systems in cloud computing environments[J]. Expert Systems With Applications,2014,4: 113-119.
4. Younis A. Younis, Kashif Kifayat, Madjid Merabti. An access control model for cloud computing[J]. Journal of Information Security and Applications,2014, 7: 22-33.
5. Yizhuo Wang, Yang Zhang, Yan Su, Xiaojun Wang, Xu Chen, Weixing Ji, Feng Shi. An adaptive and hierarchical task scheduling scheme for multi-core clusters[J]. Parallel Computing,2014, 10: 909-911.
6. Acklesh Prasad, Peter Green, Jon Heales. On governance structures for the cloud computing services and assessing their effectiveness[J]. International Journal of Accounting Information Systems,2014,154.
7. Yaser Jararweh, Moath Jarrah, Mazen kharbutli, Zakarea Alshara, Mohammed Noraden Alsaleh, Mahmoud Al-Ayyoub. CloudExp: A comprehensive cloud computing experimental framework[J]. Simulation Modelling Practice and Theory,2014, 2: 49-51.
8. Farrukh Shahzad. State-of-the-art Survey on Cloud Computing Security Challenges, Approaches and Solutions[J]. Procedia Computer Science,2014, 6: 37-45.
9. Jeffrey Dean, Sanjay Ghemawat. MapReduce: Simplified data processing on large clusters. Communications of the ACM 2008; 1: 55-56.
10. Dhruba Borthakur. HDFS Architecture. The Apache SOFTWARE foundation 2008; 32: 899-912.

Study on Video Fingerprint Algorithm Based on Luminance Structured Quality Assessment

Ning Liu[1], Qiuyun Cheng[1]

Department of Computer Science and Application, Zhengzhou Institute of Aeronautical Industry Management, Zhengzhou,China

Corresponding author is Ning Liu

Abstract
This paper focuses on research about feature extraction, fingerprint modeling and similarity matching of video fingerprinting. Three digital video fingerprint algorithms are presented in this paper. A fingerprint algorithm based on relative orientation invariant between geometric centroid is suggested. The perceptual centroid means that the same perceptual content video should have the same centroid, further, the orientation of the centroid is identical, and the relative orientation keeps unchangeable. In this method, the geometric center of every frame is used as original point and the orientation of centroid is calculated based on luminance geometric centroid. Converting orientation from the original frame rate to a fixed frame rate through temporal distance and a new orientation vector

is generated. The next, an invariant orientation is created through successive two orientation subtracting. At last, video fingerprint is generated by concatenating all of the invariant orientations. The experimental results show that the proposed algorithm is robust against lossy compression, frame rate change, resizing, rotation, cropping, global change in brightness and Gaussian white noise. The quality evalution using structural luminance comparison improves the visual effect.

Key words: VIDEO FINGERPRINTING, LUMINANCE STRUCTURED, FEATURE EXTRACTION, FINGERPRINT MATCHING, SIMILARITY MEASUREMENT

1. Introduction

In the last decade, along with the rapid development of information technology, more and more multimedia information is available, especially for video data. The proliferation of digital videos has made accessibility of video contents easier and cheaper while being the source of many problems, e.g., how to get the interested video from the vast amount of video data. At the same time, it becomes more complicated for content management because of the huge and ever increasing amount of videos [1]. Keywords-based traditional distinguishing and retrieval method cannot work efficiently anymore, it is important to find a new method for video content discrimination, retrieval, monitoring services and authentication. Video fingerprinting is a technique that can be used to solve these problems and for this reason it has become a hot research topic in recent years.

A video fingerprint is an identifier that is extracted from a piece of video content and discriminates one video from the others [2]. The process of extracting a fingerprint from the video content is referred to as fingerprinting the video or video fingerprinting. Pereptual video fingerprint means if two video clips are perceptually different, the fingerprints extracted from them should be considerably different, and if two video clips are perceptually samilar, the fingerprints extracted from them should be considerably identical [3].

With the technology development of computer and communications, broadband network, audio and video compression, and computer hardware, digital video is streaming into the common life. However, since the diversity of video content and the high-dimensionally spatiotemporal structure of video data, it becomes crucial for efficient organization, management, storing, rapid retrieval and browsing of video data, while the traditional data management and retrieval method become unsuccessful. As a result, the content-based video retrieval (CBVR) emerges. Meanwhile, with the popularity of webcams and the increasing of people to safety awareness, the video surveillance and intelligent analysis of vid}:o surveillance have become increasingly urgent demand.

2. Related technology

Recently, CBVR and intelligent analysis of video surveillance have made great achievements in many aspects, however, since some important problems remain unresolved in such field as semantic object extraction, the understanding of video content, etc., large-scale applications have not come true [4,5]. As a result, in this paper, we focus on the difficulties of content-based video analysis and retrieval-the problem of high-level semantic extraction. We propose some new frameworks and methods for low-level feature extraction, semantic object extraction, evaluation protocols, event detection and surveillance the use of medical device. (Eq. 1)

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

There are many mature visual low-level features, which are the basic unit in many tasks, but there are rare reports about how to choose visual low-level features, especially for the task of high-level semantic extraction in large-scale dataset. Therefore, in large-scale dataset we completely evaluate the performance of four different kinds of low-level features in the task of high-level semantic extraction, and then we can know how to choose the low-level features for high-level semantic extraction according to the results. After that, we find that key points feature plays an important role in high-level semantic extraction, thus focus on key points feature. From it, we discover that there is some difference and complementation when different sub-sampling rates are used in SIFT detection, therefore, we propose a multi-level SIFT (Multi-Layer-SIFT, called ML-SIFT) algorithm. Experiments on Calatech256 and Sceneclassl3 show that ML-SIFT algorithm is effective, whose performance is better than that of SIFT and SURF. (Fig.1)

Although there are many different decision-level fusion algorithms which have achieved good results, however, the performance improvement of some fiction algorithms are not very obvious, and some are helpful for only part of semantics, but it is useless for other semantics. Therefore, based on the best choice

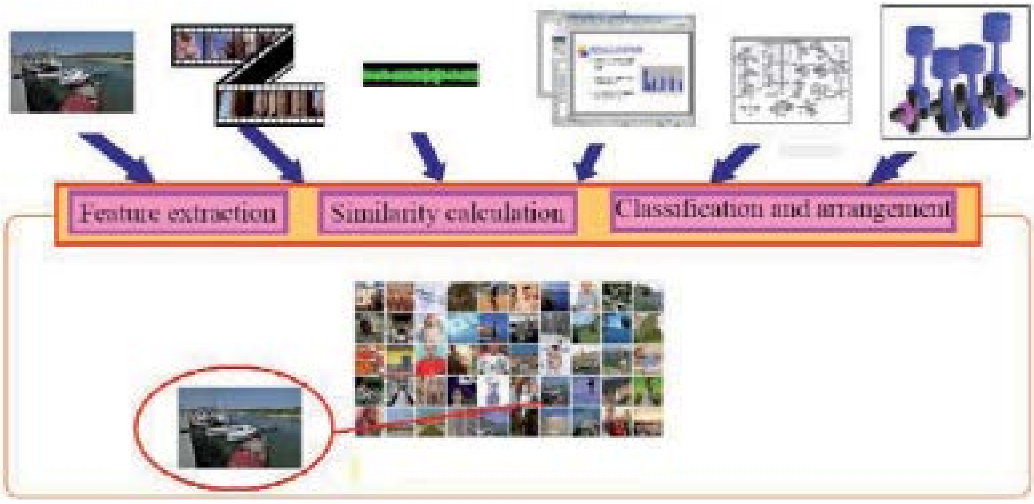

Figure 1. Multimedia Retrieval System

of hybrid fusion algorithm-BCHFA is proposed. Experiments on TRECVID 2008 show that the improvement of BCHFA is the best, and it is helpful for all semantics. (Fig.2)

What is more important, BCHFA is open to other fusion algorithms, if some fusion algorithm is excellent, it can ‘ue easily integrated into BCHFA . At the same time, as the performance of algorithms is affected by different annotations and there is some difference between different annotations, thus, based on different tagging of fusion algorithm-DTFA is proposed. Experiments on TRECVID 2008 dataset display that DTF.A is useful object-oriented high-level semantic, and the performance of more than 90% semantics can be improved. (Fig.3)

3. Fingerprint Algorithm

There are four main parts in the algorithm realization: Design and implement of the preprocessing method for the input videos. Specifically, the preprocessing consists of video segmentation, key frame and key shot extraction and the image processing of exception detection, black side removal and normalization. Design and implement features utilized for web video quality assessment which includes three

Figure 2. TinEye Retrieval System

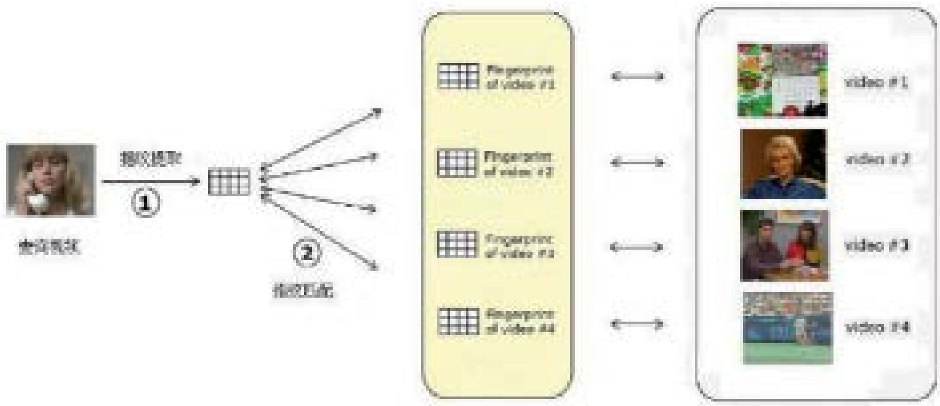

Figure 3. Video Fingerprinting System

aspects. They are spatial feature based on information in frames, temporal feature based on global motion estimation and structure feature based on the shot scale. Train and test the performance of the designed features with support vector machine (SVM) and Adaboost classifier. In this paper, the author adopts the famous library for SVM classifier and realizes the Adaboost classifier by his own hand. Gather and design a dataset for testing the performance of video quality assessment algorithm. In this dataset, there are 2498 labeled videos and the total. 1’ength is than 122 hours. Then the experiment is based on SVM and Adaboost classifiers. According to the experiment result, the merit and. Deficiency and its reason are analyzed. In the end, some comments and advises are given. (Eq. 2-5)

$$S(X,Y \mid \varepsilon)=\frac{\sum_{i=1}^{K} 1 \mid d(x_i,y_i)\le \varepsilon}{K} \tag{2}$$

Where $d = D = \sum_{i=1}^{n} | x_i - y_i |$.

$$r(X,Y)=\frac{\sum (x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum (x_i-\bar{x})^2 \sum (x_i-\bar{x})^2}} \tag{3}$$

$$\theta_{ki} = \sin^{-1}\left(\frac{c_{yki}-c_{myki}}{M_{ki}}\right) \tag{4}$$

Where

$$M_{ki}=\sqrt{(c_{xki}-c_{mxki})^2+(c_{yki}-c_{myki})^2}\ .$$

$$M_f(k,v)=\frac{1}{2\pi}\int_0^{\infty}\int_0^{\infty} f(r,\theta) r^{-iv} e^{-ik\theta} d\theta \frac{dr}{r} \tag{5}$$

Although many action recognition algorithms were proposed, the performance assessment of them does not have unified platform, what is worse, researchers have not noticed that the performance is affected by different evaluation protocols, which is detrimental to the development of related technology. In this chapter, we take MoSIFT features and SVM classifier as basic recognition algorithm, and then well assess the effect of different evaluation protocols to the performance of action recognition algorithms on widely used public dataset KTH. Experimental results show that when different cross-experimental methods are adopted, the performance of the algorithm becomes very fluctuant. when 1 times cross-experimental methods is used, the maximum fluctuation reaches 10.5%; The fluctuantion is also large when leave-one-out cross-experimental method and different times cross-experimental method is borrowed respectively, the maximum fluctuation reaches 7.926%; As for n cross-experimental methods, when the times of cross-experimental methods is equal or greater than 25, its difference can be ignored; In addition, Experiments show that different division methods for datasets have a greater influence to the performance of algorithms. (Table 1)

Tab. 1. Video Frame Formats and Bitrate

Format	Resolution	Bit/s
1080HD	1920*1080(16:9)	24883200
720HD	1280*720(16:9)	11059200
4CIF	704*576	4866048
CIF	352*288	1216512
QCIF	176*144	304128
SUB-ACIF	128*96	147456

Computational results and comparisons

Thus, we should adopt same data division methods when evaluating the performance of algorithms. If we must assess its performance under different data division methods, the leave-one-out cross-experimental method can make the error small. (Fig.4-5)

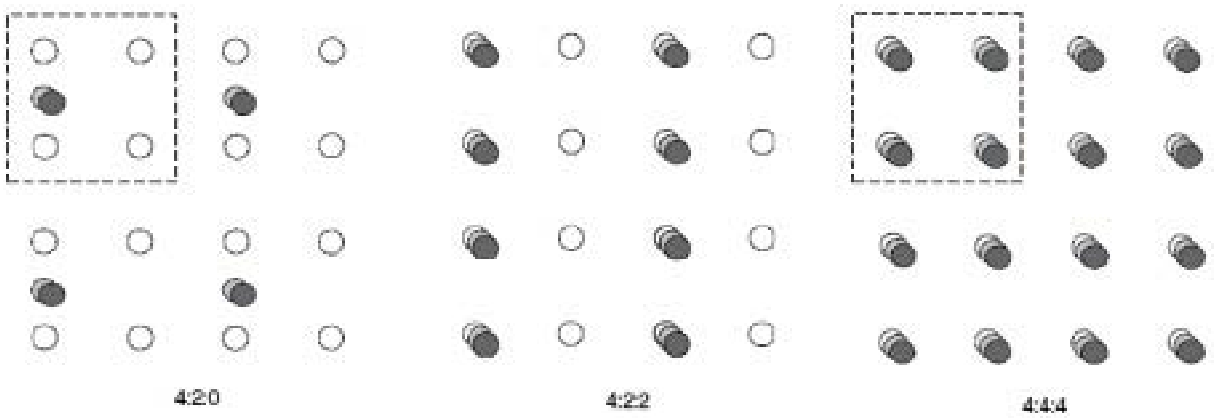

Figure 4. 4:2:0, 4:2:2 and 4:4:4 sampling patterns

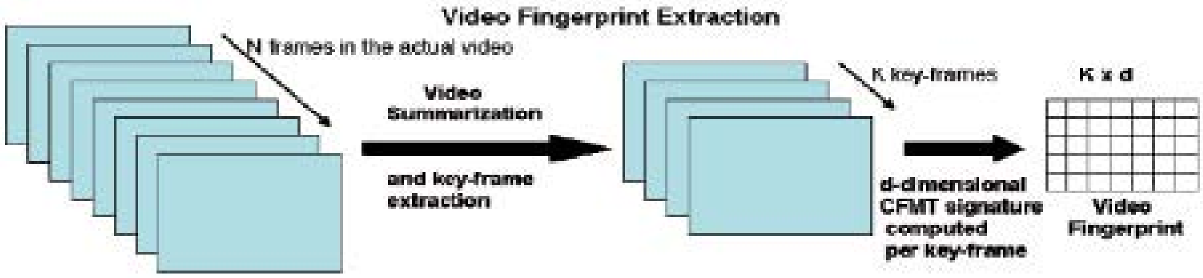

Figure 5. Generation of CFMT-Based Signature

The problem of data imbalance exists in our daily life, and it affects the performance of classifiers, thus, in this chapter, on the basis of analyzing the problems of existing algorithms for processing data imbalance, we put forward enhanced and hierarchical structure algorith}-n (EHS), which takes the algorithm of sampling, filtering and training as together, and hierarchical-level structure is integrated. Experimental results on TRECVID 2010 dataset show that the performance of EHS is much better than that of random down-sampling algorithm and ensemble down-sampling algorithm, with the increasing of numbers of level layers, its performance can improve stably, and has good stability; As for different kinds of features, the performance of EHS algorithm shows that it is robust and stable; When using EHS algorithm into TRECVID 2010 SED task, we achieve top performance in four events.

This paper introduces the basic theory and knowledge of video hash algorithm, including the data feature of a video, the definition of video hash and the evaluation of video hash algorithm. Then this paper introduces the video hash algorithm based on video shot segmentation and proposes the video hash algorithm based on the temporal-spatial visual attention weight. (Table 2)

Tab. 2. Computational Complexity Comparision Under Unchanged Frame Rate

ID	Resolution	Text	Method	Complexity
akiyo	352*288	451.164s	365.443s	0.81
forman	352*288	477.947s	377.578s	0.79
news	352*288	454.547s	368.183s	0.81
silent	176*144	224.227s	368.183s	0.84
container	176*144	225.716s	187.344s	0.83
carphone	176*144	223.443s	187.692s	0.84

4. Quality assessment

Algorithm considers that the frame content in the same shot is almost visually similar and the frame content in the different shots is quite different. Therefore, during the process of the temporally representative frame generation, only the fixed number segment can not well describe the video content and reduces the robustness and discrimination of the extracted video hash. The temporally representative frame based on the video shot segmentation is generated to obtain further video hash, which brings the video hash better performance on the description of video content, compactness, robustness and discrimination. (Eq. 6)

$$Q = \frac{1}{KI}\sum_{k=1}^{K}\sum_{i=1}^{I} l_{k,i} \qquad (6)$$

Where $l_{k,i} = \dfrac{2\mu_{k,i}v_{k,i}}{\mu_{k,i}^2 + v_{k,i}^2}$

A video hash algorithm based on temporal-spatial visual attention weight is proposed. The traditional weighting methods for video are index weighting or average weighting. Although theses methods are simple, they can not take fully into account different visual attention degrees for different video proposed weighting method generates the temporally visual weight according to eyes' visual attention change for video content change, and then obtains the temporally representative frame that can reflect the attention degree. The video hash extracted from these temporally representative not only reflects the important degree of the video attention content but also improves the performance of video hash.

5. Conclusion

In this paper, presented is a new video fingerprint base on mean, instead of using transform for further processing. Usually the video fingerprint only can discriminate one clip from others, but it is difficult to assess the video quality between the identical content video. The input video is firstly converted into a new video with a fixed frame and a fixed size. Secondly, the new video is partitioned by a macroblock(16*16). At last, fingerpring is formed by normalizing the luminance mean that is calculated from every macroblock. The experimental results show that the proposed algorithm is robust against lossy compression, resizing, frame rate change, global change in brightness and gaussian white noise. Wavelet transform is usually used in digital image processing, because it can extractes some characteristic parameters in this frequence domain. A complex wavelet is constructed based on a low-pass filter and the discrete wavelet coefficients are achieved through sampling of the continuous wavelet transform in this video sequence. The discrete wavelet coefficients are used as video fingerprints. The proposed algorithm is robust against resizing, small rotation. The experimental results show that the proposed algorithm has a low false negative rate and false positive rate.

Acknowledgements

This work is supported by the basic and frontier technology research project of Henan and Technology Department(132300410186) and the basic research project of Henan and Technology Department(2014B520067).

References

1. Wootton, C., A practical guide to video and audio compression: from sprockets and rasters to macroblocks. Taylor & Francis, 2005.
2. Staab, S., Structuring and Accessing Semantic Multimedia Data. In Advances in Multimedia Modeling. Springe r Berlin Heidelberg, 2009, pp. 2-2.
3. Esmaeili, M. M., Fatourechi, M., & Ward, R. K., A robust and fast video copy detection system using content-based fingerprinting. Information Forensics and Security, IEEE Transactions on, 2011.,6(1), pp. 213-226.
4. Law-To, J., Chen, L., Joly, A., Laptev, I., Buisson, O., Gouet-Brunet, V., & Stentiford, F., Video copy detection: a comparative study. In Proceedings of the 6th ACM international conference on Image and video retrievalACM. 2007, 7, pp. 371-378.
5. Sarkar, A., Ghosh, P., Moxley, E., & Manjunath, B. S., Video fingerprinting: features for duplicate and similar video detection and query-based video retrieval. In Electronic Imaging, 2008, 11, pp. 682-689.