Object Detection in Videos using Shot Clustering

Gunjan Gujarathi[1] Mrs. S. M. Patil[2]
[1,2]Assistant Professor
[1,2]Department of Electronics & Tele Communication Engineering
[1]SGDCOE, Jalgaon, NMU, India [2]Government College of Engineering, Jalgaon, NMU, India

Abstract— This is a new approach to object detection and tracking are the challenging task in many computer vision applications. The clustering is a natural solution to abbreviate and organize the content of video. Research on object detection suggests that momententous objects can be well represented as more feature extraction from appropriately chosen video. Tracking is usually performed in the context of higher-level applications that require the location and/or shape of the object in every frame. In proposed system use shot clustering which is based on feature extraction. Object detection and tracking in motion based videos by using the joint color and texture histogram to represent a target. Apart from conventional color histogram feature of the object are also extracted by using LBP technique to represent the object. All frames are sequentially aligned for displaying tracking result compared with traditional color histogram method. Implemented algorithm extracts effectively features in the target region which represents more robustly than tracking produced in other similar method. For future investigation is to multitracking of detected object in videos and more number of feature extractions.

Key words: Object detection, object tracking, texture, color histogram, LBP, multitracking

I. INTRODUCTION

In the recent years the rapid progress in computer technologies there has been an increase in the amount of visual information. This information is generated, stored, accessed, transmitted and analyzed. As the amount and complexity of video information grows, the need of the more intelligent video manipulating techniques becomes obvious [11]. Video is represented as a sequence of number of following in order of frames each of which has constant time interval. The smallest physical segment of video is shot which is defined as an unbroken sequence of frames. The valid frames can group together is the cluster. Since the goal of clustering is to discover a new set of categories the new groups are of interest in themselves & assessment is intrinsic.

Moving object detection and tracking is a key challenge for designing higher-level applications as monitoring traffic behavior, video surveillance and so many other different applications. These has been extensively studied by researchers in computer vision and the ITS field [6]. Object detection is the beginning for analysis of videos. It handles segmentation of moving object from stationary background object. Owing to dynamic environmental conditions such as illumination changes, shadows and waving tree branches in the wind object segmentation is a difficult and significant problem that needs to be handled well for a robust visual surveillance system. The further processing of video analysis is tracking, which simply defined as the creation of temporal correspondence among detected objects from frame to frame. This provides temporal identification of the segmented regions and generates cohesive information about the objects in the monitored area such as trajectory, speed and direction [1].

Now, many opportunities have opened for the development of applications in various areas as video surveillance, content creation, personal communications, robotics and natural human–machine interaction. Object detection in videos involves verifying the presence of an object in image sequences and possibly locating it accurately for recognition. The process of estimating over time the location of one or more objects using camera is referred to as object tracking. The main drawback of traditional systems is loss the spatial information. Moreover, similar appearance of object and background that was become invalid to distinguish them.

The proposed algorithm is to compare two successive frames and on the basis of the comparison, find the difference in the position of the detected object. Automated video surveillance computer vision system is designed to monitor the movements in an area, identify the moving objects and report any doubtful situation. The system needs to discriminate between natural entities and humans, which require a good object tracking system.

This paper organizes as follows. In section II we review previous work related to shot clustering, object detection and tracking. In section III we propose new object detection and tracking algorithm. In section IV experimental results are determined for different video data. Finally conclusions and some direction for further research will be given in section V.

II. LITERATURE SURVEY

The survey of object detection and tracking first requires key-frame and feature extraction in videos. There are various methods that detect motion and trace the moving object. Some of them are described below. The published approaches are sorted chronologically with respect to year of publication.

A. Shot Clustering

Shot clustering algorithm for object detection and tracking are different for different applications. For example, method based on cophentic criterion [10], method considering low resolution video sequences [9]. Especially, Clark F. Olson [3] proposed pose clustering object recognition algorithm based on view point consistency to determine object poses that aligned many of the object features from search images. It had been successfully applied to spacecraft pose estimation of crater matching.

B. Object Detection

Every tracking method requires an object detection mechanism either in every frame or when the object first appearance in the video. The basic method for object detection was point detectors that only interest points in images which had an expressive texture in their respective localities. Further modification in this method was background subtraction to detect moving regions for subtracting the current images to pixel by pixel from reference background images [2]. This statistical background model where each pixel was represented with its minimum and maximum intensity values. From this methods pixel ratio and spatial information with low resolution are found.

C. Object Tracking

Object tracking algorithms are divided into basic and hybrid approaches. In basic block matching method that predicts the object outline using motion vector information [4], low frame rate video that extends the standard mean shift technique using multiple kernel created on high motion areas obtained in change detection [7]. Another method used in shopping groups in stores, surveillance segments each frame into foreground regions which contains multiple people. After detection of person, appearance model based on color and edge density in conjunction with mean shift tracker was used to recover the person's trajectory [3]. The hybrid approaches was to combined two methods for example, multi mode anisotropic mean shift and particle filter to include online learning of reference object [5], use of spectrum in object tracking in digital video surveillance in [11], real time application based method that estimate the velocity for tracked object [8].

The approaches focus on domains with and unclear detected object where on object information is available and these provide big challenge. So develop a system that overcome problems defined in survey to give accuracy and reduce complexity.

III. SYSTEM DEVELOPMENT

This section introduces the system formally providing a basics of the design, discusses the ideas and implementation issues that considered with system development.

A. Block Diagram of Proposed System

Several approaches have been discussed in literature for the object detection and object tracking but still there is a problem. So develop a system can detect object and track that using joint color and texture histogram by applying mean shift algorithm where feature extraction of color and texture in video frames. In Fig. 1 shows block diagram of proposed system

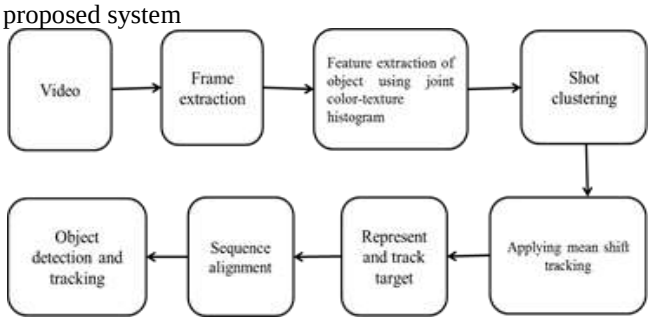

Fig. 1: Block diagram of proposed system

B. Video

Each application that benefit from smart videos has different objects for tracking and thus requires different video sequence. Owing dynamic changes in natural video scenes such as sudden illumination and weather changes, repetitive motions that causes clutter such video are preferable motion detection and then tracking. The system is initialized for feeding videos. Most of the methods are able to work on both color and monochrome video imagery.

C. Frame Extraction

The first level of shot clustering based on feature extraction is the frame extraction in videos. As shown in below Fig. 2 is the steps for video input.

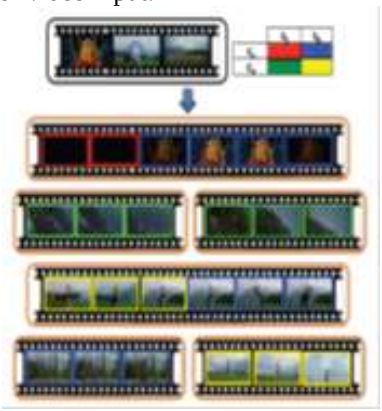

Fig. 2: Steps for video input to frame extraction

D. Feature Extraction Using Color-Texture Histogram

1) Color Feature

The use of joint color and texture features has a popular approach to color-texture analysis. The color dissimilarity measure called histogram intersection. In this distribution of image colors in RGB that reads truecolor images in RGB image type with their color quantization has value 8. The quantization method is used for obtaining three dimensional and one dimensional color distributions. As a dissimilarity measure, the histogram intersection method is utilized.

2) Texture Feature

The Local binary pattern operator labels the pixel in an image to thresholding its neighborhood with the center value and considering the result as a binary number that is binary pattern. The general version of the LBP operator is defined in equation 1.1.

$$\mathrm{LBP}_{P,R}(x_c, y_c) = \sum_{P=0}^{P-1} s(g_p - g_c)2^P \qquad (1.1)$$

- g_c- gray value of the center pixel
- x_c - center pixel
- y_c - local neighborhood
- g_p - gray values of P equally spaced pixels on a circle
- R - radius.

The function s(x) is defined in equation 1.2.

$$s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \qquad (1.2)$$

The texture model derived in this system using equation 3.1 has only gray-scale invariance. The grayscale and rotation invariant LBP texture model is obtained by the equation 1.3.

$$\mathrm{LBP}_{P,R}^{riu2} = \begin{cases} \sum_{P=0}^{P-1} s(g_p - g_c) & \text{if } U(\mathrm{LBP}_{P,R}) \leq 2 \\ P+1 & \text{otherwise} \end{cases} \qquad (1.3)$$

The superscript "riu2" is the rotation invariant "uniform" patterns have a 'U' value of at most 2. The P+1 "uniform" binary pattern occur in a circularly symmetric neighbor set of P pixels. So using this to extracted color-texture features joint them and convert into one dimensional vector this is joint color-texture histogram and apply to first frame

E. Shot Clustering

There are several types of histogram method but here use x2 is Chi-Square histogram. In this histogram method first of all frames are converted into RBG. Compute the histogram of k^{th} and $(k+1)^{th}$ frames for different three colors H_r, H_g, H_b,

$$D_f(k, k+1) = \sum_{i=1}^{3} \frac{[H(k,i) - H(k+1,i)]^2}{H(k,1)} \quad (1.4)$$

where

H_r, H_g and H_b - histogram of red, green, blue respectively.

E. Mean shift algorithm

In this system take mean shift tracking applying to all extracted frames. Features for tracking the object are color and texture. However using only color histograms in mean shift tracking has some problems. First is the spatial information of the target is lost. Second, when the target has similar appearance to the background, color histogram will become invalid to distinguish them.

In this system take iterative optimization process is initialized with the target location (y_0) in the previous frame. Using Taylor expansion around $\hat{p}_u(y_0)$ the linear approximation of the Bhattacharyya coefficient given in equation 1.5 is obtained as

$$\rho[\hat{p}(y), \hat{q}] = \frac{1}{2}\sum_{u=1}^{m}\sqrt{\hat{p}_u(y_0)}\,\hat{q}_u + \frac{1}{2} C_h \sum_{i=1}^{n_h} w_i k\left(\left\|\frac{y - x_i}{h}\right\|^2\right) \quad (1.5)$$

where

$$w_i = \sum_{u=1}^{m}\sqrt{\frac{\hat{q}_u}{\hat{p}_u(y_0)}}\,\delta[b(x_i) - u] \quad (1.6)$$

Since the first term in equation 1.5 is independent of y, to minimize the distance in equation 1.6 is to maximize the second term in equation 1.6. In the iterative process, the estimated target moves from y to a new position y1, which is defined

$$y1 = \frac{\sum_{i=1}^{n_h} x_i\, w_i g\left(\left\|\frac{y - x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n_h} w_i g\left(\left\|\frac{y - x_i}{h}\right\|^2\right)} \quad (1.7)$$

when we choose kernel g with the Epanechnikov profile is reduced to

$$y1 = \frac{\sum_{i=1}^{n_h} x_i\, w_i}{\sum_{i=1}^{n_h} w_i} \quad (1.8)$$

By using equation 1.8 the mean shift tracking algorithm finds in the new frame the most similar region to the object.

The modified thresholding method carried out and employs $LBP_{P,R}^{riu2}$ to describe the target texture features because of its low computational complexity. Calculate the LBP feature of each point in the image region, whose value is between 0 and 9. Thus an appearance model combining the color and texture is constructed and it consists of color channel and LBP texture pattern.

F. Sequence alignment

In order to find the points in the sequence of frame labels where the pattern of symbols changes and compare successive non-overlapping windows of shot labels using a sequence alignment algorithm. More specifically, given the set V of N frame, the subsequences of the original video sequence to be compare iteration i are formulated as in equation 1.9 and 1.10.

$$X_1^i = L_i L_{i+1} \dots L_{i+w-1} \quad (1.9)$$

$$X_2^i = L_{i+w} L_{i+w+1} \dots L_{i+2w-1} \quad (1.10)$$

$$i = 1, \dots, N - 2w$$

where

w - length of the window used and

L_i $i = 1, \dots, N$ - shot labels.

Using this final output of develop system to detection of object and tracking. The entire algorithm as joint color-texture histogram, mean shift tracking, LBP texture and at the end use sequence alignment. Using following algorithm to clear out develop system and shows results in next chapter.

F. Algorithm

1) Training input video file.
2) Extract frame from input video.
3) Initialization of feature parameters as color of value 8 and LBP threshold = 8.
4) Start frame of video frame00.
5) Apply color histogram and LBP texture method to frame using equation 3.3 and 3.4.
6) Apply mean shift algorithm with initial and other frames.
7) Normalize the tracking window with frames.
8) Make the tracking window to frame as rectangular parameters.
9) Apply sequence alignment algorithm and Using these two sequences compare color and texture feature after that align the frames.

Tracking result of detected object is display. The entire process of tracking the video based joint color-texture histogram is summarized in above algorithm.

IV. PERFORMANCE ANALYSIS

This section demonstrates the simulation results of proposed system obtain to compare the traditional methods. Experiment taken according to the characteristics of testing videos and compare the results, to observing result gives the comment. Performance criteria of both methods calculated as precision and recall, to show the result in graph.

A. Data

To evaluate the performance of object detection and tracking algorithm using video data set which contain high camera and object motion. All the videos in data set having different number of frames and objects for tracking which is given in table 1 as below.

Sr No.	Video	No. of frames	Color form
1	Ball	52	RGB
2	Faxpin	26	RGB
3	Lazysu	119	RGB
4	Table tennis	58	RGB
5	Torch alarm	89	RGB

Table 1: Data characteristics

B. Comparison

This implemented algorithm calculate the performance criteria to evaluate the performance of this method by using following criteria

$$\text{Recall} = \frac{N_c}{N_T} \quad (1.11)$$

$$\text{Precision} = \frac{N_c}{N_c + N_f} \quad (1.12)$$

where

N_c - number of correct tracking frames,
N_T - total number of frames and
N_f - the number of false tracking object.

C. Moderate Sequence

The first experiment is on moderate video sequence with 89 frames. In this video to track a moving torch button and compare the target locating accuracies and no. of track object for the two target representation model C1 and C2. As shown in Fig. 3 and 4 results of both models taking 2-3 frames.

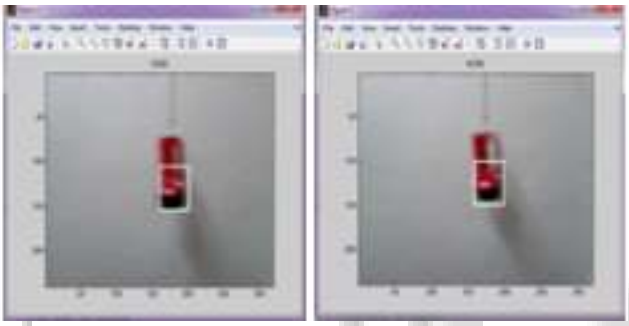

Fig. 3: Tracking results of video sequence in Torch alarm player of moving button by the target representation model C1. Frame 18, 40 are displayed.

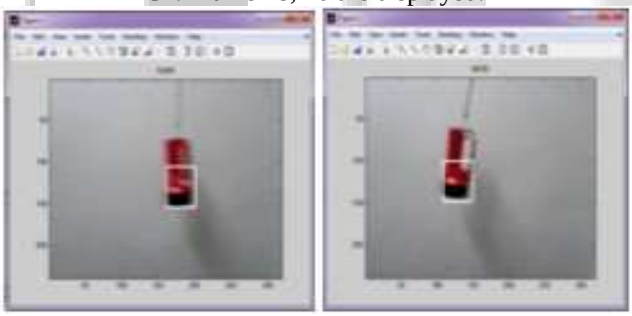

Fig. 4: Tracking results of video sequence in Torch alarm player of moving button by the target representation model C2. Frame 18, 40 are displayed.

In Fig. 3 and 4 illustrate the mean shift tracking results of two target representation models. The target location of mean shift tracking with C2 is not good. Especially, mean shift tracking with C2 even loses the object from correct target tracking. This is mainly because the proposed method only tracks the key feature points of candidate region.

D. Gray Video Sequence

The second experiment is on a video Car having gray sequence with 132 frames, where the objects to track moving car. In Fig. 5 and 6 shows the results of gray sequence video for tracking the object using C1 and C2 method.

Fig. 5: Tracking results of video sequence in Car of moving car by the target representation model C1. Frame 28, 75 are display.

Fig. 6: Tracking results of video sequence in Car of moving car by the target representation model C2. Frame 28, 75 are displayed.

In Fig. 5 and 6 illustrate the mean shift tracking results of two target representation models. The mean shift tracking result for C2 is same with above test video. After 15 frames due to interference from background C2 model loses the detected object from correct tracking. The gray sequence object tracking no matter the color of video but it gives good result as compared to above two experiments shown in graph.

E. Noise Added Video Sequence

The third experiment is to add the noise in any video given in Table 1 having no. of video sequences. So consider table tennis video sequence with 52 frames, where the objects to track playing ball and added gaussian noise is added and taken out the results. In Fig. 7 shows the results of noise added sequence for tracking the object using C1 and C2 method.

Fig. 7: Tracking results of video sequence in table tennis of tracking object is moving head using model C1 in added gaussian noise. Frame 18, 35 are displayed.

In Fig. 7 the noise added mean shift tracking results of this proposed system for target representation of C1 models. However the proposed model C1 extracts the main

target features while suppressing the background features although different noise added for same video sequence.

Video	C1		C2	
	1 – Precision	Recall	1 – Precision	Recall
Ball	0.08	0.92	0.24	0.76
Faxpin	0.2	0.80	0.54	0.46
Table tennis	0.06	0.94	0.05	0.95
Torch alarm	0	1	0.03	0.97

Table 4.2: Performance criteria of test videos

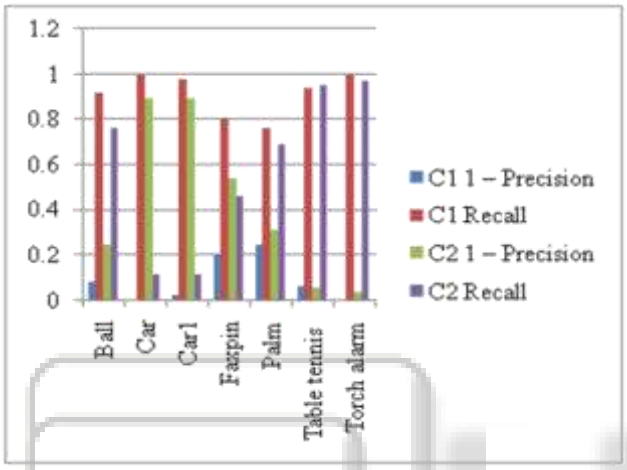

Fig. 8: Plot the Precision and Recall of different video sequence for implemented algorithm.

For performance criteria shown in above Fig. 8 the discussion can be done here as object detection and tracking achieving the high performance than original color method. This shows that the implemented method in this report satisfying the goal considerably.

As already said, video sequences having combination of low resolution, high resolution, noise added and nearly same with background. Precision and recall value are inversely and is good for proposed method. Hence here extends the future scope as method can be developed that achieves more accurate results.

V. CONCLUSION

The two main themes around which the paper presented are firstly, the benefits of using sensible and applicable videos to help the joint color-texture histogram process and secondly the ways in which object tracking can be improved by considering uncertainties in the underlying video frame parameters as part of the object tracking process. LBP operator is an effective tool to measure the spatial structure of local image texture. It reduces the computational cost and improves the robustness of target representation. The test result shows the different experiments on video frames and their quality obtained from object tracking algorithm is quite satisfactory and achieved target. The large number of frames and with high resolution in tested video there is more time for tracking depends on processor therefore computation speed will be decreases. In future multi object/person tracking is requiring using multitrack windows for number of objects and efficient methods for feature extraction.

REFERENCES

[1] Chen Xue-wen and Thomas Huang "Facial expression recognition: A clustering-based approach", Elsevier Science B.V. Pattern Recognition Letters 24, 1295–1302 May 2003.
[2] Dedeoglu Yigithan "Moving object detection, tracking and classification for smart video surveillance", a thesis submitted to the department of computer engineering and the institute of engineering and science of Bilkent University, August 2004.
[3] Haritaoglu Ismail and Myron Flickner "Detection and Tracking of Shopping Groups in Stores", IEEE 2001.
[4] Hariharakrishnan Karthik and Dan Schonfield "Fast Object Tracking Uing Adaptive Block Matching", IEEE Transactions on Multimedia Vol. 7. No. 5, Oct. 2005.
[5] Khan Zulfiqar Hasan, Student Member, IEEE, Irene Yu-Hua Gu, Senior Member and Andrew G. Backhouse, "Robust Visual Object Tracking Using Multi-Mode Anisotropic Mean Shift and Particle Filters" IEEE transactions on circuits and systems for video technology, vol. 21, no. 1, January 2011.
[6] Olson Clark F. "Pose Clustering Guided by Short Interpretation Trees" University of Washington, Bothell Computing and Software Systems.
[7] Porikli Fatih and Oncel Tuze "Object Tracking in Low-Frame-Rate Video", Mitsubishi Electric Research Laboratories, Inc., 2005 201 Broadway, Cambridge, Massachusetts 02139, TR2005-013 March 2005.
[8] Suneel A. Sai "Person or object tracking and velocity estimation in real time videos", Publications of problems & application in engineering research - paper 292 vol 04, special issue01; 2013.
[9] Superiori Luca and Olivia Nemethova "Clustering-based Object Detection for Low-resolution Video Streaming", IEEE INT. SYMP. On broadband mulitmedia systems and broadcasting, Orlando, USA, Mar. 2007.
[10] Veneau Emmanuel, R´emi Ronfard and IRISA/INRIA "From Video Shot Clustering to Sequence Segmentation" Institut National de l'Audiovisuel 4, avenue de l'Europe 94366 Bry-sur-Marne cedex, France.
[11] Vishnyakov Boris, Yuryn Viziltermand and Vladimir Knyaz, "Spectrum-based object detection and tracking technique for digital video surveillance", International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XXXIX-B3, 2012.

