CHAPTER 2 RELATED WORK

2-1 An overview of video retrieval system

As everybody knows, storage and retrieval are two main components to access video data in a visual database. In the storage process, one extracts features from videos and then use them to describe the semantics of these videos. For the purpose of efficient retrieval, these features are represented, organized and stored in the database. In the retrieval process, the system extracts the appropriate feature vector from the query data, and then the system calculates the degree of similarity (using a similarity metric) between the feature vector of the query and that of the candidate videos stored in the database to retrieve similar video data. The retrieved videos are then output to the user in the descending order based on the degree of the similarity to the query. The architecture of a generic video database system is shown in **Figure 1**. It consists of the user interface, content-based retrieval module, organization, and database management modules. Each module will be roughly described below.

1. User interface: In visual information system, user interface plays an important role for almost all of its functions (e.g., manual feature extraction, navigation, refinement). The user interface consists of a query processor and a browser to provide the interactive tools for querying and browsing the database. The query processor provides the means to query videos by some query methods. A query can be a simple keyword or a complex one such as a sketch or an object track specified by the user. After retrieving the video data that similar to the query, the browser is used to display the results. The browser allows user to further refine and navigate through the database visually.

2. Content-based retrieval module: The content-based retrieval module includes the following modules:

- Shot change detection: The first step is to decompose a video sequence into several shots.
- Key frame extraction: After a video sequence is segmented into shots, a set of key frames are selected from each shot. Then each shot will be represented by the spatial and temporal features. The spatial features mean the visual content of the key frames of a shot and the temporal feature refers to the temporal content of a shot.

- Feature extraction and representation: In this stage, features such as color, texture, etc. are extracted to describe the visual content of a still image. For video, the spatial features are generated using still image techniques [9], while the temporal features are extracted based on motion and/or camera operations within the shot [10].

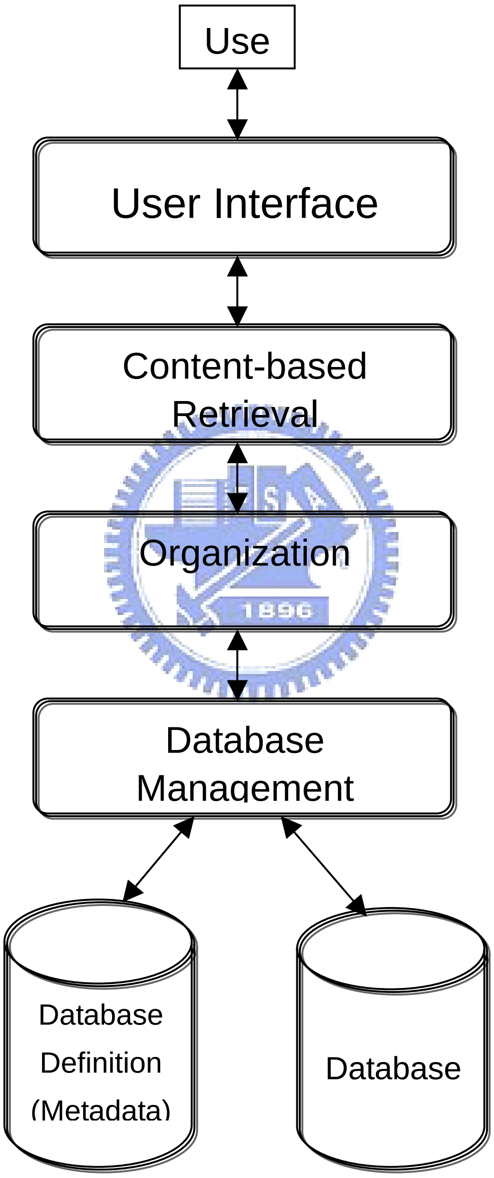

Figure 1

3. Organization: Efficient query processing needs the organization of video indices such that efficient search strategies can be used. Several indexing structures like R-tree family [11], R*-tree [12], and quad-tree [13] are commonly used. Each structure has its advantages and disadvantages. Niu et al. [14] have discussed some issues about novel indexing structures for image retrieval.

4. Database management module: The database management module provides internal level physical storage structure and access path to the database. The database management module has the following characteristics: (1) provides insulation between programs and data, (2) provides users with a conceptual representation of the data, (3) supports multiple views of the data, and (4) ensures data consistency.

2-2 Our System model

A video data consists of a sequence of frames. We first segment a video into several shots, select one or more key frames from each shot, and then extract a feature vector from each key frame. The color features are extracted to represent the key frame. A video can then be represented by a sequence of feature vectors. Therefore, computing the similarity between two videos can be transformed to the problem of computing the similarity between two sequences of feature vectors.

To retrieve the similar videos from the database, the query video is also first segmented into several shots, each of which can be represented by one or more key frames. Then, a feature vector is extracted for each key frame. Thus, the query video is also represented by a sequence of feature vectors. The sequence of feature vectors is called the query sequence. Then, we slide and match the query sequence with the subsequence of feature vectors in the feature database, and compute the similarity between them. The database video sequences with the similarity high enough are output and returned to the user. The flowchart of our approach is shown in **Figure 2**, and each component in **Figure 2** will be described in **section 3** and **section 4**.

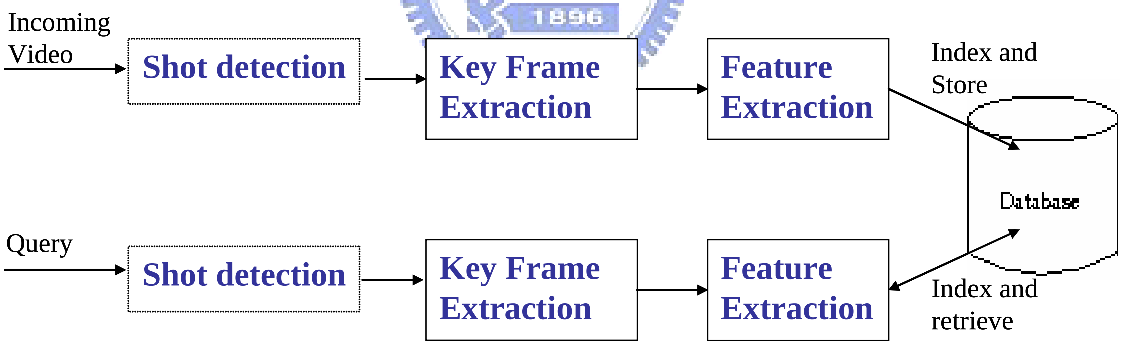

Figure 2