true color components and another for gray color components. The proposed method utilizes some of the properties of color perception of human eye. It makes use of both saturation and intensity of a pixel to determine the degree by which the pixel appears to have a true color in the red→ green→ blue → red spectrum and a gray color somewhere between black and white. A suitable weight function is chosen that provides a good estimate of the degree by which a pixel represents true color and gray color. For each pixel, one true color component and one gray color component of the histogram is updated. The performance of the proposed method on a set of images with ground truth and also on a large image database is encouraging.

Various features can be combined together by assigning suitable weights to each of them empirically or based on certain heuristics. Weights may also be determined dynamically through relevance feedback from the users. We propose a novel approach for integrating color and intensity of pixel neighborhoods in an image using a co-occurrence matrix. Properties of the HSV color space are effectively used to determine relative perception of color and gray levels of a pixel. The Integrated Color and Intensity Co-occurrence Matrix (ICICM) has four sub-matrices for capturing color-color, color-intensity, intensity-color and intensity-intensity variations between neighboring pixels. For each pixel and its neighbor, one component of each sub-matrix is updated by true color and gray color weights that depend on the saturation and intensity of the pixels. The integrated approach shows promising results in an image retrieval application.

It has been observed that, while low-level features can be efficiently used to perform retrieval in domain-specific applications, high-level information like textual annotations are needed to restrict the search space in more generic applications. We propose a method to dynamically integrate keywords with low-level features of images. Keywords can be dynamically extracted from HTML documents and images from relevant documents only can be compared using their low-level features to enhance recall and precision. The low-level features are extracted from the images and keywords are extracted from the text part of the HTML documents. Frequency of occurrence of keywords is used to determine the importance of the keyword for the associated images. The performance of the proposed method is found to be quite encouraging compared to a number of recently proposed methods.

While features can be directly extracted from images, in video processing applications, a video is first divided into a number of shots. One or more key frames are then extracted from each shot. In general, shot boundary detection is done by comparing the content of consecutive frames. Frame difference is determined by choosing a suitable distance measure. After a shot is identified, key frames are extracted which can be used as representative frames of the shot. The low-level features of frames are used both for comparing consecutive frames for shot boundary detection as well as for matching of key frames during retrieval from

video databases. Since the retrieval performance of ICICM in image retrieval was encouraging, we have extended it to video shot detection and retrieval applications as well. ICICM feature is extracted from the video frames and frame difference is determined by using a suitable similarity measure to detect the shot boundaries. While performing video retrieval for any given query frame, its ICICM feature is extracted and compared with the ICICM features stored in the feature database to return the shots for which key frames are similar to the query frame. Shot detection performance using ICICM on benchmark video databases is found to be quite good.

We have also proposed a shot-change detection method using an Object Level Frame Comparison (OLFC) technique. Objects are segmented in video frames using saturation thresholding approach from the HSV color space. The segmented objects between successive I-frames are then matched to determine frame-to-frame similarity. Shots are detected only when the similarity falls below a threshold. We have suggested a new similarity measure to determine the inter-frame similarity. Accuracy of this method is found to be quite high compared to other methods.