Proceedings of the Second International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC 2018)
IEEE Xplore Part Number:CFP18OZV-ART; ISBN:978-1-5386-1442-6

Extraction of Visual Features from Video Sequences for Better Visual Analysis

Prachi Rohit Rajarapollu
Department of Electronincs & Telecommunication Engineering
MIT AOE Pune, Maharashtra India
prrajarapollu@entc.maepune.ac.in

Vijay R Mankar
Department of Electronincs & Telecommunication Engineering
GPA, Amravati Maharashtra India
vr_mankar@rediffmail.com

***Abstract*— Video have a basic and non basic features, where basic features includes e. g. color, shape, size and non basic features include orientation of a image. Whereas Video Sequences is a series of shots/frames on a subject that are edited together to tell a story. Visual features describes the details about the image contents, which are used in various applications like, visual search, object recognition, image registration and object tracking. Many visual analysis task requires the features to be transmitted, thus it calls for the different coding algorithms to attain a target level of efficiency. Here an effort has been taken to implement a coding algorithm for local features extraction such as SIFT (Scale Invariant Feature Transform). The first stage comprises of using the SIFT algorithm property to find the 'point of interest' of an image. Further the use Kalman Filter algorithm is done as an application purpose of motion based single or multiple object detection and tracking.**

Keywords—Visual features, Extraction, SIFT, Kalman Filter, Keypoints, descriptors etc.

I. INTRODUCTION

Visual features are mainly used to represent visual contents, whose characteristics are robust and invariant for transformation. This task may be object recognition, object tracking, information retrieval etc [1]. Visual feature extraction can be done in two steps like, in step one detection of salient key points from an image is done and in step two descriptors are used to identify a patch surrounded by each key point. The system implemented in this paper can be conveniently adopted to implement a algorithm which will work on analyze then compress (ATC) anatomy. A set of visual features extracted from video has been prepared first. Encoding of all set has been done as preprocessing. After preprocessing it has been given for visual feature extraction and analysis. In traditional methods first compression and then analysis is done by sending data to central processor. Compression of visual features is one of the important tasks in VWSN (Visual Wireless Sensor Network) [5]. The main problem related to transmission of data is bandwidth, in VWSN the data which is transferred to the base station having low bandwidth which is going to affect the speed of performance. Several researchers had worked to solve this problem for the case of extraction of features and suitable methods to encode these features or to introduce suitable algorithms for feature extraction so that it will become suitable for further compression [3]. In this paper there is an overview of different algorithms for extracting features or compare different types of descriptors for video coding is given. Processing of the visual features requires more data, so the energy required for transmission of the large data is also more so this is also one of the problems regarding transmission of data directly to the base station [4]. So different algorithms are used to extract visual features and then designed encoder which encodes the data at low computational complexity.

In this paper major focus is on the extraction of local features of an image to find the point of interest (e.g., key-points and descriptors) extracted from different images/video sequences. Further Kalman filter has been used for motion detection and single and multiple object detection is done successfully.

II. ALGORITHM TO EXTRACT VISUAL FEATURES

Implemented algorithm is going to work step by step. Here use of inter frame. Intra frame coding has been used to find out the key points. Feature extraction has been done with the help of SIFT techniques [3]. Finally Kalman filter has been applied to get more and more accurate results. Following sections describes each block related to system implemented.

A. Intra-Frame Coding

In this method frame-by-frame processing has been done to extract local features from each frame [8]. As it is well known that video is moving the still images with particular speed. In intra fame processing each frame has been considered it means processing has been done on still image. Same concept has been used in video compression techniques. Video compression can be lossy or lossless depending upon the requirement and demand of application. In intra fame processing is done within the frame and there is no link with frame in video sequence.

B. Inter-Frame Coding

In inter frame coding two consecutive frames are taken into consideration. One frame has been considered as a

reference for the next one. Inter frame coding is more preferable as it helps in extracting and detecting key points relevant to each other. In inter frame coding it possible to detect the key points based on physical entities as visual contents does not change abruptly. Image patches around two key points have similar characteristics leading to similar descriptors [9].

C. Feature Extraction

Various techniques are available to extract the features from image. Image analysis is process of extracting features based on edges and shapes of an object [6]. It is desirable to convert the image in gray scale first which will make the edges more prominent and simplifies the process of edge detection [5]. Area with high frequency and fast change in gray scale is considered as a edge. Identification of this high frequency location is the edge detection. Shape is the low frequency locations. Edge is considered the external feature of an image while detection of an object is inside the image. Object detection can be done by segmentation and texture identification. In segmentation it is possible to focus on any particular object in a complete image. Identify it, process it and extract it. When focus is identification of multiple objects same can get repeated number of times based on the objects to identify. Texture is quantitative measure of coarseness of an image. Spectral or spatial domain analysis is way with which texture analysis is possible. Restoration,, image enhancement, are another type of image processing where input and output both are in the form of image.

D. SIFT - Scale Invariant Feature Transform

In computer vision visual features can be extracted by SIFT (Scale Invariant Feature Transform) by detecting and describing local features from the image. Various points has been detected which help in describing the features of an object. Based on these features it is possible to identify the exact object of interest. These features can be recorded in various ways. SIFT method is free of many of complications faced by other methods like, object rotation, scaling and flipping. In SIFT it is possible to identify an object from different position, from multiple locations in same environment.

SIFT algorithm works in four different stages like,

- Scale Space Extreme Detection

In first step of SIFT algorithm it identify the locations and scales of an object which available at different views in an image. Scale space function is the best option for identification. Gaussian function has been used for assumption and equalization. The scale space is defined by the function: $L(x, y, s) = G(x, y, s) * I(x, y)$ Where * is the convolution operator, $G(x, y, s)$ is a variable-scale Gaussian and $I(x,y)$ is the input image.

- Key-point Localization

In key-point localization main focus is on identification of edges. All those points are identified which have low contrast or high frequency on an edge. By calculating Laplacian it is possible to get the exact values, which called image interpolation.

- Orientation Assignment

Next step followed in SIFT algorithm is assignment of consistent orientation to the key-points based on local image properties. Key-point descriptors can be identified which are invariant to rotation. As per the requirement of Gaussian smoothed image L Key-point scale has to be selected. From above gradient magnitude, m has been calculated. With the help of this magnitude it is possible to calculate orientation histogram, representing as highest peak. By keeping this peak as a reference point and any other local peak within 80 percent of the height can be used to create a keypoint. After identification of such point a parabola has been drawn this will fit 3 to 4 histogram values closest to each peak.

- Keypoint Descriptor

By using above local gradients data keypoint descriptors can be created. As per the orientation of keypoint gradient information is adjusted and then weighted by a Gaussian with variance of 1.5 * keypoint scale. With the help of this data a set of histogram has been created over a window centered on the keypoint. These are SIFT keys used to identify an object from its nearest neighbors [10].

- KLT – Karhunen Loeve Transform

The Karhunen Love theorem represents a stochastic process as an infinite linear combination of orthogonal functions, analogous to a Fourier series representation of a function on a bounded interval. This theorem is related to Principal Component Analysis (PCA) technique specifically used for processing the image in various applications.
Assume that C is the covariance matrix of the population and A is the matrix representing eigenvectors of matrix C. Matrix A is arranged in such a way that rows of this matrix will represent eigenvectors of matrix C. First row will include largest eigenvector values and last row will include smallest eigenvector values of matrix C. This can be represented in mathematical way as follows; this is a Karhunen Love theorem.

$$y = A = (x - mx)$$

- Kalman Filter

Kalman filter is used to find out Linear Quadratic Estimation (LQE). Kalman filter calculates a new more accurate variable based on statistical noise or any other inaccuracies from a set of measurements. Bayesian inference and estimating a joint probability distribution over the variables for each time frame is base for calculation of more accurate variable. Object tracking is one of the most prominent applications in the field of computer vision. Tracking of exact object from moving video is the most challenging task [2] in applications like, surveillance, military guidance etc. Kalman filter is one of the solutions to filter out unwanted portion and track the required object accurately.
This challenge can be solved by performing the task in two phases,

- Detect moving object frame wise

- Track the same object on time scale

Kalman filter can been used to perform above task. Kalman filter specifically used to in applications where object tracking and detection is critical issue like navigation, monitoring of vehicles, military applications etc.
In Kalman filter analysis it is assumed that errors are Gaussian distributed. Following two steps has been followed in Kalman filter processing, prediction and processing. Kalman filter produces estimates of the current state variables, along with their uncertainties and these estimates are updated using a weighted average, with more weight being given to estimates with higher certainty. Kalman filter works in real time analysis based on current input and previously calculated states, no past information is required.

III. SYSTEM IMPLEMENTATION

As shown in system block diagram figure 01 video is given as an input. Video is got converted into 'n' number of frames. As visual features [7] are transmitted using bandwidth limited network encoding of the signal is essential one. Thus coding involves several processes like Encoding, Transformation, Quantization and Decoding. Then further we apply the KLT (Karhunen Loeve Transform) compression technique for compression of the image and further the SIFT algorithm to get the desired result which is stored in the Database of the system.

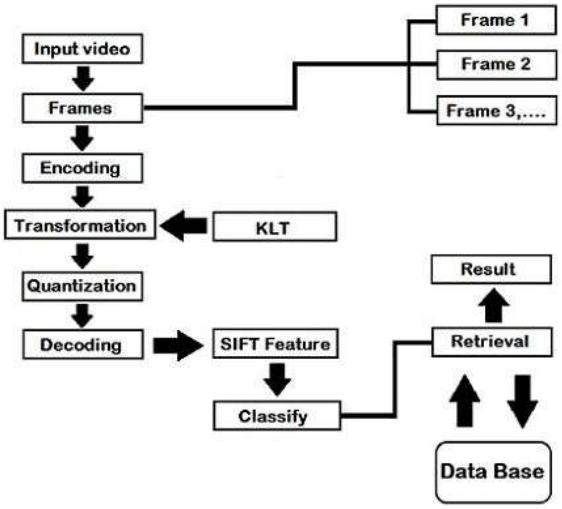

Fig. 01 System block diagram

Video has been given as an input and following steps has been followed.
Input video : the given data or input.
Frame : It is a sequence of an image extracted from a single video. Necessity of frame division is for easily extracting the information.
Encoding : It is a process encryption. It also arranges the frame into a specialized manner for efficient transmission or storage.
Transformation : Is a function block that maps one set to another after performing some operation. It converts the input image in another domain equivalent. Here, KLT compression technique is used in transformation block which removes the redundancies by de-correlation of given data to get the desired results.
Quantization : It is a lossy compression technique which divides the signal/data into small parts.
Further, Quantization is used to reduce the grey levels for more efficient result.
Decoding : It a process of decryption technique. Here we extracted the features of all the frames by SIFT feature technique. After extracting the features from SIFT technique, this features of an image is undergoing the process of Classification. Further, the data is retrieved and goes to the Database of the system.

IV. RESULT

After converting the input video into 'n' number of frames it gets converted into Gray Scale Image for feature extraction process as shown in fig. 2. A single Frame is used for analysis process. Using the SIFT (Scale Invariant Feature Transform) coding algorithm we have analyzed a single frame and observed its Local features such as:

- Key-points
- Descriptors

Further using the Kalman Filtering algorithm property extraction of key-points and descriptors is done as shown in fig. 03. we have also performed an application based function on a pre-recorded video content for its Motion based multiple object detection and tracking is shown in fig 4. The multiple object detection is done results are shown in fig. 05

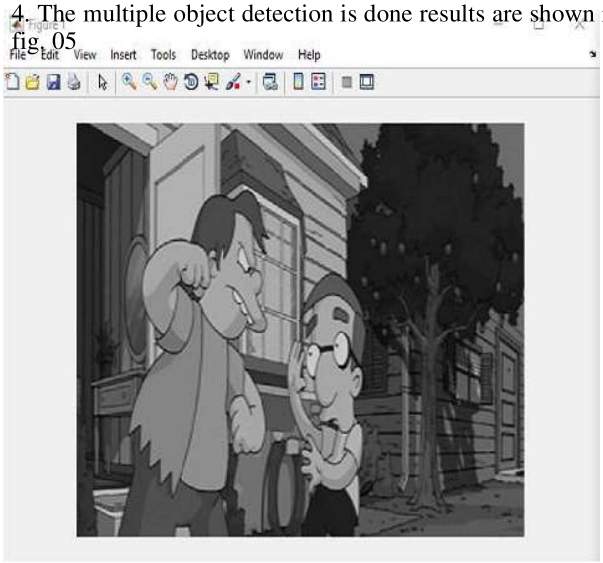

Fig. 02 Conversion of image to gray scale as preprocessing

Now, at the second stage of our work, we have performed an application based function on a given pre-recorded video or even we can perform it on a real time based video so thus by using the Kalman filter algorithm which gives the results as shown in figure 04.

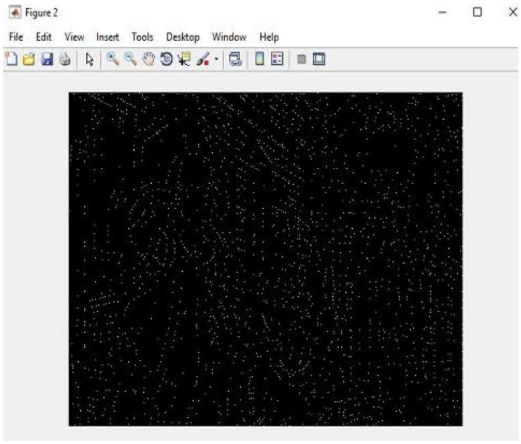

Fig. 03 Extraction of Key-points and descriptors

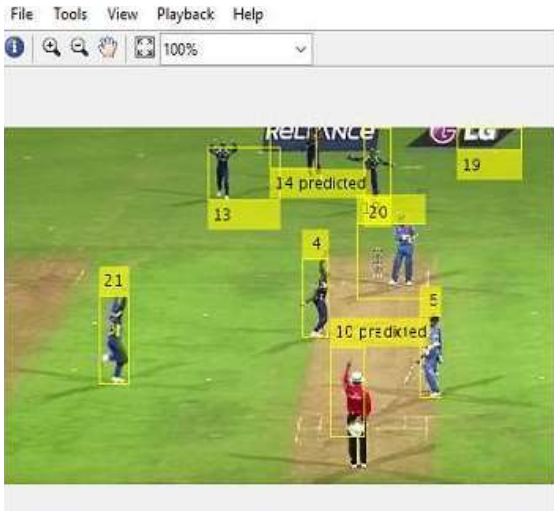

Fig. 04 Multiple object detected and tracked

Here, Input video is divided into 'n' number of frames and further it gets converted into Gray Scale image for feature extraction process. A single frame is used for analysis process. Now, using the SIFT (Scale Invariant Feature Transform) coding algorithm we have analyzed a single frame and observed its Local features such as:

i. Key-points
ii. Descriptors

Here, we are using the Kalman Filter algorithm coding to perform an application based function on a pre - recorded video content for its Motion based multiple object detection and tracking.

V. CONCLUTION

This paper presents the techniques to extract local visual features from a frame. Inter frame and intra frame techniques has been used to exploit the redundancy along with temporal dimension. Coding efficiency has been improved significantly by coding mode decision scheme. By using implemented algorithm it is possible to find out any type of local visual features which will plays a major role. We conclude that Analysis-then-Compression (ATC) is much better than Compressed-then-Analyzed (CTA) method. And also conclude that by comparing different algorithms of compression that binary feature extraction algorithms are more suitable than SIFT and SURF algorithm. We have also concluded that by using the Kalman Filtering algorithm coding we have performed an application based function on a pre-recorded video content for its motion based multiple object detection and tracking.

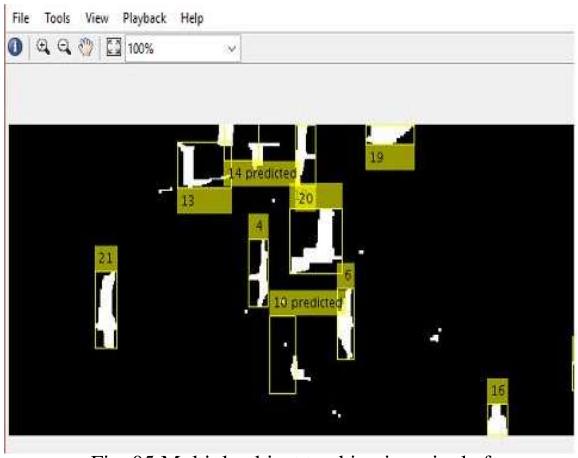

Fig. 05 Multiple object tracking in a single frame

REFERENCES

[1] M. Karpushin, G. Valenzise and F. Dufaux, "Local visual features extraction from texture+depth content based on depth image analysis," *2014 IEEE International Conference on Image Processing (ICIP)*, Paris, 2014, pp. 2809-2813.

[2] M. Oelsch, D. V. Opdenbosch and E. Steinbach, "Survey of Visual Feature Extraction Algorithms in a Mars-like Environment," *2017 IEEE International Symposium on Multimedia (ISM)*, Taichung, 2017, pp. 322-325.

[3] L. C. Chiu, T. S. Chang, J. Y. Chen and N. Y. C. Chang, "Fast SIFT Design for Real-Time Visual Feature Extraction," in *IEEE Transactions on Image Processing*, vol. 22, no. 8, pp. 3158-3167, Aug. 2013.

[4] Z. Shi and Y. Wan, "A color-to-gray conversion based on visual feature extraction of JND," *2016 International Conference on Progress in Informatics and Computing (PIC)*, Shanghai, 2016, pp. 345-349.

[5] E. Eriksson, G. Dán and V. Fodor, "Algorithms for distributed feature extraction in multi-camera visual sensor networks," *2015 IFIP Networking Conference (IFIP Networking)*, Toulouse, 2015, pp. 1-9.

[6] Z. Zhang, F. Liu and W. Qu, "Review of the visual feature extraction research," *2014 IEEE 5th International Conference on Software Engineering and Service Science*, Beijing, 2014, pp. 449-452.

[7] X. Shen, J. Wang, Q. Yang, P. Chen and F. Liang, "Feature based inter prediction optimization for non-translational video coding in cloud," *2017 IEEE Visual Communications and Image Processing (VCIP)*, St. Petersburg, FL, 2017, pp. 1-4.

[8] M. S. R. Sajib and S. M. Tareeq, "A feature based method for real time vehicle detection and classification from on-road videos," *2017 20th International Conference of Computer and Information Technology (ICCIT)*, Dhaka, 2017, pp. 1-11.

[9] Z. Wang, L. Wang, Y. Wang, B. Zhang and Y. Qiao, "Weakly Supervised PatchNets: Describing and Aggregating Local Patches for Scene Recognition," in *IEEE Transactions on Image Processing*, vol. 26, no. 4, pp. 2028-2041, April 2017.

[10] W. Zhang, D. Liu, Z. Xiong and J. Xu, "SIFT-based adaptive prediction structure for light field compression," *2017 IEEE Visual Communications and Image Processing (VCIP)*, St. Petersburg, FL, 2017, pp. 1-4.