

Extracting Features Using Improved Techniques From Encoded Video For Searching

Saba B.Sayed¹, Dr.Nuzhat F.Shaikh²

Student¹, Head Of Department²

Modern Education Society's College of Engineering Pune, India^[1,2]

sbsayed17@gmail.com¹,

nfshaikh@mescoepune.org²

Abstract: There is a need for the transportation of visual content from a particular node to a computation unit, on a network limited by bandwidth for distributed visual analysis applications, like mobile visual search or Visual Sensor Networks (VSNs). Proposed system used technique of video compression to transfer information over a communication network. The system is used for key point detection, descriptor calculation, and feature matching. The proposed system provides key points to the server as side data and to retrieve only the relevant points from the encoded video. We used only first frame for key point encoding to address different types of changes in video content, which is helpful for increase the speed of encoding and decoding. In our approach for encoding, User choose the video from dataset and extract frame from video then extract key point of selected frame and encode the string or key point on extracted frame then reconstruct or compress the video and upload it on server. For decoding, Whenever user enter search query on browser, server receive the search query. The server then selects video which is stored on server and extract frame from video then extract key point or encoded data from selected frame using decoding then compare the search query and decoded data if it is true the send the video to user. Present pair wise matching and image retrieval experiments conducted to evaluate the performance of the proposed approach using the Stanford dataset of videos. The results show that the proposed approach offers significantly improved feature matching and image retrieval performance at a given bitrates.

Keywords: Coding, 64bit encoder, key points, matching, prediction, retrieval, SIFT.

I. Introduction

In many computer vision algorithms, the retrieval of features from images or videos is a key component. This process identifies features such as rotation, scale, illumination, shift, etc that does not vary or change. These extracted features are repetitively compared with those key elements stored in a database to find similar features. In the extraction process, initially detection of keypoints are done, and after that, descriptors that are local, situated around these keypoints are computed from the image patches. Existing system presented a key point encoding approach for still images. Applying this approach directly to individual frames in an image sequence would significantly increase the bit rate. To address this issue, same as the conventional inter-frame prediction scheme in video coding, we propose several key point prediction approaches that significantly decrease the quantity of keypoints to be encoded and thus the bit rate required for the side information. Either the client or the server can perform feature extraction. For feature extraction at client-side, features are compressed and uploaded to the server. It is expected that this compression will create only small distortions compared with the uncompressed features. For feature extraction at server-side, the compression of image/video takes place and are send to the server. In this case, it is required to reduce the effects of the compression on the feature extraction performed at the server. Ideally, the features extracted from a compressed image or video should be identical, or more similar, to the features extracted from the uncompressed version.

II. REVIEW OF LITERATURE

The focus is on improving extraction process from compressed video in [1]. In the feature extraction process, initially detection of keypoints are done, and after that, descriptors that are local, situated around these keypoints are computed from the image patches. Existing system presented a key point encoding approach for still images. Applying this approach directly to individual frames in a image sequence would significantly increase the bit rate, to address this issue. The focus is on improving feature extraction from compressed images in [2]. Since compact pictures are sent through an internet network for examination at a destination, the action at the destination follows extracting features and comparison. Compression of picture have an undesired results on feature matching task. To overcome this problem, they proposed to sent the key points, as extra data to destination, and acquire only the feature points from the compressed images. Hence, an idea been proposed to effectively encode the areas, scales, and orientations of key points taken from the actual picture. Another approach focuses on choosing important yet fragile keypoints as side data, further decreasing the volume of data. In [3] they concluded that for video/image compression and preserving the features, the compression artifacts easily affects the keypoints detection, while the calculation of descriptor is more robust. The authors of [5] and CDVS [6] proposed ways to deal with encoding the areas of key points utilizing sum based context-based arithmetic coding. [5] with [6] proposed to encrypt areas, orientations and scales of features with extensive scale. It does encoding of down sampled picture having constant QP, and just a less keypoints with extensive value are preserved, because the features of small scale are not needed in this situation. Thus, down sampled image and the encrypted key points are transmitted to the server. In [8] author showed in an image, the computation of feature descriptors are done considering a repeatable keypoint from gradients of a canonical patch. A detector for keypoint focuses to permit effective interframe coding of canonical patches. In [10] focuses on Hybrid-Analyze-Then-Compress concept, an effective paradigm tailored to distribute visual analysis tasks. Such model exploits a joint pixel and local feature-level coding architecture, leading to significant bitrate savings. It focuses on increasing the coding efficiency of both, the keypoint location and the descriptor enhancement layer modules and at extending the approach to different classes of local features. In [16] they concentrates on strategy for keypoint location, depiction and coordinating. It proposes BRISK algorithm as adaptive having quality of high performance at low computational cost as compared to SURF. In the paper [19] focuses on descriptors of binary which is regarded as alternatives having low-complexity to today's descriptors such as SIFT. Binary string is utilized to represent the descriptor where each bit results as a pair-wise comparison of smoothed pixel values selected correctly considering each keypoint in a particular area. It introduces two functions. First, a designing an coding scheme that allows an ordering of the descriptor internally that reduces the bits required to portray it. Second, computation of selection strategies that is used for recognizing which pair-wise comparisons can be utilized when creating descriptor. The lossless coding of parallel descriptors problem is focused in this paper. In such context, the most suitable comparison is selected as a function of rate. These mentioned methods of selection can be used with any binary descriptor. In this paper, results are represented using BRISK and FREAK descriptors. Annotated database for patches of image retrieved is used for evaluation. In this way, aim is on the descriptor design independent from the given descriptor. Outcome conclude that a correct selection of the pairs, along the correct reordering before lossless coding, allows to enhance the discriminative power. For example, in the case of BRISK, the false positive rate is minimized from 76% to 58%, when the true positive rate is equal to 95% and 128 pair wise comparisons are used, which are lossless compressed using about 80 bits on average.

III. SYSTEM OVERVIEW

Existing systems focuses on key points that are provided to the server as side data and from compressed video only descriptors are extracted. Locations, scales, and orientations of keypoints retrieved from the original image are efficiently encoded. Existing system presented a key point encoding approach for still images. Applying this approach directly to individual frames in a image sequence would significantly increase the bit rate, to address this issue, similar to the conventional inter-frame prediction scheme in video coding, we propose several key point prediction approaches that significantly decrease the quantity of keypoints to be encoded. In mobile visual analysis scenarios, server is used to analyze the video compressed and transmitted through the communication network. Compression of image take lot of time and has adverse effect on each frames which is extracted from video.

In proposed work user chooses the video from dataset and extract frame from video then extract key point of selected frame and encode the string or key point on extracted frame then reconstruct or compress the video and upload it on server. For decoding, whenever user enter search query on browser, the server receive the search query. The server then select video which is store on server and extract frame from video, then extract key point or encoded data from selected frame using decoding then compare the search query and decoded data. If it is correct it sends the video to user. We proposed a Content Based Video Retrieval (CBVR) system which gives query image to the system and list down all videos consisting of query image.

Original SIFT keypoints from a video are encoded and these key points and compressed video is transmitted together to the server. Instead of encoding of keypoints according to scale, locations and orientation, a particular string is encoded which is useful for selection of relevant videos from the database.

IV. SYSTEM ARCHITECTURE

Figure 1. shows the system architecture.

Input: Input given to the system are:- Video in any format.

Process: For encoding, User choose the video from dataset and Extract frame from video then extract key point of selected frame and encode the string or key point on extracted frame then reconstruct or compress the video and upload it on server. For decoding, Whenever user enter search query on browser then server receive the search query then server select video which is store on server and Extract frame from video then extract key point or encoded data from selected frame using decoding then compare the search query and decoded data if it is true the send the video to user.

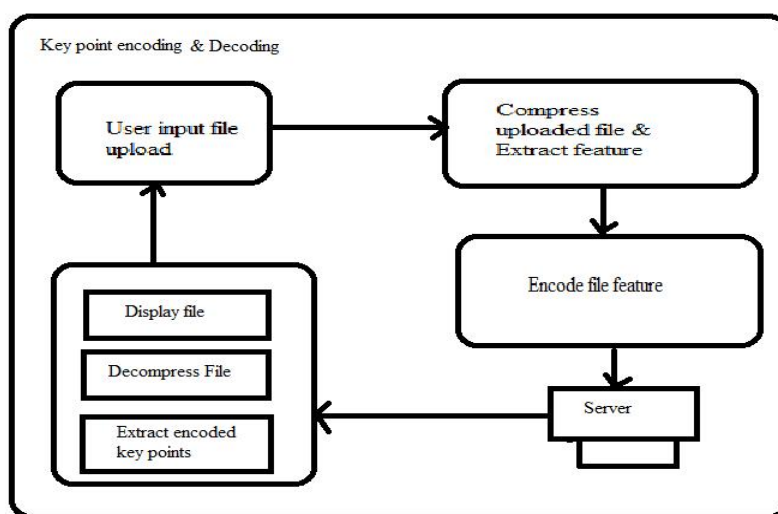


Fig. 1 Proposed System Architecture

Output: Video related search query and we check or test the result is accurate or not.

We use 64 bit encoder/decoding technique for encoding and decoding string in an image.

Algorithm : 64Bit Encoder/Decoder

- 1) The input stream of bytes is divided into 3 bytes blocks.
- 2) Divide 24 bits of every 3-byte block into 4 groups of 6 bits.
- 3) Map each combination of 6 bits to 1 printable character, in the view of the 6-bit value utilizing the Base64 character set map.
- 4) If the last 3-byte block has only 1 byte of information, pad 2 bytes of zero (\x0000). In the wake of encoding it as a normal block, supersede the last 2 characters with 2 equal signs (==), so the decoding process knows 2 bytes of zero were padded.
- 5) If the last 3-byte block has only 2 bytes of input info, pad 1 byte of zero (\x00). In the wake of encoding it as a normal block, supersede the last 1 character with 1 equal signs (=), so the decoding process knows 1 byte of zero was padded.

Carriage return (\r) and new line (\n) are embedded into the output character stream. They will be ignored by the decoding process.

V. PERFORMANCE RESULTS AND ANALYSIS

Our test is based on comparison of Proposed System and Similar System based on algorithms used for extracting features and its retrieval application. Tabe 1 shows comparisons of various algorithms used in existing systems.

Table 1: Comparison of Proposed System and Similar System based on algorithms used for feature extraction and its retrieval application

Algorithm	Approach	Features used	Retrieval Application
Color-Texture Classification	Segment images into Regions	Color, Texture Features	Perceptual importance
Multi-Modal Content Based Browsing and Searching Methods	Peer2Peer retrieval Systems	Key Frame Extraction, Shape Feature	Object Based
Character Identification	Object retrieval of movie	Segmentation	Object Based
Semantic Video Retrieval	Automatic Audio categorization	Audio	music, speech, etc
Recognizing Object in Video Sequences	Sequential classification methods for blob	Kalman filter applied over blob	mage Object based
Automated Scene Matching	Image matching	Wavelets, Color Descriptors	3D scene in a movie
Content Based Face Retrieval	Image matching	self-organizing maps (SOMs) and user feedback Visual clustering	Face Recognition
Proposed System	Encoded content based Video Retrieval System	Frame Extraction, Base64 Encoding and Decoding, H.265 for Video Construction and Compression.	Encoded Content Based

The following are the results of implementation of the system.

We analyze our system using precision and recall strategy.

Precision and recall are the basic measures used in analyzing the strategies for searching.

$$Precision = \frac{\text{No. of similar video retrived}}{\text{Total no. of video encoded}} \quad Recall = \frac{\text{No. of similar video retrived}}{\text{Total no. of videos in dataset}}$$

Table 2: Table shows results based on precision and recall.

Video Category	No. of Videos	Precision %	Recall %
Songs	11	80	59.95
Funny	9	85	58.66
Speech	10	90	66
Cricket	10	83	49
Movie	8	85	81.25
Drama	8	87	77.03
Football	7	90	100

The Following graph shows the result based on precision and recall. The number of videos are taken as input from the database and their precision and recall values are compared and analysed.

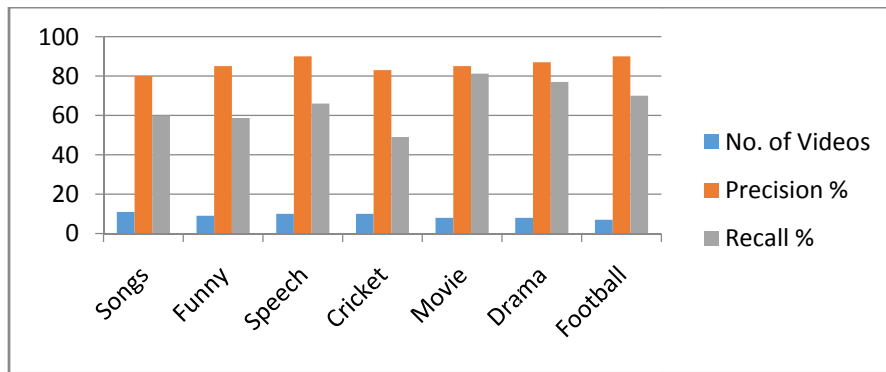


Fig 2: Graph of (Precision(%) and recall(%) based on Proposed System)

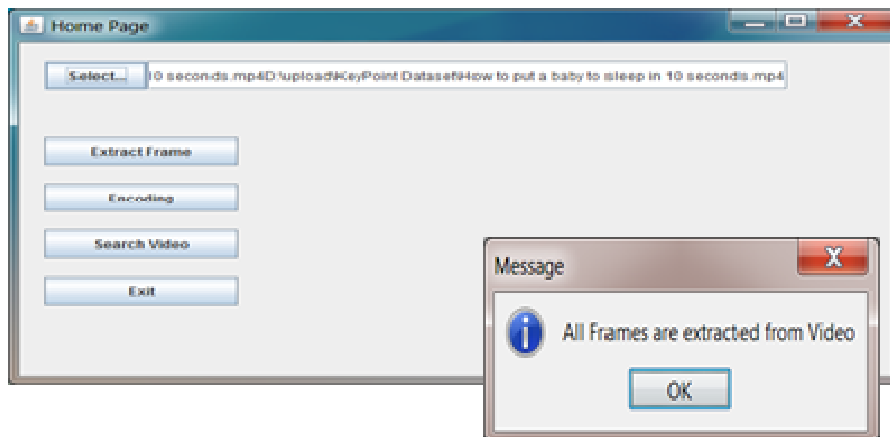


Fig 3: User selects the video and frames are extracted from it.

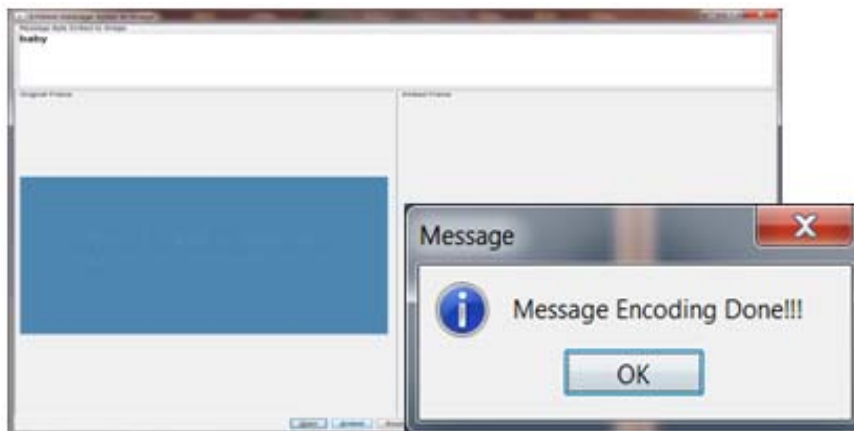


Fig 4: Message is encoded on selected frame.



Fig 5: The searching of video is done based on the string encoded on frame

CONCLUSION

Various techniques and methods are been studied for extracting images from the video and then extracting various features from the images .Discovering various methods and techniques for encoding the original keypoints from a video and send them along with the compressed video to the server. In our approach for encoding, user chooses the video from dataset and extract frame from video and encode the string or key point on extracted frame then reconstruct or compress the video and upload it on server. For decoding, whenever user enter search query on browser, the server receive the search query. The server extract key point or encoded data from selected frame using decoding then compare the search query and decoded data. If it is true it sends the video to user.

This system discovers various methods and techniques for encoding the string as keypoints from a video which can be used as to search relevant video on the server.The method which we focused in this paper is Base64 for video compression and encoding/Decoding. This system is developed for the purpose of retrieving the video from the Multimedia Database by using efficient algorithms to increase the performance of the system which is difficult in traditional video retrieving system. Proposed system can be enhanced to reduce time required to search all relevant videos on the server. The matching performance of image query given to the Content based Video Retrieval System can be improved by matching spatial features like color, shape and texture and temporal features like motion and audio.

REFERENCES

- [1] J. Chao and E. Steinbach, "Keypoint encoding and transmission for improved feature extraction from compressed video at low bitrates", in Proc. IEEE Int. Conf. Multimedia Expo, Jan. 2016, pp. 2439.
- [2] . Chao, E. Steinbach, and L. Xie, "Keypoint encoding and transmission for improved feature extraction from compressed images", in Proc. IEEE Int. Conf. Multimedia Expo, Jun.Jul. 2015, pp. 16.
- [3] J. Chao, R. Huitl, E. Steinbach, and D. Schroeder, "A Novel Rate Control Framework for SIFT/SURF Feature Preservation in H.264/AVC Video Compression", IEEE Transactions on Circuits and Systems for Video Technology (accepted for publication), 2014.
- [4] <http://mpeg.chiariglione.org/standards/mpeg-7/compact-descriptors-visual-search>
- [5] Chen, Tsai Sam, G. Takacs, V. Chandrasekhar, M. Makar, R. Grzeszczuk, and B. Girod, "Improved coding for image feature location information", in Proc. SPIE, 2012, vol. 8499.
- [6] H. Yue, X. Sun, F. Wu, and J. Yang, "SIFT-Based Image Compression", in IEEE International Conference.
- [7] M. Fischler and R.C. Bolles, "Random sample consensus:A paradigm for model fitting with applications to image analysis and automated cartography, Communications of the ACM", vol. 24, no. 6, pp. 381395, 1981..
- [8] M. Makar, S. S. Tsai, V. Chandrasekhar, D. Chen, and B. Girod, "Interframe coding of canonical patches for low bit-rate mobile augmented reality", Int. J. Semantic Comput., vol. 7, no. 1, pp. 524, 2013.
- [9] D. Lowe, "Distinctive Image Features From Scale-Invariant Keypoints," International Journal of Computer Vision, vol.60, pp. 91-110, Nov. 2004.
- [10] L. Baroffio, M. Cesana, A. Redondi, M. Tagliasacchi, and S. Tubaro, Hybrid coding of visual content and local image features, CoRR, 2015 [Online]. Available: <http://arxiv.org/abs/1502.07828>.
- [11] L. Baroffio, M. Cesana, A. Redondi, and M. Tagliasacchi, "Compress-then-analyze vs. analyse-then-compress: Two paradigms for image analysis in visual sensor networks", in IEEE International Workshop on Multimedia Signal Processing (MMSP) 2013, Pula, Italy, September 2013.
- [12] S. S. Tsai, D. Chen, G. Takacs, V. Chandrasekhar, M. Makar, R. Grzeszczuk, and B. Girod, "Improved coding for image feature location information", in Proc. SPIE, 2012, vol. 8499.
- [13] "Compact descriptors for visual search: Context and objectives," in [ISO/IEC/JTC1/SC29/WG11/N11530], (2010).
- [14] S. Leutenegger, M. Chli, and R. Y. Siegwart, BRISK: Binary robust invariant scalable keypoints, in Proc. Int. Conf. Comput. Vis., Nov.2011,pp. 25482555.
- [15] K. Mikołajczyk and C. Schmid. "A performance evaluation of local descriptors". In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2003.
- [16] "Evaluation framework for compact descriptors for visual search," in [ISO/IEC/JTC1/SC29/WG11/N12202], (2011).

- [17] M. Makar, V. Chandrasekhar, S. S. Tsai, D. Chen, and B. Girod, Interframe coding of feature descriptors for mobile augmented reality,IEEE Aug 2014.
- [18] A. Redondi, L. Baroffio, J. Ascenso, M. Cesana, and M. Tagliasacchi, "Rate-accuracy optimization of binary descriptors, in Proc. IEEE Int. .2013, pp. 29102914.
- [19] L. Baroffio, J. Ascenso, M. Cesana, A. Redondi, and M. Tagliasacchi,"Coding binary local features extracted from video sequences", in Proc.IEEE Int. Conf. on Image Process., Oct. 2014, pp. 27942798.
- [20] S. S. Tsai, D. Chen, G. Takacs, V. Chandrasekhar, M. Makar, R. Grzeszczuk, and B. Girod, "Improved coding for image feature location information", in Proc. SPIE, 2012, vol. 8499.
- [21] H. Yue, X. Sun, J. Yang, and F. Wu, "Cloud-based image coding for mobile devices toward thousands to one compression", IEEE Trans. Multimedia, vol. 15, no. 4, pp. 845857, Jun. 2013.