ISSN:0975-9646

Anju P S et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (1) , 2015, 152-158

Video Copy Detection Using F-Sift and Graph Based Video Sequence Matching

Anju P S#1, Soumya Varma*2, Vince Paul#3, Sankaranarayanan P N#4

[1]*Student, Computer Science and Engineering Department, Calicut University, Sahrdaya College of Engineering and technology*

[2]*Assistant Professor, Computer and Engineering Department, Sahrdaya College of Engineering and technology*

[3]*Head of Department, Computer and Engineering Department, Sahrdaya College of Engineering and technology*

[4]*Assistant Professor, Computer and Engineering Department, Sahrdaya College of Engineering and technology*

Abstract— **Content based copy detection aims to detect copies that of a given media. As with the growth of technology more and more media contents are available in the internet. This large number of copies leads to violation of digital rights. So we need an effective and efficient method to detect duplicated media contents. An auto dual threshold method is used to eliminate redundant video frames of a video segment which will reduce non necessary matching of video frames. Then used local features of F-SIFT for video content description. Flip-invariant SIFT (or F-SIFT), that preserves the original properties of SIFT while being tolerant to flip like transformations. Since matching computational cost of F-SIFT is very large, so uses 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.**

Keywords— **FSIFT Feature, graph, SIFT Feature, SVD, and graph Based matching**

I. INTRODUCTION

With the quick growth in the internet and multimedia technology, we are able to access and store huge volumes of video data easily. That is huge volumes of videos are transmitted, searched and stored on the internet. Some statistics of the YouTube shows that, there are about 100 hrs of user generated videos are uploaded to YouTube every minute. According to BBC motion gallery, it contains over 2.5 million hours of professional video contents. Among these huge volumes of videos there exist a large numbers of duplicated and near duplicated videos. It is reported that about 27% videos in a video search results obtained from YouTube, Google & yahoo videos are duplicated or near duplicated copies of a popular version. For particular queries, the redundancy can be as high as 93%.A duplicate video means we can divide it into two Duplicate Videos and Nearly Duplicated Videos. Duplicated Video will be extracted video copies that can be easily detected. Near Duplicated video copies are transformed video clips and detection of such copies is challenging. So we can define a video copy as, it is a segment of video derived from another video usually by means of various transformations such as addition, deletion, modification and cam coding. According to the definition of video copy in TRECVID2008 tasks, A video V1, by means of various transformations such as addition, deletion, modification(of aspect, colour, contrast, encoding, and so on),cam cording, and so on, is transformed into another video V2,then video V2 is called a copy of video V1. In content based copy detection task of TRECVID 2008, 10 Transformations are defined.

T1. Cam-cording; T2. Picture in picture; T3. Insertions of pattern: Different patterns are inserted randomly: captions, subtitles, logo, sliding captions; T4. Strong re-encoding; T5.

Change of gamma; T6, T7. Decrease in quality: Blur, change of gamma (T5), frame dropping, contrast, compression (T4), ratio, white noise; T8, T9. Post production: Crop, Shift, Contrast, caption (text insertion), flip (vertical mirroring), Insertion of pattern (T3), Picture in picture (the original video is in the background); T10. Combination of random five transformations among all the transformations described above. Fig 1 shows image Examples for 10 transformations,

Fig 1 These videos come from MUSCLE –VCD-2007 and TRECVID 2008

This kind of editing and duplication of video data may lead to violation of digital rights. Also users are often frustrated when they need to spend significant amount of time to find videos of interest, they have to go through number of duplicated or nearly duplicated videos that are streamed over the internet before arriving at an interesting one. This duplicated video leads to wastage of storage space. To avoid this situation that is getting overwhelmed by a huge amount of repeating copies of the same video in search, we need an efficient and effective copy detection and elimination which is essential for effective search, retrieval and browsing.

Application of this video copy detection includes, Copy Right Enforcement allows doing detection and elimination of video copies. And Usage Tracking allows keeping track of when this video clips aired and how many times. Also people will be more satisfied if the near duplicated videos or duplicated videos in search results returned from video search engines can be merged into one. And video sharing websites need only to store one video as representative of all its copies to greatly reduce the storage. The main idea behind video copy detection system is to find whether a queried video is a copy of a video from the video data set. A copy can be created by various transformations. Based on a study some complex transformations are difficult to detect. For detecting such kind of copies local features of SIFT is normally valid. But SIFT is not flip invariant. Nowadays we can observe flip or flip invariant transformations due to artificial flipping, opposite capturing viewpoint, or symmetric patterns of objects. However matching based on F-SIFT local features of each frame within two videos leads to high computational cost. This paper mainly focus on feature extraction using F-SIFT [16] with enhanced security, a method for reducing redundancy in frames, Feature set matching with SVD,and a graph method for video similarity checking.

II.TECHNIQUES USED

There are two basic approaches to address the issue of copyright protection – water-marking and content-based copy detection. In the first approach, watermark/non-visible information is embedded into the content and later, if required, this embedded information is used for establishing the ownership. But watermarking is not applicable for the video sequences already in circulation without any such embedded information.

On the other hand, in content-based approach, no additional information is inserted. It is said that “Video itself is the watermark”. So in this approach unique signatures (features) are derived from the content of the video itself. Such signatures are also extracted from the questioned video and are compared with those of the original media stored in the database. Performance of a video copy detection scheme relies on the suitable signature of the frame sequence and also on the sequence matching scheme. The system must be robust to the presence of various distortions adopted by the copier. Such attacks/distortions may be broadly classified as photometric attack (change in brightness/contrast, contamination by noise) and post-production attack (change in display format, logo insertion etc.). To handle such attacks, attempts have been made to design robust signatures or different post-processing techniques are adopted. A content based video copy detection system consists of two major modules namely, fingerprint generation and sequence matching technique.

A.Fingerprint

Fingerprint can be defined as perceptual features for short summaries of a multimedia object. The goal of video fingerprinting is to judge whether two video have the same contents even under quality preserving distortions like resizing, frame rate change, lossy compression. Sequence matching technique detects whether a query sequence is copied version of referenced one or not based on their fingerprints. Fingerprint must be robust so that the fingerprint of a copied video (degraded to whatever extent possible) and the original one should be similar. On the other hand, the perceptually different video sequences should have different fingerprints. Thus, the selected fingerprint should meet two diverging requirements. Based on the features they extract, fingerprint extraction algorithms can be classified into five, colour-space-based, temporal, spatial, Spatio-temporal and texture. In the colour based extraction method, colour information from the histogram of the colour is extracted. In temporal characteristics of video over time are extracted. Texture fingerprints refers to surface characteristics and appearance of an object given by the size, shape, density, arrangement, proportion of its elementary parts. But these features alone are not suited for video copy detection. In Spatial fingerprints are features derived from each frame or from a key frame. Fingerprint or descriptor of a video sequence can be broadly categorized as global or local one. Global ones are derived from the whole video sequence or from a subset of sequence. They are generally not very robust as a change in part of the image may cause it to fail as it will affect the resulting descriptor.

Local descriptors are computed for each individual frame in the video. Local descriptors are transformed into concise form to generate the global fingerprint. Methods based on Local descriptors have better detection performance in various complex transformations. Multiple local descriptors are used to match an image and this is more robust as not all the descriptors need to match for the comparison to be made. This makes them more robust to changes between the matched images. SIFT is a good example of Local descriptors. A wide variety of frame level features have been tried by the researchers to generate the fingerprints. Colour histogram is very widely used one. But, it lacks in terms of discriminability as the spatial distribution of colour is not retained in the histogram.

B.Sequence Matching

The query video sequence and reference sequences in the database are to be matched on the basis of extracted signature. Varieties of matching schemes have been tried by the researchers that can be broadly classified as (i) dense matching technique and (ii) sparse matching technique. Dense scheme considers all the frames for comparison. But, a sparse technique deals with representative frames (key frames) only. Hence, a sparse technique is faster but a dense one is more robust.

In dense matching technique, query video sequence (sq) is matched with the sub-sequence (sr) of same length taken from the reference video sequence. Different sr is obtained by shifting the start position of the sub-sequence in reference video. If the distance between sq and the most similar sr is less than a pre-defined threshold then sq is taken as a copy. Selection of threshold is very crucial. Moreover, the query video sequence cannot exceed the length of the reference video sequence.

Global descriptors like temporal and spatio-temporal measures incorporate frame level comparison and reflect the dissimilarity between sq and sr. But in case of local descriptor based system, the distance between sq and sr is to be computed by combining the distance between the feature vectors of corresponding frames in sq and sr. Depending on the features, measures like Euclidean distance, histogram intersection are widely used. Several key frame based schemes (sparse technique) have been reported. In such technique, selection of key frame is an important task. Sparse technique deals with representative Frames only. Sparse technique is faster but a dense one is more robust. In order to cope up with the attacks/deformations incorporated in the copied version, researchers have mostly focused on the robustness in designing the signatures of the video sequence and also on the tolerance allowed by the matching strategies.

III. RELATED WORKS

In the feature extraction process, descriptors extracted from the video and image fall into two categories as global and local descriptors. Global descriptors are derived from the whole video sequence, such descriptors like ordinal measure, have poor performance with local transformations such as shift, cropping and cam coding. While local descriptors are derived from representative frames only and is robust to both local and global transformations. Hessian Based STIP descriptor (spatio temporal), Harris interest point descriptors which are comes under local descriptors but behaves poorly with addition of salt and pepper noise and also not invariance to scale and affine changes. Harris interest point detector is based on the Eigen values of second moment matrix. Another widely used local descriptor is SIFT, which has best copy recognition performance. The hessian based descriptor SURF [17], is another scale invariant and robust local descriptor with similar performance as SIFT, and similar speed as the Harris descriptor, which is faster than SIFT.

In the matching process, most recent methods utilizes index structure like hash structures and tree structures. In most of methods, features are extracted from each key frame and similarity search carrying out.

An early method based on colour histogram proposed by satoh.yeh and ching uses a method that extracts a markov stationary feature (MSF), Extended HSV colour histogram. Another method proposed in [1] by x.wu, c.w.Ngo was based on global colour histogram descriptor but detection complicated copies was not satisfactory. In [2] by Matthijs Douze,Herv´eJ´egou and Cordelia Schmid proposed a detection method with spatio temporal post filtering based on an asymmetric sampling strategy to select fixed number of frames and uses Hessian interest point detector in combination with automatic scale selection. In [3], zhang proposed a near duplicate detection by stochastic attributed relational graph matching. Such image to image comparison provides robustness and discrimination but prevent the use of large databases. In [4] by Alexis Joly, Olivier Buisson, and Carl Frélicot proposed a copy retrieval using distortion based probabilistic similarity search. Such similarity search is robust to clutter and occlusion but not invariance to scale and affine changes. In [5] by Zi Huang, Heng Tao Shen, Jie Shao, Bin Cui proposed a method to transform a video stream into a one dimensional video distance trajectory monitoring the continuous changes of consecutive frames with respect to a reference point, which is further segmented and represented by a sequence of compact signatures called linear smoothing functions (LSFs).LSF adopts compound probability to compute the sequence similarity search for near duplicate detection. In [6] by Vishwa Gupta , Parisa Darvish Zadeh Varcheie , Langis Gagnon , and Gilles Boulianne proposed a nearest neighbour mapping algorithm for mapping test frame to the query frame and count the number of frames of query that match the frames in the test segment. In [7] bertini et.al present a clip Matching algorithm that are video fingerprint based on standard MPEG-7 descriptors. In [8] Law t et.al and

Joly et.al adopted Harris corner point or feature points in video frames. In [9] Satoh et.al detects duplicate using trajectory characteristics of the feature points. Similarly zhpou et.al [10] proposed a shot based interest point selection approach for near duplicate search. In this paper uses F-SIFT [16] instead of SIFT. In [11] proposed local descriptors such as SPIN Image, RIFT which are flip invariant. However both descriptors are sensitive to scale changes. And also not as discriminative as SIFT.F-SIFT enriches SIFT with flip invariant property.

IV. PROPOSED METHOD

As we know, content based copy detection is an effective approach for video copy detection. Below figure 2 shows the framework of content based copy detection. The framework consist an online part and an offline part. During offline part the key frames are extracted from the reference video database. Features are extracted from the Extracted key frame and are saved in the feature database. During online part key frames are extracted from the query video. Features are extracted from the key frames and a similarity search has been done with the feature database and matching results are analysed and results are returned. In the proposed method uses F-SIFT for feature extraction.

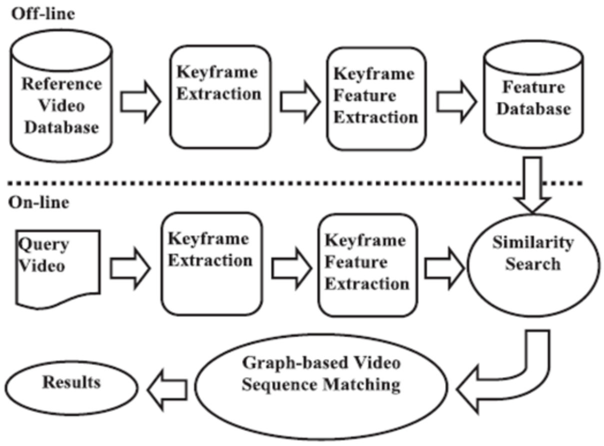
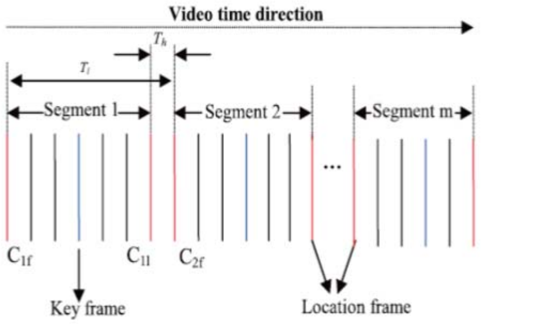


Fig 2 Framework of CBCD



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.

Fig 3 Key Frame Extraction

Features are extracted from the key frames and a similarity search has been done with the feature database and matching results are analysed and results are returned. In the proposed method uses F-SIFT for feature extraction. Nowadays we can observe flip or flip invariant transformations due to artificial flipping, opposite capturing viewpoint, or symmetric patterns of objects. F-SIFT is similar to SIFT, but it performs well against flip transformations.

However matching based on F-SIFT local features of each frame within two videos leads to high computational cost like that of SIFT. So it uses,

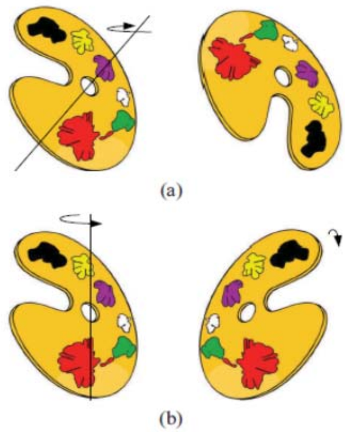
- A dual threshold method to eliminate redundant video frames.
- SVD-Flip Invariant SIFT for matching feature sets.
- Graph-based video sequence matching is used to match the query video and reference video.

A.Auto Dual Threshold Method

Auto dual threshold method is used to eliminate temporally redundant video frames. So feature Extraction and sequence matching need not to be carried out using all the video frames. This methods cuts continuous video frames into video segments by eliminating temporal redundancy of visual information of continuous video frames. This is an effective way of eliminating non necessary matching and extracts certain key frames from video segment, matching is performed using these selected frames. This method has the following characteristics. First, two thresholds are used. One threshold is used for detecting abrupt changes of visual information of frames and another issued for gradual changes of visual information of frames. Second, the values of two thresholds are determined adaptively according to video content. Three frames are extracted using auto dual threshold, the first frame, key frame and last frame. Key frame is determined using average feature value of all the frames within the segment. Here the key frame is used for matching purpose. First and last frames are used for determining the segment location. Auto dual threshold method for video segment is shown in the figure 3.

B.Flip Invariant SIFT Feature Extraction

Here, proposes to use F-SIFT instead of SIFT for feature extraction from the selected key frame. SIFT preserves original properties of SIFT while being tolerant to flips. Flip transformation can happen along arbitrary axis. However it is easy to imagine that any flip can be decomposed into as a flip along a predefined axis followed by a certain degree of rotation as shown in figure 4. The flip invariant descriptor can be made by normalizing a local region before feature extraction through rotating the region to predefined axis and then flipping it along the axis. Prominent solution for flip invariance is to determine whether flips should be performed before extracting local features from the region. F-SIFT uses curl computation, curl is mathematically defined as a vector operator that describes the infinitesimal rotation of vector field. In this case curl is defined in 2D discrete vector field. In this case the curl at a point is the Cross product on the first order partial derivatives along X and Y direction respectively.



Standardizing arbitrary flip (a) to as a horizontal flip followed by (b) rotation.

Fig 4 Standardizing arbitrary flip

The dominant curl along the tangent direction can be defined by,

$$c = \sum_{x,y\epsilon I} \sqrt{\frac{\partial I(x,y)^2}{\partial x} + \frac{\partial I(x,y)^2}{\partial y}} \times \cos\theta$$

Where

$$\frac{\partial I(x,y)}{\partial x} = I(x-1,y) - I(x+1,y)$$

$$\frac{\partial I(x,y)}{\partial y} = I(x,y-1) - I(x,y+1)$$

Where θ is the angle from direction of the gradient vector to the tangent of the circle passing through (x, y).

Generally possible direction of C is either clockwise/anti clockwise indicated by its sign. Flipping of vector field along an arbitrary axis causes the change of its sign. Normalization is performed by flipping regions whose sign are counter clockwise. That is solution of whether to flip a region prior to feature extraction is based on sign of C. The robustness can be further enhanced by assigning higher weights to vectors closer to region center as following,

$$c = \sum_{x,y \epsilon I} \sqrt{\frac{\partial I(x,y)^2}{\partial x} + \frac{\partial I(x,y)^2}{\partial y}} \times \cos\theta \times G(x,y,\sigma)$$

Where the flow is weighted by a Gaussian kernel G of size σ equal to the radius of local region1. That is, F-SIFT generate descriptors as following. Given a region rotated to its dominant orientation, Equation of C is computed to estimate the flow direction of either clockwise or anti-clockwise. F-SIFT ensure flip invariance property by enforcing that the flows of all regions should follow a predefined direction indicated by the sign of C in Eqn.

For regions whose flows are opposite of the predefined direction, flipping the regions along the horizontal (or vertical) axis as well as complementing their dominant orientations are explicitly performed to geometrically normalize the regions. SIFT descriptors are then extracted from the normalized regions. In other words, F-SIFT operate directly on SIFT and preserves its original property. Selective flipping based on dominant curl analysis is performed prior to extracting flip invariant descriptor. The only overhead involves is the computation of these equation but is cheap to calculate. The feature extraction and matching performance of SIFT (left) and F-SIFT (right) is shown in figure 5.

Comparing the matching performance of SIFT (left) and F-SIFT (right) under flip transformations. (X/Y) shows the numbers of match pairs (X) against the number of key points (Y).For illustration purposes, not all matching lines are shown. (a) Scale. (b) Scale (c) Scale+flip. (d) Scale+flip.Figure a&b for transformations involving no flip.F-SIFT has similar performance as SIFT. When flip happens F-SIFT performs stronger than SIFT. Figure c-d shows that numbers of matching pairs recovered by F-SIFT is much more than SIFT.

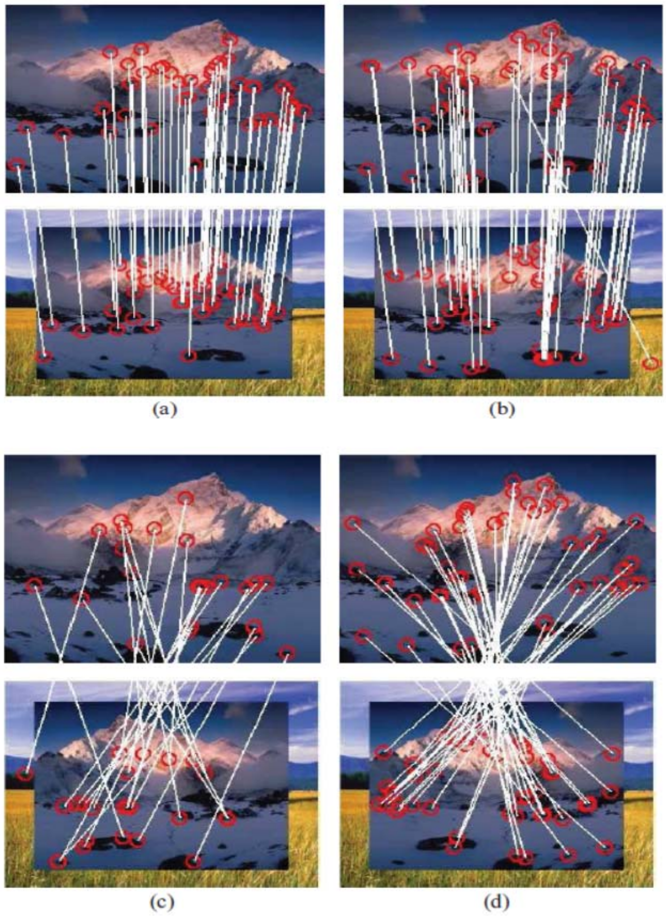


Fig 5 Matching performance of SIFT and F-SIFT

C.SVD-Flip invariant SIFT matching

A comparative study of the performance of various feature descriptors showed that the F-SIFT descriptor is more robust than others with respect to rotation, scale changes, view-point change, local affine transformations and flip like transformations.SVD [12] method has been widely used in pattern recognition, signal processing and other fields. We use SVD method to measure the similarity between two F-SIFT feature point sets. The objective of SVD algorithm is to match two point sets and compute similarity between two images. F-SIFT feature local points sets of one image can be represented as matrix.

Let A and B are two images having F-SIFT feature point sets m and n respectively. Steps involved in calculating the matching is as follows;

Step1: $A^{N\times M} = A_1, A_{2.....}A_n$

$B^{N\times M} = B_{1,}B_{2.......}B_n$

Where A, B are the feature point set matrix of image A & B.

A_i = (1,2,3…m)

$B_i = (1,2,3 \ldots n)$

Represent F-SIFT feature points in image A and B, dimension of A & B is N (N=128).

Step2: A d-dimensional linear subspace of A and B is represented by an orthonormal basis matrix,
$P_A \epsilon A^{N\times d}$ &$P_B \epsilon B^{N\times d}$ such that
$AA^T \cong P_A K_A P_A^T$ & $BB^T \cong P_B K_B P_B^T$
Where K_A and K_B the Eigen value dialogue matrices of the d largest Eigen values.
P_A And P_B the eigenvector matrices of the d largest Eigen values.

Step3: Make the singular value decomposition for
$P_A^T P_B \in R^{d\times d}$ That is $P_A^T P_B = USV^T$.

d. Graph Method for Video Matching With F-Sift

Graph-based video sequence matching method that reasonably utilize the video's temporal characteristics. The method presented as follows.

Step1: Segment the video frames and extract features of the key frames, perform auto dual threshold method to segment the video sequence and then extract the F-SIFT features of the key frames.

Step2: Match the query video and target video, assume that $Q_c = C_1^Q, C_2^Q, C_3^Q \ldots. C_4^Q$ & $T_c = C_1^T, C_2^T, C_3^T \ldots. C_4^T$ be the segment sets of query video and target video. For each C_i^Q in the query video, compute similarity Sim (C_i^Q,C_j^T) and return k largest matching result sets.

Step3: 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.

Step4: 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. A dynamic programming 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.

Step5: Output the result of detection. For each vertex of the matching result graph, it has more than one path or no path. 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, 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 vertices of the path, and $M_{ij}$is the vertex in the path.

V. CONCLUSION

This paper proposes a framework for content based copy detection and video similarity search. The proposed CBCD method uses local features of F-SIFT to describe video frames. F-SIFT has similar performance as that of SIFT, it performs stronger against flip like transformations compared to SIFT. Since the number of local features extracted using F-SIFT is very large which causes high computational cost. So that here used an auto dual threshold method to eliminate redundant video frames and an SVD matching method to compute the similarity between query video and target video. 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. Graph based matching has several advantages like it can find the best matching sequence in many messy matches and it has high copy location accuracy.

VI. FUTURE WORKS

Future work includes enhancing the security to the fingerprint. Also enhancing the feature extraction speed by implementing with other advanced feature Extraction methods. It also includes study of methods in SVD to compare the speed in matching computation.

ACKNOWLEDGEMENT

I express my deepest thanks to “Mrs.Soumya Varma” the mentor of this work for guiding and correcting various documents of mine with attention and care. She has taken the pain to go through this work and make necessary corrections as and when needed, I deeply express my sincere thanks to Mr. Vince Paul, HOD of computer Science and Engineering Department for his whole hearted support and our coordinator Mr.Sankaranarayanan P N who helped me throughout this work. I also extended my heartfelt thanks to my family and well-wishers.

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,_ IEEE Trans. Multimedia, vol. 11,no. 2, pp. 196-207,Feb. 2009.
[2] M. Douze, H. Je´gou, 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, June 2010.
[3] D.-Q. Zhang and S.-F. Chang, _Detecting image near-duplicate by stochastic attributed relational graph matching with learning,_ in ACMInt. Conf. Multimedia, New York, 2004
[4] 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
[5] Z. Huang, H.T. Shen, J. Shao, B. Cui, and X. Zhou, _Practical Online Near-Duplicate Subsequence Detection for Continuous Video Streams,_ IEEE Trans. Multimedia, vol. 12,no. 5, pp. 386-397, Aug. 2010.
[6] Vishwa Gupta, Parisa Darvish Zadeh Varcheie, Langis Gagnon, Gilles Boulianne content-based video copy detection using nearest-neighbor mapping_ Centrede recherche informatique de Montr´eal (CRIM).
[7] Mei-Chen Yeh, Kwang-Ting Cheng _Video Copy Detection by Fast Sequence Matching, July 10.2009
[8] 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.
[9] Shi Chen , Jinqiao Wang , Yi Ouyang , Bo Wang Qi Tian , H. Lu _ Multi-level trajectory modeling for video copy detection_ Inst. of Autom., Chinese Acad. of Sci., Beijing, China DOI:

[10] X. Zhou, X. Zhou, L. Chen, A. Bouguettaya, N. Xiao, and J.A. Taylor,_An E_cient Near-Duplicate Video Shot Detection Method Using Shot-Based Interest Points, IEEETrans. Multimedia, vol. 11, no. 5, pp. 879-891, Aug. 2009

[11] S. Lazebnik, C. Schmid, and J. Ponce, _A sparse texture representation using local a_ne regions,_ IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, no. 8, pp. 1265_1278, Aug.2005.

[12] V. Klema and A. Laub, _The Singular Value Decomposition: Its Computation and Some Applications,_ IEEE Trans. Automatic Control, vol. AC-25, no. 2, pp. 164-176, Apr. 1980

[13]] E. Delponte, F. Isgro' , F. Odone, and A. Verri, _SVD-Matching Using Sift Features, Graphical Models, vol. 68, no. 5, pp. 415-431,2006

[14] F. Dufaux, _Key Frame Selection to Represent a Video,_ Proc. IEEE Intl Conf. Image Processing, vol. 2, pp. 275-278, 2000.

[15] K. Mikolajczyk and C. Schmid, _A performance evaluation of local descriptors,_ IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, no. 10,pp. 1615_1630, Oct. 2005

[16] Flip-Invariant SIFT for Copy and Object Detection_ Wan-Lei Zhao and Chong-Wah Ngo, Member, IEEE, IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 22,NO. 3, MARCH 2013

[17] SURF: Speeded Up Robust Features Herbert Bay1, Tinne Tuytelaars2, and Luc Van Gool12 1 ETH Zurich bay, vangool@vision.ee.ethz.ch 2 Katholieke Universities

[18] K. Mikolajczyk and C. Schmid, _Scale and a_ne invariant interest point detectors,_ Int. J. Comput. Vis., vol. 60, no. 1, pp. 63_86, 2004.

[19] R. Ma, J. Chen, and Z. Su, _MI-SIFT: Mirror and inversion invariant generalization for SIFT descriptor,_ in Proc. Int. Conf. Image Video Retr., 2010, pp. 228_236.

[20] PCA-SIFT: A More Distinctive Representation for Local Image Descriptors Yan Ke1,Rahul Sukthankar2,1,ykeS