An Effective Approach for Video Copy Detection and Identification of Misbehaving Users

S.Sujatha

Dhanalakshmi Srinivasan College of Engineering and Technology

Mamallapuram, Chennai, Tamilnadu, India

Abstract— **Fast development in the field of multimedia technology has become easier to store and access large amount of video data. This technology has editing and duplication of video data that will cause to violation of digital rights. So, copy rights security becomes a critical problem for the huge volume of video data. This has led to the requiring Video Copy Detection has been actively learned in a large range of the multimedia applications. The Auto dual-threshold method is utilized and it is segmenting the videos into segments with the content of homogeneous and then the key frame is extract from every segment. The SIFT features are extracted from that segments of key frames. Then Propose an SVD-based technique to match two video frames with the SIFT point set descriptors. To obtain the video sequence matching result propose a graph- based method. It is used to convert the video sequence into identifying the longest path in the frames to identify the video matching-result with time constraint. As the Proposed Work, suppose if the query video is matched then server identifies the unauthorized user and block that identified user at particular time. I.e. based on timestamp, again the same user queries the same video then that particular user timing will increases (i.e. 2 in to previous block time). Here we additionally provide opportunity to all users' i.e. (user's tries to login 3 attempts). Suppose if attempt of user is crossed the limit then that particular user will discard from the network.**

Keywords— **Video Copy Detection, SVD-SIFT, Keyframes, Features, Graph Based Sequence method, TimeStamp**

I. INTRODUCTION

The main objective of the project is to detect whether the query video frames are a copy of a video from the train video database or not

- ✓ Auto dual Threshold is used to eliminate the redundant frame.
- ✓ SVD-SIFT features are used to compare the two frames features sets points.
- ✓ Graph-based video sequence matching is used to match the query video and train video.
- ✓ If the query video is matched then server identifies the unauthorized user and block that identified user at particular time. I.e. based on timestamp.

With the rapid development of multimedia technologies and media , the copyrighted materials become easily copied, stored, and distributed over the Internet. This situation, aside from enabling users to access information easily, causes huge piracy issues. One possible solution to identify copyrighted media is watermarking. Digital watermarking was proposed for copyright protection and fingerprinting. The basic idea is to embed information into the signal of the media (audio, video, or photo). Some watermarks are visible (e.g., text or logo of the producer or broadcaster), while others are hidden in the signal, which cannot be perceived by human eye. Today all DVD movies, video games, audio CDs, etc. have fingerprints that prove the ownership of the material. As a disadvantage, watermarks are generally fragile to visual transformations (e.g., re-encoding, change of the resolution/bit rate). For example, hidden data embedded on a movie will probably be lost when the clip is compressed and uploaded to a video sharing web site. Besides, temporal information of the video segments (e.g., frame number, time-code) is also important in some applications. Watermarking technique is not designed to be used for video retrieval by querying with a sample video clip.

Definition of copy video: A video V1, by means of various transformations such as addition, deletion, modification (of aspect, color, contrast, encoding, and so on), camcording, and so on, is transformed into another video V2, then video V2 is called a copy of video V1.

Content-based copy detection (CBCD) is introduced as an alternative, or in fact, complementary research fields to watermarking approach. The main idea of CBCD is that the media visually contains enough information for detecting copies. Therefore, the problem of content-based copy detection is considered as video similarity detection by using the visual similarities of video clips.

Server

Keyframes are extracted from the reference video database and features are extracted from these keyframes. The extracted features should be robust and effective to transformations by which the video may undergo. Also, the features can be stored in an indexing structure to make similarity comparison efficient.

Client

Query videos are analyzed. Features are extracted from these videos and compared to those stored in the reference database. The matching results are then analyzed and the detection results are returned.

Based on the study, in these transformations, picture in picture is especially difficult to be detected. And for detecting this kind of video copies, local feature of SIFT is normally valid. However, matching based on local features of each frames in two videos is in high computational complexity. In this paper, we focus on detecting picture in picture and propose twin-threshold segmentation; feature set matching, and graph-based sequence matching method.

II. . RELATED WORK

An early method based on colour histogram intersection is proposed by Satoh. Yeh and Cheng use a method that partitions the image into 4 regions, and extracts a Markov stationary feature (MSF)-extended HSV colour histogram. Basharat et al. present a video-matching framework using spatio-temporal segmentation. A set of features (colour, texture, motion, and SIFT descriptors) is extracted from each segment, and the similarity between two videos is computed with a bipartite graph and Earth Mover's Distance (EMD).

TABLE 1

LIST OF TRANSFORMATIONS USED IN THE CBVCD TASK

#	Transformation details
T1	Camcording
T2	Picture-in-picture Type 1
T3	Insertion of patterns (15 different patterns)
T4	Strong re-encoding (change of resolution, bitrate)
T5	Change of gamma
T6	Combination of 3 transformations amongst: blur, gamma, frame dropping, contrast, compression, ratio, noise (A)
T7	Combination of 5 transformations amongst (A)
T8	Combination of 3 transformations amongst: crop, shift, contrast, caption, flip, insertion of pattern, picture-in-picture Type 2 (original video is behind) (B)
T9	Combination of 5 transformations amongst (B)
T10	Combination of 5 transformations amongst all the transformations from 1 to 9

Wu et al. propose that specific types of visual features (i.e., texture, intensity, motion, gradient, frequency, interest point) should be used for different types of transformations by a video near-duplicate video matching system. The methods based on points of interest and their trajectories are popular in this field. Joly et al. present a technique for content-based video identification based on local fingerprints. Local fingerprints are extracted around interest points detected with Harris detector, and matched with an approximate nearest neighbors search. In the same authors focus on the retrieval process of the proposed CBCD scheme by proposing statistical similarity search (S3) as a new approximate search paradigm. In, Joly et al. present distortion-based probabilistic approximate similarity search technique (DPS2) to speed-up conventional techniques like range queries and sequential scan method in a content-based copy retrieval framework. Zhao et al. extract PCA-SIFT descriptors for matching with approximate nearest neighbour search, and train SVMs to learn matching patterns. Law-To et al. present a video indexing approach using the trajectories of points of interest along the video sequence. They compute temporal contextual information from local descriptors of interest points, and use this

information in a voting function for matching video segments. Ren et al. employ a similar technique by taking into account spatial and temporal changes of visual words constructed by SIFT descriptors and bag-of-words approach. Williams et al. propose a video copy detection method based on efficiently matching local spatiotemporal feature points with a disk-based indexing scheme. In general, extracting and matching points of interest are costly operations in terms of computation time.

There are also promising copy detection techniques based on the similarity of temporal activities of video clips. Mohan presents a video sequence matching technique that partitions each frame into 3 x 3 images and computes its ordinal measure to form a fingerprint. The sequences of fingerprints are compared for video similarity matching. Kim and Vasudev use ordinal measures of 2 x 2 partitioned image and consider the results of various display format conversions, e.g., letter-box, pillar-box.

Some video similarity detection methods take the advantage of visual features that can be directly extracted from compressed videos. Ardizzone et al. use MPEG motion vectors as an alternative to optical flows, and show that the motion-based video indexing method they propose does not require a full. There are numerous descriptors for near-duplicate image or video detection available in the literature. Global statistics, such as color histograms, are widely used to efficiently work with a large corpus. These global descriptors are, in general, efficient to compute, compact in storage, but insufficiently accurate in terms of their retrieval quality. Alternatively, local statistics, such as interest points calculated with local descriptors, were proposed in. This description type is relatively invariant and, thus, robust to image transformations such as occlusions and cropping. However, local descriptors require more storage space and matching between them is computationally more complex. In the video domain, both global and local descriptors have been extended to incorporate temporal information. Law-To et al. presented a comparative study for video copy detection and concluded that, for small transformations, temporal ordinal measurements are effective, while methods based on local features demonstrate more promising results in terms of robustness. However, Thomee et al. conducted a large-scale evaluation of image copy detection systems and reached a somewhat different conclusion. Their chosen method that used interest points performed poorly due to its inability to find similar sets of points between copies. They concluded that either a simple median method or the retina method performs the best. To design a practical copy detection system which meets the scalability requirements, a compact, frame-level descriptor that retains the most relevant information, instead of just sets of interest point descriptors, is desirable. Furthermore, frame level descriptors are readily integrated into fast detection frameworks such as the one presented in. decomposition of the video, and thus, it is computationally efficient. Bertini et al. present a clip-matching algorithm that use video fingerprint based on standard MPEG-7 descriptors. An effective combination of color layout descriptor (CLD), scalable color descriptor (SCD), and edge histogram descriptor (EHD) forms the fingerprint. Fingerprints are extracted from each clip, and they are

compared using an edit distance. Sarkar et al. use CLD as video fingerprints and propose a non-metric distance measure to efficiently search for matching videos in high-dimensional space.

Hampapur and Bolle made a comparative analysis of color histogram-based and edge-based methods for detecting video copies. Another study by Hampapur et al. compares motion direction, ordinal intensity signature, and color histogram signature matching techniques. As a result of this study, they conclude that the techniques using ordinal features outperform the others. State-of-the-art copy detection techniques are evaluated in the comparative study by Law-To et al. Compared descriptors are categorized into 2 groups: global and local. Global descriptors use techniques based on the temporal activity, spatial distribution and spatio-temporal distribution. Local descriptors compared in their study are based on extracting Harris interest points for keyframes with high global intensity of motion (AJ), for every frame (ViCopT), and interest points where image values have significant local variations in both space and time. It is stated that no single technique is optimal for all applications; but ordinal temporal measure is very efficient for small.

For the identification of content links between video sequences, supporting the range of applications described above, Content-Based Copy Detection (CBCD) is a very relevant tool. Actually, most of the recent video mining developments just mentioned are CBCD-related methods. By copies we understand potentially transformed versions of original video sequences. The transformations belong to a large family and their amplitude varies significantly (e.g. Fig. 2).

Fig. 1. A framework of video copy detection system.

But CBCD methods that are robust to a wide range of transformations are also computationally expensive, and the cost of Video Mining by content-based Copy Detection (VMCD inthe following) is even higher.

Fig. 2. Copy(left) and Original Content(Right)

III. AUTO DUAL-THRESHOLD METHOD

An auto dual-threshold method to eliminate redundant video frames. This method cuts continuous video frames into video segments by eliminating temporal redundancy of the visual information of continuous video frames.

This method has the following two characteristics. First, two thresholds are used. Specifi-cally, one threshold is used for detecting abrupt changes of visual information of frames and another for gradual changes. Second, the values of two thresholds are deter-mined adaptively according to video content. The auto dual-threshold method to eliminate the redundant frames is shown in Fig. 3.
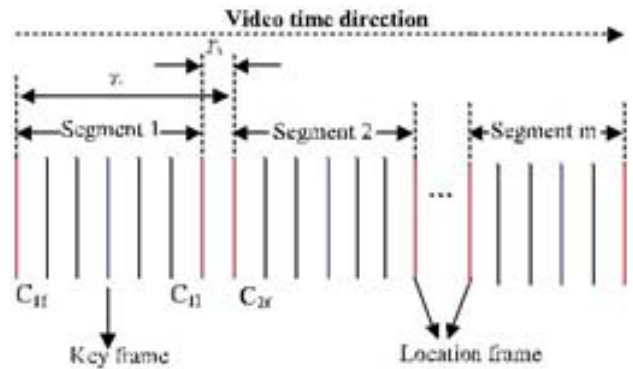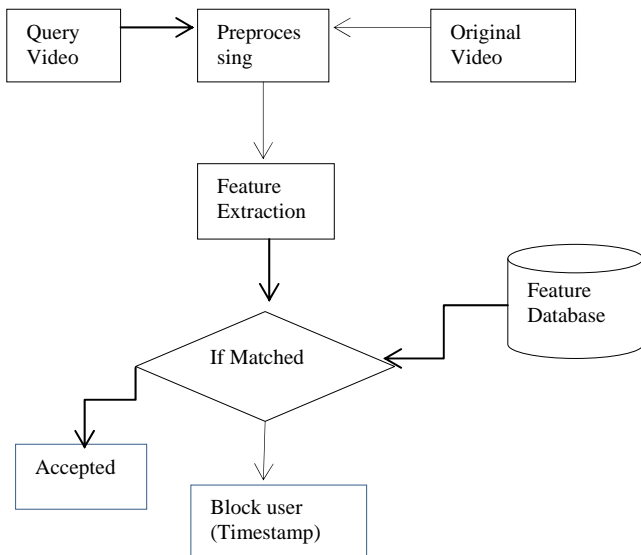

Fig. 3. Auto dual-threshold method to eliminate redundant video frames, select keyframes. C_{1f} denotes the first frame of Segment 1, C_{1l} the last frame of the Segment 1; C_{2f} denotes the first frame of the Segment 2

IV. SVD-SIFT MATCHING

In this section we discuss the use of the SIFT descriptor in the SVD-matching algorithm. As mentioned in the previous section SVD-matching presented in [16] does not perform well when the baseline starts to increase. The reason for this behavior is in the feature descriptor adopted. The original algorithm uses the grey level values in a neighborhood of the keypoint. As pointed out in Section 2 this description is too sensitive to changes in the view-point and more robust descriptor have been introduced so far. A comparative study of the performance of various feature descriptors showed that the SIFT descriptor is more robust than others with respect to rotation, scale changes, view-point change, and local affine transformations.
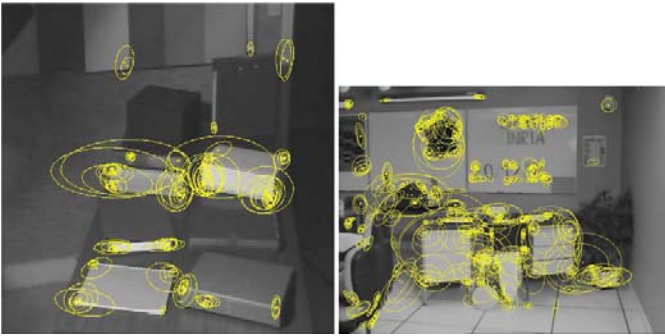
Fig. 4. Examples of features extracted. The ellipse around the feature points represents the support area of the feature.

In the same work, cross-correlation between the image grey levels returned unstable performance, depending on the kind of transformation considered. The considerations above suggested the use of a SIFT descriptor, instead of grey levels. The descriptor is associated to scale and affine invariant interest points [27], briefly sketched in Section 2. Some examples of such key points are shown in Fig.4.

IV.GRAPH METHOD FOR VIDEO SIMILARITY CHECKING

The graph-based video sequence matching method for video copy detection. The method is presented as follows:

Step 1: Segment the video frames and extract features of the key frames. According to the method described in Section 3, we perform the dual-threshold method to segment the video sequences, and then extract SIFT features of the key frames.

Step 2: Match the query video and target video.
Assume that
Q_c ¼ fC_1^Q; C_2^Q; C_3^Q; . . . ; C_m^Qg and T_c ¼ fC_1^T ; C_2^T ; C_3^T ; . . . ; C_n^T g are the segment sets of the query video and target video from Step 1, respectively. For each C_i^Q in the query video, compute the similarity sim(C_i^Q; C_j^T), and return k largest matching results. K ¼ _n, where n is the number of segments in the target set, and _ is set to 0.05 based on our empirical study.

Step 3: Generate the matching result graph according to the matching results. In the matching result graph, the vertex M_{ij} represents a match between C_i^Q and C_j^T . To determine whether there exists an edge between two vertexes, two measures are evaluated.

- ✓ **Time direction consistency:** For M_{ij} and M_{lm}, if there exists (i-1)*(j-m) then M_{ij} and M_{lm} satisfy the time direction consistency.

- ✓ **Time jump degree:** For M_{ij} and M_{lm}, the time jump degree between them is defined as

$$\Delta t_{lm}^{ij} = \max(| t_i - t_j |, | t_j - t_m |)$$

If the following two conditions are satisfied, there exists an edge between two vertexes:
1. The two vertexes should satisfy time direction consistency.
2. The time jump degree $\Delta t < \tau$ (τ is a preset threshold based on our empirical study).

Condition 1 indicates that if the query video is a copy deriving from the target video, then the video subsequence temporal order between query video and target video must

be consistent, which is reasonable in real application. If Condition 1 is satisfied, Condition 2 is used to constrain the time span of two matching results between the query video and the target video. If the time span exceeds a certain threshold, it is considered that there does not exist certain correlation between the two matching results. This method is similar to the probability model in [14].Also, as an example, the matching results in can be converted into a matching result graph. Obviously, the matching result is a directed acyclic graph. In the graph, in Case 1, because of violating the condition of time direction consistency, it does not exist an edge between $M_{2;29}$ and $M_{3;26}$. For Case 2, although it meets time direction consistency, the time jump between $M_{4;30}$ and $M_{5;70}$ exceeds the threshold, so it also does not exist an edge between $M_{4;30}$ and $M_{5;70}$. For each vertex of the matching result graph, it may have more than one path or no path. For example, for vertex $M_{1;29}$, $M_{1;76}$, $M_{2;76}$, it has not any path to other vertexes (or say the path is the vertex itself).

Step 4: Search the longest path in the matching result graph. The problem of searching copy video sequences is now converted into a problem of searching some longest paths in the matching result graph. The dynamic programming method is used in this paper. The method can search the longest path between two arbitrary vertexes in the matching result graph. These longest paths can determine not only the location of the video copies but also the time length of the video copies.

Step 5: Output the result of detection. For each vertex of the matching result graph, it has more than one path or no path. As in Fig. 6, for the vertexes $M_{1;29}$, $M_{1;76}$, and $M_{2;76}$, they have no path to other vertexes, or only have path to the vertex itself. For $M_{1;26}$, four paths are available. Accordingly, we need to combine these paths that overlap on time. Then, we can get some discrete paths from the matching result graph; it is thus easy to detect more than one copy segments by using this method. For each path, we use (3) to compute the similarity of the video

$$sim(path) = \frac{\sum_{k=1}^{m} sim_k(M_{ij})}{m} \log(1+m)$$

where m is the number of vertexes of the path, M_{ij} is the vertex in the path, $sim(M_{ij}) = sim(C_i^Q, C_j^T)$. According to the start point and end point of the path, we can obtain the time stamp of the two copies.

V.TIMESTAMP TO BLOCK THE MISBHEVING USERS

Timestamp
A timestamp is the time at which an event is recorded by a computer, not the time of the event itself. In many cases, the difference may be inconsequential: the time at which an event is recorded by a timestamp (e.g., entered into a log file) should be close to the time of the event.
This data is usually presented in a consistent format, allowing for easy comparison of two different records and tracking progress over time; the practice of recording timestamps in a consistent manner along with the actual

data is called **time stamping**. The sequential numbering of events is sometimes called time stamping.

Timestamps are typically used for logging events or in a sequence of events (SOE), in which case each event in the log or SOE is marked with a timestamp. In filesystems, timestamp may mean the stored date/time of creation or modification of a file.

TimeStamp for Video copy detection

Server monitors each and every users query videos. It will continuously monitor the query videos of each and every user's communication. When monitoring the server will identify the copied frames from the input queries i.e. the matching frame result is been verified from the database which is already been trained. It identifies the unauthorized user and block that identified user at particular time. I.e. based on timestamp, again the same user queries the same video then that particular user timing will increases (i.e. 2 in to previous block time). Here we additionally provide opportunity to all users' i.e. (user's tries to login 3 attempts). Suppose if attempt of user is crossed the limit then that particular user will discard from the network.

Advantages of Timestamp

> Server will identify misbehaving user in the proposed system that is based on timestamp and block that identified user at particular time.

> Additionally give opportunity to all user's i.e. user's tries to login 3 attempts and block that identified user at particular time.

VI. EXPERIMENTS

Feature Extraction for Video Copy Detection

In video copy detection, the signature is required to be compact and efficient with respectto large database. Besides, the signature is also desired to be robust to various coding

variations. In order to achieve this goal,many signature and featureextraction methods are presented for the video identification and copy detection tasks[11] [12] [13] [14] [15] [16].As one of the common visual features, color histogram is extensively used in videoretrieval and identification [12] [11]. [12] applies compressed domain color features to form compact signature for fast video search. In [11], each individual frame is represented by four 178-bin color histograms in the HSV color space. Spatial information

is incorporated by partitioning the image into four quadrants. Despite certain level of success in [12] and [11], the drawback is also obvious, e.g. color histogram is fragile to color distortion and it is inefficient to describe each individual key frame using a color histogram as in [12]. Another type of feature which is robust to color distortion is the ordinal feature. Hampapur et al. [13] compared performance of using ordinal feature, motion feature and color feature respectively for video sequence matching. It was concluded that ordinal

signature had the best performance. The robustness of ordinal feature was also proved in [14].As a matter of fact, many works such as [3] and [14] also incorporate the combined feature in order to improve the performance of retrieval and identification. Generally, the selection of

ordinal feature and color feature as signature for copy detection task is motivated by the following reasons:

(1) Compared with computational cost features such as edges, texture or refined color histograms which also contain spatial information (e.g. *color coherent vector* applied in

[15]), they are inexpensive to acquire (2) Such features can form compact signatures [] and retain perceptual meaning.(3) Ordinal features are immune to global changes in the quality of the video and also contain spatial information, hence are a good complement to color features [14].

Ordinal Feature Description

In our approach, we apply Ordinal Pattern Distribution (*OPD*) histogram proposed in [13] as the ordinal feature. Different from [26], the feature size is further compressed in this paper, by using more compact representation of I frames. Figure 2 depicts the operations of extracting such features from a group of frames. For each channel c =Y, Cb, Cr, the video clip is represented by *OPD* histograms as:

$$H_c^{OPD} = (h_1, h_{2,\ldots\ldots}, h_{l,\ldots\ldots} h_N) 0 \le h_i \le 1 and \sum_i h_i = 1$$

Here $N= 4! = 24$ is the dimension of the histogram, namely the number of possible patterns mentioned above. The total dimension of the ordinal feature is $3 \times 24 = 72$.

Color Feature

For the color feature, we characterize the color information of a GoF by using the cumulative color information of all the sub-sampled I frames in it. For computational simplicity,

Cumulative Color Distribution (*CCD*) is also estimated using the DC coefficients from the I frames. The cumulative histograms of each channel (c=Y, Cb, Cr) can be defined as:

$$H_C^{CCD} = \frac{1}{M} \sum_{i=b_k}^{b_{k+M}-1} H_i(i) j = 1, \ldots B$$

where H_i denotes the color histogram describing an individual I frame in the segment. M is the total number of I frames in the window and B is the color bin number. In this paper, $B = 24$ (uniform quantization). Hence, the total dimension of the color feature is also $3 \times 24 = 72$, representing three color channels.

CONCLUSION

In this research propose a framework for content-based copy detection and video similarity detection. The proposed framework Based on the analysis, we use local feature of SIFT to describe video frames. Since the number of SIFT points extracted from a video is large, so the copy detection using SIFT features has high computational cost. Then, we use a dual-threshold method to eliminate redundant video frames and use the SVD-based method to compute the similarity of two SIFT feature point sets. After that graph based video sequence matching method are utilized for matching the each frame from the video sequence Thus, detecting the copy video becomes finding the longest path

in the matching result graph are obtained. Suppose if the result of the frame is matched, i.e the matching frame result is been verified from the database which is already been trained. It identifies the unauthorized user and block that identified user at particular time. I.e. based on timestamp, again the same user queries the same video then that particular user timing will increases (i.e. 2 in to previous block time).

REFERENCES

[1] X. Wu, C.-W. Ngo, A. Hauptmann, and H.-K. Tan, "Real-Time Near-Duplicate Elimination for Web Video Search with Content and Context", Feb. 2009.

[2] O. Ku¨cu¨ ktunc, M. Bastan, U. Gu¨du¨ kbay, and O ¨. Ulusoy, "Video Copy Detection Using Multiple Visual Cues and MPEG-7 Descriptors", 2010.

[3] J. Law-To, C. Li, and A. Joly, "Video Copy Detection: A Comparative Study," Proc. ACM Int'l Conf. Image and Video Retrieval, pp. 371-378, July 2007.

[4] X. Zhou, L. Chen, A. Bouguettaya, Y. Shu, X. Zhou, and J.A. Taylor, "Adaptive Subspace Symbolization for Content-Based Video Detection," IEEE Trans. Knowledge and Data Eng., vol. 22, no. 10, pp. 1372-1387, Oct. 2010.

[5] A. Joly, O. Buisson, and C. Frelicot, "Content-Based Copy Retrieval Using Distortion-Based Probabilistic Similarity Search," IEEE Trans. Multimedia, vol. 9, no. 2, pp. 293-306, Feb. 2007.

[6] H. Liu, H. Lu, and X. Xue, "SVD-SIFT for Web Near-Duplicate Image Detection," Proc. IEEE Int'l Conf. Image Processing (ICIP '10), pp. 1445-1448, 2010.

[7] TRECVID 2008 Final List of Transformations, http://www-nlpir. nist.gov/projects/tv2008/active/copy.detection/final.cbcd. video.transformations.pdf, 2008.

[8] N. Guil, J.M. Gonza´lez-Linares, J.R. Co´zar, and E.L. Zapata, "A Clustering Technique for Video Copy Detection," Proc. Third Iberian Conf. Pattern Recognition and Image Analysis, Part I, pp. 452- 458, June 2007.

[9] K. Sze, K. Lam, and G. Qiu, "A New Key Frame Representation for Video Segment Retrieval," IEEE Trans. Circuits and Systems Video Technology, vol. 15, no. 9, pp. 1148-1155, Sept. 2005.

[10] J. Yuan, L.-Y. Duan, Q. Tian, S. Ranganath, and C. Xu, "Fast and Robust Short Video Clip Search for Copy Detection," Proc. Pacific Rim Conf. Multimedia (PCM), 2004.

[11] S. Cheung and A. Zakhor, "Efficient video similarity measurement with video signature," In *IEEE Trans. on Circuits and System for Video Technology*, vol. 13, pp. 59-74, 2003.

[12] M.R. Naphade et al., "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, 2000.

[13] A. Hampapur, K. Hyun, and R. Bolle., "Comparison of Sequence Matching Techniques for Video Copy Detection," In *SPIE. Storage and Retrieval for Media Databases 2002*, vol. 4676, pp. 194-201, San Jose, CA, USA, Jan. 2002.

[14] Junsong Yuan et al., "Fast and Robust Search Method for Short Video Clips from Large Video Collection," in *Proc. of ICPR'04*, Cambridge, UK, Aug. 2004.

[15] R. Lienhart et al., "VisualGREP: A Systematic method to compare and retrieve video sequences," In*SPIE. Storage and Retrieval fro Image and Video Database VI*, Vo. 3312, 1998.

[16] J. Oostveen et al., "Feature extraction and a database strategy for video fingerprinting," In *Visual 2002, LNCS 2314*, pp. 117-128, 2002.