

Near Duplicate Video Retrieval using Spatio Temporal Approach with Multifeature Mechanism

Varsharani Bhosale¹, Mrs.D.A.Phalke²P.G. Student, Department of Computer Engineering, D.Y.Patil College of Engineering, Akurdi, Pune, India¹Associate Professor, Department of Computer Engineering, D.Y.Patil College of Engineering, Akurdi, Pune, India²

ABSTRACT: A large number of videos are uploaded and shared on social websites in each single day. There are a large number of near-duplicate videos (NDVs) on the web are generated in different ways. The near-duplicate video retrieval and localization achieved by using a spatio temporal pattern-based methodology under the hierarchical filter-and-refine structure. Dynamic programming is used to find the near duplicate video ranks and location of near duplicate segment in reference videos, which based on m-patterns of video called as mPDP. But the problem with the spatio temporal approach is that video copies applied with picture-in-picture transformation, super-slow motion, or high speed fast forwarding cannot be accurately retrieved and localized. Thus a multi-feature mechanism instead of using single feature is involved in spatiotemporal filter and refine approach to improve the effectiveness of system.

KEYWORDS: Near-duplicate video localization, near-duplicate video retrieval, video copy detection, m-pattern dynamic programming, pattern based index tree.

I. INTRODUCTION

With the popularity of social media, there has been exponential growth in videos available on the net. Users can obtain web videos easily, and distribute them again with some modifications. Users upload thousands of new videos each day on video sharing website like YouTube and the daily video views are over the millions. Among these huge volumes of videos, there exist large numbers of duplicate and near-duplicate videos. It becomes important to manage these videos in an automatic and efficient way[4].

NDVR system, including three major components, signature generation, signature management, and signature search. First, a video is divided into a sequence of key frames extracted by time sampling or shot boundary detection algorithms like histogram based, motion vector based algorithms. These key frames are represented by their visual features, such as color histogram, local points, and local binary pattern. The sequence of the key frames features is then considered as the signature of the video. To find out NDVR, the system needs to compare the similarity between the signatures of the query video and each dataset video and return the dataset videos which are most similar to the query example. Normally for a given database video, in the signature generation component, low-level features are first extracted. Based on this video localization can be further performed to generate compact and distinctive video signatures to access or perform the retrieval. Now a day's video databases grow rapidly and the sizes get larger and larger[9].

The spatiotemporal filter PI-tree based on I-patterns fast filters out most of non-near-duplicate videos to reduce the search space improves the execution speed. Since -patterns, instead of frames, are regarded as basic units in dynamic programming, the mPDP algorithm achieve efficient near-duplicate video localization and re-ranking[1]. The proposed system of multifeature mechanism filters out most of non-near-duplicate videos which remarkably improves the retrieval precision.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

II. RELATED WORK

NDVR have been successfully considered in several kinds of authentic application. On the basis of different kinds of surveys and relative studies various methods that are accountable for the development of large numbers of NDVs are found in [6], [7]. Different methods utilizing different features and related methods have been recommended for NDVR lately. The present methods on NDVR can be generally divided into two categories, first is international features based methods and another is local that is local features based methods. A local function based methods have furthermore sucked in much research concern. Features, for example colour, surface and shape removed at the key-frame level are further portioned into several area designs, and are especially appropriate for recuperating NDVs with complex kinds. The well-known local function based methods incorporate key-frames based local function identification approach [8]. As an option to international function based methods most of the NDVR methods highlight the fast determining proof of NDVs [1].

Frame-level techniques need considerably high calculations cost. As spatiotemporal techniques are more beneficial and efficient than frame-level techniques, they are still not appropriate for web-scale programs. Video-level techniques are quicker but cannot accomplish NDVL [9]

Using frame-level local signatures, changes in the aspects of the videos like the angle of viewport when recording the video can be properly recognized where low matching speed is expected. It suits to applications that are not focus to speed but aims to find videos that are captured from different viewports about the same scene. A frame-level local signature represents the local information on individual frames with local features like local keypoint descriptors. Such usage of local information in NDVR is very closely related to the near-duplicate image retrieval task. Local keypoints contain so much information that makes it not easy practical to directly perform a search[6].

Video-level signatures are generally working well in applications where the NDVs include changes of the video colors like color, contrast, encoding, etc. as well as the temporal orders of frames, and response speed is a key requirement. The abstraction of video colors makes it tolerant to minor color changes, while the dropping of frame orders leads to robustness to any NDV that is created with temporal changes. A video-level signature represents a whole video with a single signature. Video-level signature is considered as an efficient form of representing videos and widely used in many works[4].

By introducing temporal information into the video signatures, spatio-temporal signatures are often more robust to heavy changes in the colors or in the video geometry e.g. video color is shifted and video is cropped. However, it is more sensitive to temporal changes in the video, because such changes disrupt the very essential information based on which it is generated. Spatio-temporal signatures represent videos with spatial and temporal information. In recent years, spatio-temporal signature has drawn attention in NDVR for its better invariance to viewpoint changes compared to global signature-based techniques and for its relatively better efficiency compared to local signature techniques[8].

Existing work on near-duplicate video identification can be grouped into two categories: global feature based methods and local feature based methods. Global feature based methods extract frame level signatures to model the information distributed in spatial, color and temporal dimensions. The similarity between query video and database video is calculated as the matching score of the signature sequences. Due to temporal editing, such as inserting in or cutting out short clips, dynamic programming method is often employed for locating a short query video clip in a long target video. The state-of-the-art, the existing global feature based methods using sequence matching algorithm is still very time-consuming. Local feature based methods are very suitable for copyright, since they can handle various challenging distortions (strong cropping, insertions of large advertisements, picture-in-picture, etc.). However, the time and storage complexity is unacceptably high for a practical realtime applications[11][12].

III. SYSTEM OVERVIEW

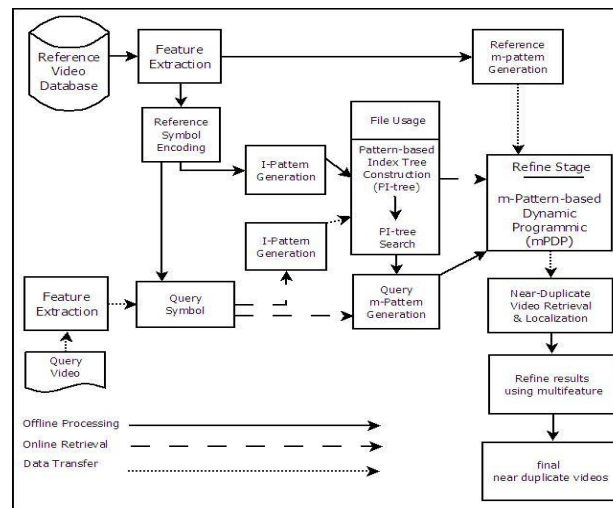


Fig.1: System Architecture

1. A spatiotemporal pattern-based approach, which is capable of all fully NDVR, partial NDVR, and NDVL, under the hierarchical filter-and-refine framework. Given a reference database and a query video, aim at not only retrieving both fully and partial near-duplicate videos but also localizing precise positions of near-duplicate segments. Firstly, low-level features of video frames are symbolized, and sequences of symbols are collected to form index patterns (I-pattern) and keyframe patterns (m-pattern) by sliding windows of different sizes. In the filter stage, a spatiotemporal indexing structure utilizing index patterns, termed Pattern-based Index Tree (PI-tree), is used to fast filter out non-near-duplicate videos and to retain more partial near-duplicate videos than video-level methods. The PI-tree is substantially more efficient than frame-level methods. In the refine stage, extended m-Pattern-based Dynamic Programming (mPDP) to localize near-duplicate segments and to re-rank the results of the filter stage. The influence of time shift misalignment can be alleviated by the time-shift m-pattern similarity measurement. From the derived dynamic programming and backtracking tables, the diagonal property of a near-duplicate segment block is used to discover the temporal relations between the query video and a reference video[1].
2. Multifeature Extraction: Simultaneously utilizing the multiple features is helpful for disambiguation. For example, local signature is less robust to the changes in frame rate, video length, captions and global signature is sensitive to changes in contrast, brightness, scale, rotation, camera viewpoint, and so on. It is difficult, if not impossible, to find a single visual feature which is robust to all types of variations. It is therefore particularly important to exploit multiple features for the sake of NDVR, where various variations and distortions could take place between NDVs. To this end, multiple features (i.e., global and local visual features) of videos are used in spatiotemporal filter and refineframework which are more accurate to represent video content.

Histogram of Oriented Optical Flow: Histogram of oriented optical flow features used to signify individual actions. These novel features are outside of the dimensions of the moving person as well as the route of movement. Extraction of HOOFF features does not need any prior individual segmentation or qualifications subtraction. First, optical flow is calculated at every frame of the screen. Each flow vector is binned according to its main position from the horizontally axis and calculated according to its scale. Lastly, the histogram is stabilized to sum up to 1[7].

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

Color change pattern histogram : Color histogram is used to find out change in the motion. Color histogram value is calculated for each frame and then calculate difference between values of color features of query video and videos in dataset. To calculate color features system divides image into blocks and take average of color intensity from each. Each block will return a value as a feature, so number of blocks is equal to number of color features. This process is applied on each frame. By extracting these feature proposed system will generate more accurate results than existing system.

Texture feature: Texture features is extracted from each extracted frame and key frame. It reflects the texture pattern for selected frame. It is having feature values, each value is having its own meaning, like contrast, correlation, entropy etc. The feature gabor wavelet transform the gabor wavelet of grey level values is computed and gabor wavelet is applied with remaining scale and orientation and then the values are updated.

IV. EXPERIMENTAL RESULTS

The system is built using Java framework (version jdk 8) on Windows platform. The Netbeans (version 8.1) and openCv (version 2.4.10) is used as a development tool. The system doesn't require any specific hardware to run, any standard machine is capable of running the application.

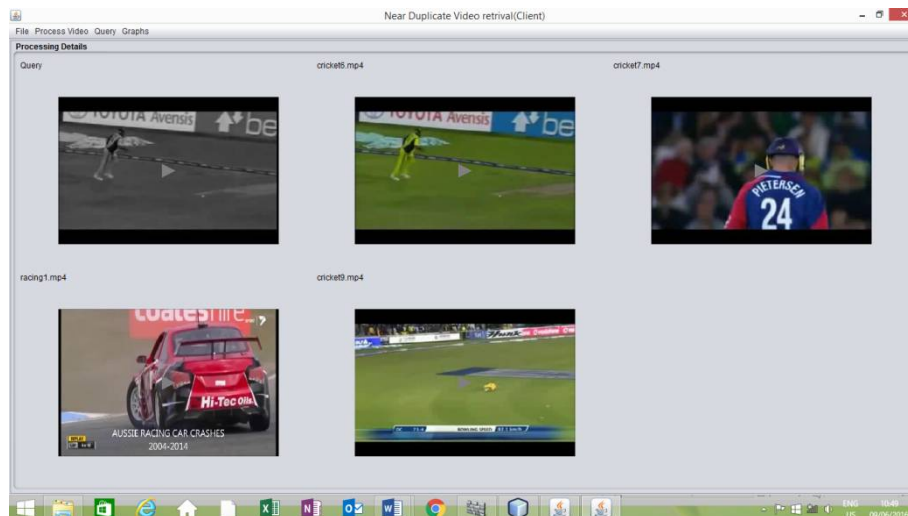


Fig2. Retrieved near duplicate for given query video

The results for the given query video with multifeature mechanism are shown in above figure. For the given query video the frames and the multi features for the respective frames are extracted using which the pattern set based index tree as well as the queue pool is generated and using the near duplicate score the near duplicate videos are retrieved and localised.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

Table 1: Accuracy Comparison using Precision

Cases of QueryVideo 's	Existing System	Proposed System
Video1	0.81	0.90
Video2	0.79	0.85
Video3	0.72	0.78
Video4	0.80	0.87

Above table shows comparison between precision of existing system using k-means clustering with hoof feature only and proposed system using with multifeature mechanism. Using the above values the precision graph is plotted that makes the comparison between the existing system and the proposed system. The precision of proposed system is greater than existing system.

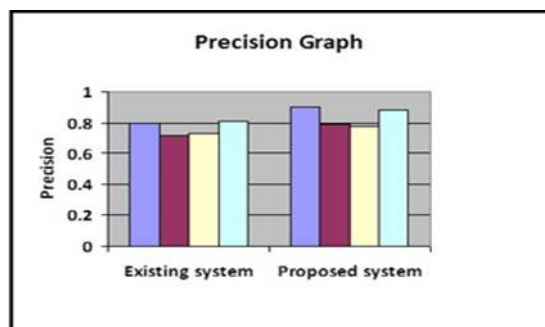


Fig. 3. Recall Graph for query video's. shows that precision of our proposed system is higher than existing system.

Generated dataset includes reference videos is created by our own, using videos downloaded from various video sharing websites like Youtube Yahoo! Video's, and Google video's. etc. The four different query videos for various type of transformations found in near duplicate video retrieval are created using which following recall values are calculated.

Table 2: Accuracy Comparison using Recall

Cases of QueryVideo 's	Existing System	Proposed System
Video1	0.81	0.90
Video2	0.79	0.85
Video3	0.72	0.78
Video4	0.80	0.87

Above table shows comparison between recall of existing system of spatiotemporal approach with hoof feature only and proposed system with multifeature mechanism. Using the above values of given in the recall graph is plotted that makes the comparison between the existing system and the proposed system. The recall of proposed system is greater than existing system.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

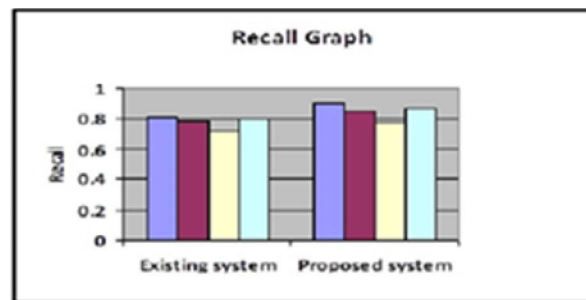


Fig. 4. Recall Graph for query video's shows that precision of out proposed system is higher than existing system.

V. CONCLUSION

The results of multi feature evaluations shows that adopting the local feature such as histogram of change pattern color, gabor feature with the global feature histogram of oriented optical flow instead of only one features can improve the effectiveness of the proposed approach. The significance and efficiency of the proposed multifeature mechanism is evaluated and verified by the experiments. The experimental results demonstrate that the proposed system which is the combination of Spatio temporal filter and refine approach and multifeature mechanism, not only achieves comparable precision on fully NDVR but also outperforms other state-of-the-art methods on partial NDVR and NDVL in terms of precision and recall in most cases.

The efficiency and scalability of the proposed framework could be further improved by adapting this work onto cloud.

REFERENCES

- [1] Chou, Chien-Li, Hua-Tsung Chen, and Suh-Yin Lee. "Pattern-Based Near Duplicate Video Retrieval and Localization on Web-Scale Videos." *Multimedia, IEEE Transactions on* 17.3, 382-395, 2015.
- [2] 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.
- [3] J. Song, Y. Yang, Z. Huang, H. T. Shen, and R. Hong, Multiple feature hashing for real-time large scale near-duplicate video retrieval, in *Proc. 19th ACM Int. Conf. Multimedia*, pp. 423-432, Nov. 2011.
- [4] X. Wu, A. G. Hauptmann, and C. W. Ngo, "Practical elimination of near-duplicates from web video search," in *Proc. 15th ACM Int. Conf. Multimedia*, pp. 218-227, Sep. 2007.
- [5] C. L. Chou, H. T. Chen, Y. C. Chen, C. P. Ho, and S. Y. Lee, "Near-duplicate video retrieval and localization using pattern set based dynamic programming," in *Proc. 2013 IEEE Int. Conf. Multimedia Expo*, pp. 1-6, Jul. 2013.
- [6] M. Douze, H. Jégou, C. Schmid, and P. Pérez, "Compact video description for copy detection with precise temporal alignment," in *Proc. ECCV*, pp. 522-535, 2010.
- [7] [13] R. Chaudhry, A. Ravichandran, G. Hager, and R. Vidal, Histograms of oriented optical flow and Binet-Cauchy kernels on nonlinear dynamical nonlinear dynamical systems for the recognition of human actions," in *Proc. 2009 IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 1932-1939, Jun. 2009.
- [8] P. P. W. Chan, M. R. Lyu, and R. T. Chin, "Copyright protection on the web: A hybrid digital video watermarking scheme," in *Proc. 13th Int. World Wide Web Conf. Alternate Track Papers Posters*, pp. 354-355, May 2004.
- [9] J. Liu, Z. Huang, H. Cai, H. T. Shen, C. W. Ngo, and W. Wang, Near duplicate video retrieval: Current research and future trends, *ACM Comput. Surveys*, vol. 45, no. 4, pp. 123, Aug. 2013
- [10] L. Zheng, G. Qiu, J. Huang, and H. Fu, Salient covariance for near duplicate image and video detection, in *Proc. 2011 IEEE Int. Conf. Image Process*, pp. 2537-2540, Sep. 2011
- [11] Cherubini, R. de Oliveira, and N. Oliver, "Understanding Near-Duplicate Videos: A User-Centric Approach," in *Proc. ACM Multimedia*, pp. 35-44, 2009.
- [12] 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*, pp. 371-378, 2007.