A Video Copy Detection System based on Human Visual System

Yu Bai[1], Li Zhuo[1], YingDi Zhao[1] and Xiaoqin Song[2]
[1]Singal & Information Processing Laboratory, Beijing University of Technology, Beijing, China
[2]College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China

Keywords: Itti Attention Model, K Neighbor Search, Near-duplicate Detection, Surfgram.

Abstract: The technology of near-duplicate video detection is currently a research hot spot in the field of multimedia information processing. It has great value in the areas such as large scale video information indexing and copyright protection. In the case of large-scale data, it is very important to ensure the accuracy of detection and robustness, in the meanwhile improving the processing speed of video copy detection. In this respect, a HVS(Human Visual System)-based video copy detection system is proposed in this paper.This system utilizes the visual attention model to extract the region of interest(ROI) in keyframes, which extracts the Surfgram feature only from the information in ROI, rather than all of the information in the keyframe, thus effectively reducing the amount of the data to process. The experimental results have shown that the proposed algorithm can effectively improve the speed of detection and perform good robustness against brightness changes, contrast changes, frame drops and Gaussian noise.

1 INTRODUCTION

With the rapid development of multimedia technology,videos are widely used in business, entertainment and many other fields.The management of these data becomes a challenging task(Wang, 2010).Online users can download, upload and modify the videos in convenience,which also broughts many problems.Such as illegal modification of the online video(near duplicate video).To protect intellectual property rights, it's important to detect these near duplicate videos.

Nowadays most of the domestic and international researches focus on finding a variety of complex feature extraction method to improve the detecion accuracy. However, in practical, the more important problem is how to significantly improve the speed of the detection system while maintaining high detection accuracy and robustness at the same time.

Currently the mainstream technology of video copy detection is the content-based video copy detection methods. Content-based video copy detection technology can be grouped into two categories: global feature-based methods and local feature-based methods.

Global feature-based methods extract frame-level signatures to model the information distributed in spatial and color dimensions. The keyframe is divided into 64 regions of the same size and for each region Y component is further extracted as the video features (Wu, 2006). But this method is unable to resist strong size transformation attack. The OM (Ordinal Measures) feature is applied to image copy detection (Kim, 2003). However, though it is robust against global changes, local changes will disrupt the relative relationship between the image blocks, which makes this method fail to work.

In comparison to the global feature-based methods, local feature-based method has stronger robustness. However, the time consumption and computational complexity is unacceptably high for the practical application. For example, Douze presented a video copy detection system based on local features (Douze, 2008), in which the local descriptors are over 400 in each frame. In a 4-minute video with fast changing scenes, the keyframe may be more than 400 and the local descriptors may be more than 160K, which would result in high computational cost on keyframes matching and poor real-time performance.

In order to improve detection efficiency,we propose a novel system based on human visual in this paper.Unlike existing video copy detection systems, the proposed method only deals with ROI(region of interest) in keyframes and then extracts features to determine video similarity, which can effectively reduce the complexity of the

Bai Y., Zhuo L., Zhao Y. and Song X..
A Video Copy Detection System based on Human Visual System.
DOI: 10.5220/0004292107920795
In *Proceedings of the International Conference on Computer Vision Theory and Applications* (VISAPP-2013), pages 792-795
ISBN: 978-989-8565-47-1

algorithm while ensuring the robustness.

2 VIDEO COPY DETECTION SYSTEM BASED ON HVS

The architecture of our approach contains two processes: feature extraction and similarity comparison. First keyframes are extracted, and then visual attention model is employed to extract ROI, followed by a Surfgram feature extraction to represent the video content which can effectively reduce the amount of data to process. In the video similarity comparison, this paper uses BID+ (bit-difference) approximate nearest neighbor search algorithm to improve the matching speed. The following sections will describe the preprocessing steps in detail.

2.1 Video Feature Extraction

As the information of video data is very large, it is critical to decide which video features should be used to represent the video content. In order to improve the speed of the copy detection system, keyframes are used to represent video content.

2.1.1 Keyframe Extraction

In order to reduce data redundancy, we first extract the keyframes and employ the method of abrupt shot change detection (Hou, 2009). Then we extract keyframes between the consecutive abrupt shots uniformly, specifically, a keyframe is extracted from about every 30 to 100 frames in our experiments.

2.1.2 the ROI Extraction

It has been found that the HVS has some certain selectivity. This selectivity in HVS indicates the eye movements and form the focus of attention or ROI.VAM (Visual Attention Model) is based on the HVS. Therefore, the visual attention model can be used to extract ROI and reduce the amount of information to be processed.The most classic VAM is Itti Model proposed by Laurent Itti (Itti, 1998). Itti model extracted from the input image with many features, such as intensity, color, direction, forming conspicuity maps of features. The saliency map is a linear combination of three conspicuity maps. Conspicuity maps are based on the center-surround differences.For this reason, the region beyond a certain threshold T in saliency map is regarded as the ROI. We use T=0.6 as the threshold.After the extraction of keyframes,ROI further reduces the amount of data and improve the processing speed.

2.1.3 Surfgram Feature Extraction

In this paper, we propose Surfgram feature to represent the content of ROI because the feature is fast and robust. SURF(Speeded Up Robust Features) is several times faster than SIFT and can be robust against brightness changes, contrast changes, Gaussian noise (Bay, 2008).Therefore, we extract SURF from ROI and generate a frame-level histogram of visual codewords based on SURF to construct the Surfgram feature.

As there are a lot of feature points in an image, SURF features are of large amount of data. Therefore, Surfgram feature is adopted only to represent the content of ROI. The extraction steps are as follows:

- Extracting SURF features from ROI.
- Using the K-means clustering approach to partition SURF features into 200 clusters.
- Computing the histogram of the features. The number of the SURF features in class 1 to class n is counted to obtain the histogram of SURF features, which is the Surfgram feature.

2.2 Surfgram Features Matching and Scoring

When comparing the similarity of Surfgrams from the query video and reference video, if the similarity between them is greater than a certain threshold, then the query video is identified as a near-duplicate of the reference video. In order to speed up feature comparison, we use the BID+ for Approximate Nearest Neighbor to index all features (Cui, 2005).We define the similarity measure score as follows:

$$score = \frac{\sum_{i=1}^{n} q_i}{Q_i} + 30 \times \frac{1}{\sum_{i=1}^{n} D_i} \quad (1)$$

where q_i is the matched query frame, n is the total number of matched query frames, Q_i is the total number of keyframs in reference video, D_i is the distance between the query frame and matched frame in the reference video.

3 EXPERIMENTAL RESULTS

To evaluate the performance of the proposed

algorithm, a database of 50 videos (collected from TRECVID2011) is created. Videos in the database were attacked to generate query videos.Table 1 shows the attacks and their parameter values, where U is the average luminance value and G(0,u) is Gaussian function.

Table 1: Attack parameter values used in this paper.

Attacks	Effect	Parameter values
Brightness	Output=Input+A U	A=0.35
Contrast	Output=U+(Input-U)(1+P)	P=0.7
Frame drops(f)	f% of the frames are randomly dropped	f=30
Noise	Output=Input+G(0,u)	u=20
PIP	position:top left	scale:16.6%
Ratio	compression	scale:20%

Table 2 shows the experimental results.Our experimental results have shown that the proposed algorithm performs good robustness against brightness changes, contrast changes, frame drops and Gaussian noise. Fig.1(a) is one of the keyframes in query video and Fig1.(b) is the result. Although they have slightly different contents,they are both extracted from the same video.

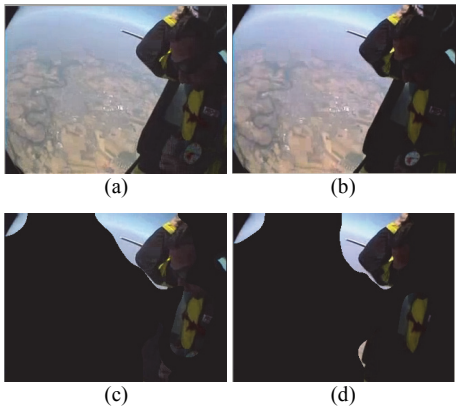

Figure 1: Example of query clip and result. (a) original image.(b)Contrast attack.(c)ROI of (a).(d) ROI of (b).

Table 2: Results of the proposed algorithm.

Attacks	Accuracy rate(the number of correct/total number)
Brightness	76%
Contrast	78%
Frame drops	82%
Noise	92%
PIP	50%
Ratio	52%

The reason of the good performance is mainly resulted from the SURF feature which is robust against brightness changes, contrast changes, frame drops and Gaussian noise.

The attacks for which our method shows less ideal performance are PIP and Ratio.Fig.2(b) and Fig.2(d) shows that ROI of parts of the keyframes are very different.And this increases the differences of Surfgram features.

Feature extraction is the most time-consuming computation part.The time cost of feature extraction depends not only on the single feature extraction time, but also depends on the number of the features. SURF feature is faster than SIFT (Apostol, 2010). In addition, we use Itti model to extract ROI, which reduce the number of features. The number of features in the first video of the database with Itti Model is 7261. While the total number of features is 40092 without Itti Model (Apostol, 2010). The average time cost to extract Surfgram features from keyframe in this paper is 554.3ms, while the average time (Apostol, 2010) is 1367.1ms, which is as 2.46 times the duration as our proposed algorithm.Therefore, the proposed algorithm performs much better in the aspect of time cost compared to the algorithm (Apostol, 2010).

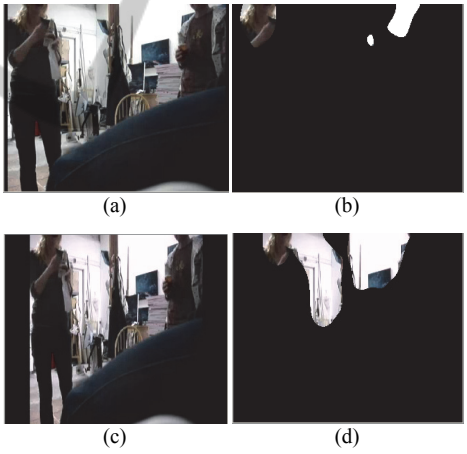

Figure 2: Comparison of ROI. (a) original image.(b)ROI of the original image.(c)Ratio transform of the original image.(d)ROI of the ratio image.

4 CONCLUSIONS

In this paper, we propose a video copy detection algorithm based on HVS. Experimental results have shown that the proposed algorithm is very fast and robust against brightness, contrast change, frame drops and Gaussian noise attack, but sensitive to PIP and ratio attack. In the future work, we will focus on the robustness of Itti Model and Surfgram features.

ACKNOWLEDGEMENTS

The work in this paper is supported by Program for New Century Excellent Talents in University (No.NCET-11-0892), Doctoral Fund of the Ministry of Education, the National Natural Science Foundation of China (No.61003289,No.61100212) , the Natural Science Foundation of Beijing (No. 4102008), the Excellent Science Program for the Returned Overseas Chinese Scholars of Ministry of Human Resources and Social Security of China, Scientific Research Foundation for the Returned Overseas Chinese Scholars of MOE, Youth Top-notch Talent Training Program of Beijing Municipal University, the Fundamental Research Funds for the Central Universities (No.NS2012045).

REFERENCES

Meng Wang., 2009. Unified Video Annotation Via Multi-Graph Learning. *IEEE Transactions on Circuits and Systems for Video Technology*.

Wu Ming-Ni., Lin Chia-Chen., Chang Chin-Chen., 2006. A Robust Content based Copy Detection Scheem. *Fundamenta Informaticae*.

Kim, C., 2003. Content-based Image Copy Detection. *Singal Processing: Image Communication*.

M, Douze., A, Gaidon., H, Jegon., M, Marszatke., and C, Schmid., 2008. Inria-Learrs Video copy detection system. *In TRECVID*.

Y, Hou., H, Z, Hou., 2009. Shot segmentation method based on intensity histogram frame difference. *Computer Engineering and Applications*.

L, Itti., C, Koch., Emst Niebur., 1998. A Model of Saliency-Based Visual Attention for Rapid Scene.Analysis. *IEEE transactions on pattern analysis and machine intelligence*.

Bay, H., Tuytelaars, T., Van Gool, L., 2008. Speeded-Up Robust Features(SURF). *Computer Vision and Image Understanding*.

B, Cui et al., 2005. Exploring bit difference for approximate KNN search in high dimensional databases. *Proceedings of the 16th Australasian database conference*.

N, Apostol., H, Matthew., S, John R., 2010. Design and evaluation of an effective and efficient video copy detection system. *2010 IEEE International Conference on Multimedia and Expo*.