ISSN:0975-9646

Nearest Duplicate Video Retrieval with Multiple Features using Temporal Correlation

Shital Bhirud[#1], Prof. P. S. Desai[#2]

[#]Department of Computer , Savitribai Phule University of Pune
Smt. Kashibai Navale College of Engineering, Pune, India

Abstract— **Online video database is increasing exponentially; a technique is needed for the Near-Duplicate Video Retrieval (NVDR). There are many programs for this strategy, in trademark protection, video labelling, video recording tracking etc. Though, lot of methods has been developed for NDVR over the past several years. To represent the video single feature is not sufficient as fast increase in video manipulation techniques. May other methods use multiple features but they were not able to face extensive quantity of videos. Thus, these methods were having less usage in real life programs. Our method is inspired by and improves upon recent work on Multiple Feature Hashing (MFH) designed to used multiple features to interpret the image instead of single feature may not represent the image exactly. Our Evaluation of the system shows that our proposed method substantially improves upon previously proposed methods in time efficiency, accuracy and scalability. To achieve the same or more efficient search quality temporal correlation used same hashing technique as basic MFH method while reducing number of keyframes by time between them. And object detection provides more accuracy.**

Keywords— ***M**ultiple feature hashing Near-duplicate videos (NDVs), Temporal correlation, Video copy detection, Video retrieval.***Introduction**

I. INTRODUCTION

Day by day use of web has increased to an excellent level. Internet searching methods are also designed very rapidly. Because of video capturing devices, and video modifying software as such methods, the variety of video clips is constantly on the increase at very quick rate. Everyday millions of video clips are uploaded daily online. Millions of video clips are submitted daily on a website like YouTube.

Thus huge amount of near-duplicate video clips (NDVs) are existing on the Web. These are generated from various kinds of ways like simple reformatting, to different products, changes, versions, and mixes of different effects. In many novel applications, such as trademark administration, video labelling, video recording usage tracking, video data source cleansing, cross-modal divergence recognition, video result re-ranking, and so on, the presence of massive NDVs enforces strong demand for efficient near-duplicate video recovery (NDVR).

A huge analysis is done in the domain of NDVR. There are many frameworks available for the NDVR. Those are mentioned in [1], [5], [6] and [13]. One cannot directly process the video. Thus using time testing or shot-boundary recognition methods, the video is separated into a series of key frames. Visible functions, such as color histogram, local Binary Pattern (LBP), etc represents the key frames. This series of key frames is considered as a signature of the video. The NDVR system then differentiate the two signatures that is trademark of the query video and the trademark of the video clips those are existing in the data source [4], [5]. Already analysis is done to accomplish real-time NDV recovery. It uses whole video with a single and international trademark. But this strategy did not confirm efficient for longer video clips. Another strategy of pair-wise framework connection between two video clips is designed to measure the near-duplicate connection. Use of multiple features for NDVR has been shown to be very efficient [5]. For example, local signature is less effective to the changes in frame rate amount, video length, etc. and global signature is delicate to changes in color contrast, brightness, scale, camera perspective, and so on.

As the numbers of video clips are increasing day by day with an exponential speed, the scalability of NDVR criteria has become very important issue of analysis. Along with the scalability, complexity of video is also increasing. Effective listing framework is needed for quick look for over a huge data source. But it requires real-time reaction. Many indexing methods are suggested like tree-structures [11], dimensionality reduction [9], hashing [7], etc. In an indexing framework how to maintain the near-duplicate connection among the video clips is an interesting part to see.

Motivation

In the traditional methods of Near Duplicate Video Retrieval, researchers have used many different methods. In many techniques they have used single feature to represent the video, but it is not feasible to characterize the whole video by single feature, since complexity of today’s videos is increasing daily. Hence it is important to develop NDVR techniques for handling multiple complex features of video. Also, in practice, it is necessary to get fast results than pre-processing. Hence, the time for searching similar videos in dataset should be less.

II. RELATED RESEARCH

A. NDVR

NDVR have been effectively considered in several types of authentic application. On the basis of different types of surveys and comparative studies various methods that are responsible for the development of huge number of NDVs are found in [2], [3] and [16]. Different strategies utilizing different functions and related techniques have been suggested for NDVR recently. The present techniques on NDVR can be generally separated into two groups, first is global features centered strategies and another is regional that is local features centered strategies. A regional feature centered strategies have likewise drawn in much research concern. Features, for example shade, surface and shape eliminated at the key-frame level are further portioned into several area models, and are especially suitable for recovering NDVs with complicated types. The well-known regional feature centered strategies integrate key-frames centered regional function recognition approach [4]. As an option to global feature centered strategies most of the NDVR strategies emphasize the fast identifying evidence of NDVs [1], [6]. In these works, video clips are verbal to by traditional international functions. These strategies perform well in concern of just about identical video clips. Very good example, in [1], the developers receive HSV to speak to their key-frames and further produce a function trademark by cumulating all the key-frames in the function. This reflection reaches fast restoration speed and high perfection in their dataset.

B. *Hashing*

Hashing is an essential strategy to accomplish quick similarity search. Many hashing techniques have been suggested. Hashing methods can be usually separated into two primary categories: random projector screen centered hashing techniques and device learning that is machine learning based hashing techniques. The distinction between these two categories lies in the creation of the hash features.

Random projector centered hashing techniques use unique vectors (with some particular submission, e.g., conventional Gaussian distribution) as the hashing functions. Locality sensitive hashing (LSH) [7], [8] is certainly one of the most representative random projector screen centered hashing techniques. LSH uses a group of locality delicate hash features consisting of linear projection over unique guidelines in the function area. The intuition behind is that for at last one of the hash features, nearby information factors have great possibility of being hashed into the same basket. However, in exercise, many hash platforms have to be designed for a higher look for precision, which takes in a large quantity of storage. To decrease the variety of hash tables, multi-probe LSH [9] is suggested. It can acquire the same look for great quality with much less platforms. Tao et al. [10] has recently suggested the concept of LSB-forest to further improve the efficiency. The hash requirements are showed as one-dimensional Z-order principles and listed in the -tree. Multiple trees developing an LSB-forest can be designed to enhance the search quality. However, since the hash features in all these hashing methods are arbitrarily produced and information separate, they are not very efficient. To accomplish an appropriate precision, it usually needs lengthy hash requirements in each hash desk. However, with the increase of the rule duration, the accident possibilities decreases very quickly, then many platforms are required to get an excellent recall.

C. *Multiple Feature Fusion*

The key issue lying in several function combinations is the recognition of the likeness or connection between two views talked to by several functions [5]. Late fusion techniques and early fusion techniques are the conventional methods for multiple source fusion. The late blending techniques [11], first create individual outcomes from unique functions, and subsequently negotiate these outcomes together by different techniques. The strategies don't consider the connection among functions. Furthermore, late fusion is computationally more costly for preparing. The early fusion systems attempt to join several functions at the data stage [12]. In reference [13] system used the regional descriptors for lightweight representations. Though, this method is very efficient for the replication recognition, the lightweight generation process may cause some information loss.

D. *Temporal Correlation*

A spatial connections and temporary connections methods are used commonly in image and video processing programs like [14], [15]. Both these functions are used in the techniques like [14], to evaluate the likeliness between the two video clips. The temporal correlations are used to identify the likeness between the two supports. We have used this technique in our suggested system, to produce the unique key supports. Once the key frames are produced, the evaluation between two video clips is relatively easy. In reference [6], two different techniques are developed for real time NDV clip detection; one is bounded coordinate system (BCS) which ignores temporal information and another is frame symbolization (FRAS) which considers temporal and symbol order.

In this paper, the temporal correlation is included in the system proposed by [5]. Due to temporal correlation between frames, system optimization is done which will result in increased performance.

III. Proposed System

A. *System Architecture*

We can divide our system into two phases like training phase and testing phase. Figure 1 shows the system architecture of the proposed system.

1) *Training Phase*

In training phase we take number of videos as input to the system, we extract frames from each video, we will get number of frames from this, but there is possibility that we can find various similar key frames. So we will take out unique key frames from collection of frames by applying temporal correlation. We have to extract various features of extracted key frames. For best retrieval of the search we are using hashing function for each keyframe so that it will make it easy to search.

2) *Testing Phase:*

Here in this phase, first we need to input video for searching purpose, we will extract frames from input video, then apply temporal correlation as we had applied in training phase. After that we will extract different features form that key frames. Hashing will be performed. At the end system will search dataset for similar videos and will retrieve such videos at the end.

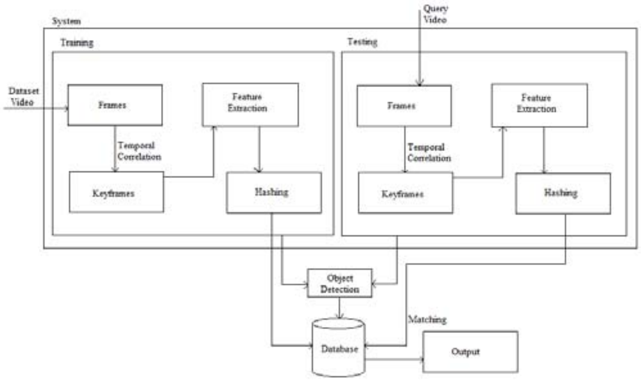

Fig 1: This system architecture shows all components of the system in both phases as well as object detection.

3) *Object Detection*

In this work object detection plays very important role of increasing accuracy. Here we locate keypoints first and then object is detected. If the frame to be matched with will contain same object then it will returned as related video.

B. Mathematical model

Set Theory

Let, System S is represented as:

S = {D, E, C, F, N, M, P, Q}

1) *Framing*

Consider, D is a set of frames, D= {f1, f2…} Where, f1, f2... are number of frames extracted from training video.

2) *Hash code generation:*

Let E is a set for hash codes, E = {c1, c2, c3...} where, c1, c2... are hash codes of each frame of the training video.

3) *Extracts keyframes:*

Let, C is a set for keyframes C = {e1, e2, e3...} where, e1, e2… are the number of keyframes of training video used for finding near duplicate videos.

4) *Query video framing:*

Let, S is a set of query video frames,
F= {s1, s2, s3...} where, s1, s2, ... are number of frames extracted from query video.

5) *Query video Hash code generation:*

Let N is a set for hash codes, N = {d1, d2, d3...} where, d1, d2... are hash codes of each frame of the query video.

6) *Extracts keyframe from query video:*

Let, M is a set for keyframes, M = {l1, l2, l3...} where, l1, l2, l3… are the numbers of keyframes of query video.

7) *Matching Hash code:*

Let, P is a set for matched hash code,
P = {n1, n2, n3...} where, n1, n2… are the number of keyframes of query video.

8) *Matched videos (Near duplicate videos):*

Let, Q is a set for near duplicate videos Q = {x1, x2, x3...} where, x1, x2… are the number of matched videos to query video.

C. Results

Following table shows the average training time and average detection time taken by all three systems for similar query video with no orientation change.

TABLE 1
Average time required for systems

Type of System	Training	Detection
Without Keyframes	182037	644937
With Keyframes	133495.7	132612
Using Temporal Correlation	84848.3	52561

Above table and below chart show that NDVR system without keyframes takes much more time because features are extracted for all frames which is very heavy process. It also shows that NDVR system with keyframes takes less time than time taken for system without keyframes but it needs more time than system using temporal correlation because temporal correlation reduces the number of keyframes.

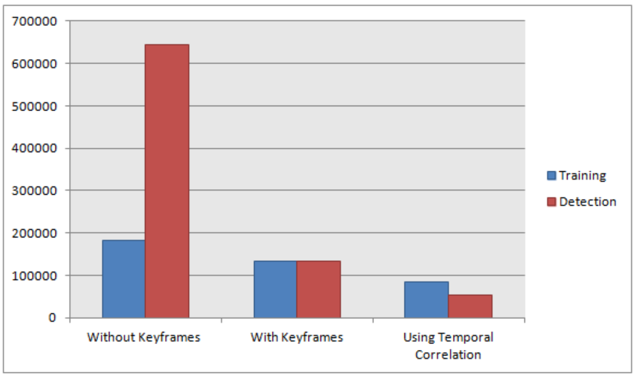

Fig 2: Above figure shows bar chart showing average training time and average time taken for detection of nearly duplicated video for three systems that are 1)NDVR without keyframes 2)NDVR with Keyframes and 3)NDVR using temporal correlation between keyframes, for same input.

IV. Conclusions

Here we proposed the method for nearest duplicate video retrieval to incorporate the temporal correlation between key frames within the same video with the multiple features hashing method. Due to consideration of temporal correlation method with the multiple features hashing, the performance of the system for detecting nearest duplicate video is improved. This system can provide similar videos in the dataset in less time as compared to the system without temporal correlation since it reduces the load of unnecessary feature extraction of all frames of the video.

REFERENCES

[1] X. Wu, C.-W. Ngo, and A. G. Hauptmann, "Practical Elimination of Near-Duplicates from Web Video Search", In ACM MM, 218–227, 2007.

[2] M. Cherubini, R. de Oliveira, and N. Oliver, "Understanding Near-Duplicate Videos: A User-Centric Approach," in Proc. ACM Multimedia, 2009, pp. 35–44.

[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. CIVR, 2007, pp. 371–378.

[4] K. Mikolajczyk and C. Schmid, "A Performance Evaluation of Local Descriptors," IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, no. 10, pp. 1615–1630, 2005.

[5] J. Song, Y. Yang, Z. Huang, H. T. Shen, and J. Luo, "Effective Multiple Feature Hashing for Large-Scale Near-Duplicate Video Retrieval", IEEE Transactions on Multimedia, vol. 15, no. 8, December 2013.

[6] H. T. Shen, X. Zhou, Z. Huang, J. Shao, and X. Zhou, "UQLIPS: A Real Time Near-Duplicate Video Clip Detection System," in Proc. VLDB, 2007, pp. 1374– 1377.

[7] A. Andoni and P. Indyk, "Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions," Commun. ACM, vol. 51, no. 1, pp. 117–122, 2008.

[8] M. Datar, N. Immorlica, P. Indyk, and V. S. Mirrokni, "Locality-sensitive hashing scheme based on p-stable distributions," in Proc. Symp. Computational Geometry, 2004, pp. 253–262.

[9] Q. Lv, W. Josephson, Z. Wang, M. Charikar, and K. Li, "Multi-probe LSH: Efficient indexing for high-dimensional similarity search," in Proc. VLDB, 2007, pp. 950–961.

[10] Y. Tao, K. Yi, C. Sheng, and P. Kalnis, "Efficient and accurate nearest neighbor and closest pair search in high-dimensional space," ACM TODS, vol. 35, no. 3, 2010.

[11] M.Wang, X.-S. Hua, R. Hong, J. Tang, G.-J. Qi, and Y. Song, "Unified Video Annotation Via Multi-graph Learning," IEEE Trans. Circuits Syst. Video Technol., vol. 19, no. 5, pp. 733–746, 2009.

[12] C. Snoek, M. Worring, and A. W. M. Smeulders, "Early Versus Late Fusion in Semantic Video Analysis," in Proc. ACM Multimedia, 2005, pp. 399– 402.

[13] M. Douze, H. Jégou, C. Schmid, and P. Pérez, "Compact video description for copy detection with precise temporal alignment," in Proc. ECCV, 2010, pp. 522–535.

[14] M. Douze, H. Jegou, and C. Schmid, "An image-based approach to video copy detection with spatio- temporal post-filtering," IEEE Trans. Multimedia, vol. 12, no. 4, pp. 257–266, 2010.

[15] Wang, W., Farid, H.,"Exposing Digital Forgeries in Interlaced and De- interlaced video", IEEE Transactions on Information Forensics and Security 2(3), 438–449 (2007).

[16] Liu, J., Huang, Z., Cai, H., Shen, H. T., Ngo, C. W., and Wang, W. 2013. Near-duplicate video retrieval: Current research and future trends. ACM Comput. Surv. 45, 4, Article 44 (August 2013), 23 pages. DOI: http://dx.doi.org/10.1145/2501654.2501658