Robust Video Hashing Based on Temporally Informative Representative Images

Mani Malek Esmaeili and Rabab K. Ward, Member, IEEE

Abstract-- This paper proposes a video hashing method for the purpose of digital rights management, media monitoring and tracking the distribution of illegally copied video. The proposed algorithm is applied to temporally representative images of a video. The resulting hashes are found experimentally to be highly robust to a range of attacks including noise, rotation, time shift, and frame dropping.

I. INTRODUCTION

Users of consumer electronics, mainly digital cameras, upload tens of thousands of videos on video sharing websites every day. This raises the concern regarding copyright protection of these videos. Previously proposed copyright protection approaches such as watermarking are not feasible solutions for individual users since they need extra information to be inserted in a video which may not be possible at the video owner side. Moreover, watermarking is not applicable on the already existing videos. Content-based multimedia hashing, also known as multimedia fingerprinting, has recently been introduced as an alternative to watermarking [1]. Its aim is to identify a video or an image based on its content. This is done by extracting some features from the media that conform only to the content of that specific media and that can be used to distinguish it from other media. Although many papers on image hashing have been published in the literature, video hashing is still in its preliminary research phase. Existing video hashing approaches mostly extend the hashing techniques developed for still images. Some studies directly apply these image-hashing techniques to every frame of the video [2-4] and some apply them only to key frames of the video [1], [5]. The drawback of the former approach is the large dimensionality of the generated hash that makes database searching computationally costly. The latter approach is sensitive to frame dropping and noise [8]. Neither of the above approaches considers the temporal information of the video during the feature extraction phase. As the temporal information distinguishes a video sequence from a series of random images, utilizing this information is expected to result in better performance.

We propose a new approach that incorporates temporal information into images derived from a video sequence. The sequence is divided into video segments and the frames of every video segment are linearly combined to generate a temporally informative representative image (TIRI) of that segment. This provides the image hash with both the spatial and the temporal information of the video segment. We show below that this simple procedure can increase the performance of the hashing algorithm. Another approach that considers both temporal and spatial information of the video is proposed by Coskun et al. in [6]. They consider a video segment as a three-dimensional matrix and use the low-frequency coefficients of two 3D cosine-based transforms, the Discrete Cosine Transform (DCT) and the Random Basis Transform (RBT), to generate the hashes for that segment. The hashing techniques proposed in [6] exhibit very good performance under a variety of attacks, including added Gaussian noise, changes in brightness/contrast, and geometrical attacks. DCT outperforms RBT in terms of robustness and uniqueness of the hashes.

In this paper we first adapt the 3D-DCT hashing method in [6] so it can be applied on the proposed 2D TIRIs and we then introduce a new hashing algorithm that we explain later in section II. Both algorithms generally yield better results than the DCT-based hashing algorithm in [6]. In the next section, we briefly explain about how to generate TIRIs of a video. We also propose our new hashing algorithm in this section. In Section III we demonstrate the simulation results. We compare the hashing method in [6] with both our new hashing method and that of applying [6] on our proposed TIRIs and show the high performance of new algorithms in terms of robustness and uniqueness. Conclusions are discussed in Section IV.

II. THE PROPOSED METHOD

This section describes the details of the proposed video hashing approaches. First it is explained how TIRIs are generated from a video sequence. The hashing method in [6] and how to adapt it for our proposed TIRIs are also explained briefly. The proposed hashing algorithm is presented in details at the end of this section.

A. Temporally Informative Representative Images

To make the algorithm resistant to scaling and frame-rate changing, a pre-processing step is first applied. This step consists of time re-sampling and spatial resizing of the video into fixed $W \times H \times F$ pixels/second, where $W \times H$ is the frame size and $F$ is the frame rate. After this pre-processing step, the video is segmented into fixed short segments. This is done simply by putting frames in a particular time range together. Then for each segment, TIRI, the temporally informative representative image, is generated as follows: let $I_{x,y,k}$ be the luminance value of the $(x,y)^{th}$ pixel on the $k^{th}$ frame in a segment of $L$ frames. The pixels of the TIRIs are obtained as:

$$a_{x,y} = \sum_{k=1}^{L} \beta^k I_{x,y,k}$$

(1)

The DCT-based hashing algorithm in [6] uses the low frequency 3D-DCT coefficients of the 3D video data. To make
this algorithm applicable to our approach we used low frequency 2D-DCT coefficients of the TIRIs. For comparison purposes we have used the same number of coefficients in both cases. To generate the binary hash each coefficient is replaced by 0 if it is less than the median values of all the coefficients and 1 otherwise [6]. In the next section we propose a new hashing algorithm to be applied on TIRIs.

B. Proposed Hashing Algorithm

Our proposed algorithm segments TIRIs into overlapping blocks of size $M \times M$. Then two features are extracted from each block, one from the horizontal direction and another from the vertical direction. The horizontal feature, $\alpha$, from a block is extracted as:

$$\alpha = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} f_{i,j} \cos \left( \pi \left( x + 0.5 \right) / M \right)$$

(2)

The vertical feature $\beta$ is extracted similarly. Let $K$ be the total number of the overlapping blocks in a TIRI, then $(\alpha_1, \alpha_2, ..., \alpha_K)$ and $(\beta_1, \beta_2, ..., \beta_K)$ will be the generated features. To generate the binary hashes we use the same approach mentioned in section II.A. The final hash is generated by concatenating the two binary vectors. In the next section we compare the performance of our algorithm with the 3D-DCT approach in [6] and that of its 2D version applied to TIRIs.

III. SIMULATION RESULTS

To determine the parameters of the algorithms, we ran some simulations on a training set of videos available from [7]. We chose $\beta=0.6$ and $M=32$ pixels with a 50% of overlapping between adjacent blocks. The hash lengths and detection thresholds were also chosen accordingly for all the methods so as to generate the best results on the training set. To evaluate the performance of the proposed hashing algorithms, a test set of 200 videos all taken from the ocean [8] was used. Then attacks were individually mounted on the videos. The mounted attacks included added Gaussian noise, changes in brightness/contrast, rotation, temporal shift, spatial shift, and frame dropping. The hashes of the attacked versions were then checked against the hashes of the original test set to see if the correct video can be detected by the algorithms. The rate of correct hits with the data-base, true positive rate (TPR), and the precision of the algorithms were computed. Table I compares the performance of the algorithms. All three algorithms were completely robust against the changes in brightness and contrast, not listed in the table. It can be seen from Table I that both TIRI based algorithms proposed in this paper are highly robust to many of the attacks. The newly proposed algorithm has the highest TPR of 88% amongst the three methods. On average, both TIRI based algorithms have a high precision of over 99% compared to 98% of the algorithm in [6].

Proposed algorithms are slightly less robust to severe geometric attacks such as rotation and shift. However, it should be noted that these attacks degrade the perceptual quality of the video and render them as practically unusable.

Both TIRI-based algorithms are significantly more robust to unintentional attacks that occur frequently, i.e. the time shift, representing the time misalignment of the segments, and the frame dropping that happens during transmission. In terms of computational time the proposed algorithms run faster than the algorithm in [6] by a factor of 6 within our simulation setup.

IV. CONCLUSION

The detection of whether a video is an unauthorized copy of another video is very important for any video copyright application and forms a main concern in many video sharing websites. In this paper two robust hashing algorithms for video copy detection are proposed. Both algorithms extract binary hashes from temporally informative representative images (TIRIs) of a video sequence. Having a low computational load the algorithms are robust to a wide range of signal processing and geometric attacks. Simulations run on a database of 200 videos show the high discriminating ability of the algorithms. We also aim to evaluate the performance of the mentioned algorithms in the presence of other attacks such as cropping and logo insertion.

REFERENCES


