Signal Processing: *Image Communication* 29 (2014) 788–806

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

journal homepage: www.elsevier.com/locate/image

Adaptive weighted fusion with new spatial and temporal fingerprints for improved video copy detection

Semin Kim [a], Jae Young Choi [a], Seungwan Han [b], Yong Man Ro [a,*]

[a] *Image and Video Systems Lab, Department of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Yuseong-Gu, Daejeon 305-701, Republic of Korea*
[b] *Cyber Security-Convergence Research Laboratory, Electronics and Telecommunications Research Institute (ETRI), 218 Gajeongno, Yuseong-gu, Daejeon 305-700, Republic of Korea*

ARTICLE INFO

Article history:
Received 24 January 2013
Received in revised form
14 April 2014
Accepted 2 May 2014
Available online 23 May 2014

Keywords:
Video copy detection
Video sequence matching
Modality fusion
Video fingerprint
Weighted adaptive fusion
Spatial and temporal information

ABSTRACT

In this paper, we propose a new and novel modality fusion method designed for combining spatial and temporal fingerprint information to improve video copy detection performance. Most of the previously developed methods have been limited to use only pre-specified weights to combine spatial and temporal modality information. Hence, previous approaches may not adaptively adjust the significance of the temporal fingerprints that depends on the difference between the temporal variances of compared videos, leading to performance degradation in video copy detection. To overcome the aforementioned limitation, the proposed method has been devised to extract two types of fingerprint information: (1) spatial fingerprint that consists of the signs of DCT coefficients in local areas in a keyframe and (2) temporal fingerprint that computes the temporal variances in local areas in consecutive keyframes. In addition, the so-called temporal strength measurement technique is developed to quantitatively represent the amount of the temporal variances; it can be adaptively used to consider the significance of compared temporal fingerprints. The experimental results show that the proposed modality fusion method outperforms other state-of-the-arts fusion methods and popular spatio-temporal fingerprints in terms of video copy detection. Furthermore, the proposed method can save 39.0%, 25.1%, and 46.1% time complexities needed to perform video fingerprint matching without a significant loss of detection accuracy for our synthetic dataset, TRECVID 2009 CCD Task, and MUSCLE-VCD 2007, respectively. This result indicates that our proposed method can be readily incorporated into the real-life video copy detection systems.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Recently, a huge amount of video contents have been created and distributed by internet users, mainly due to the rapid development of the Internet and video content devices. However, a high number of video contents are illegally copied from the corresponding original video contents. The illegal copy problem causes the infringements, and wasteful usage of storage space and network bandwidth [1]. In order to overcome this problem, currently existing solutions fall into two categories: (1) video contents are tagged with text labels; (2) watermarks are embedded into video frames in a direct way, as is described in [2]. However, one critical limitation of the former approach is that text labels can be easily changed by users [1]. The users can eliminate text labels from video contents. In addition, the latter approach has one disadvantage that

* Corresponding author. Tel.: +82 42 350 3494; fax: +82 42 350 7619.
E-mail addresses: resemin@kaist.ac.kr (S. Kim), jygchoi@kaist.ac.kr (J.Y. Choi), hansw@etri.re.kr (S. Han), ymro@ee.kaist.ac.kr (Y.M. Ro).

http://dx.doi.org/10.1016/j.image.2014.05.002
0923-5965/© 2014 Elsevier B.V. All rights reserved.

video contents could be easily modified by embedding watermarks; hence, they are usually susceptible to visual transformations [3].

To cope with above-mentioned problems, considerable research efforts have been made to develop content-based video copy detection (CBCD) techniques, aiming to effectively protect video contents from the infringement of copyright [1–46]. The objective of CBCD techniques is to automatically judge whether a video in question contains any contents originated from the original video [4]. To that end, video is generally converted into the video signatures to increase the compactness and discrimination ability [5]. Video signatures developed so far include spatial fingerprints [6–18], temporal fingerprints [19–23], and spatio-temporal fingerprints [24–29]. However, it should be pointed out that most of the previous CBCD methods have been limited to using only spatial fingerprints as video signatures [4].

There has been limited but increasing amount of work on the combined use of spatial and temporal information for the purpose of CBCD [19–33]. The results in these works indicate that temporal information can play a crucial role in differentiating the original videos from the illegally copied videos. Also, temporal information has been found to be robust to changes in spatial transformations (such as caption insertion and shifting) [24]. In [19], a video was divided into four sub-regions and computed mean luminance of each region. Then, the frame ordinals of each region were computed and these ordinals were used as video fingerprints. In [24], video tomography (slices) was extracted by cutting video domains (height, width, and timeline). This tomography included temporal variances and can be denoted edges. Thus, video fingerprints were extracted by computing edges from video tomography. In [25], 3D DCT coefficients were computed since video consists of three domains. Thus, temporal variances were transformed into DCT domain, and several ACs were selected as video fingerprints. In [26], a video was projected into 2D image according to timeline. Then, the 2D image divided overlapped sub-blocks, and 2D DCT coefficients were computed. Next, two coefficients of each sub-block were extracted and concatenated. Finally, these concatenated coefficients were used as video fingerprints.

Following the aforementioned studies, it is natural to expect better CBCD performance by combining spatial and temporal information than the case of using only color or texture information. However, at the moment, how to effectively combine both spatial and temporal information for the purpose of CBCD still remains an open problem. The aim of this paper is to propose a new CBCD framework, which effectively integrate spatial and temporal information, aiming to enhance CBCD performance. The main contribution of our paper is as follows.

- It is reasonable to assume that combining various fingerprint modalities is more useful for obtaining better performance in video copy detection than using only spatial or temporal fingerprints. To this end, we proposed a novel way of extracting spatial and temporal features (based on DCT coefficients) well-suited for our fusion solution, as well as a new way of measuring the temporal variances (such as differences in consecutive video frames) from video shots at hand.
- It has been generally believed that the effectiveness of temporal features for CBCD depends largely upon the temporal variances embedded in video [2–4]. To address this issue, in the proposed method, the amount of temporal variances is used to determine the importance (significance) of temporal fingerprints. To be specific, temporal variance information can be applied to assign the weights, aiming to combine spatial and temporal modalities for the video match of CBCD. Differing from the previous work, the advantage of our method lies in its ability to adaptively determine the weights to be used for effectively combining spatial and temporal modalities. This is accomplished by considering the degree of discriminating power of temporal features. In our method, the discriminatory power is computed based on the temporal strength of videos to be matched. Then, spatial and temporal modalities are efficiently and effectively fused by means of weighted-sum model, leading to a considerable improvement in detection.
- Extensive and comparative experiments have been carried out to evaluate our proposed CBCD method. For this, public and popular video database (DB), named TRECVID 2009, was used. We selected 50 videos (23 h) as reference videos. By applying a shot boundary detection algorithm to reference videos, we obtained 9358 video shots as reference shots. In addition, ten videos were randomly selected as query videos, and 1675 shots were extracted from the query videos. We selected six video transforms: blurring, brightness change, caption insertion, cropping, flipping, and framerate change. Thus, a total of 10,050 shots were generated as query shots. With our database, video copy detection test was implemented comparing previous video fingerprint algorithms [5,6,11,12,25,26,31,52,56]. It has been found that our adaptive-weight fusion algorithm can considerably improve *F*1-Scores on query test set, compared to the previous fusion algorithms [30–33]. In addition, from our test, we could save about 39%, 25.1%, and 46.1% video matching time with very small loss of *F*1-Score for our synthetic dataset, TRECVID 2008 CCD Task dataset [36], and MUSCLE-VCD 2007 dataset [34], respectively. Further, the proposed spatial and temporal fingerprints show superior performance, compared with existing temporal fingerprints [5,31], spatio-temporal fingerprints [25,26], and modality fusion methods [30–33]. In addition, the effectiveness of our method has been successfully evaluated with the public queries of TRECVID 2009 CCD Task and MUSCLE VCD 2007.

The rest of this paper is organized as follows. Section 2 briefly reviews the existing video fingerprint methods. Section 3 presents the overview of the proposed modality fusion method. Sections 4 and 5 describe the way of extracting the proposed fingerprints and adaptive modality fusion, respectively. Section 6 outlines the experimental results compared with previously developed video

Download English Version:

https://daneshyari.com/en/article/536886

Download Persian Version:

https://daneshyari.com/article/536886

Daneshyari.com