International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 5 Issue: 11 139 – 142

# Efficient and Robust Detection of Duplicate Videos in a Database

Ragho Soni R.
Department of Computer Science and Engineering
M. S. Bidve Engineering College, Latur
*e-mail: sonirragho@gmail.com*

Shah H.P.
Department of Computer Science and Engineering
M. S. Bidve Engineering College, Latur
*e-mail: shahhemali@gmail.com*

**Abstract**— In this paper, the duplicate detection method is to retrieve the best matching model video for a given query video using fingerprint. We have used the Color Layout Descriptor method and Opponent Color Space to extract feature from frame and perform k-means based clustering to generate fingerprints which are further encoded by Vector Quantization. The model-to-query video distance is computed using a new distance measure to find the similarity. To perform efficient search coarse-to-fine matching scheme is used to retrieve best match. We perform experiments on query videos and real time video with an average duration of 60 sec; the duplicate video is detected with high similarity.

***Keywords***-*Video fingerprintingt; color layout descriptor; distance mesure; vector quantization.*

*****

## I. INTRODUCTION

With the fast development of technology and increasing use of the widespread accessibility of ADSL and the World Wide Web, people can easily find and upload tons of videos on the internet. There exist too many duplicated and transformed video clips online and some of them may be illegally copied or broadcasted, so database and copyright management have become big issues nowadays.

Copyright infringements and data piracy have recently become serious concerns for the ever increasing online video database management. Videos on commercial sites e.g., www.google.com, www.YouTube.com, www.metacafe.com, are mainly textually tagged. These tags are of little help in monitoring the content and preventing illegal upload. There are two approaches to detect such infringements that are watermarking and Content-Based Copy Detection (CBCD). The watermarking approach tests inserting distinct pattern in video to decide if it is copyrighted. The other approach CBCD detects the duplicate by comparing the fingerprint of the query video with the fingerprint of the original video.

A video "fingerprint" is a feature extracted from the video that should represent the video compactly, allowing faster search without compromising the retrieval accuracy. The bottleneck of watermarking is that the inserted marks are likely to be destroyed or distorted as the format of the video get transformed or during the transmission so noise robustness of the watermarking schemes is not ensured in general [1], where as the video fingerprint extraction of the Content-Based Copy Detection can be mostly unchanged after various noise attacks, that's why video fingerprinting approach has been more successful.

Duplicate video is derived from only one database video it consists entirely of a subset of frames in the original video. The individual frames may be further processed. The temporal order of the frames may also be altered. In [2], a set of 24 queries searched in YouTube, Google video and yahoo video, 27% of the returned relevant videos are duplicates or nearly duplicate to the most popular version of video in the search result. In [3], each web video in the database is reported to have an average of five similar copies. Also, for some popular queries to the yahoo video search engine, there were two or three duplicates among the top ten retrievals [4].

## II. RELATED WORK

**Feature representation:** In the feature extraction process, feature extracted from the video and image fall into three categories as global image, keypoint based, and Entire Video based Features. Many technique use global features for a fast initial search to find duplicates using color histogram computed over the entire video [2] for coarse search and keyframe-based features are use for a more refined search. A comparative study for video copy detection methods can be found in [5].

*Global Image Features* are derived from sets of time-sequential video keyframes. A combination of MPEG-7 features such as the Scalable Color Descriptor, Color Layout Descriptor (CLD) [6] and the Edge Histogram Descriptor (EHD) has been used for video-clip matching [7], using a string-edit distance measure.

*Keypoint based feature* technique described in [8] by Joly, this includes a key-frame detection, an interest point detection in these key-frames, and the computation of local differential descriptors around each interest point. Interest points are ideal for matching applications because of their local uniqueness and their high information content. In [9]The SIFT descriptors by Lowe use the Divergence of Gaussian (DoG) interest point detector which better handles significant changes in the scale of images. The SIFT descriptors then encode the image gradients and their orientations around the points into a 128-dimensional histogram. The PCA-SIFT descriptors simply apply PCA on Lowe's SIFT descriptors, reducing their dimensionality to 36. From the SIFT family this scheme is called $EFF^2$ as these descriptors are computed EFFiciently and yield EFFective search results. Overall, local descriptor schemes handle rotations, translations of objects in images, changes in color and to some extent compression and scale changes.

*Entire video based features* derived from the whole video sequence, such descriptors like ordinal measure, have poor performance with local transformations such as shift, cropping and cam coding.

**Indexing method:** a number of indexing techniques have been used for speed up the detection process .In [10] For the videntifier system the NV-Tree [9], which is able to perform accurate approximate High-dimensional nearest neighbor

queries in constant time, independent size of the video descriptor collection. In [11] author evaluate approximate search paradigm, called *Statistical Similarity Search* (*S*3) in a complete CBCD scheme based on video local fingerprints. The proposed *statistical query* paradigm relies on the distribution of the relevant similar fingerprints. Joly [8] use an indexing method based on the Hilbert's space filling curve principle and a simple search method is used to find closest derived key in the database.

**Hash-based Index:** Locality Sensitive Hashing (LSH) [12], have been effectively useful for similarity indexing in vector spaces and string spaces under the Hamming distance (a well-liked approximate for L2 distances and in this proposed distance function is non-metric). LSH formalism is not applicable for analyzing the behavior of these tables as index structures DBH is a hash-based indexing method that is *distance based*. Consequently, DBH can be applied in arbitrary (and not necessarily metric) spaces and distance measures, Whereas LSH cannot.

**Final Duplicate Confirmation:** From the top ten Nearest Neighbors, the system is supposed to return answer whether or not it is a duplicate of an already existing copyrighted database video. In[5][13],the partial result must be post-processed to compute a similarity measure and to decide if the more similar video is copy of copyrighted video using voting strategy. A robust voting algorithm utilizes trajectory information, spatio temporal registration, and label of behavior indexing to make a final decision.

## III. IMPLEMENTATION DETAILS

### A. *System Architecture*

In Fig. 1, denotes system architecture of duplicate video detection. This system works in two phases, offline phase and online phase. *Offline phase* (model related) consists of the un-quantized model fingerprint generation, VQ design and encoding of the model signatures, and computation and storing of appropriate distance matrices. *Online phase* (query related) consist query video is decoded, sub-sampled, key frames are identified, and features are computed per keyframe. It obtain k-means based compact signatures, perform VQ-based encoding on the signatures to obtain sparse histogram-based representations, compute the relevant lookup tables, and then perform two-stage search to return the best matched model video.

In this paper the database video referred as original or model videos. Model video is decoded and sub-sampled at a factor of number to get (T) frames and feature (P) is extracted from per frame. To summarize feature, we perform k-means clustering and save the cluster centroid as fingerprint (F).The duplicate video detection task return best matching model fingerprint for query fingerprint. The model video to query video distance is computed using distance measure. Two phase procedure is used for search; coarse search is used for to return top K Nearest Neighbors, refine search returns best match for given query video. The final module decides whether the return video is duplicate video or not.

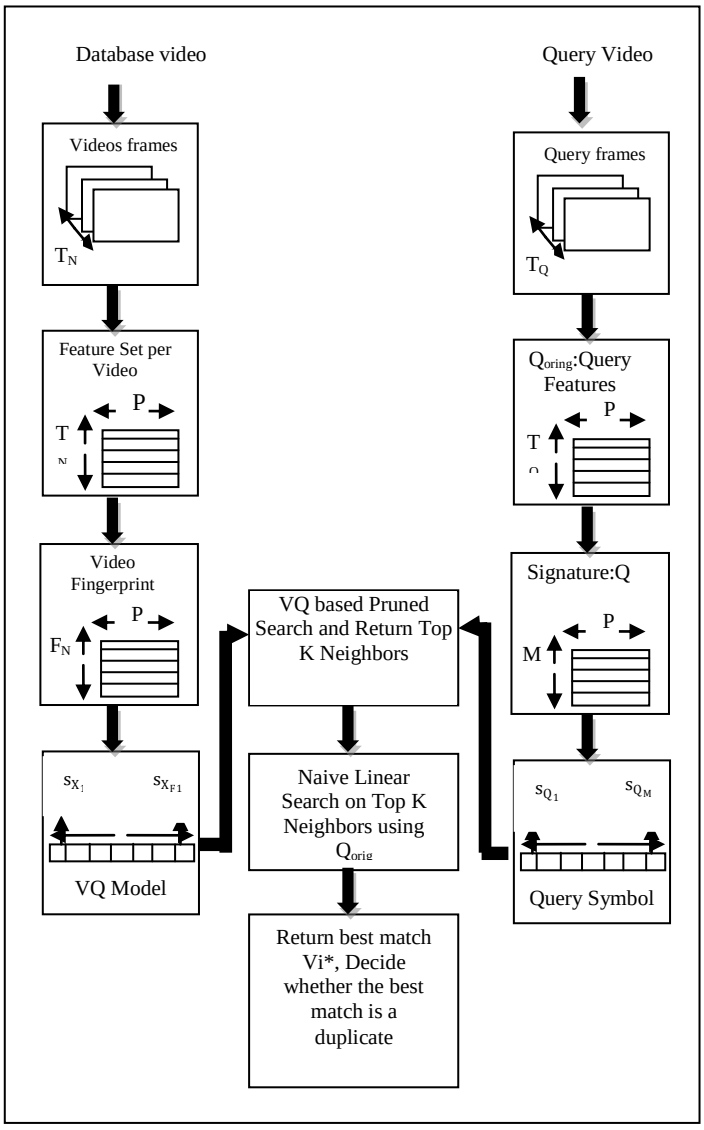


Figure 1. System Architecture

### B. *Feature Extraction*

*1) CLD:* The video is decoded frames are sub-sampled at a factor of value to obtain frames and a vector dimensional feature is extracted per frame using CLD technique. The CLD is very compact and resolution invariant representation of color for high speed image retrieval and it has been designed to efficiently represent spatial distribution of color. The extraction process of this color descriptor is obtained by converting the frame to 8*8 into 64 blocks to guarantee the invariance to resolution or scale. After the image partitioning stage, a single representative color is selected from each block, on averaging, along each(Y/Cb/Cr) channel. Y/Cb/Cr is transformed by 8x8 DCT, so three sets of 64 DCT coefficients are obtained. A zigzag scanning is performed with these three sets of 64 DCT coefficients, The DC and first few AC DCT coefficients for each channel constitute the CLD feature dimension. Color Layout Descriptor (CLD) [6] achieved higher detection accuracy than other candidate features. To summarize feature, we perform k-means based clustering and store the cluster centroids as its fingerprint. The number of

clusters is fixed at a certain fraction of video frames. Therefore, the fingerprint size varies with the video length. K-means based clustering generally produces compact video signatures, perform VQ based encoding on the signatures to obtain sparse histogram-based representations, compute the relevant keyframes, and then perform two-stage search to return the best matched model video.

*2) OCS:*Perception of color is usually not best represented in RGB. A better model of HVS is the so-call opponent color model. In [14], the opponent histogram is a combination of three 1-D histograms based on the channels of the opponent color space. The intensity is represented in channel O3 and the color information is in channels O1 and O2. Due to the subtraction in O1 and O2, the offsets will cancel out if they are equal for all channels (e.g. a white light source). Therefore, these color models are shift-invariant with respect to light intensity. The intensity channel O3 has no invariance properties. The histogram intervals for the opponent color space have ranges different from the RGB model.

### C. Distance Mesure

The duplicate detection task is to retrieve the best matching model video fingerprint for a given query fingerprint. The model-to-query distance is computed using a non-metric distance measure between the fingerprints. The distance function to compare a ($F_i \times p$) sized model fingerprint $X_i$ with the ($M \times p$) sized query signature Q is $d(X_i, Q)$.

$$d(Xi, Q) = \sum_{k=1}^{M}\{\| X_j^i - Q_k \Theta\} \qquad (1)$$

Where $\| X_j^i - Q_k \|$ refers to the L1 distance between $X_j^i$, the $j^{th}$ feature vector of $X^i$ and $Q_k$ , the $k^{th}$ feature vector of Q. Previous technique effectively depend on a single query frame and model video frame, and errors occur when this query (or model) frame is not representative of the query (or model) video. In distance function (1), d ($X^i$,Q) is computed considering all the "minimum query frame-to-model video" terms and hence, the effect of one (or more) mismatched query feature vector is compensated. 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.

### D. Video Matching

To perform efficient search, we propose a two phase procedure. The first phase is a coarse search to return the top-K nearest neighbors (NN) from all the N model videos. The second Phase uses the unquantized features for the top-K NN videos to find the best matching video $V_{i*}$.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, the first pass compares the model fingerprint $X^i$ with the query signature Q, and returns the K nearest videos. The second pass finds the best matched video Vi from these K videos, based on the larger query signature $Q_{orig}$.

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 (LBG) algorithm [10].

*1) Algorithm*

Step1: Query video is converted to number of frames.

Step2: Feature is extracted from frames of query videos by using CLD method.

Step3: The model-to-query distance is computed using distance measure

Step4: Coarse search return top k-NN from all the model videos

Step5: refined search find the best matching video $V_{i*}$

Step6: decides whether the query is indeed a duplicate derived from $V_{i*}$.

*Duplicate confirmation*, After finding the best matched video $V_{i*}$, to confirm whether it is indeed a duplicate use threshold approach on the model-to-query distance.

## IV. Experimental Result

We describe the duplicate detection experiment for feature comparison. A database contains 100 videos, worth about 1 hours of video content. 20 original videos are used to generate the query videos. We use various image processing and noise addition operations to generate the queries. gamma correction, JPEG compression at quality different factors, varying frame rates, We have experimented with different query lengths, as a duplicate can be constructed as a subset of a model video full-length, 35% of the original model video length.

We also perform experiment on video queries collected form YouTube. These *web videos* are identical or approximately identical videos close to the exact duplicate of each other, but different in file formats, encoding parameters, photometric variations (color, lighting changes), editing operations (caption, logo and border insertion), different lengths, and certain modifications (frames add/remove).

TABLE I. Comparison of the Similarity of CLD and OCS

| Similarity value | | |
|---|---|---|
| ***Input videos*** | ***CLD*** | ***OCS*** |
| Video1 | 65.21 | 76.19 |
| Video 2 | 79.5 | 87.14 |
| Video 3 | 69.28 | 74.78 |
| Video 4 | 71.0 | 81.0 |
| Video 5 | 64.70 | 87.5 |

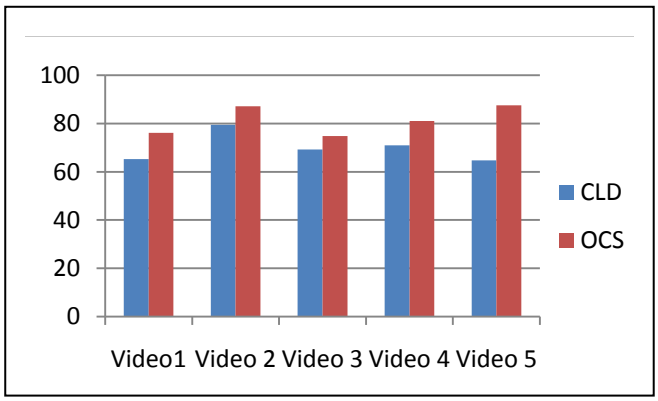


Figure 2. Similarity between videos with CLD and OCS

## V. CONCLUSION

This paper can deal with various kinds of video transformations, such as video compression, blurred, video cutting. As well, two feature extraction methods are used for extract feature from video, and find the similarity between videos using distance measure, and indexing method used to speed up the matching process. We retrieve the duplicate video, an average detection accuracy of over 97% when the query video as a noisy subset of the original video, and 80% detection accuracy when the query videos are real time videos.

## REFRENCES

[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] X. Wu, A. G. Hauptmann, and C. Ngo. Practical elimination of near-duplicates from web video search. In Proceedings of the 15th International Conference on Multimedia, pages 218–227. ACM, 2007.

[3] S. Cheung and A. Zakhor. Estimation of web video multiplicity. In Proc. SPIE–Internet Imaging, volume 3964, pages 34–36, 1999.

[4] 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.

[5] 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. of CIVR, pages 371–378. ACM, 2007.

[6] 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.

[7] M. Bertini, A. D. Bimbo, and W. Nunziati. Video clip matching using MPEG-7 descriptors and edit distance. In Proc. of CIVR, pages 133–142, 2006.

[8] A. Joly, C. Frelicot, and O. Buisson. Robust content-based video copy identification in a large reference database. In Int. Conf. onImage and Video Retrieval, pages 414–424, 2003.

[9] H. Lejsek, F. H. Asmundsson, B. Jonsson, and L. Amsaleg. NV-tree: An efficient disk-based index for approximate search in very large high-dimensional collections. IEEE Transactions on Pattern Analysis and Machine Intelligence, 99(1), 2008.

[10] K. Dadason, H. Lejsek, F. Asmundsson, B. Jonsson, and L. Amsaleg. Videntifier: identifying pirated videos in real-time. In Proc. Of the 15th International Conference on Multimedia, pages 471–472. ACM, 2007.

[11] A. 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.

[12] V. Athitsos, M. Potamias, P. Papapetrou, and G. Kollios. Nearest neighbor retrieval using distance-based hashing. Proc. of ICDE, pages 327–336, April 2008.

[13] A. 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.

[14] Koen E. A. van de Sande and Theo Gevers and Cees G. M. Snoek. Evaluation of Color Descriptors for Object and Scene Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, volume 32, pages1582 -1596.