International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Efficient and Robust Detection of Duplicate Videos in a Large Database: A Survey

Soni R. Ragho[1], C. S. Biradar[2]

[1]SRTM University, Department of CSE, M.G.M's Engineering College, Nanded, Maharashtra, India,

[2]Visvesvaraya Technological University, Department of CSE, B. E. C. College of Engineering, Bagalkot, Karnataka, India

Abstract: ***A duplicate video detection method is based on video signature and Fingerprint. Robust or compact frame based descriptor and color layout descriptor used to extract features from images that are constructed from video to create fingerprints which are encoded by vector quantization. Mapping process, measure the distance to find the similarity between query video and video in video database. Then apply a coarse-to-fine matching scheme to check the duplicate video. Vector quantization based model signature retained after dataset pruning to reduce search time for the nearest duplicate video by using pre-computed distance measures tables and discard the near duplicate video using just the partially computed distance from the video to query video. This survey gives an introduction to feature extraction using CLD and explores the various search algorithms, along with the pruning method.***

Keywords: Duplicate detection, video fingerprinting, Color layout descriptor, Non-metric distance, Data pruning, and Vector quantization

1. Introduction

With the high-speed development of technology and growing use of the widespread accessibility of ADSL and the World Wide Web, people can simply find and upload bags of videos on the Internet. There exist lots of duplicated and distorted video clips online and some of them may be illegally copied or broadcasted, so database and exclusive rights management have become large issues these days. Videos on commercial sites e.g., youtube.com, vimeo.com, metacafe.com, hulu.com, veoh.com are mainly textually tagged. These tags are of little help in monitoring the content and preventing exclusive rights infringements [1]. There are two general techniques has been used to detect duplicate video, and detect such infringements,

1. Digital watermarking
2. Content-based copy detection (CBCD)

The digital watermarking method determines the existence of a watermark in a video to decide if it is copyrighted. The other method CBCD finds the duplicate by comparing the fingerprint of the query video with the fingerprints of the original videos from database. Content-based copy detection schemes extract a small number of important features from the original media, called fingerprints or signatures of the video [2]. The same signatures are extracted from the test media stream and compared to the original media signature according to a dedicated distance measure to decide if the test stream contains a copy of the original media. A complete overview of duplicate detection framework is shown in Figure 1.

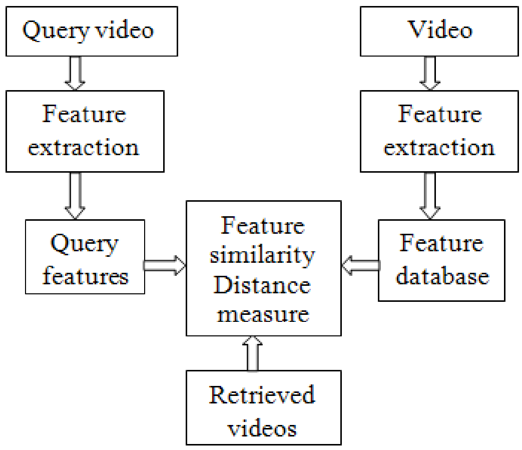

Figure 1: Block diagram of the proposed duplicate detection framework

Videos on commercial sites rely on the comparison of metadata or textual tags associated with the videos. This method relies on human participation to provide an interpretation of the video content so as to construct tags associated with the video. Though, the ever increasing occurrence of large video databases has resulted in the improvement of algorithms to enhance and replace textually tag based video retrieval with content based video retrieval that is content based copy detection (CBCD). A non-metric distance function is used for duplicate video detection, when the query video is a noisy subset of an original video. It performs better than other expected distance measures [1].For the Vector Quantization based signatures retained after dataset pruning, we decrease the search time for the top-K candidates by using appropriate pre-computed distance tables and by dumping many non-candidates using just the partially computed distance from the original video fingerprint to the query video fingerprint.

A dataset pruning method, based on our distance measure in the space of Vector Quantization encoded fingerprints, then it returns the top-K nearest neighbors (NN) still after pruning.

Joly O. Buisson [3], get significantly high pruning than that provided by distance based hashing methods, train on distance function. Both these methods involve an intuitive extension of the mathematical definition of a distance between two objects [1].

2. Related Work

2.1 Feature Extraction

There are three most common characteristics upon which images or videos are compared in content based image retrieval algorithms are
1. Color
2. Texture
3. Shape

Color features are defined subject to a particular color space or model. A number of color spaces have been used in journalism, such as RGB, LUV, HSV and HMMD. It is generally believed that human visual systems use texture for recognition and interpretation. In general, color is usually a pixel property while texture can only be measured from a group of pixels. Shape is known as an important cue for human beings to identify and recognize the real-world objects, whose purpose is to encode simple geometrical forms such as straight lines in different directions [4].

We perform duplicate video detection with various kinds of features is used- Global image features and Entire video based feature for a fast initial search for prospective duplicates, and keypoint based feature are employed for a more refine search. CLD, CFMT [5], Localized Color Histogram (LCH) [6] and EHD [7], are the techniques for extract feature from frame, image, or video. The LCH feature divides the image into a certain number of blocks and the 3D color histogram is computed per block.

In this describe the CLD feature vector and also provide some instinct as to why it is highly appropriate for the duplicate detection problem. The CLD capture the spatial distribution of color in an image, CLD signature or fingerprint [8] is obtained by partitioning the image into 8× 8 blocks, a single representative color is selected from each block on averaging, Once the tiny image logo is obtained, the color space translation between RGB and YCbCr is applied along every channel. The DCT (Discrete Cosine Transform) is computed for all image. The DC and (first 5 in zigzag scan order) AC DCT coefficients for each channel represent the 18-dimensional CLD feature.

In [9], there are several approaches for color feature extraction from image which showed in Table1.

Table 1: Color Method for Feature Extraction

Sr. No.	*Color Of Method*	*Pros.*	*Cons.*
1	Histogram	Simple to compute, intuitive	High dimension, no spatial info, sensitive to noise
2	CM	Compact, robust	Not enough to describe all colors, no spatial info
3	CCVs	Spatial info	High dimension, high computation cost
4	Correlogrm	Spatial info	Very high computation cost, sensitive to noise, rotation and scale
5	DCD	Compact, robust, perceptual meaning	Need post-processing for spatial info
6	CSD	Spatial info	Sensitive to noise, rotation and scale
7	SCD	Compact on need, scalability	No spatial info, less accurate if compact

2.2 Distance measure

In [1], proposed distance measure to compare a model fingerprint X^i with the query signature Q is denoted by d (X^i, Q) [1] (1). This distance is the sum of the best-matching distance of each vector in Q with all the vectors in X^i. In (1), $\| X_j^i - Q_k \|$ refers to the L1 distance between X_j^i, the jth feature vector of X^i and Q_k is the kth feature vector of Q. Note that d (·, ·) is a quasi-distance.

$$d(X^i, Q) = \sum_{K=1}^{M} \{\| X_j^i - Q_k \|\} \quad (1)$$

What is the motivation behind this distance function? We suppose that each query frame in a duplicate video is a tampered or processed version of a frame in the original model video. Therefore, the summation of the best-matching distance of each vector in Q with all the vectors in the signature for the original video (X^i) will yield a small distance. Hence, the model-to-query distance is small when the query is a (noisy) subset of the original model video. Also, this definition accounts for those cases where the duplicate consists of a reordering of scenes from the original video.

Dynamic time warping (DTW) is commonly used to compare two sequences of arbitrary lengths. The proposed distance function has been compared to DTW in [10]; where it is shown that DTW works well only when the query is a continuous portion of the model video and not a collection of disjoint parts. This is because DTW considers temporal constraints and must match every data point in both the sequences. Hence, when there is any mismatch between two sequences, DTW takes that into account (thus increasing the effective distance), while the mismatch is safely ignored in distance formulation.

2.3 Search Algorithm

In this survey paper a two-phase approach for fast duplicate retrieval. The distance measure (1) is used in search algorithms for copy detection. Initial, we discuss a naive linear search (NLS) algorithm in Section 2.3.1. Search techniques based on the vector quantized representation of the fingerprints that achieve speedup through suitable lookup tables are discussed in Section 2.3.2. Algorithms for further speedup based on dataset pruning are presented in [1].

2.3.1Naive Linear Search(NLS)

A naïve linear search method is to compute all the *N* model video- to query video distances and then find the best match. This set of *N* distances is denoted by *a* (2).We speedup the coarse search by removing various computation steps involved in *A*.

$$A = \{d(X^i, Q)\} = \{\sum_{K=1}^{M} \min_{1\le j\le Fi} \| X_j^i - Q_k \| \}_{i=1}^{N} \quad (2)$$

The NLS algorithm implements the two-pass method without any pruning. In the first pass, it retrieves the top-K candidates based on the smaller query signature Q by performing a full dataset scan using an ascending priority queue L of length K. The precedence queue is also used for the other coarse search algorithms in this section to keep track of the top-K NN candidates. The k^{th} entry in L holds the model video index ($L_{k,1}$) and its distance from the query ($L_{k,2}$). A model signature is inserted into *L* if the size of *L* is less than *K* or its distance from the query is smaller than the largest distance in the queue. In the second pass, NLS computes the distance of the K candidates from the larger query signature Q_{orig} so as to find the best matched candidate [1].

2.3.2VQ and Acceleration Techniques

When the feature vectors are vector quantized, an inter-vector distance reduces to an inter-symbol distance, which is fixed once the VQ codevectors are fixed. Hence, we vector quantize the feature vectors and represent the signatures as histograms, whose bins are the VQ symbol indices. For a given VQ, we store and pre-compute the inter-symbol distance matrix in memory. Describe the VQ-based signature creation [1]. Using the CLD features extracted from the database video frames, a VQ of size *U* is constructed using the Linde–Buzo–Gray algorithm [11]. The distance *d* (·, ·) (1) reduces to *dVQM*(·, ·) (2) for the VQ-based framework, where D is the inter-VQ codevector distance matrix (2).

2.4 Duplicate confirmation

From the top retrieved candidate, the duplicate detection system has to validate whether the query has indeed been consequent from it. The keyframes for a copy video can generally be matched with the corresponding frames in the original video using suitable spatio-temporal registration methods. In [12], the approximate NN results are post-processed to compute the most globally similar candidate based on a registration and vote strategy. In [13], Law-To et al. use the interest points proposed in [14], for trajectory building along the video sequence. A robust voting algorithm utilizes the trajectory information, spatio-temporal registration, as well as the labels computed during the off-line indexing to make the final retrieval decision. In our duplicate detection system, we have a "distance threshold based" and a registration-based framework in [1], to determine if the query is actually a duplicate derived from the best-matched model video.

2.5 Challenges in video detection

A video clip can be encoded in different formats depending on the function. We perform various modifications on the query frames; a randomly chosen subset of 1200 videos to generate 18 duplicates per video [1], Different formats can give increase to a number of distortions, such as blurring in image or video, resizing, change in saturation, cropping, shift in hue, and change in brightness, compression in the picture. There are various types of signature extraction methods depend on the color and images block information in the videos, such as color histograms and coherence vectors, and due to the artifacts above a sometimes wrong detection occurs [2].

Additional to the point distortions from different formats, there are some other factors make difficult to determine copy video, such as sub sequence of frame, missing frame, noise during storage transmission, blurring, image compression are common distortions [1],[2]. Almost detected factors come from the construction of a photocopy video, for example, zooming or changing the contrast, inserting words or logos or watermark, just cutting a small part of a movie, and changing the background of the original video, changes in color video into gray-level or even combining several video clips into a new video.

3. Conclusion

The purpose of this survey paper is to provide an overview of the functionality of content based copy detection system. We empirically selected CLD for fingerprinting as it was robust to the duplication attacks. This distance measure has high computational complexity as it computes the distances between all model video to query video keyframe pairs. We evaluated approximate search paradigm based on dataset pruning to reduce video retrieval time. But above mentioned techniques have some limitations that, however, a query video contains portions of multiple videos, the similar asymmetric distance will not be efficient.

References

[1] Anindya Sarkar, Vishwarkarma Singh, Pratim Ghosh, B. S. Manjunath, and Ambuj Singh, "Efficient and Robust Detection of Duplicate Videos in a Large Database",2011.
[2] Wei-Lun chao, Introduction to video fingerprinting, Gradute institute of communication Engineering, NTU.
[3] Joly, O. Buisson and C. Frelicot. Statistical similarity search applied to content-based video copy detection. Int. Conf. on Data Engineering Workshops, page 1285, 2005.
[4] Rishav Chakravarti, Xiannong Meng. A Study of Color Histogram Based Image Retrieval, Department of Computer Science Bucknell University Lewisburg, PA 17837
[5] S. Derrode and F. Ghorbel. Robust and efficient Fourier-Mellin transforms approximations for gray-level image reconstruction and complete invariant description.

Computer Vision and Image Understanding, 83(1):57–78, 2001.
[6] L. Liu, W. Lai, X. Hua, and S. Yang. Video Histogram: A Novel Video Signature for Efficient Web Video Duplicate Detection. Lecture Notes in Computer Science, 4352:94–103, 2007.
[7] Won, D. Park, and S. Park. Efficient use of MPEG-7 edge histogram descriptor. Etri Journal, 24(1):23–30, 2002.
[8] E. Kasutani and A. Yamada. The MPEG-7 color layout descriptor: a compact image feature description for high-speed image/video segment retrieval. In Proc. of ICIP, volume 1, pages 674–677, 2001.
[9] Dong ping Tian, A Review on Image Feature Extraction and Representation Techniques. International Journal of Multimedia and Ubiquitous Engineering Vol. 8, No. 4, July, 2013
[10] L. Rabiner and B. H. Juang. Fundamentals of Speech Recognition. Prentice Hall Signal Processing Series, NJ, 1993
[11] Y. Linde, A. Buzo, R. Gray, et al. An algorithm for vector quantizer design. IEEE Transactions on communications, 28(1):84–95, 1980.
[12] Joly, O. Buisson, and C. Frelicot. Content-based copy retrieval using distortion-based probabilistic similarity search. Multimedia, IEEE Transactions on, 9(2):293–306, Feb. 2007.
[13] Joly, O. Buisson, and C. Frelicot. Statistical similarity search applied to content-based video copy detection. Int. Conf. on Data Engineering Workshops, page 1285, 2005.
[14] A. Joly, C. Frelicot, and O. Buisson. Robust content-based video copy identification in a large reference database. In Int. Conf. on Image and Video Retrieval, pages 414–424, 2003.

Author Profile

S.R. Ragho received the B.E. degree in Computer Science and Engineering from M.G.M's Engineering College, Nanded in 2012. Now, she is pursuing Master's in Engineering (Computer Science and Engineering) from M.S. Bidve Engineering college, Latur, SRTM University Nanded, Maharashtra.

C.S. Biradar received the B.E. and M.E. degrees in Computer Science & Engineering, from B. L. D. E. A College of Engineering, Bijapur in 2011 and from B. E. C. College of Engineering, Bagalkot in 2013, respectively. He is now with M.S. Bidve Engineering College, Latur (M.S.) as Assistant Professor since 2013.