Video Content Identification Using Video Signature

Article · July 2017

2 authors:

Tejashri Shinkar

Dinesh Bhagwan Hanchate
Vidya Pratishthan's, College of Engineering, Baramati

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue VII, July 2017- Available at www.ijraset.com

Video Content Identification Using Video Signature

Tejashri Shinkar[1], Dinesh Hanchate[2]
[1,2]VPKBIT, Savitribai Phule Pune University, Baramati, Dist-Pune, Maharashtra, India

Abstract: ***The volume of video data is increasing day by day. To detect a required video segment from unrelated video is a very tough challenge. Most of the available methods are found near copy detection of videos, but video duration is small. There are very few methods are available which address the problem of finding video segment from larger unrelated video. This approach provides a video signature based technology, which precisely detects a video segment form larger unrelated video. Video signature (VS) is nothing but the one type of fingerprint of video content which provides robustness and high performance for video editing operations. The System proposes frame level extraction of features to create video signature. To reduce complexity video signature is used as a key feature and comparisons between signatures will be performed to detect query video. The VS is suitable for desktop level and web scale deployment level, it can be used in right management, distribution management, content based linking etc.***

Keywords: ***Content Description, Content Identification, Video Signature, Fine Signature, Coarse Signature, Content fingerprint.***

I. INTRODUCTION

In the mass media industries; the data volume is growing enormously. There is also advancement in the telecommunication and in the internet data transfer. Also the infrastructure of multi-platform content delivery is advanced which accelerate the transmitting speed and increased volumes. The most of the videos are manipulated or illegal versions of existing media. Video copy detection is the key issue over the internet which is the fastest media for propagation of information. Video copy detection is the process of detecting the copied or edited video version by comparing and analyzing data with the original video data. Video contents are generally managed using metadata like keywords, thumbnail images, preview videos, etc. Generally, this provides a significant volume of metadata search results through the increase in the data volume and the circulation of video contents.

Watermarking and video fingerprinting/signature are two ways which are useful for video copy detection. Watermarking is performed by inserting or embedding an unrelated or distinct pattern in a video clip. But the bottleneck of watermarking is embedded patterns are destroyed or distorted during transmission. Signature/fingerprint techniques use content based signature for video copy detection. After media distribution also signature can be extracted from content based data. Due to this, nowadays signature technology attracts more and more attention. So we propose this method in our proposed method. System proposed a technology that identifies video content automatically and efficiently by managing it as metadata. The proposed system focuses on the first to design tools for video fingerprinting that will provide high robustness over common editing operations like cropping, labelling, morphing etc. Also detect the particular video content that will be embedded in longer unrelated video.

Video identification technology analyses each frame of the video content and extracts a descriptor which is unique. This unique descriptor is called video signature which is used to identify identical video scenes.

High precision and high speed video identification are achieved through the video signature without embedding ID information in the content.

The Video Signature Tools, which standardizes an interoperable descriptor for video identification. Content identification can be performed using mainly two steps:1) Signature Extraction and 2)Signature Matching. There are two types of signature extraction: a) Extraction of Fine signature and b) Extraction of coarse signature. For fine signature extraction query video is divided into frames. Intensity values are calculated from the sub regions of the frames. Differences of average intensity values are calculated. Then ternerization will be applied over these different values. The adaptive threshold value is used. Through his frame signature is calculated. Representative value will be selected from the difference of intensity values is called as Confidence score. Frame signature and confidence score collectively form fine signature. Through the fine signature coarse signature will be generated. Bag-of-words model is applied on fine signature that will create coarse signature.

For the matching purpose L1 distance between two frame signatures will be calculated. If L1 distance is smaller than pre-set threshold then it is considered as a matching pair.

II. PREVIOUS METHODS

Section II includes the different designing techniques of video signature/fingerprint. Survey mainly focuses on two issues. One is to detect a particular piece of video in the longer unrelated video using video signature and second is video signature can be used for video copy detection where videos are nearly similar. Video signatures are extracted using different types of features like local, global, block based etc.

Hampapur et. El. [2] studied over the first issue that identification of video content from larger unrelated video. For that purpose he proposed a survey of different types of signatures. From his study there are three types of signatures:

A. Ordinal Signature

It is a block based signature [8][9] which provides efficient result for short queries.

B. Motion Signature

It is also a block based signature where the patch of the image is selected as a center and minimum SAPD (Sum of absolute Pixel Differences) is calculated.

C. Color Histogram Signature

In this YUV histogram considered as the signature. The distance measure between two frames is histogram intersection. The matching factor will be calculated by NHI (Normalized Histogram Intersection) [10].

Law-To et al. [3] proposed a paper over second issue that video copy detection for nearly similar videos. It proposes different types of video copy detection techniques using different descriptors and voting functions. There are two types of descriptors global and local. In the global descriptor temporal, ordinal and temporal-ordinal type of signature measurement can be performed. While in local descriptor spatial, temporal and spatio-temporal information is considered for signature. For more specific search, according to criteria like operations performed over video are, resize, rotation particular voting function will be selected. For that purpose author proposed a CBCD technique [3] in detail.

In [5] Shen, presents a duplicate video clip detection system. The videos are near about similar. The author focuses on a comparison between two detection schemes. In a first scheme technique is used as a bounded coordinate system (BCS), while second one utilizes frame symbolization. The experiment was performed on TV commercials where the BCS system gives better result comparatively other one. So the final conclusion is system is better for matching short clips but not efficient for localizing the clip embedded into larger unrelated video.

In [6], another duplicate video clip detection system is presented by Shang. In this system also video clips are nearly similar. Two problems are tried solve here. One is how to represent video in a compact way and second is how to match frames efficiently. To describe spatial information two different strategies are used. First is automatic and based on CE (Conditional entropy) while the second one is heuristic and based on LBP (Local Binary Pattern). The given figure depicts the ordinal measure where the image is divided into blocks. The dimension of the block is 3*3. An average of gray level values in each block is computed. The set of average gray level values (intensities) is arranged in ascending order and allocate a rank to each one. Color degradation is the robust factor to ordinal measure. This method efficiently finds out near duplicate web video retrieval, but not successful in locating segment in larger video.

The author [16] of this proposed inverted file based similarity search for fingerprint matching. The method was proposed for detection of near duplicate videos. So the videos considered are nearly similar. Each fingerprint is divided into small non overlapping blocks, called as words. These words are useful for creating an inverted file from a database of fingerprints which is represented as a table. From the table word position inside a fingerprint is represented horizontally and possible values of the words are shown vertically.

III.PROPOSED SYSTEM

The Proposed system is designed to detect a small piece of video content (V_Q) from larger unrelated video (V) using video signature (V_s).

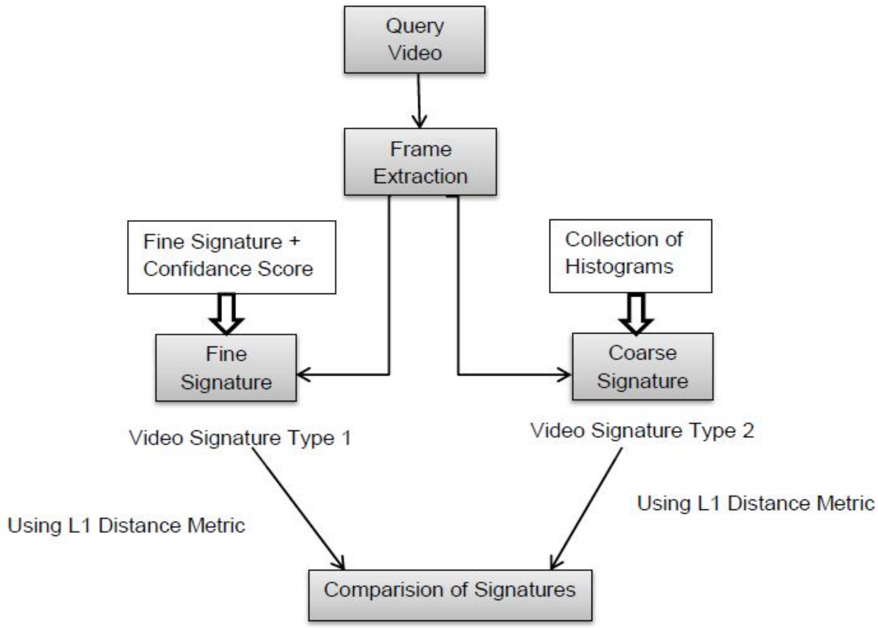

Fig. 1 Block diagram of the Proposed system

The proposed system is divided into the following steps

A. Signature Extraction from Video
1) Extraction of fine signature
2) Extraction of coarse signature

B. Matching of Signature
1) Matching using Fine signature
2) Matching using coarse signature

Signature Extraction from video

The proposed system uses two types of signatures. One is the fine signature as the video signature and another one is coarse signature as the video signature. Extraction of Fine signature F_s

Fine signature is composed of two components, one is frame signature and second is confidence score.

1) Frame Signature
2) For frame signature extraction query video V_Q is divided into n number of frames f_R.

$$V_Q = \{f_{R1}, f_{R2}, \dots\dots\dots\dots f_{Rn}\}$$

Where,

V_Q :Query Video f_R: Frames

n: 1 to N (Number of frames)

a) . From sub-regions of f_R extract intensity values.

$f_R I_i$ is the intensity values of each frame. It is the given by equation:

$$f_R I_i = \{f_R I_1, f_R I_2, \dots\dots\dots\dots f_R I_n\}$$

b) Calculate the average of intensities *Ai.*

$$A_i = \frac{\sum_{R=1}^{r} f_{Ri} I_R}{r}$$

c) Calculate differences of average intensities D_{AFI}.

$$while (l < n)$$
$$D_{AFI} = |A_{i+1} - A_i|$$

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue VII, July 2017- Available at www.ijraset.com

$$l \leftarrow l + 1$$
$$i \leftarrow i + 1$$

d) Ternary quantization is applied over these differences. So ternary vector will be generated of values (0, 1, 2).

e) Frame signature values are selected using given formula:

$$Xi = \begin{cases} 2, & \text{if } (Ai) > ThDi \\ 1, & \text{if } |Ai| \leq ThDi \\ 0, & \text{if } (Ai) < -ThDi \end{cases}$$

Where,

Xi: Number of elements from frame signature $FR_{S.}$

Ai: Average intensity

ThDi: Adaptive threshold.

f) Frame signature FR_S is represented as :

$$FR_S = \{X_1, X_2, \dots \dots \dots \dots X_n\}$$

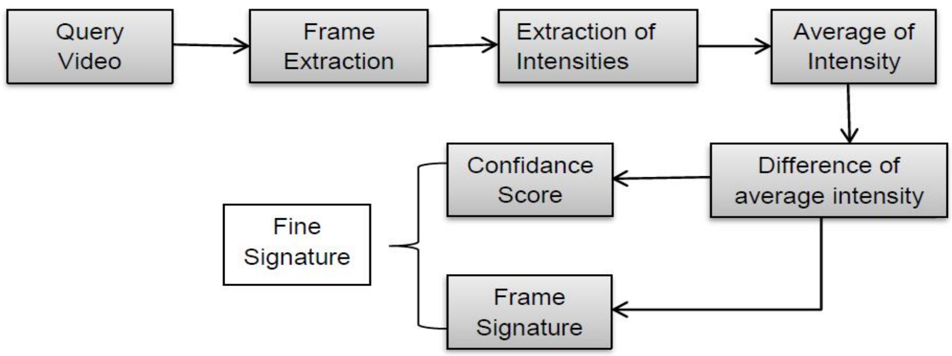

Fig 2. Block diagram of fine signature extraction

IV. CONFIDENCE SCORE

The confidence score represents the reliability of the frame signature. Low value of confidence score implements frames containing the flat and featureless images.

$$for(l \leftarrow 1 \; to \; n, l \leftarrow l + 1)$$

$$S_E \leftarrow sort(D_{AFl})$$

Sorted elements are represented as:

$$S_E = \{D_{AF1}, D_{AF2}, \dots \dots \dots \dots . D_{AF(n-2)}\}$$

Median of the sorted elements given by the formula:

$$M_S = \frac{(n - 1)}{2}$$

Confidence score is given by the formula

$$Cs = \min(\lfloor M_S . 8 \rfloor, 255)$$

Where,

Cs: Confidence score

Ms: Median of sorted elements

V. EXTRACTION OF COARSE SIGNATURE

To generate coarse signature bag-of-words approach will be applied over frames.

A. Bag of Words Approach

1) Video is a collection of frames /images represented as:

$$V = (f_1, f_2, \dots \dots \dots \dots . f_N)$$

Where,

f: frame extracted from the video

2) For computing local extrema convert image to gray scaled image.

$I_{RGB\ to\ G} = (0.299 * R) + (0.587 * G) + (0.114 * B)$

Where,

R:Red color value from frame

G: Green color value of frame

B: Blue color value of frame

3) For creating blur image apply a Guassian filter given by the formula:

$$G(x,y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-\sigma^2}{2\sigma^2}}$$

Where,

G(): Guassian filter

x,y : co-ordinates of the frame

σ : deviation value

4) scale-space computation is given by formula:

$$L(x,y,\sigma) = G(x,y,\sigma) * f(x,y)$$

Where,

$G(x, y, \sigma)$:variable scale guassian

f(x,y): frame of video

5) Difference of guassians given by the formula:

$$D(x,y,\sigma) = L(x,y,K\sigma) - L(x,y,\sigma)$$

Where,

L:scale-space of frame

K: Number of scaling

6) Local extrema given by the formula:

$$L_{E=}\max(D(x,y,\sigma), M_p)$$

Where,

$D(x,y,\sigma)$: DoG(Difference of Guassians)

M_p: Neighboring points

7) K-means algorithm for vocabulary generation (histogram Generation):

i) Initialize *P* i.e. number of clusters and centroids.

ii) To find out the nearest neighbor calculates distance between centroid and pixel of the images. It is given by the formula:

$$d = \| f(x,y) - C_p \|$$

iii) Cluster the pixel using minimum distance.

iv) Generate the histogram *h* for these clustered elemnts.

VI. MATCHING OF SIGNATURE

1) *Matching using fine signature*: Matching segments between two videos can be found by comparing frame signatures sequentially. Distance between two frames can be calculated using L1 distance metric given by the formula:

$$ds = \sum_{x=20,y=19}^{x=0,y=0} \sqrt{(R1_{x,y} - R2_{x,y}) + (G1_{x,y} - G2_{x,y}) + (B1_{x,y} - B2_{x,y})}$$

Where,

R1,R2: Red pixels of first image and second image respectively
G1,G2: Green pixels of first image and second image respectively
B!,B2: Blue pixels of first image and second image respectively

2) *Matching using Coarse Signature:* As we know histograms are generated in the coarse signature using bag-of-words model. The generated histograms are in image format. Distance between two histograms computed using L1 distance metric.

VII. EXPERIMENTAL RESULTS

The proposed system doesn't require any specific dataset. Experiments can be performed over films, documentary, cartoons, video songs, any visual data. The proposed system is focused on longer video data, so no time limitation. For efficient result 20s to 30s time is required. But system will provide result for longer video also.

The given table shows the total number of frames fetched from video and number of frames matched from that video. According to that percentage success rate will be calculated as mentioned in the table.

TABLEI

Number of frames extracted	Number of Frames matched	Matched Intervl Time	Percentage Success rate(%)
8	6	12s	75
10	7	8s	78
15	12	10s	81.63
6	4	7s	72.66
13	10	11s	76.92

VIII. APPLICATION

A. Rights Management and Monetization

To detect copyright infringement it is useful for content owners and to identify content owner is the requirement for content consumers. So it is useful for both content owners and consumers.

B. Distribution Management

It is useful in organization by applying video fingerprint i.e signature database that contains sensitive data and automatically detects and stops accidental transmission of this content via email, unauthorized copy of an external device or other.

C. Video-Content based Linking

Video content can become a linking content like a text and it can be used to infer an association with other pages.

D. Database Management and Duplication

The system is useful for high volume owners and creators like studios and archives also personal video libraries to manage the data and avoid duplication.

IX. CONCLUSION

As we know the mass media database is increased day by day. The large number of manipulating and duplicate video clips is uploaded over the internet. Existing techniques, mostly focused on matching of short video clips as discussed in section II. There are some methods present with watermarking having some drawbacks. The proposed system focuses on video signature technology, which is unique and provides robustness over video editing operations. A particular piece of video content is accurately detected through video signature from larger unrelated video. No any specific dataset required for experiments. Results are generated for films, cartoons, sports video, political videos, etc. Technique is useful over the internet, desktop level. This technique is mainly useful for longer video clips. Future work can be performed by applying different compression techniques or clustering techniques to reduce the complexity.

REFERENCES

[1] Paschalakis et. al, "MPEG-7 Video Signature Tools for Content Identification" in IEEE Transactions On Circuits And Systems For Video Technology, Vol. 22, No. 7, July 2012.
[2] A. Hampapur, K. Hyun, and R. Bolle, "Comparison of sequence matching techniques for video copy detection", in Proc. Conf. Storage Retrieval Media Databases, 2002, pp. 194-201.
[3] J. Law-To, L. Chen, A. Joly, I. Laptev, O. Buisson, V. Gouet- Brunet, N. Boujemaa, and F. Stentiford, "Video copy detection: A comparative study," in Proc. 6^{th} ACM Int. Conf. Image Video Retrieval, Jul. 2007, pp. 371-378.
[4] Z. Liu, T. Liu, and B. Shahraray, "AT&T research at TRECVID 2009 content-based copy detection," in Proc. TRECVid Workshop.
[5] H. T. Shen, X. Zhou, Z. Huang, J. Shao, and X. Zhou, "UQLIPS: A real-time near-duplicate video clip detection system," in Proc. 33rd Int. Conf. Very Large Data Bases, Sep. 2007, pp. 1374-1377.
[6] L. Shang, L. Yang, F. Wang, K.-P. Chan, and X.-S. Hua, "Real-time large scale near-duplicate web video retrieval," in Proc. ACM Int. Conf. Multimedia, Oct. 2010, pp. 531-540.
[7] A. Sarkar, V. Singh, P. Ghosh, B. S. Manjunath, and A. Singh, "Efficient and robust detection of duplicate videos in a large database," IEEE Trans. Circuits Syst. Video Technology, vol. 20, no. 6, pp. 870-885, Jun. 2010
[8] D. Bhat and S. Nayar, “Ordinal measures for image correspondence,” in IEEE Transactions on Pattern Analysis and Machine Intelligence, 20 Issue: 4, pp. 415–423, April 1998.
[9] R. Mohan., “Video sequence matching.,” in Proceedings of the International Conference on Audio, speech and Signal Processing, IEEE Signal Processing Society, 1998.
[10] M. Y. M. Naphade and B.-L. Yeo, “A novel scheme for fast and efficient video sequence matching using compact signatures.,” in Proc. SPIE, Storage and Retrieval for Media Databases 2000, Vol. 3972, pp. 564–572, Jan. 2000.
[11] L. Chen and F. W. M. Stentiford, “Video sequence matching based on temporal ordinal measurement”, Technical report no. 1, UCL Adastral, 2006.
[12] A. Joly and C. Frelicot, “Content-based copy detection using distortion-based probabilistic similarity search”, IEEE transaction on Multimedia, 2007.
[13] S. Maji, A. Berg, and J. Malik,” Classification using intersection kernel support vector machine is efficient”, In Proc. CVPR, 2008
[14] J.X. Wu, and J. M. Rehg. Beyond the Euclidean distance: Creating effective visual codebooks using the histogram intersection kernel. In Proc. ICCV, 2009.
[15] Chou, Chien-Li, Hua-Tsung Chen, and Suh-Yin Lee. "Pattern-Based Near Duplicate Video Retrieval and Localization on Web-Scale Videos." IEEE Transactions on Multimedia,17.3 ,382 395, 2015.
[16] Mani Malek Esmaeili, Mehrdad Fatourechi, and Rabab Kreidieh Ward, Fellow, IEEE” A Robust and Fast Video Copy Detection System Using Content-Based Fingerprinting”, IEEE Transactions On Information Forensics And Security, Vol. 6, No. 1, March 2011