Volume 4, Issue 4, April 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

New Trends in Object Motion Analysis

Prerna Rajput
M.Tech,Dept of CSE,
Graphic Era University
Dehradun, India

Vikas Tripathi
Assistant Professor, Dept of CSE ,
Graphic Era University
Dehradun, India

Abstract: This paper presents various algorithms for motion detection, object classification, object tracking and action recognition. It is an essential benchmark in the fields of video surveillance, traffic monitoring, border security etc. A novel approach is presented for detecting vehicles and pedestrians based on haar features. Haar feature value is calculated by using the rectangle integral, each rectangle area is multiplied by its weights and results are added together. Object detection using haar classifier provides robust and effective results.

Keywords- Motion detection, shadow elimination, haar classifier, object tracking, action recognition.

I. INTRODUCTION

Recently in computer vision, video processing is gaining more and more interest rather than single images. The aim of human activity recognition system is to develop a system that can automatically predetermine activities from recorded video sequences. Video surveillance area covers applications such as parking lots, banks, shopping malls where one or more subjects are being tracked and monitored for special action. It is one of the major advancement in the field of computer vision. However, it remains a challenging problem for computers to achieve robust action recognition due to blurred background, camera motion, self-occlusion, view point changes and shadowing. A blurred background introduces irrelevant information about region of interest, making harder to isolate. Camera motion introduces ambiguous motion patterns observed in the image plane. In addition human actions can also be observed only partially due to self-occlusions, thus actual region of interest can be reduced. The making of video surveillance software "smart" requires fast, robust and reliable algorithms for moving object detection, classification, tracking and activity recognition.

II. MOTION DETECTION

Motion detection in video surveillance system can be termed as the ability of the system to detect movement and capture it. It is be defined as a software-based monitoring algorithm that when detects motion will signal the camera mounted for surveillance to start capturing we can also call it activity detection. An advance motion detection surveillance system can analyse the type of motion to see if it warrants an alarm.

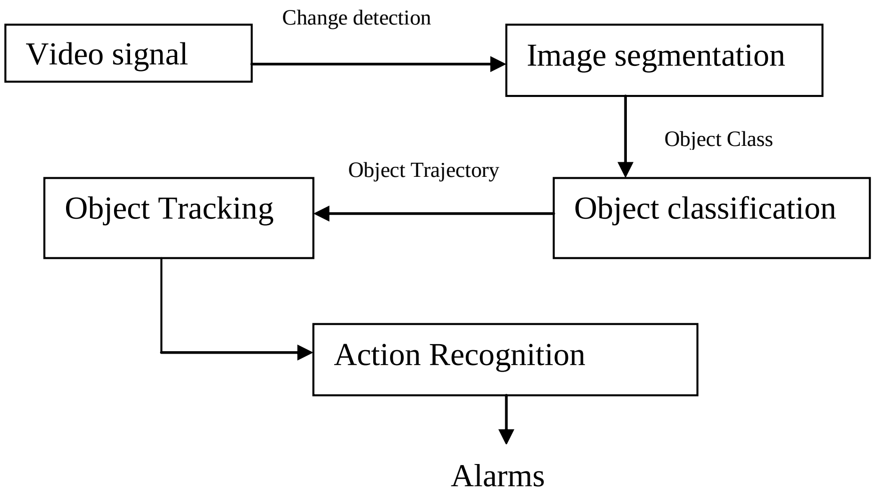

Fig 1 System Architecture

In other words, motion detection aims at segmenting regions in equivalence to moving objects such as vehicles, humans, birds, etc. This is the very first step of almost every automated visual surveillance applications. Due to dynamic changes in natural scenes such as weather changes, repetitive motion detection is a difficult problem to process reliably. Techniques used for moving object detection are background subtraction, statistical methods, temporal differencing and optical flow[14].

Background subtraction is a simplest used technique for motion segmentation for static background. It aims to detect moving regions by differencing the intensity of current image pixel by pixel from a reference background image. The pixels where the difference is above a threshold are classified as foreground. Once the reference is computed, then it will be updated with each new frame. But this technique is more sensitive to noise.

The simplest background model is median filtering approach[1], in which background estimate is computed as the median of all the pixel values stored in a buffer at each pixel location.

In [2] Taking the median value of the pixel color, which is far more robust than the mean over a sequence of images. The median, threshold values are then used to create the difference image. Let F^N_{jk} represent a series of N collected images and (j,k) the pixel location. The resulting background denoted as B_{jk} is calculated as follows.

$$B_{jk} = \text{Media}(F^1_{jk}, F^2_{jk} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots F^N_{jk})$$

Once the background primal sketch and threshold data are constructed, the difference image is taken from a scene, the pixels that form the difference image are referred as outliers [3], denoted as D_{jk} is calculated as:

$$D_{jk} = \begin{cases} 1, & if\ |F^i_{jk} - B_{jk}| > T_{jk} \\ 0, & otherwise \end{cases}$$

Background subtraction method inspired statistical methods for dynamically updating statistics of the pixels that correspond to the background image processing. Statistical methods generally use dense based Representation, which means require prior knowledge, works well with multiple objects and able to handle occlusions and temporal stopping. Statistical methods that make use of the features of individual pixels or group of pixels to create more advanced background models. Foreground pixels are classified by comparing each pixel's statistic of the current background model.

The w4[4] presented a statistical background model where each pixel is classified by its minimum (M) and maximum (N) intensity values and maximum intensity difference (D) between consecutive frames obtained during initial period when no moving objects present in the scene. A pixel in the current image is identified as foreground if it satisfies:

$$|M(x,y) - I_t(x,y)| > D(x,y) \text{ or } |N(x,y) - I_t(x,y)| > D(x,y)$$

Pfinder[5] presented a statistical model in which subject was represented by numerous blobs with individual shape and color statistics.

The temporal differencing approach attempts to detect moving regions by making use of the pixel wise differences of consecutive frames(two or three) in an image sequence. Temporal differencing is highly adaptive to dynamic environments, however, it fails to extract entire relevant feature pixels.

Lipton et al.[6] described a two-frame differencing method where tracking the absolute difference between the current and the previous frame and a threshold function was used to identify the change. The pixels that satisfy the following equation are referred as foreground.

$$|I_t(x,y) - I_{t-1}(x,y)| > r$$

Optical flow based motion segmentation uses the flow vectors of moving objects to detect moving region in an image over time[7]. This approach is basically used to represent coherent motion of points between image sequences. They can detect moving objects in video sequences even from a moving camera however, flow computation methods are complex and hardware dependent[7].

Bregler[8] described, each pixel was represented by its optical flow. These connected flow vectors were grouped to form blobs having articulated motion and characterized by a mixture of multivariate Gaussians.

VSAM[9] has developed a hybrid algorithm which combines an adaptive background subtraction algorithm with three-frame differencing technique for motion segmentation. This hybrid approach is fast and robust for detecting motion in image sequences.

After image segmentation shadow detection is equally important for obtaining relevant information. Generally shadow is processed by using shadow based model c1c2c3[10] to detect and eliminate the shadow. This model is defined as:

$$\{c1, c2, c3\} = \left\{ arctan\left[\frac{R}{\max(G,B)}\right], arctan\left[\frac{G}{\max(R,B)}\right], arctan\left[\frac{B}{\max(R,G)}\right] \right\}$$

Where c1c2c3 model is dependent on sensor and surface reflection coefficient. Thus the c1c2c3 value is same at same point of background in shadow as well as in light conditions. The steps of elimination by 3C models are as follows:

(a) Calculate the c1c2c3 model of foreground detected by the i^{th} frame , as c1(x,y,i), c2(x,y,i), c3(x,y,i).

(b) Calculate the c1c2c3 model of background frame recording as c1(x,y,b),c2(x,y,b),c3(x,y,b).

(c) In the neighborhood window W(x,y,i) whose center is $v_i(x,y)$ detect the change of c1c2c3 at the same pixel between foreground and background. Set the size of window to q=(2N+1)(2M+1),and height and width of window is(2N+1) and (2M+1).

$$\Delta c1(x,y,i) = \frac{l}{q}\sum_{n=-N}^{N}\sum_{m=-M}^{M}|c1(x+n,y+m,i) - c1(x+n,y+m,b)|, l = 1,2,3$$

If the value is less than the threshold, the pixel is classified to shadow, otherwise it belong to foreground. Set the shadow region to black.

III. OBJECT CLASSIFICATION

Moving objects detected in video corresponding to different objects in real word such as humans, vehicles etc. it is very necessary to identified the type of a detected objects in order to track and analyse activities. Mainly there are two approaches for object classification[7] shape based and motion based.

3.1 Shape-Based Classification

Different description used in shape-based classification scheme are the bounding box, area, points, silhouette, blobs and gradient are available for classifying moving regions.
VSAM[9] makes use of the object's silhouette contour, length, area to classify moving objects into four groups: human, human group, vehicle and clutter using neural network classifier. Dispersedness is used as the classification metric and it is defined in terms of objects area and contour length as follows

$$\text{Dispersedness} = \frac{perimeter^2}{area}$$

Brodsky et al.[11] uses a Redial Basis Function (RBF) classifier which has a similar architecture like a three layer back propagation network. The normalized gradient image of the detected object region is the input to the classifier.

3.2 Motion Based Classification

The method presented in [12] is based on the temporal self similarity of the moving object. As an object that exhibits periodic motion. This has been used as a strong cue for distinguishing moving objects.
A.J.Lipton presented a method that makes use of local optical optical flow analysed the detected object regions[13]. Residual flow is used to analysed rigidity & periodicity of moving objects such as vehicle would present little residual flow whereas non-rigid such as human would present high average residual flow by using this cue, human motion can be distinguished from other objects.
In general, a person's position or size scale on the image is unknown. The system thus needs to evaluate the entire image in different scales. There are multi scale descriptors, such as the Haar-like features, or the Histograms of Orientations (HOG) in first case, or by cascades of boosted classifiers in the second. Low computational cost method was given by Viola and Jones[14] detecting objects in images, this is based on four key concepts:

a) Haar features i.e rectangular features
b) An integral image which is used for rapid feature detection
c) The Ada-boost machine learning method which combines many weak classifiers to develop one strong classifier
d) A cascaded classifier clubs many features effectually.

A Haar Classifier [15] is used for object detection can also detect various types of objects such as vehicles, humanbeing, flying objects etc. It is an approach which is based on machine Learning where a cascade function is trained from a lots of positive and negative images and then used to detect objects in other images. Then they extract features from it, Each feature has a single value obtained by differencing sum of pixels under white rectangle from sum of pixels under black rectangle. For each feature calculation, need to find under white and black rectangle. In this algorithm, first load the required classifier then load the input video/image, convert it into gray scale mode, now find the objects in the image. If the object are found, it returns the positions of the detected objects as Rect(x,y,w,h).

Fig 2 Object Detection

IV. OBJECT TRACKING

The goal of tracking is to establish correspondence of tracking is to establish correspondence of objects & object parts between consecutive frames of video tracking has been a difficult task to apply in crowded places due to inaccurate segmentation of objects. Common problems are long shadows, partial &full occlusion of objects with each other & with stationary items in the scene. Tracking over time involves matching objects in consecutive frames by using features like

lines, rectangle, blobs, etc. There are two common approaches in tracking objects one is model based &other based on correspondence matching.
In general, geometric structure of human body can be represented as stick figure which is obtained by means of axis transfer or distance transform, 2-D contour, in such description, human body sentiments all analogous to 2-D ribbons or blobs, volumetric models in which to depict the human body structure 3-D models such as elliptical cylinders, cones sphere etc are used[16].
Struffer etal[17] describes a linearly predictive multiple hypothesis tracking algorithm. The algorithm uses size & positions of objects for maintaining a set of kalman filters for motion estimation. Additionally Kalman filters are also used for occlusion handing & trajectory prediction.

V. ACTION RECOGNITION

Action Recognition is a very important area in computer vision and other fields. In Action recognition is defined as "looking at people" is to analyse and interpret human motion.
An efficient real-time approach for representing human motion using a compact "Motion History Image". MHI is basically used for action recognition, motion analysis. *Md Atiqur Rahman Ahad, J K Tan, H Kim S Ishikawa* in their paper[18] mentioned that motion history image approach is a view based temporal method which is robust for action recognition ,motion analysis. They mentioned that the main limitations of MHI were due to large variations, self-occlusion due to motion, motion overlapping or multiple repetitions. These problems are still unsolved.
James W Davis in their paper [19] described a real time computer vision approach for recognizing human movements based on motion pattern. In their orientation histogram approach to become "motion orientation" histograms, where the directions of motion are accumulated in a histogram format and used for recognition. These histograms separate and localize regions of motion for better description of the movement. Quantitative results show that the method can easily discriminate between different human movements and is extendible to variable length motion.
Gary R Bradski and James W Davis[20] presented a fast and simple method for representing motion from gradient using timed motion history image in successively layered silhouettes. They also presented a novel method of motion segmentation based on segmenting layered motion region that are meaningful connected to movements of the object of interest, together with silhouette pose recognition provides a useful tool for gesture and motion recognition.
Abdunnaser Diaf,Rachid Benlamri ,in their paper [21]described a new method that is Motion Intensity Image for representing and recognizing human motion in videos. This is basically a appearance based template matching approach which is simple, concise and expressive. It aligns the human silhouette of each background-subtracted binary image to a reference point and then forms a single intensity image of these silhouettes. An eigenvector-based linear dimension reduction technique has been used for matching temporal templates against stored instances of views of known motions. This method shows a high level of performance.

VI. CONCLUSIONS

In this paper various algorithms for motion detection, object classification, object tracking and action recognition have been discussed, which strongly demonstrated that video surveillance system deal with complex human movements. Median filtering algorithm would be suitable for monitoring an indoor scene environment, however, same algorithm is not proper choice for outdoor scene environment. Even though there have been many advancement in computer vision yet some challenges still needs to be worked upon. Some of these challenges are physical placement of cameras, installation cost, unfavourable weather and lighting condition etc.

REFERENCES

[1] R. Cutler and L. Davis, "*View-based detection*," in Proceedings Fourteenth International Conference on Pattern Recognition, 1998.
[2] S.Arseneau, and J.R Cooperstock, "*Real-Time Image Segmentation for Action Recognition*," in IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM'99),1999.
[3] Y.H. Yang, M.D. Levine, "*The background primal sketch: an approach for tracking moving objects*," Machine Vision Application,1992.
[4] I. Haritaoglu, D. Harwood, and L.S. Davis, "*W4: A real time system for detecting and tracking people*," In Computer Vision and Pattern Recognition, 1998.
[5] C.R. Wren, A. Azarbayejani, T. Darrell, A.P. Pentland, "*Pfinder: real-time tracking of the human body*," IEEE Trans. Pattern Anal. Mach. Intell.,1997.
[6] A. J. Lipton, H. Fujiyoshi, and R.S. Patil, "*Moving target classification and tracking from real-time video*," In Proc. of Workshop Applications of Computer Vision, 1998.
[7] L. Wang, W. Hu, and T. Tan, "*Recent developments in human motion analysis*," in Pattern Recognition, 2003.
[8] C. Bregler,"*Learning and recognizing human dynamics in video sequences*," Proceedings of the IEEE CS Conference on Computer Vision and Pattern Recognition,1997.
[9] R. T. Collins, A. J. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y.Tsin, D. Tolliver, N. Enomoto, O. Hasegawa, P. Burt, and L.Wixson, "*A system for video surveillance and monitoring*," Carnegie Mellon Univ.,Pittsburgh, PA, Tech. Rep., CMU-RI-TR-00-12, 2000.
[10] J. cao, D. Zhang*Dong XuFang Wang, "*Background subtraction based human region extraction method,*" 3rd International Conference on multimedia technology (ICMT),2013.

[11] T. Brodsky et al., "*Visual Surveillance in Retail Stores and in the Home*,"Video-Based Surveillance Systems. Kluwer Academic Publishers, Boston, 2002.

[12] R. Cutler and L.S. Davis, "*Robust real-time periodic motion detection, analysis and applications*," In IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000.

[13] A. J. Lipton., "*Local application of optic flow to analyze rigid versus non-rigid Motion*," Technical Report CMU-RI-TR-99-13, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, December 1999.

[14] P.Viola, M.Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," Conference on Computer Vision and Pattern Recognition (CVPR), 2001.

[15] P.I.Wilson, J.Fernandez,*"Facial Feature Detection Using Haar Classifiers," in* JCSC21,2006.

[16] W. Hu, T. Tan, L.Wang, and S. Maybank, "*A survey on visual surveillance of object motion and behaviors*," in IEEE Transaction on Systems, Man and Cybernetics—Part C: Applications and Reviews, 2004

[17] C. Stauffer and W.E.L. Grimson, "*Learning patterns of activity using real-time tracking*," IEEE Pattern Recognition and Machine Intelligence, 2000.

[18] Md. A.R. Ahad , J. K. Tan , H. Kim , S. Ishikawa, *"Motion history image: its variants and applications*," Springer in Machine Vision and Applications,2012.

[19] J. W. Davis,"*Recognizing movement using motion histogram,"*M.I.T Media Laboratory Perceptual Computing , 1998.

[20] G.R. Bradski and J. W. Davis, "*Motion Segmentation and pose recognition with motion history gradients*," Springer in Machine Vision Applications,2002.

[21] A. Diaf, R.Benlamri, "*A novel eigenspace-based method for human action recognition*," International Conference on Digital Information Management ICDIM,2010.

