Duplicate video and object detection by video key frame using F-SIFT

Prof. Bere Sachin.S.
Department of Computer Engineering,
Dattakala Group of Institution Faculty ofEngineering,
Savitribai Phule Pune University, Pune, India
sachinbere@gmail.com

***Abstract*— To explain and detect different features in images scale-invariant feature transform can be used effectively. From starting, a set of reference images SIFT important points of objects are extracted and stored in a database. An object in a new image can be recognized by individually balancing each feature from the new image to this database and then finding features for candidate matching. As a valuable local SIFT can be utilize as a solution point descriptor for its invariance to lighting, scale, and rotation changes in images. Since SIFT is not flip invariant, flip invariant SIFT is planned. These F-SIFT is established to identify large scale duplicate videos, object finding as well as recognition. It requires to take out all the frames from query video and videos in dataset for similarity matching, time complexity of f-SIFT is more, So to remove such limitation we have projected dual threshold technique. Our method will eliminate redundant video frames by applying auto dual threshold method. So there will be no necessity to execute the extraction of features and matching of sequence with all video frames. Unnecessary frames are detached by making segments of video. Only the key frames are extracted for matching purposes. Here we are using two thresholds. The first is for identifying direct changes of visual information of extracting frames and other second for detecting usual changes of visual information of extracting frames. Threshold values are decided as per the information of the video. This system, extracting total three frames like first frame, last frame and key frame from video segment. By using the average feature value of all the frames in the segment, key frames are decided. For similar propose a key frame is used and remaining two frames are used to detect the segment location.**

Keywords—key frames, SIFT, video segmentation, scale invariant feature transform

I. INTRODUCTION

With the rapid growth in the multimedia technology and web, the method is able to access and stock up huge numbers of video information rapidly. These huge numbers of video clips are transmitted, investigated and stored on web. Some figures of video search website YouTube shows that, there are about plenty of users produced video clips are submitted to YouTube every minute. As per the statement of BBC motion gallery, it comprises above 2.5 million hours of video contents. From this huge figure of video clips there live a large numbers of duplicated and near copied video clips. As per information, near about 27% video clips in videos search outcome obtained from yahoo, Youtube and Google video clips are copied near copied duplicates of a well-known version. For the meticulous queries, the redundancy can be high as 94%. The duplicate video can be Separated into two types nearly copied Videos and duplicated Videos. Duplicate video will be taking out video clips duplicates that are easily noticed. Near duplicated video can be transformed videos clips and acknowledgement of this type is very challenging. So we can describe video clips copy as, it is a segment of video derived from an additional video clip usually through a variety of transformations such as adding, removal, alteration and cam coding. There is necessitating recognizing such duplicate videos for patent propose. Scale invariant feature convert will be used to take out the different features of videos clips. The beauty of SIFT is mainly because of its invariance to a variety of picture transformations like: scaling, displacements, rotation and lighting changes of pixels. SIFT is normally calculated over a local silent region which is situated by rotated and multi- scale detection to its leading direction. The descriptor will be invariant to both rotation as well as scale. In addition, due to rotation and spatial partitioning it is insensitive to lighting, small pixel dislocation and color. But the fact is that SIFT is not flip invariant.

Flipping video is one of the frequently used activities to generate replica videos. There are two types of flip processes vertical flipping and horizontal flipping. Vertical flipping is used regularly since it will not affect alteration into the content of video. Also the video of the similar object taken from reverse direction can flip videos. To avoid this restriction F-SIFT is established. It develops the SIFT with flip invariant attribute. It also can be used for finding and recognition of related objects from duplicate videos.F-SIFT needs an extract all the frames of the query video and videos in dataset. So the time complexity for copy finding is much more. So we have established method which can assist to decrease this time complexity.

Our method will get rid of redundant video frames by applying auto dual threshold method. So there will be no requirement to perform extraction of features and matching of chain with all video frames. Redundant frames are removed by making segments of video. Only the key frames are extracted for matching proposes. Here system using two thresholds. One is for identifying instant transformation of visual information of extracted frames and other for identifying regular changes of visual information of extracted frames. Threshold values are decided as per the content of videos. System removing three frames like first frame, last frame and key frame from video section. By depending on average feature value of all the frames in the segments, key frames are to be determined. For matching propose key frame is used and remaining frames are used to detect segment location.

978-1-5386-5257-2/18/$31.00 ©2018 IEEE

II. RELATED WORK

The descriptor is build by programming pair-wise relationship within a frame. Even if image property change due to transformation, this descriptor employs the inside assembly of a video frame, which creates it well-built to hits based on signals such as blurring, color changes as well as contrast enrichment to sure attacks such as scaling of frames.

Law-To et al. presented a comparative research for video copy determined and detection that for temporal, small changes, ordinal measurement will be efficient, while methods

based on the local features of demonstrate more appealing consequences in circumstances of robustness[2]. However, Thomee et al. conducted a assessment of large-scale image second copy recognition systems and accomplished a what different abstract. Their chosen technique that used interest factors conducted badly due to its lack of capability to find related sets of factors between duplicates [3]. They determined that either a simple average technique or the retina practice works the best. To propose a practical duplicate recognition program which assures the scalability specifications, a lightweight, frame-level descriptor that maintains the most suitable information, instead of just places of interest point descriptors, is suitable [4].

In addition, frame level descriptors are simply included into speedy recognition frameworks such as the one provided in [5]. Actually HOG [6] descriptors and SIFT Descriptor are both well-made histograms will be used in object discovery and sorting tasks. In those tasks, to accomplished robustness against objects rotation and flipping is of great importance. Lot of extensions of SIFT [7] descriptors can be planned to tackle such type of transforms. RIFT [8] achieved rotation and flip invariance by separating a section on the log polar direction instead using 4 × 4 grids, which, however is less distinctive than original SIFT. In contrast, MIFT [10] and FIND [9] conserve the distinctiveness while they are obtained by sorting new SIFT descriptors according to their relative magnitude. That is invariant under rotation and flip. Similarly, FSIFT [11] conclude the direction of SIFT depends on the leading curl related with the local section, and executes selective flipping on the region before descriptor computation. Contrary to those techniques, which preserve the original SIFT properties that is MI-SIFT [12] applies direct on flip- invariant converts to SIFT to generate flip-invariant descriptors.

MI-SIFT [14], functions straight on SIFT while be transforming to a latest descriptor which is flip invariant. This is realized by noticeably determining the categories of function of which are cluttered placed due to flip function. The MI-SIFT brands 32 of such symbolize and categories each group with four instants which are flip invariant. However, the descriptor depending on instance is not discriminative. As reported in [14], this results in more than 10% of associated performance degradation than SIFT when no-flip transformation occurs.

In [15], the authors tracks HSV to characterize their key frames and further produce videos clip signature by cumulating all the key frames in it. This reflection achieves fast improvement speed as well as high exactness in their dataset. However, a restriction is that global function based method usually happen to less resourceful in managing video duplicates with levels of modifying cosmetics [16]. At the same time, global function centered method relies deeply on the chosen function types.

The method planned will eliminate redundant video frames by apply auto dual threshold technique. So there will be no requiring executing extraction of features and matching of succession with all video frames. Redundant frames are eliminated by making segments of video. Only the important frames are take out for matching proposes. At this point, the system using two thresholds. One is for recognizing instant changes of visual information of extracted frames and other for detecting usual variations of visual information of extracted frames. Threshold values are determined as per the content of video. The system extracting three frames like first frame, last frame and key frame from video segment. By using average feature value of all the frames in the segment, key frames are decided. For matching propose key frame is used and remaining tow frames are used to identify segment position.

III. IMPLIMENTATION DETAILS

a. *System Architecture*

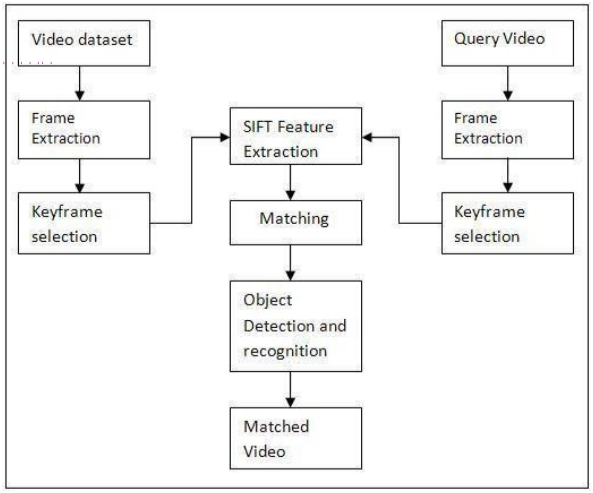

Fig 1: System Architecture

Above figure (Fig: 1) indicates method construction of our projected system.

This method works in two stages training phase and testing phase. In training phase video dataset is taken as input. Dissimilar frames from all those videos are extracted sequentially. Similar frames for the set of frames are neglected and only the key frames are taken out.

Same process is done for input query video in testing phase. From those extracted key frames SIFT features are extracted for matching purpose. Similar objects are detected and recognized from matching frames. And finally similar matched videos are detected.

b. *Algorithm*

Step1: Query video is converted to number of frames.

Step2: Key frames are extracted from frames of query videos by using auto dual threshold method.

Step3: SIFT Features are extracted from each key frame.

Step4: Matching of query video key frames is done with the original video.

Step5: Detection and recognition of copied object is done.

Step6: Matched video is extracted from the database.

c. Mathematical model for proposed work

The lower and upper thresholds, T_L and T_U, are premeditated according to both *Rn k* and *Oa*(*n*) as following:

$$R_k^n = \frac{N_n}{N_n}, n = k+1, \dots, k+N$$

Let $f_1 \dots f_n \dots f_{Num}$ indicate the frames of the video and f_k be a key-frame. f_n indicate a frame among f_k and f_k+N,

$$T_L = \begin{cases} T_{ref} + \dfrac{R_k^n - T_{ref}}{4} & \overline{O_a} < 0.4 \\ T_{ref} + \dfrac{R_k^n - T_{ref} + 2\overline{O_a}}{4} + 0.1 & \overline{O_a} \geq 0.4 \end{cases} \quad \dots (1)$$

$$T_U = \begin{cases} T_{ref} + \dfrac{R_k^n max - T_{ref}}{2} & \overline{O_a} < 0.4 \\ T_{ref} + \dfrac{R_k^n max - T_{ref} + 2\overline{O_a}}{2} + 0.1 & \overline{O_a} \geq 0.4 \end{cases} \quad \dots (2)$$

$$\overline{O_a} = \frac{Num}{N+1} \sum_{n=k}^{k+N} \frac{dO_a(n)}{dn} / \sum_{n=1}^{Num} \frac{dO_a(n)}{dn} \quad \dots (3)$$

Where $dO_a(n)$ d_n is the derivative of the accumulative occlusion area, T_{ref} is an empirical parameter. $O_a(n)$. Rn k_{max} is the maximum R_n k for the key-frame f_k.

d. Experimental Setup

The implementation of this system is done by using Java framework (version jdk 7) on Windows platform.

For the development tool system has Net beans (version 7). There is no special hardware requirement for this system.

IV.RESULTS AND DISCUSSION

Fig 2: Time required for training videos

Above figure (Fig: 2) compares time required for training in existing system and our proposed system. The system can see that time required for training in our proposed is less than that of existing.

TABLE 1: TIME REQUIRED FOR DUPLICATE VIDEO AND OBJECT RECOGNITION.

Video	Existing System	Proposed system
1	3.2	3
2	5.3	4.9
3	7.3	6.5
4	9.2	8.1
5	10.7	9.3

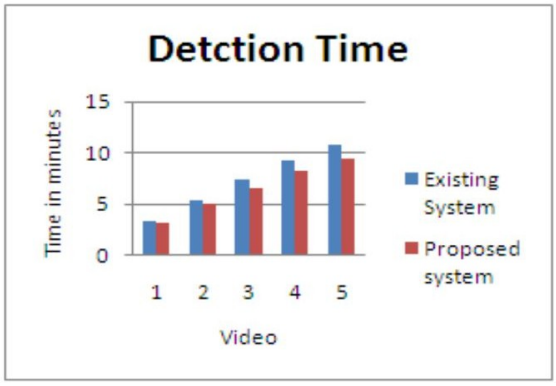

Fig 3: Time required for copy detection

Above figure (Fig: 3) denotes time required for duplicate video detection.

V. CONCLUSION

To find out duplicate videos from large video dataset numbers of systems are developed but those systems are infected by some limitations. As SIFT is not invariant to flip system introduce F-SIFT which extracts key frames from query and videos in dataset to find duplicate videos and it finds and recognized similar objects in it for detection of duplicate video. Here, system using two thresholds. One is for identifying immediate changes of visual information of extracted frames and other for detecting regular changes of

visual information of extracted frames. Threshold values are decided according to the content of video. Time complexity for duplicate video and object detection is less than other existing systems.

REFERENCES

[1] M.-C. Yeh and K.-T. Cheng, "A compact, effective descriptor for video copy detection," in Proc. Int. Conf. Multimedia, 2009, pp. 633–636.

[2] 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 Proceedings of the ACM International Conference on Image and Video Retrieval, pages 371-378, 2007.

[3] B. Thomee, M. J. Huiskes, E. Bakker, and M. S. Lew. Large scale image copy detection evaluation. In Proceedings of the ACM International Conference on Multimedia InformationRetrieval, pages 59-66, 2008.

[4] S. Poullot, M. Crucianu, and O. Buisson. Scalable mining of large video databases using copy detection. In Proceedings of the ACM International Conference on Multimedia, pages 61-70, 2008.

[5] Navneet Dalal and Bill Triggs, "Histograms of oriented gradients for human detection," in Proceed- ings of IEEE Computer Vision and Pattern Recognition (CVPR), 2005.

[6] David G Lowe, "Distinctive image features from scale invariant key points," International Journal of ComputerVision (IJCV), vol. 60, no. 2, pp. 91–110, 2004.

[7] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce,"A sparse texture representation using local affineregions,"IEEE Transactions on Pattern Analysis and Machine" 2005.

[8] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce,"A sparse texture representation using local affine regions,"IEEE Transactions on Pattern Analysis and Machine" 2005.

[9] Xiaojie Guo and Xiaochun Cao, "FIND: A neat flip invariant descriptor," in Proceedings of IEEE International Conference on Pattern Recognition (ICPR), 2010.

[10] Xiaojie Guo and Xiaochun Cao, "MIFT: A framework for feature descriptors to be mirror reflection invariant," Image and Vision Computing, vol. 30, no. 8, pp. 546– 556, 2012.

[11] WL Zhao, CWNgo, et al., "Flip-invariant SIFT for copy and object detection," IEEE Transactions on Image Processing, vol. 22, no. 3, pp. 980–991, 2013. Rui Ma, Jian Chen, and Zhong Su, "MI-SIFT: mirror and inversion invariant generalization for sift descriptor," in Proceedings of ACM International Conference on Image and Video Retrieval (CIVR), 2010.

[12] Sreeraj M Asha S, "State-of-the-art: Transformation invariant descriptors," International Journal of Scientific and Engineering Research (IJSER), vol. 4, pp. 1994–1998, 2013.

[13] A. F. Smeaton, P. Over, and W. Kraaij, "Evaluation campaigns and TRECVid," in Proc. Int. Conf. Multimedia Inf Retr., 2006, pp. 321–330.

[14] X. Wu, A. G. Hauptmann, and C.-W. Ngo, "Practical elimination of near-duplicates from web video search," in Proc. ACM Multimedia, 2007, pp. 218–227.

[15] X.Wu, C.-W. Ngo, A.G. 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, 2009.