Content based Video Retrieval using Enhance Feature Extraction

Dipika H Patel
Department of Computer Science and Engineering
L. J. Institute of Technology
Ahmedabad 382210, Gujarat, India

ABSTRACT
Videos are a powerful and communicative media that can capture and present information. In recent times, large video databases are created because of the advancements in many video acquiring devices and Internet. A reliable system is needed to automate the process of this large amount of data. Content-based video retrieval has attracted extensive research during the decades. There are various models used for video retrieval. Content Based Video Retrieval is one model for retrieval of videos. Different users have different results in their minds. These lead to the process of selecting, indexing and ranking the database according to the human visual perception. This paper reviews the recent research in content based video retrieval system. Also the paper focus on video structure analysis, like, frame extraction from video, key frame extraction, feature extraction using SURF, similarity measure, video indexing, and video browsing. This system retrieves similar videos based on local feature detector and descriptor called SURF (Speeded-Up Robust Feature). For image convolution SURF relies on integral images. In SURF we use Hessian matrix-based measure for the detector and a distribution-based descriptor. SURF can be computed and compared much faster with respect to repeatability, uniqueness and robustness. SURF is better than previous proposed methods as SIFT, PCA-SIFT, GLOH, etc. Finally the future scope in this system is specified.

Keywords
Frame extraction, Video retrieval, Feature extraction, Feature matching, SURF, C-SURF, Video browsing.

1. INTRODUCTION
An important research issue in multimedia databases is fast and robust content-based video retrieval (CBVR) in large video collections. Videos have the following characteristics: 1) much richer content than individual images; 2) huge amount of raw data; and 3) very little prior structure [10]. These characteristics make the indexing and retrieval of videos quite complex. In the past because of small database, indexing and retrieval have been based on text. Nowadays level of these databases has become much larger and content-based indexing and retrieval is required. The content of a video can be represented using either global features or local features. Local features are widely utilized in a large number of applications, e.g., object categorization, video retrieval, robust matching, and robot localization. [16]. A local feature consists of a feature detector and a feature descriptor. Here we use SURF for local feature detector and descriptor.

Paper is organized as follows: in section II, related works are briefed. Section III introduces the methodology used in the proposed system. Section IV explains the algorithms used in the implementation of different modules. Finally, future scope and work is concluded in section V.

2. RELATED WORK
"Content-based" means that the search will be based on the actual content of the video. The term 'Content' here refer to the features such as color, shape, texture of the video. The content based approach focuses on the retrieval of videos by their similarity matching based on its video content. This content can be represented by either: global feature or local feature [4]. Global descriptors detail the overall content of the image but with no information about the spatial distribution of this content. Local descriptors relate to particular image regions and, in combination with geometric properties of these latter, express also the spatial arrangement of the content.

Among the various local feature descriptor, SURF has proven to be stable, reliable and has high efficiency in information retrieval. S. Huang, C. Cai, F. Zhao in [17], used SURF feature descriptor for extracting the contents of images for wooden image retrieval.

Various methods are proposed for automatic detection of frames from video. The simplest approach is to extract frame per second in a video.

From the various proposals made on key frame identification, here we extract key frames based on the change of color, texture and other visual information of each frame [18][19][20][21], when this information changes significantly, the current frame is key frame. The basic idea is that the first frame is selected as the new frame, and is viewed as reference frame, then the back frames are compared with the reference frame in order, the k-th frame do not become the new key frame until the distance between the k-th and (k-1)-th frame exceed a specific threshold.

Next, from the extracted key frame features are detected at specific location using SURF detector. These features are matched using SURF descriptor. Using C-SURF we can add color information in the existing SURF detector and descriptor.

3. METHODOLOGY
The content based video retrieval system is outline in Fig 1.

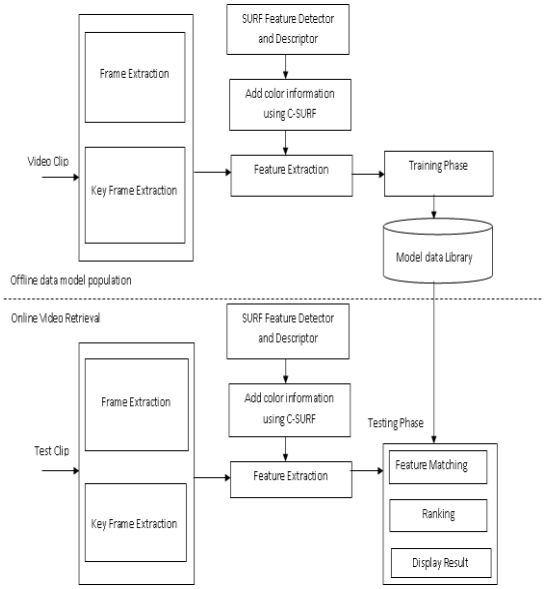


Fig 1 Generic framework for visual content-based video indexing and retrieval [4]

Here during the offline stage input video clip is undergoes a preprocessing phase, which includes frame extraction and Key Frame extraction modules. During this preprocessing stage, the input video gets converted into a set of key frames. From this identified key frames, SURF feature descriptor is extracted. Also color information is added using C-SURF. These extracted features are passed into the training phase and stored in a model data library.

On video retrieval, for a given test clip, its features are computed. During the event of feature matching, the videos in the model data library are ranked based on its similarity to the test clip. From the list of similar videos the highest ranking videos are retrieved.

4. IMPLEMENTATION

A generic CBVR system consists of four phases: frame extraction, Key Frame Extraction, Feature Extraction and Feature Matching. The algorithm used in each of the four phases is described in the following subsections.

4.1 Frame Extraction

Here we use opencv library with visual studio. After reading the whole video, frame can be extracted using fps (frame per second) function. For 32 second of video we can get around 540 frames.

4.2 Key frame extraction

There are great redundancies among the frames in the same Shot; therefore, certain frames that best reflect the shot contents are selected as key frames [12], [13], [14], [15] to succinctly represent the shot.

Key frame can be extracted using the threshold value. [1]
Steps:

I. Collect the frames from input video

II. Convert frame color from RGB to HSV

III. Compute histogram for each channel (H-H, S-S, V-V)

IV. Normalize the histogram.

V. Compare histogram difference using chi-square method

$$d(H_1, H_2) = \sum_I \frac{(H_1(I) - H_2(I))^2}{H_1(I)}$$

VI. If histogram difference is greater than threshold value, then save it as a key frame otherwise discard that image.

Flow chart for key frame extraction is as below:

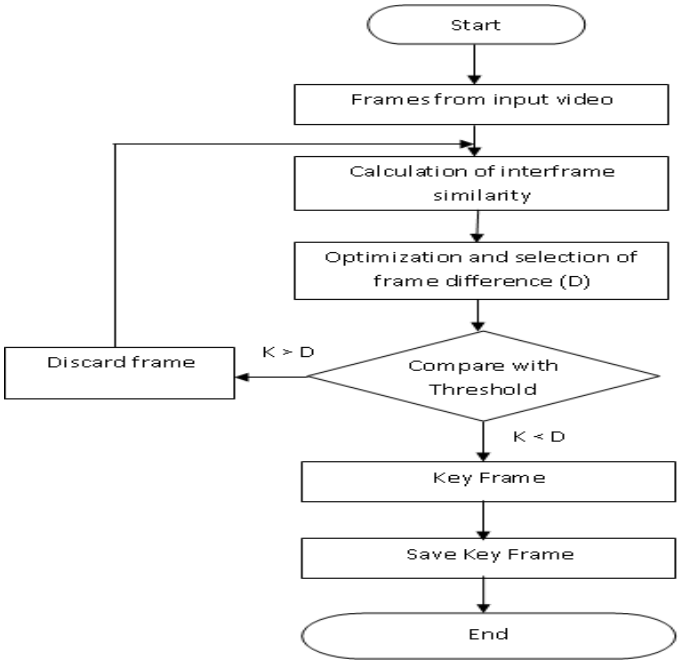


Fig 2 Flow chart of Key Frame Extraction

4.3 Feature Extraction using SURF [6] [13]

For feature extraction SURF detector and descriptor is used.

First of all 'interest point' are selected at distinctive location in a key frame, such as blobs, corners, and T-junctions.

4.3.1. Interest Point Detection

Hessian-matrix approximation is used for interest point detection. Integral images are used here.

4.3.1.1 Integral Images

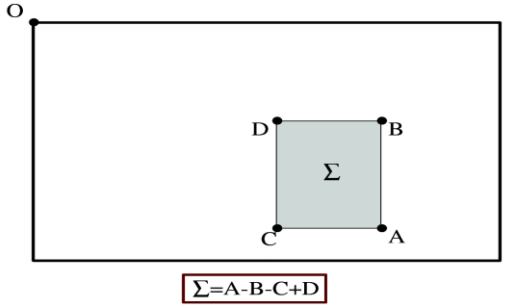


Fig 3 Integral image calculation [6]

Integral Image or summed area tables is an intermediate representation of the image. It contains the sum of intensity values of all pixels in input image *I* within rectangular region formed by origin O = (0, 0) and any point X = (x, y). It provides fast computation of box type convolution filters.

4.3.1.2 Hessian Matrix Based Interest Points:

Here blob like structure is detect where the determinant is maximum.

- Scale Space Representation:

Interest points need to be found at different scales. The scale space is analyzed by up-scaling the filter size rather than down scaling the image size.

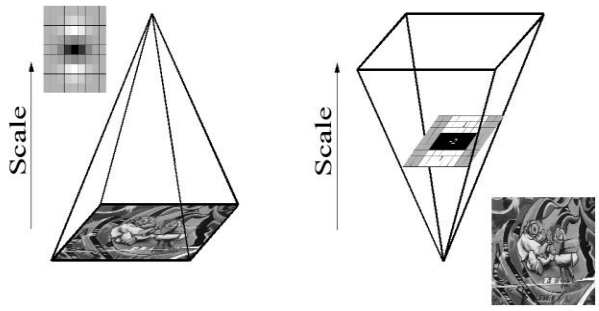


Fig 4 Instead of iteratively reducing the image size (left), the use of integral images allows the up-scaling of the filter at constant cost (right)[6]

4.3.2. Interest Point Descriptions and Matching:

They build on the distribution of first order Haar wavelet responses in x and y direction.

- First step consists of fixing a reproducible orientation based on information from a circular region around the interest point.
- Then, we construct a square region aligned to the selected orientation and extract the SURF descriptor from it.

Finally, features are matched between two images.

4.3.3. C-SURF: colored speeded up robust features: [7]

- Color is an important component for objects recognition.
- This will adds the color information into the scale-and rotation-invariant interest point detector and descriptor, coined C-SURF (Colored Speeded-Up Robust Features).
- The first three stages are the same with SURF. In the last stage after calculating the Harr-Wavelet response we also calculate three factors namely $\sum r\ (x, y)$, $\sum g\ (x, y)$, $\sum b\ (x, y)$ for each sub-region.
- The figure below explains how pure gray-based geometric description can cause confusion between two different features.

Fig 5 An example that illustrates the neglecting of color information may confuse the two magnified corners [7]

4.4 Add color information using C-SURF

Color is an important component for objects recognition.

Algorithm

I. Extracting interest points by using the Hessian matrix

II. Finding the location as well as scale of the interest points.

III. Assigning Orientation.

IV. Adding color information to SURF descriptor

4.5 . Feature Matching

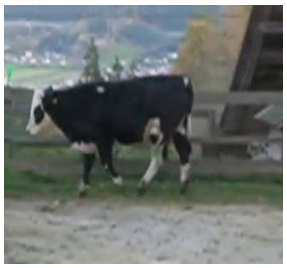
Features are matched between stored model data video and test video using distance calculation. Matched video are ranked and display as a result.

5. RESULT AND DISCUSSIONS

The experimental result of the system is presented in this section. For computation random video is selected. The size of video is 1000 KB to 1500 KB. The duration of these videos is about 30 seconds to 50 seconds. From the training set consisting of 100 videos a number of test clips are constructed. The test set consists of 40 random clips belonging to all groups. The query clips contains relevant and irrelevant videos. Thus the system is tested using clips belonging to both training set and test set.

Initially around 540 frames are extracted using opencv library. Using function fps(frame per second) we can calculate per second number of frames, here it is 15.

Next key frame can be extracted using threshold comparison. Key frame are frame which represent the consecutive frames with no or minor changes. For our video some sample key frames are presented in figure 6 (a)-(d).



(a)

(b)

(c)

(d)

Fig 6 Key Frames from total number of frames

Feature point can be extracted using SURF detector from a single frame which will be use for matching. From our one of the key frame extracted feature points are as under:

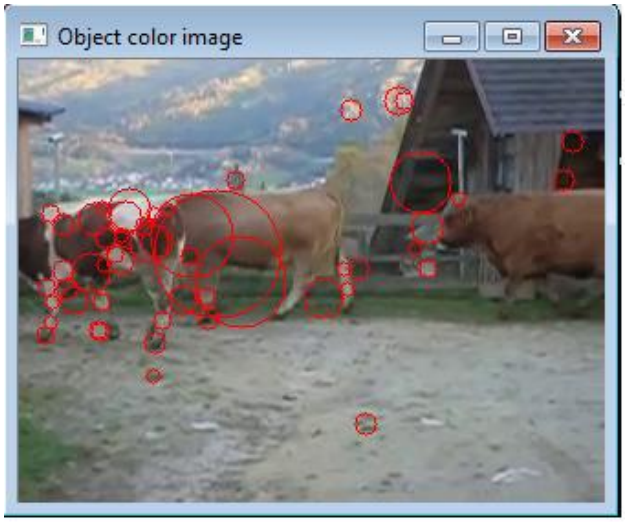


Fig 7 Feature point extraction from key frame

Now, extracted feature points are matched between key frames of test video clip and trained video clip. The figure below shows the matched feature point of two key frames. After then matching video is rank and display to user as a search result.

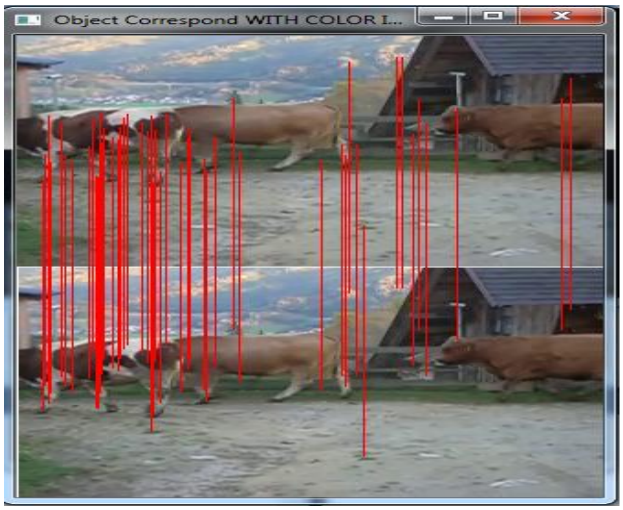


Fig 8 Feature point matching

The following outcome shows the obtained videos by query video as an input.

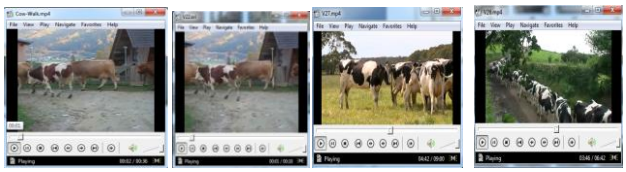


Fig 9 (a) Query Video (b) Retrieved Videos as a Result

6. FUTURE DEVELOPMENTS

Although a large amount of work has been done in visual content-based video indexing and retrieval, many issues are still open and deserve further research, especially in the following areas [10].

1) Motion Feature Analysis: The effective use of motion information is essential for content-based video retrieval. To distinguish between background motion and foreground motion, detect moving objects and events, combine static features and motion features, and construct motion-based indices are all important research areas.

2) Hierarchical Analysis of Video Contents: One video may contain different meanings at different semantic levels.

7. CONCLUSION

It is concluded form the paper that for key frame extraction threshold based comparison algorithm gives good performance. For feature extraction SURF (Speeded-Up Robust Feature) outperforms the other. SURF provides scale- and rotation-invariant detector and descriptor. Repeatability, distinctiveness and robustness are unique features of SURF. The main drawback of SURF is that both detector and descriptor not use color information. C-SURF adds color information to the existing SURF method.

8. REFERENCES

[1] Yarmohammadi, H.; Rahmati, M.; Khadivi, S., "Content based video retrieval using information theory," Machine Vision and Image Processing (MVIP), 2013 8th Iranian Conference on , vol., no., pp.214,218, 10-12 Sept. 2013

[2] Dyana, A.; Subramanian, M.P.; Das, S., "Combining Features for Shape and Motion Trajectory of Video Objects for Efficient Content Based Video Retrieval," Advances in Pattern Recognition, 2009. ICAPR '09. Seventh International Conference on , vol., no., pp.113,116, 4-6 Feb. 2009

[3] Chattopadhyay, C.; Das, S., "STAR: A Content Based Video Retrieval system for oving camera video shots," Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), 2013 Fourth National Conference on , vol., no., pp.1,4, 18-21 Dec. 2013

[4] Asha, S.; Sreeraj, M., "Content Based Video Retrieval Using SURF Descriptor," Advances in Computing and Communications (ICACC), 2013 Third International Conference on , vol., no., pp.212,215, 29-31 Aug. 2013

[5] Jianshu Chao; Al-Nuaimi, A.; Schroth, G.; Steinbach, E., "Performance comparison of various feature detector-descriptor combinations for content-based image retrieval with JPEG-encoded query images,"Multimedia Signal Processing (MMSP), 2013 IEEE 15th International Workshop on , vol., no., pp.029,034, Sept. 30 2013-Oct. 2 2013

[6] Herbert Bay; Andress Ess, Tinne Tuytelaars,Luc Van Gool, "Speeded-Up robust features (SURF)" Vol. 110, No. 3, pp. 346--359, June 2008.

[7] Jing Fu, Xiaojun Jing, Songlin Sun, Yueming Lu, Ying Wang," C-SURF: Colored Speeded Up Robust Features", International Conference, I SCTCS 2012, Beijing, China, Volume 320, pp 203-210. May 28 – June 2, 2012

[8] Jin Zhao; Sichao Zhu; Xinming Huang, "Real-time traffic sign detection using SURF features on FPGA," High Performance Extreme Computing Conference (HPEC), 2013 IEEE , vol., no., pp.1,6, 10-12 Sept. 2013

[9] H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: Speeded Up Robust Features", Computer Vision–ECCV 2006.

[10] Weiming Hu; Nianhua Xie; Li Li; Xianglin Zeng; Maybank, S., "A Survey on Visual Content-Based Video Indexing and Retrieval," Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on , vol.41, no.6, pp.797-819, Nov. 2011

[11] Y.-F. Ma, X.-S. Hua, L. Lu, and H.-J. Zhang, "A generic framework of user attention model and its application in

video summarization," *IEEE Trans. Multimedia*, vol. 7, no. 5, pp. 907–919, Oct. 2005.

[12] K. W. Sze, K. M. Lam, and G. P. Qiu, "A new key frame representation for video segment retrieval," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 9, pp. 1148–1155, Sep. 2005.

[13] B. T. Truong and S. Venkatesh, "Video abstraction: A systematic review and classification," *ACM Trans. Multimedia Comput., Commun. Appl.*, vol. 3, no. 1, art. 3, pp. 1–37, Feb. 2007.

[14] D. Besiris, F. Fotopoulou, N. Laskaris, and G. Economou, "Key frame extraction in video sequences: A vantage points approach," in *Proc. IEEE Workshop Multimedia Signal Process.*, Athens, Greece, Oct. 2007, pp. 434–437.

[15] D. P. Mukherjee, S. K. Das, and S. Saha, "Key frame estimation in video using randomness measure of feature point pattern," *IEEE Trans.Circuits Syst. Video Technol.*, vol. 7, no. 5, pp. 612–620, May 2007.

[16] Jing Li _, Nigel M. Allinson , "A comprehensive review of current local features for computer vision", Elsevier Neurocomputing, 2008 Elsevier B.V.

[17] S. Huang, C. Cai, F. Zhao, "An Efficient Wood Image Retrieval using SURF Descriptor", Proc.International Conference on Test and Measurement, 2009, pp.55- 58, doi: 978-1-4244-4700-8/09.

[18] Zhang Y J, Lu H B, "Hierarchical video organization based on compact representation of video units". Proc. Workshop on Very Low Bitrates Video'99, 1999: 67-70.

[19] Calic J, Izquierdo E, "Efficient key-frame extraction and video analysis". Proceeding of the International Conference on Information Technology: Coding and Computing (ITCC'02), 2002: 28-33.

[20] He Xiang, Lu Gung-ho, "Algorithm of key frame extraction based on image similarity". Fujian Computer, 2009, 5: 73-74.

[21] Ding Hong-li, Chen Huai-xin, "Key frame extraction algorithm based on shot content change ratio". Computer Engineering, 2009, 13: 225-231.