STM JOURNALS

**Journal of Artificial Intelligence Research & Advances**
ISSN: 2395-6720 (Online)
Volume 6, Issue 2
www.stmjournals.com

# Decisive Video Fingerprinting Using Visual Regions

***Somesh Bhinda**, Shruti Bijawat***
Department of Computer Science Engineering, Poornima Institute of Engineering and Technology, Jaipur, Rajasthan, India

***Abstract***

*In this research paper we are trying to detect a video and generate a fingerprint of it which is unique to that video; then this fingerprint of video will help in detecting or comparing with other videos to find out whether it is original or not or is the video existing in our database. This video fingerprinting technology will help us in copyright management and to curb piracy, it can detect whether the video is fake or not. We are going to use a famous technique saliency map to generate and compare fingerprints of two videos.*

***Keywords:*** *Visual regions, video fingerprinting, DRM, copyrights, feature detection*

****Author for Correspondence*** E-mail: 2015pietcssomesh@poornima.org

## INTRODUCTION

Video fingerprinting is a technique used to uniquely identify the video. So, in order to do that, it finds the fingerprint using visual regions using saliency map; what saliency map does is that it extracts features from the video and generates a fingerprint. This technique can be used in many ways, there are many applications of video fingerprinting like detecting fake videos or to detect whether the video is edited or morphed from the original or to detect if someone is using a video without permission or the video belongs to someone else. Apart from finger printing, many techniques are already present in the market but they are no that efficient like DRM (Digital Right management). It is easily breakable and there also exists watermarking technique but it is also easy to break; if video is cropped or encoded in another format, then this technique fails. The fingerprint is supposed to satisfy three conditions [1]:

1) Robustness: It should be robust, like fingerprint should be able to detect the video when it was cropped or brightness was changed or aspect ratio was changed.
2) Accuracy: It should produce accurate results; if we take two videos, the fingerprint generated should be unique and different.
3) Efficiency: the extracted fingerprint should be efficient enough to be searched in the database fast and accurately.

Past work can be divided into two categories [2]: First method is using full video and the second method is based on single frame extracted from the video or a video sequence like spatio-temporal transform coefficients [3]. Previous techniques like randomized first order video technique summarize video by small set of representative feature vectors [4]. The disadvantage of this technique is that it cannot accurately detect the video if the video is distorted and shorter than the original one. Recently, more attention is placed on this kind of approaches. The following video fingerprinting is in this class: the differential block luminance [5], centroid of gradient orientations (CGO) [1], robust color histogram descriptors [6], the radial hash (RASH) [7] which uses radial projection of the image pixels, and so on. These can extract and compare fingerprints of variable length videos; these techniques are robust and have lossy compression, frame rate change etc. In this study, we are trying to generate fingerprints using visual regions by slightly changing the algorithm to get robustness, accuracy and efficiency.

The rest of the paper is divided as follows: Next part of the paper shows the video fingerprinting extraction model; followed by video matching approach and database strategy, experimental results and performance evaluation and lastly the conclusion [5].

  

## VIDEO FINGERPRINTING EXTRACTION MODEL

This is our overall model or you can say architecture for extracting fingerprints from the video, it comprises of six steps in order to generate the video fingerprint (Figure 1).

In the first step, we need input video frames. So, in order to get that we need to decode our video into a standard format and resize it to CIF format (350×282) to make it stable or to eliminate all the noise part because not all the videos are same, they are on different formats and of different sizes; all these things are done in our preprocessing of the image; this is the essential step in order to get the fingerprint from the desired video.

In the next step, we extract visual feature map such as color, orientation and intensity which are computed from the given input; this will help in removing modality dependent differences, and these maps are then normalized and combined to make the saliency map which is unique to every video. The block diagram Fig. 1 shows how to generate a saliency map; this map will be in M×N block, where N are rows and M are columns then an average value of this is calculated to generate the fingerprint from this map.

### Saliency Map

Visual attention a characteristic in human visual system considered as the best system. Most system are emphasized on this system because it follows a bottom up approach; saliency map is based on this approach, simple feature maps can detect discontinuities in intensity, orientation and color and then combine to make a map [8]. Itti and Koch proposed an applied system for detecting human attention regions [9]. We follow their work to get saliency map in three steps:

1.) Compute visual feature map for color, intensity and orientation;
2.) Normalize it; and
3.) Combine these to make unique map.

Pixels represent the attention of the map. If the value of the saliency map is very high, the region is more attention getting; we can quantize this value to get the desired fingerprint. In the first step, we follow the Itti's approach [9]. Using set of linear center surround operations we get the feature map. All the maps are from different modalities and unrelated range, so before combining into a unique map, we normalized it to remove such noises.

$$DoG(x,y) = \frac{c_{ex}^2}{2\pi\sigma_{ex}^2} e^{-\frac{x^2+y^2}{2\sigma_{ex}^2}} - \frac{c_{inh}^2}{2\pi\sigma_{inh}^2} e^{-\frac{x^2+y^2}{2\sigma_{inh}^2}} \quad (1)$$

After this, all the maps are normalized into a single unique saliency map, color intensity and orientation; these three of them are combined using the same method.

$$S = \frac{1}{m}\sum_{i=1}^{m} N(X_i) \quad (2)$$

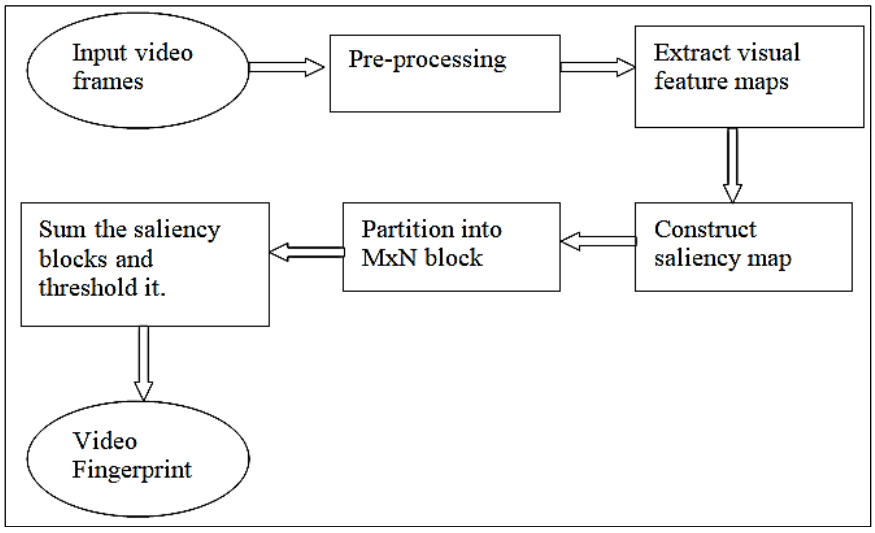


***Fig. 1:*** *Extraction Model of Fingerprint.*

## FINGERPRINT COMPARISION AND RETRIEVAL STRATEGY

After getting or generating the required unique fingerprint of the video we need a fast technique to retrieve, compare and give the output from the input video to the user. As the videos are very large in size, we also need a way to compare only fingerprints instead of videos, we can use pre-computed database of fingerprints of video to compare instead of generating the fingerprint of the video. At the time of comparing this way we only need to generate fingerprint of the input video, we also need to set some minimal amount of video to be input in order to get accurate and efficient results, so we need to fix the duration of input video to 5, 10 or 20 sec video. A window is used which is slightly greater than the minimal unit to query the video. We can also set a look up table LUT which is a faster way to search results in the database. This LUT table contains all the possible entries of the video fingerprints which position or point at the videos [10].

## EVALUATION

In our evaluation, author takes 4-times the same 50 min video (900–1100 i-frames per video) with different time offsets, encodings, and video overlay (logo); and the author checks:

(1) Test for accuracy called independent test. In this, two perceptually different videos should be distinguished using both fingerprints, if not, then the case is "false case".
(2) This test is to check the robustness of the algorithm. It should able to detect the video using the fingerprint when the video is slightly changed or distorted or the encoding is changed or the brightness is changed; if not able to detect, the case is missed case.

In our result, what we found is that compared to CGO, our algorithm is 1.5% more accurate and robust while detecting and comparing the fingerprints of the video even when the video is changed or distorted; and when we change the logo overlay maximum to 20%, it still able to detect the video using the fingerprint generated using our visual region process (Table 1).

***Table 1:*** *Evaluation of Video Distortions.*

| Video Distortions | CGO[2} | Our Method |
|---|---|---|
| Resolution CIF | 1.02% | 0.14% |
| Resolution QCIF | 1.67% | 0.09% |
| Frame-rate 4fps | 1.01% | 1.72% |
| Frame-rate 5fps | 0.97% | 1.20% |
| Color to gray | 0.86% | 1.28% |
| Logo overlay by 5% | 18.8% | 0.40% |
| Logo overlay by 20% | 34.9% | 0.40% |

## CONCLUSIONS

In this paper, we reviewed that how visual areas can be used to generate the fingerprints which are not even accurate but robust enough to detect video when the conditions are changed. We also discussed the comparing and retrieval strategy of the video from the database which is fast and efficient enough to detect or retrieve video from the database. We also see that saliency map is effective and robust algorithm to generate unique fingerprints to the video. In future, we would like to make it more accurate and efficient. Due to the single frame-based algorithm, its accuracy is sometimes less when the input frame video is not long enough to get the fingerprint which is unique.

## REFERENCES

1. Lee S, Yoo CD. Robust Video Fingerprinting for Content Based Video Identification. *IEEE Trans Circuits Syst Video Technol*. Jul 2008; 18(7): 983–988p.
2. Radhakrishnan R, Bauer C. Content-Based Video Signatures Based on Projections of Difference Images. *IEEE 9th Workshop on Multimedia Signal Processing*. Oct 2007; 341–344p.
3. Coskun B, Sankur B, Memon N. Spatio-Temporal Transform Based Video Hashing. *IEEE Trans Multimedia*. Dec 2006; 8(6): 1190–1208p.
4. Cheung SS, Zakhor A. Efficient Video Similarity Measurement with Video Signature. *IEEE Trans Circuits Syst Video Technol*. Jan 2003; 13(1): 59–74p.
5. Radhakrishnan R, Bauer C. Robust Video Fingerprints Based on Subspace Embedding. *Proc ICASSP 2008*. Apr 2008; 2245–2248p.

6. Ferman AM, Tekalp AM, Mehrotra R. Robust Color Histogram Descriptors for Video Segment Retrieval and Identification. *IEEE Trans Image Process*. May 2002; 11(5): 497–508p.
7. Roover CD, Vleeschouwer CD, Lefebvre F, *et al.* Robust Video Hashing Based on Radial Projections of Key Frames. *IEEE Trans Signal Process*. 2005; 53(10): 4020–4037p.
8. Itti L, Koch C. A Comparison of Feature Combination Strategies for Saliency-Based Visual Attention Systems. *Proc SPIE, Human Vision and Electronic Imaging IV (HVEI'99)*. Jan 1999; 3644: 473–482p.
9. Itti L, Koch C, Niebur E. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Trans Pattern Anal Mach Intell*. Nov 1998; 20(11): 1254–1259p.
10. Oostveen J, Kalker T, Haitsma J. Feature Extraction and a Database Strategy for Video Fingerprinting. *5th Int Conf on Visual Information Systems*. 2002; LNCS 2314: 117–128p.

**Cite this Article**
Somesh Bhinda, Shruti Bijawat. Decisive Video Fingerprinting Using Visual Regions. *Journal of Artificial Intelligence Research & Advances*. 2019; 6(2): 41–44p.