Research on Key Technologies of Content-based Video Retrieval

Xiu Hu

Jingchu University of Technology, Jingmen, Hubei, China

Keywords: Video retrieval, Content based technology, Key levels

Abstract: How to achieve fast, effective and convenient query retrieval has gradually become a highly popular research topic in the field of video research. It is difficult for video to express its features in precise language, because its data has rich information, so the traditional method of manually inputting the attributes of video elements has great disadvantages. Under such circumstances, video content features and video retrieval techniques that can be automatically extracted by computers have become an urgent issue to be solved. In recent years, the rapid development of content-based video retrieval technology has been able to meet the requirements of the query plan, which is mainly divided into shot detection, key frame extraction, video matching, video retrieval and other steps, according to different retrieval targets, extract video The key frames and features of the system can achieve the purpose of fast retrieval.

1. Introduction

At present, through the text-based video retrieval method, the video search function of these video websites can be realized. As information is highly inflated and the Internet is widely popularized, some videos and news with their own text information can be retrieved from their names and main contents, which can basically meet the user's query requirements. Nowadays, due to the popularity and popularity of the Internet, various videos have a lot of manually marked information used to describe various attributes of the video and the content expressed therein. The annotation of various text information makes it easier for users to search for the video clips they need. With the annotations of different people, each video has a lot of different information and contains enough content. However, due to the large amount of information, errors often occur in the process of text retrieval, and you need to find what you need from large number of searched files. However, if the text information of the video can be marked with accurate vocabulary as much as possible, the efficiency of text retrieval can be improved. Today, when the text retrieval technology is mature, people are accustomed to use the text retrieval method to search for the required video, so the most widely used text-based retrieval method is currently used. The emergence of the Internet has brought great convenience to mankind, especially after the realization of resource sharing, but in the face of this vast amount of resources, which ones are of value to themselves? With the rapid development of network and multimedia technologies, people are rapidly moving towards the information society since the 1990s. Today, when video information is highly inflated, how to achieve fast, effective and convenient query and retrieval has gradually become a highly popular research topic in the field of video research.

2. Structured Video Content

The video stream is composed of thousands of image frames. Frames are the smallest unit of video. If each frame is processed separately, the efficiency of indexing and retrieval will be very low. Fortunately, video is usually composed of a large number of logical units or blocks, and we call these blocks video clips. A shot is a short sequence of adjacent frames, which depicts the same scene, representing a camera action, an event or a continuous action. Any video is connected by shots, shots are the basic unit of video retrieval. In addition, several sets of shots that are semantically related and temporally adjacent can be combined into a scene, which can express the high-level abstract semantics contained in the video. The process of video structuring is to divide a

  DOI: 10.25236/iwmecs.2020.054

video frame sequence stream into several segments according to the development of the plot. These segments can be divided into several levels of hierarchical structure, and indexes are established separately. General video data can be divided into several levels of video-scene-shot-key frame. Users can quickly understand the content of the entire video by browsing the video catalog, instead of browsing all the image frame sequences in sequence.

When the content of the video plot changes, a lens switch will appear. Due to the different connection methods of video shots, there are two kinds of switch between shots: abrupt change and gradual change. Sudden change is the simplest switch between two shots, without transition. Gradient refers to the process of gradually transitioning from one lens to another without obvious lens jumps, including fading in and out, blending and scrubbing. Shot edge detection is to divide the original continuous video stream into long and short shot units, which provides the basis for subsequent video analysis and processing. A lot of work has been done in this area at home and abroad. Early work mainly focused on the detection of lens mutations. In recent years, more analysis of the lens gradual change has been made, and each different special transformation is also regarded as a different lens, And specializes in how to detect the occurrence of these special effects lenses. The current video lens segmentation technology is mainly based on the changes reflected by the video data of the lens when the lens is switched. Since the change between adjacent frames in a shot will not be very large, the feature difference between them will always be limited to a certain threshold. However, when the shot changes suddenly, the two adjacent frames before and after the change point usually show a large amount of change in content. If the feature difference exceeds a given threshold, it means that a segmentation boundary occurs.

Video shot segmentation not only needs to detect abrupt changes between shots, but also segment the gradient. When the gradation of adjacent frames in a shot occurs, the difference is higher than the difference in the shot, but much lower than the lens threshold. In this case, a threshold is not enough, because, in order to capture these boundaries, the threshold must be greatly reduced, which may produce many false detections. In order to solve the above problems, the researchers proposed a double comparison technology that can detect both normal lens shear and lens gradation, that is, use two difference thresholds: the threshold Tb is used to detect normal lens shear; the threshold TS It is smaller and is used to detect frames that may appear in the gradient. In addition to the lens gradual change that can produce a detection boundary problem that is impossible to detect with a simple quantitative measure, the problem of “false segmentation” of the lens boundary can also be caused by camera positioning and zooming. This problem can be detected using motion features such as optical flow.

3. Video Indexing and Retrieval

Content-based video retrieval mainly relies on visual features and spatiotemporal features in video data to measure the similarity of these features. The usual search method submits example videos and queries for similar videos. The following introduces several video indexing and retrieval methods.

This is the commonly used video indexing and retrieval method. The main content of r frame capture lens can extract and index the characteristics of r frame based on color, shape and texture. In the retrieval process, the query is compared with the index or feature vector of frame r. If the r frame is similar or relevant to the query, it is submitted to the user. If the user finds that the r frame is related, he can watch the video segment it represents by playing. The indexing and retrieval methods based on r frames regard the video as a collection of static images, ignoring the time or dynamic information contained in the video. The video index and retrieval method based on dynamic information can make up for this deficiency. Dynamic information is usually derived from optical flow or dynamic vectors. The parameters commonly used for dynamic indexing are: a. Dynamic content. It measures the behavioral content of the video. For example, a talking head video has very little dynamic content, while a violent explosion or vehicle collision generally has very high dynamic content. b. Dynamic consistency. This is a function of time and an indicator of the dynamic smoothness of the video. For example, a smooth positioning lens has a high dynamic

smoothing value, while a video with staggered positioning has a low dynamic smoothing value. c. Dynamic positioning. It captures positioning dynamics (camera movement from left to right or right to left). A smooth positioning lens has a higher dynamic positioning value than a zoom lens. d. Dynamic tilt. It is an indicator of the vertical motion component of a video stream. Positioning the lens has a lower dynamic tilt value than video with a lot of vertical motion. The above dynamic parameters are related to the entire video stream or video footage.

The main disadvantage of lens-based video indexing and retrieval is that although the lens is the smallest unit of a video sequence from the perspective of cinematography, it does not directly use content-based representation. The content can change continuously in a single shot, or it may remain almost unchanged in a series of consecutive shots. How to determine the “content change” is a key problem to be solved based on content indexing and retrieval methods. Any given scenario is a complex collection of parts or objects. The location and physical properties of an object and its interaction with other objects determine the content of the scene. The object-based indexing and retrieval method is to segment all objects from the video stream and use the information of each object for indexing. This indexing strategy should be able to capture changes in the content of the entire video stream. In still images, segmentation and recognition of objects are usually very difficult, but in video streams, objects move as a whole, so we can combine moving pixels together to form an object. Use this idea to accurately segment the object. By tracking the motion of the segmented object, a description of the motion can be constructed for later video shot retrieval.

The main disadvantage of video retrieval based on feature matching is that the feature lacks semantic information, which makes users feel inconvenient when explaining the query of video data. To this end, researchers have proposed annotation-based retrieval. Annotation is a set of semantic attributes related to a specific video segment, which can capture the advanced content of the video. Annotations can be obtained in the following three ways: a. Manually explain and annotate the video. This is a time-consuming task, but this method is still widely used because the current technology does not automatically describe advanced video content. The manual annotation process can be simplified in two ways. One is to provide a definite framework for manual input, and the other is to make full use of the domain knowledge of specific video types to semi-automate annotation. b. Many videos have relevant copies and subtitles that can be directly used for video indexing and retrieval. c. If you can't get the subtitle, you can use speech recognition to extract the vocabulary words, and then use these vocabulary words for indexing and retrieval. However, there are still many difficulties with this method, because speech and non-speech are usually mixed in the channel, and there is background music and noise in the speech signal, so the recognition rate is low. A video is composed of a sequence stream of thousands of image frames, but it is not simply a simple accumulation of the visual content of image frames. What is important is that the plot evolves as the images change. The video contains rich information, which has both high-level semantic features, visual characteristics of the underlying image frames, and plot characteristics of time and space development, so video retrieval is very difficult to define and implement. Because of the huge data volume and rich performance content of the video, it is difficult to capture all the content of the video with a single feature or technology. The practical application system should use the comprehensive method of the above technology. Furthermore, the indexing and retrieval system is likely to be application-dependent, and strengthen certain aspects according to application requirements.

4. Conclusion

This article analyzes and explains the concept, system structure, key technology, and typical system of content-based video retrieval. Content-based video retrieval is the current research hotspot of information retrieval, involving many fields such as image understanding, artificial intelligence and database technology. Many valuable technologies have been studied. Although these technologies are a big step towards content-based automated video management, they generally only deal with low-level features. However, advanced features such as time events and interactions between objects in the video are still difficult to identify and extract. In order to obtain

automated video indexing and retrieval based on advanced features and concepts, more research is needed.

Acknowledgment

Scientific research project of Hubei Provincial Department of education, project name: Research on Key Technologies of content-based video retrieval, No.b2018242; Science and technology research project of Jingmen science and Technology Bureau, project name: Research on the application of shot segmentation algorithm based on mutual information of block image in video retrieval, No. 2018ydky071; Hubei University excellent young and middle-aged science and technology innovation team plan project, No.t201923.

References

[1] Wu Yongjun. Content-based video retrieval technology. 2008 New Internet Media New Technology Seminar, 2008.

[2] Jin Hong, Zhou Yuanhua. Video processing technology based on content retrieval. Journal of Image and Graphics, 2000.

[3] Pu Xiaoge. Summary of research on key technologies of content-based video retrieval. Information Science, no. 3, pp. 464-469, 2010.

[4] Zhang Tingting. Review of Research on Key Technologies of Content-based Video Retrieval. Journal of Agricultural Library and Information Sciences, no. 12, pp. 55-60, 2009.

[5] Zhang Hongde, Liu Yu, Tang Bo. Research on Content-based Video Retrieval Technology. Television Technology, no. 6, pp. 30-33, 2001.