Scientific Journal of Information Engineering
November 2017, Volume 7, Issue 2, PP. 77-84

The View of Copyright Protection Technology for Video Gene Recognition*

Quan Jinlan [1], Cui Jizhe [1†]

1.College of Economic & Management YANBIAN University, Yanji 133002, China

†Corresponding Author email: cuijizhe@foxmail.com

Abstract

Copyright is the core competitiveness of cultural industry, the future cultural influence much depends on a standardized copyright management system and supporting information processing technology. Through the research on the performance evaluation system of TRECVID video gene recognition technology and analysis of relevant technology development status in China, this article explores the achievements in TRECVID video gene identification and analysis technology and the deficiencies of related technologies applied in copyright protection. The theoretical analysis of video gene recognition technology found that the technologies in the field of TRECVID video gene identification and analysis technology can effectively retrieve the illegal use of video content and completely protect the copyright of video copyright holders. China's domestic technology has also been shaped to effectively support video Copyright protection needs. Finally, this paper gives the suggestions of standardizing the market of video copyright protection in China and the technical fields of video copyright protection that need to be strengthened with the analysis of the structure.

Keywords: Video Gene Analysis Technology; Video Retrieval; TRECVID; Video Copyright

1 INTRODUCTION

After entering the information age, copyright protection of audiovisual content should be strengthened through effective measures in the copyright management system. The domestic video gene recognition technology research is still in the development stage, resulting that when YouTube, Daily motion and other video sites clear most of the infringing content on the site, China's potato, Youku is increasingly becoming a "black hole" of copyright [1]. Therefore, it has important scientific meaning and application value to focus on the systematic research of copyright protection technology of video gene recognition class and plan the technical performance evaluation mode suitable for the current network environment in our country. The traditional retrieval method is text retrieval, mainly based on some text message, this technique is not completely suitable for video retrieval[2]. The video content is not exchanged if the video's logo information is replaced, and the traditional retrieval method is undetectable, but the TRECVID detection can effectively detect the video content that the text retrieval can't retrieve, thereby protecting the video owner copyright. This article intends to through TRECVID video gene analysis technology performance evaluation system and related domestic technical analysis to discover the extent of technology development related to the TRECVID detection technology in the country. And through the theoretical discussion of video gene recognition technology, we can search the infringement of video copyright effectively and locate the development possibilities of these technologies in the field of copyright protection.

2 TRECVID SUMMARY

TRECVID started with video tracking in TREC in 2001. TREC is a competitive seminar that uses text material for

*Fund support：By the Jilin Provincial Department of Education, "The 12th Five year" scientific and technological research projects to support funding (Jilin educational research contract the word [2015] No. 38)

information search. After the discussion of video tracking technology of TREC2001 and TREC2002, from 2003 TREC developed into TRECVID seminar independently. The TRECVID Symposium is dominated by NIST and DTO and is committed to becoming an international authority for video retrieval projects. NIST provides a lot of additional information in TRECVID. In addition, NIST provides low-level features (visual features, text features, etc.) obtained from images, programs using SVM's identifiers, and a package's accuracy evaluation program.

Because of the coverage, comprehensiveness and consistency of the data set used in the TRECVID evaluation, the authority of the evaluation results has been recognized. TRECVID has become an authoritative conference in the field of video search, which is of great significance for the development of video search.

Since 2003, the TRECVID Video Search Contest has been held once a year. The main tasks include: Shot boundary detection, Camera motion Detection, Rushes summarization, High- level feature extraction, semantic-based video search, Surveillance event detection and Content- based copy detection. In 2010, Event detection in internet multimedia was added to the TRECVID evaluation mission, replacing the previous semantic-based video search with two new video search tasks: Semantic-based Known-item search and Instance search based on image examples. In 2010, TRECVID added the most popular "web video" at the moment, and these videos are applied to both known - item search and semantic indexing tasks. China's Tsinghua University, Peking University, Chinese Academy of Sciences Institute of Automation, Beijing University of Posts and Telecommunications, FUDAN University, Beijing JIAOTONG University and HUAZHONG University of Science and Technology have all made good results in the TRECVID evaluation.

3 TRECVID VIDEO GENE ANALYSIS TECHNOLOGY SYSTEM

This article mainly from the TRECVID video gene recognition analysis of Shot boundary detection，High- level feature extraction，semantic-based video search，Surveillance event detection and Known- item search to conduct technical performance evaluation.

3.1 SBD (Shot Boundary Detection) Monitoring

SBD monitoring mainly with the correctness, recall and testing three factors as the main evaluation project analysis. Because video information is contained in different media formats and the existing processing and recognition capabilities based on any single media format or feature have limitations, there is a need to analyze these multimedia content differently in video retrieval, get the multi-modal information that describes the video change, and then get the final video retrieval result through the effective fusion method. The integration and interaction of multiple modalities in video also play an important role in solving the problem of semantic gap between low-level features and high-level semantics. In the literature, Lu Bo has tested the TRECVID data and found that the accuracy of the semantic understanding obtained by the fusion analysis of different types of multimedia data is higher than that of the single type of multimedia data based on statistical models such as support vector machines, mixed Gaussian and hidden Markov chains, nearly 35%.

Zhao Jinlong [3] proposed the design based on BoW model in the literature, and realized the object sample picture search algorithm, tested and researched on the TRECVID INS 2013 video retrieval database. This model enhances the visual vocabulary of the ability to distinguish. In 2014, using the same data as the TRECVID INS 2013, the three local features were cascaded and the proposed object-based image search system was used to enhance the significance of searching for the target object.

Cen Shusheng [4] proposed the key techniques of image retrieval based on local features in the literature, improve the retrieval accuracy, and also time-efficient. The 2013 TRECVID INS evaluation examined the system's application. Because in some cases, the goal is not very clear, if the whole picture is used for retrieval, the background picture will occupy the majority, so the method of weighting different regions will improve the significance of the retrieval. Huang Chong [5] put forward the semi-supervised reordering algorithm in the literature and tested it with TRECVID INS 2012 as the experimental data. The results show that the algorithm can ensure the accuracy and recall rate increase and improve the quality of the retrieval effectively. Li Xuan [1] put forward the key technologies of the video repetitiveness detection system in the literature and tested using the database of mission CBCDs in TRECVID 2011.

The system mainly includes key frame extraction and lens segmentation, feature extraction, feature matching. Video repetitive detection is a challenging research topic. In addition, video sequence level video repeatability detection algorithm is also proposed, and the method of optimizing histogram and the matching method based on diagonal method are proposed to obtain the video sequence level. After the TRECVID 2011 CBCD public database test shows that based on the diagonal method can better detect video duplication. The literature has good results for both accuracy and recall.

Tag-based text retrieval method can't detect the copyright infringement issues, so Zhang Yang [6] in the literature proposed content-based similar audio and video detection. The evaluation data are all from the data of TRECVID in 2011, and the test results have a high degree of correct matching.

Wu Siyuan [7] proposed content-based repeated audio and video detection in the literature. This paper proposed a time pyramid matching algorithm, the individual experimental results of each feature and the experimental results after fusion. Kong Weiting [8] proposed the video semantic concept detection based on topological independent component analysis and Gaussian mixture model in the literature. The GMM super vector was sent to support vector machine for video semantic concept detection. Through the analysis of TRECVID data in 2012, it was found that the accuracy of video semantic detection is improved.

Lin Huihuang [9] proposed a convolution neural network acceleration method based on pre-decision. This method makes full use of the redundancy of CNN features and discrimination of multi-layer features, and establishes a cascade classifier based on CNN, which achieves the effect of acceleration. Wang Weiwei [10] proposed the lens boundary detection algorithm based on SIFT and macroblock types in the literature. The feature information used by the algorithm can be directly extracted, convenient and quick. Using TRECVID data detection, proved that the detection of high efficiency and speed.

Qu Youjia [11] proposed the key frame extraction algorithm based on SIFT feature in the literature, and proposed a new key frame discrimination method. Then, this algorithm was combined with the SIFT feature point extraction algorithm to design a key frame extraction algorithm based on SIFT feature. Testing through the TRECVID database shows that key frame extraction is better.

3.2 HLFE (High Level Feature Extraction) Special Content of Gene Analysis

HLFE is a genetic analysis technique that detects the specific content and meaning of a video. Video semantic modeling is using low-level video visual features or voice features to analyze the high-level semantic information includes in the video. Video semantic information modeling plays a very important role in video segmentation, summarization, indexing, video content analysis and retrieval. Among the TRECVID International Conference on Video Retrieval hosted by NIST, USA, video semantic information modeling is one of the major evaluation tasks.

Lu Bo [12] in the literature for the image and video data, using TRECVID released semantic concept extraction system to extract semantic features. This system extracts 2617 semantic concepts defined by TRECVID. The approach is to use an associative, multi-concept label learner and a multi-instance based kernel model to obtain semantic concepts by passing multiple labels between adjacent concepts.

Zhou Baiqing [13] proposed a video feature extraction method based on visual features, which can be run directly in the compressed domain to realize the progressive generation of video digests, and the video digests can be generated online. Through the TRECVID test in 2007, the methods proposed in the literature can produce high-quality video summaries.

Mao Yuanhui[2] proposed a semantic concept detection system and participated in the TRECVID2010 high-profile Korean version of the test. The system constructs the relationship between low-level features and high-level semantics. Using the evaluation results of the underlying visual features as weights, the results of each classifier are weighted fused and the test results of high-level semantic concepts are obtained. The test found that the system to improve the detection performance.

3.3 Search

Detection is the ultimate goal of TRECVID, it is based on the HLFE extraction results, detection and search of the matching lens evaluation system. Each question consists of a description of the natural statements and visual examples of the visual images online.

The simple video retrieval method based on low-level visual features obviously can't meet the needs of people. Therefore, more and more video retrieval systems begin to develop the direction of video semantic analysis and multi-modal information fusion, thus forming semantic video analysis and retrieval of opportunity concept. At present, the most mature research on video retrieval system mostly attempts to synthetically analyze both the bottom visualization features and the high-level semantic features to get the video clips that meet the user's query requirements, and then obtain the final query results through effective multi-modal information fusion.

In the fusion phase, the semantic concepts in the previous paragraph are fused with the results of the concept detector to correct the semantic concepts. The test results of TRECVID find that it can effectively solve the problem of indefinite semantic concept expression in video information and improve the performance of semantic video detection.

On the basis of audio research, Wang Lezi [14] put forward a set of audio and video copy detection system in the literature, and participated in the 2011 TRECVID copy detection. In the aspect of high-level semantic retrieval, this system overcomes the problem of "semantic gap", establishes the connection between the bottom visual similarity and the semantic similarity based on random walk algorithm by using the tactics of human-computer interaction and puts forward an interactive retrieval system. And it participated in the 2012 TRECVID Instance Search Contest.

In the video annotation optimization based on probability calculation, ZHONG Cencen[15] passed the test of TRECVID2006-2008 dataset and comparison with other methods shows that the algorithm can effectively reflect the semantic content of the data and improve the performance of video annotation optimization more effectively. On this basis, the literature proposes a context-based audio and video annotation model that can give limitations of probabilistic modeling and labeling method based on probabilistic potential semantic analysis in video annotation. The TRECVID 2010 test shows the efficiency of this model test.

3.4 Surveillance Event Detection in Monitor Video

The SED includes seven event detections, which are hug, drop, run, call, detach, and gather. According to the number of people involved in the time, the first five events are classified as individual events and the last two events are classified as group events [16].

Wang Menglai [16] put forward CNV-based monitoring video event detection in the literature. In the detection of 2015 TRECVID SED, the paper conducted the testing of the remaining 6 events except for telephone calls and achieved good results, proving The effectiveness of the program.

Huang Yuhui [17] in the 2012 TRECVID evaluation of SED detection method is proposed based on block running events, and selects the spatial and temporal characteristics of optical flow curl feature points and random forest classifier based on. In the evaluation of TRECVID SED in 2013, a pedestrian event detection algorithm based on pedestrian detection and tracking was proposed. The spatiotemporal features based on dense trajectories and cascade SVM classifier were selected. In the 2014 TRECVID SED evaluation, a running event monitoring algorithm based on pedestrian detection tracking and forward-backward motion histories was proposed. Through the comparison of the previous algorithms, it is proved that the performance of the algorithm is improved and it is feasible.

3.5 KIS (known - item search) Video Search

The most popular "network video" that TRECVID joins features video content, style, video quality, resolution and video encoding and decoding methods are very diverse, the video duration is limited to 10 seconds to 5 minutes, while each video has a human-tagged meta-data that includes a description of the keywords and video content.

Zhang Kaiqi [18] proposed in the literature cross-media search key technologies, can use the video images, audio, text

and other multimedia information to complete the video search. The system in the literature is suitable for the search of mass network video, and is evaluated in the TRECVID KIS video search data set. The system has achieved good results and proved the effectiveness of the system.

Table 1 discusses the types of algorithms discussed in the five parts of the TRACVID rating system discussed in this article.

TABLE 1 CONTENT-MODEL/TECHNOLOGY INTRODUCTION

MODEL	TECHNOLOGY INTRODUCTION
SBD(Shot Boundary Detection)monitoring	Based on BoW model
	Key Techniques of Image Retrieval Based on Local Features
	Semi-supervised reordering algorithm
	Video Repetitive Detection System Key Technologies
	Tag-based text retrieval method
	Content Based Similar Audio and Video Detection
	Video Semantic Concept Detection Based on Topology Independent Component Analysis and Gaussian Mixture Model
	Convolution Neural Network Acceleration Method Based on Pre - decision
	Lens Boundary Detection Algorithm Based on SIFT and Macroblock Types
	Key frame Extraction Algorithm Based on SIFT Features
HLFE(High Level Feature Extraction) special content of gene analysis	Semantic concept extraction system
	Video abstract extraction based on visual features
	Semantic Concept Detection System
Search	Audio and video copy detection system
	Video tagging optimization based on probability calculation
Surveillance event detection in monitor video	CNN-based monitoring video event detection
	Block-based method to detect running events
KIS(known - item search)video search	Cross-media search key technologies

4 RELEVANT TECHNOLOGY DEVELOPMENT IN CHINA

In 2015, Zhao Jinlong[3] team at Beijing University of Posts and Telecommunications designed and implemented a sample search system based on BoW model. Research and improvement of feature extraction, similarity measure, weighting of feature items based on foreground information, multi-query fusion and query expansion are carried out. It was tested on the video database of TRECVID INS 2013. Experimental results show that the proposed method improves the retrieval precision significantly compared with the traditional method. And in the TRECVID INS 2014 of 22 teams, ranked sixth, verify the effectiveness of the algorithm. In the same year, Zhang Kaiqi[18], Beijing University of Posts and Telecommunications, in semantic information extraction, extract the conceptual semantics of audio information, text information in images and human faces, common objects and so on, and export them in text form, index video meta-information and video semantic information respectively, completing video retrieval and sorting. The system is tested in the TRECVID KIS video search task data set with MAP value of 0.265, which proves the effectiveness of the system.

Wang Menglai[16] studied the research on the task of incident detection in complex surveillance video in TRECVID-SED evaluation and proposed a method based on CNN cascaded network to detect pedestrian head instead of detecting the whole pedestrian, so as to realize the detection effect compared with the most advanced pedestrian detection method; The method of event detection scheme proposed by CNN for detecting individual events with key gestures can make complex 4D space event detection achieved in 2D space; The scheme based on

trajectory analysis to detect group events and Gaussian process regression are introduced into the trajectory parameter equation fitting so that the tracking algorithm does not have to be limited to the assumption that the trajectory needs to satisfy the linear or quadratic function and is more suitable for describing the actual movement of the target. And in the TRECVID-SED 2015 international assessment, good results were obtained, confirming the effectiveness of the program.

Peking University Professor Peng Yu Xin's team in the 2016 TRECVID international evaluation proposed computer self-search and interactive search, two projects both won the first place in the competition. This is the excellent result attained again after Professor Peng Yu Xin's team won the number one place for many consecutive years since 2009.

In the TRECVID International Assessment concluded in November 2016, the team led by Professor Hu Ruimin achieved excellent results in the case retrieval mission. The team achieved the best result with an average detection rate (MAP) of 0.758 out of 30 official search queries and the team has fully entered the top tier of international video retrieval.

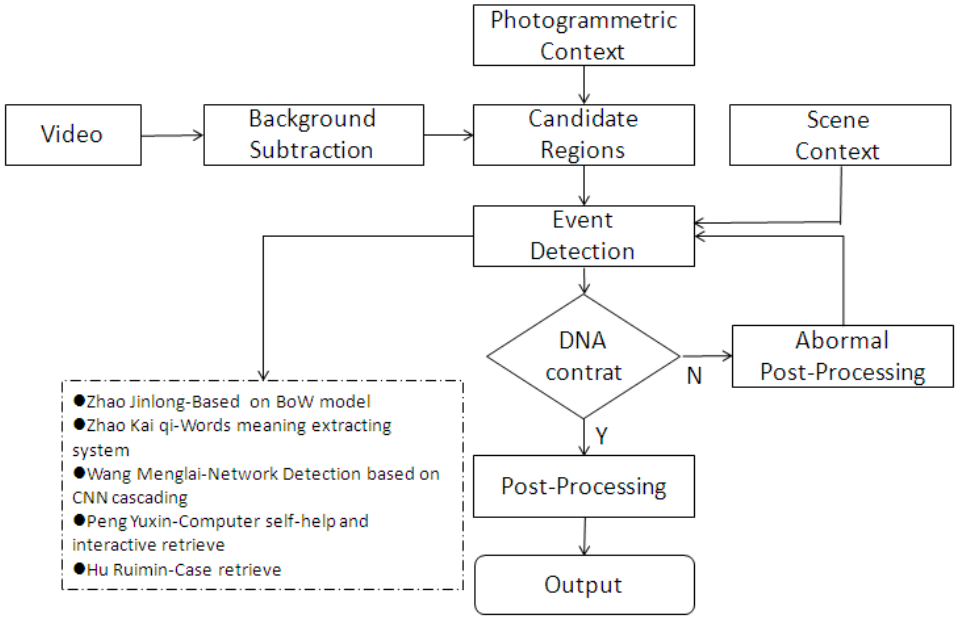

FIGURE 1 THE TECHNOLOGY APPLICATION LOGISTIC DIAGRAM OF VIDEO

TRECVID video gene recognition technology is based on the video detection, and then reduce the background (refers to the process of subtracting the background image) processing, according to the upper and lower semantic environment, select the candidate area, and detect the scene content of the event. If the genetic test is successful, it will be handled according to the normal channel; otherwise, it will be handled according to the abnormal channel and the event detection will be restarted.

5 CONCLUSION AND INSPIRATION

In this paper, through TRECVID Video Gene Analysis on Shot boundary detection, High-level feature extraction, semantic-based video search, Surveillance event detection, and Known-item search conducted a technical performance evaluation of the domestic situation and analysis of the following conclusions:

- TRECVID video gene identification analysis technology has been developed in the country through advanced image search and image decoding technology to provide high-quality, open degree of technical specifications. Through the modular, scalable and consistent to reflect the copyright protection market demand and reduce its development costs. The result of TRECVID certification can be used as a guide to product development, presenting customers with objective and trustworthy data. In the domestic video gene identification technology, an object example search system based on the BoW model and a CNN-based cascade network detection have

completed the video gene recognition technology from different perspectives.

- Domestic Tsinghua University, Peking University, Beijing University of Posts and Telecommunications and other teams have achieved good results in the field of video gene analysis technology and realized a breakthrough in the field of feature extraction algorithms to improve the detection accuracy, and speed up the detection speed.
- Most of the domestic TRECVID video gene recognition and analysis technologies are based on shot boundary detection, high-level feature extraction, semantic-based video search, Surveillance event detection and Known-item search to study. Camera motion detection, Rush summarization, Content-based copy detection, Event detection in internet multimedia and other technical gaps in the field of research should be strengthened, forming a more accurate, faster, and better functional module video gene recognition technology.

TRECVID video gene analysis technology, by applying technology to the field of copyright protection, provides the value and profit to the video owner and protects the copyright of the video copyright holders in our country, which is very important for guiding our country to actively participate in the video development.

REFERENCES

[1] Li Xuan, The Research on the Algorithm of Content-Based Video Copy Detection[D]. Tianjin, Civil Aviation University of China, 2013: 35-49

[2] Mao Yuanhu, Content Based Video Retrieval[D]. Beijing, Beijing University of Posts and Telecommunications, 2012: 5-12

[3] Zhao Jinlong, Key Technology OF Large Scale Cross-Media Data Retrieval[D]. Beijing, Beijing University of Posts and Telecommunications, 2015: 19-28

[4] Cen Shusheng, Image Retrieval with High Level Semantics[D]. Beijing, Beijing University of Posts and Telecommunications, 2014: 9-16

[5] Huang Chong，Semi-Supervised Content-Based Image Retrieval[D]. Beijing, Beijing University of Posts and Telecommunications, 2013: 14-30

[6] Zhang Yang, Content-Based Audio Video Searching and Commercial Detection[D]. Beijing, Beijing University of Posts and Telecommunications, 2014: 25-51

[7] Wu Siyuan, A Mulit Modal Conteit-Based Copy Detection Approach[D]. Beijing, Beijing University of Posts and Telecommunications, 2013: 24-40

[8] Kong Weiting, Video semantic detection based on topographic independent component analysis and Gaussian mixture model[J]. 2016: 4-6

[9] Lin Huihuang, Abstract[D]. Beijing, Beijing University of Technology, 2016: 28-40

[10] Wang Weiwei, Video Shot Boundary Detection Based on Mpeg Compressed Domain [D]. Hebei, YANSHAN University, 2016: 30-41

[11] Qu Youjia，ABSTRACT[D]. Beijing, Beijing JIAOTONG University, 2015: 15-29

[12] Lu Bo, Study on Techniques of Multi-Modality Media Information Retrieval[D]. Shenyang, Northeastern University, 2012: 15-26

[13] Zhou Baiqing, Online video abstract extraction based on visual features in compressed domain[J]（Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition)）, 2016: 3-6

[14] Wang Lezi, Research of Content-Based Large-Scale Audio and Video Data Search[D]. Beijing, Beijing University of Posts and Telecommunications, 2012: 7-12

[15] Zhong Cencen, Research on Context-Based Audio and Video Annotation[D]. Beijing, Beijing Jiaotong University, 2014: 87-91

[16] Wang Menglai, Surveillance Event Detection Based on CNN[J]. Automation Journal, 2016, 42(6): 892-903

[17] Huang Yuhui, Event Detection Algorithm of Surveillance Video [D]. Beijing, Beijing University of Posts and Telecommunications, 2014: 5-7

[18] Zhang Kaiqi, Key Technologies on Cross-Media Retrieval[D]. Beijing, Beijing University of Posts and Telecommunications, 2015: 9-15

AUTHORS

CUI Jizhe is a professor of Dept. of Information Management and Systems at YANBIAN University in China. He received the Undergraduate course Degree from Northeast Normal University, Major in Mathematics Education in 1996. Received the Ph.D. degree from Sangmyung University, major in Computer Science in 2007. His research interests are digital watermarking, copyright management, Data analysis and decision making, Data Modeling, and BigData Analysis.

Quan jinlan, graduate of technical economics and management in YANBIAN University. Major in project management and logistics supply chain management. E-mail: quanjinlan@foxmail.com.