

Development of Human Tracking System For Video Surveillance

Debmalya Sinha#1, Gautam Sanyal#2

#Department of Computer Science and Engineering, National Institute of Technology,
Durgapur, India.

¹debmalya.nit@gmail.com

²nitgsanyal@gmail.com

Abstract

Visual surveillance in dynamic scenes, especially for human and some objects is one of the most active research areas. An attempt has been made to this issue in this work. It has wide spectrum of promising application including human identification to detect the suspicious behavior, crowd flux statistics, and congestion analysis using multiple cameras.

In this paper deals with the problem of detecting and tracking multiple moving people in a static background. Detection of foreground object is done by background subtraction. Detected objects are identified and analyzed through different blobs. Then tracking is performed by matching corresponding features of blob. An algorithm has been developed in this perspective using Angular Deviation of Center of Gravity (ADCG), which gives a satisfying result for segmentation of human object.

Keywords

Tracking, Visual Surveillance, Blob, Center of Gravity (CG) and Feature Extraction.

1. INTRODUCTION

As an active research topic in Computer Vision, Visual Surveillance in dynamic scenes attempt to detect, recognize and track certain object from image sequences and more generally to understand the human or any object behavior. The aim of this research is to develop an intelligent surveillance system for tracking human in dynamic scenes. It has wide range of potential applications such as security issue in important installation, traffic Surveillance in expressways, to measure the crowd flux in railway station, airports etc.

In surveillance system considerable amount of work has been carried out by researchers [1]. Technology has reached a stage where video camera may be affordable in public and private areas [2] for keeping track of movement of any human or object. This paper has presented a vision-based system for accurate segmentation and tracking of moving objects in cluttered and dynamic outdoor environments, surveyed by a single fixed camera. Each foreground Object of Interest (OI) has been segmented and shadows/highlights removed.

The video surveillance system usually has two major components, one is detecting moving object the other one is to tracking them in sequence from video images. The accuracy of these components largely affects the accuracy of overall surveillance system. Detecting moving regions in the scene and separating them from background image is a challenging problem. In the real world, some of the challenges associated with foreground object segmentation are illumination changes, shadows, camouflage in color, dynamic background and foreground aperture [3]. Foreground object segmentation can be done by three basic approaches: frame differencing, background subtraction and optical flow. Frame differencing technique does not require any knowledge about background and is very adaptive to dynamic environments [4], but may suffers from the problem of foreground aperture due to homogeneous color of moving object. Background subtraction can extract all moving pixels, but it requires perfect modeling. It is extremely sensitive to scene changes due to lighting and movement of background object. Optical flow, one of the robust technique to detect all moving objects, even in the presence of camera motion, but it may be computationally expensive and may have limited application.

Object can be represented as,

Points: The object is represented by a point, that is the centroid shown in (Figure 1(a)) In general, the point representation is suitable for tracking objects that occupy small regions in an image.

Primitive geometric shapes: Object shape may be represented by a rectangle, ellipse, shown in (Figure 1(c), (d), etc. Though the primitive geometric shapes are more suitable for representing simple rigid objects. However they may also be used to represent nonrigid objects.

Object silhouette and contour: Contour representation may be used to define the boundary of an object (Figure 1(g), (h)). The region inside the contour is called the silhouette of the object (see Figure 1(i)). Silhouette and contour representations are suitable for tracking complex nonrigid shapes or objects.

Articulated shape models: Articulated objects are composed of body parts with different joints. For example, the human body is an articulated object with torso, legs, hands, head, and feet connected by joints. In order to represent an articulated object, one can model that constituent part by integrating different graphical shape like cylinders or ellipses as shown in Figure 1(e).

Skeletal models: Object skeleton can be extracted by applying medial axis transform to the object silhouette. This model is commonly used as a shape representation for recognizing objects. Skeleton representation can be used to model both articulated and rigid objects (Figure 1(f)).

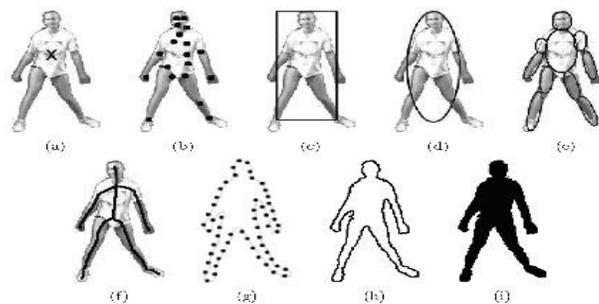


Fig. 1: Object Representation

It is difficult to get a background model from the video because background information keeps always changing by different factors like illumination, shadows etc. The static background is considered for analyzing the object in this paper. Background subtraction method is used for detecting moving object, because it gives maximum number of moving pixels in a frame.

Object tracking methods usually divided into four groups [9], they are:

1. Region-based tracking
2. Active-contour-based tracking
3. Feature-based tracking
4. Model-based tracking

It is not so easy because of some of the problems, which generally occur during tracking. Occlusion handling problem i.e. overlapping of moving blobs has to be dealt carefully. Other problems like lighting condition, shaking camera, shadow detection, similarity of people in shape, color and size also pose a great challenge to efficient tracking.

Giovani Garibotto, Carlo Cibeï [4] proposed a new solution for 3D scene analysis in Security and Surveillance applications. It is based on binocular stereovision using prediction verification paradigm. Adaptive change in motion detection is performed to detect moving object in the scene. There are many reviews on image segmentation: Pal and Pal [5], which does not go details into the algorithms, but which classifies segmentation technique, discuss advantages and disadvantages of each class of the segmentation method and contain exhaustive list references to the literature up to the early 1990's.

Object Segmentation

Most of the work on foreground object segmentation is based on three basic methods, namely frame differencing, background subtraction and optical flow. Only background subtraction requires modeling of background. It is faster than other methods and can extract maximum features pixels. In [4], Collins et al. used a hybrid of frame differencing and background subtraction for effective foreground segmentation. Researchers usually use Gaussian [7], a mixture of Gaussian [8], kernel density function [6] or temporal median filtering techniques for modeling background [9]. Assuming that surveillance is taken at the scenario, which is Static background. Object extraction i.e. foreground segmentation is done by background Subtraction [5]. Building a representation of the scene called the background model and then finding deviations from the model for each incoming frame can achieve object detection. Any significant change in an image region from the background model signifies a moving object. Usually, a connected component algorithm is applied to obtain connected regions corresponding to the objects. This process is referred to as the background subtraction.

An alternate approach [5] for background subtraction is to represent the intensity variations of a pixel in an image sequence as discrete states corresponding to the events in the environment. For instance, for tracking cars on a highway, image pixels can be in the background state, the foreground (car) state, or the shadow state. In the context of detecting light on and off events in a room, Stenger et al. [2001] use HMMs for the background subtraction. They have reported that advantage of using HMMs in certain events, which are hard to model correctly using unsupervised background modeling approaches, can be learned using training samples.

1.1 Tracking

A feature-based object-tracking algorithm requires useful feature selection, feature extraction, feature matching and proper handling of object's appearance and disappearance. Object Entry and Exit in a scene was proposed by Stauffer [10]. Most of the works on tracking use a prediction on features in the next frame and compare the predicted value with estimated value to update the model. Usually a model like Kalman filter [2] is used for prediction. Techniques like Euclidean distance function [2] successfully used by Xu, Collins et al. used a correlation function for matching regions in motion. Comaniciu [14] proposed a mean-shift technique to calculate most probable target position. They calculated similarity of objects by constructing histograms of target model and target candidates. Similarity is expressed by a metric derived from the Bhattacharyya coefficient.

1.2 Tracking System

Our surveillance activity goes through three phases. In first phase the target is detected in each video frame. Segmentation using background subtraction is generally used to identify any moving object in the scene, but some time due to some environmental factors such as light condition, camera position detected object are splitted into more than one blob. While acquiring target proposed methodology namely Angular Deviation of Center of Gravity (ADCG) which could be useful to combining the splitted blob and grouped into a single object.

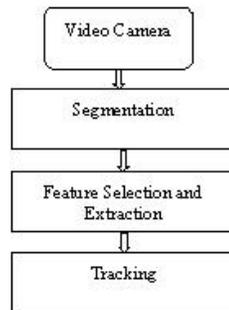


Fig. 2: Block Diagram of the Tracking System

In second phase, feature extraction is done for matching. The blob feature such as, Center of Gravity, size of the blob is extracted and used for tracking the people.

Lastly in third phase, the detected target is tracked through a sequence of video frames using the blob feature.

2. SOLUTION METHODOLOGIES AND MATCHING

The coherent pixels are grouped together as image blob by region growing approach, using any seeded pixel [12]. The approach used in this paper is similar to their region growing approach, but different in terms of number of regions and selecting seed pixel. We try to grow one region at a time until all connected neighboring pixels are taken into the account and then start growing another region

2.1 Blob extraction:

is the main procedure for shape definition. Most of the existing blob extraction algorithms such as are based on binary images. This indicate that neighborhood has to be predefined, but this is not possible in many practical application where gray scales images are utilized.

2.2 Size of blob:

Region-growing approach is used to select the image blobs as $B_1, B_2, B_3 \dots B_n$. Using coordinate geometry we can get the dimension of each blob i.e. length and width of each blob, which will be used to calculate the area of each blob. Size of the blob is represented as total number of pixels in the blob. Dimension of blob B_1 can be given as w_1 and l_1 . So, the area of blob B_1 is $Area_{B1} = w_1 \times l_1$. Similarly, Area of all detected blobs can be calculated as $Area_{B2} \dots Area_{Bn}$. Arranging the area of the blobs in decreasing order we will identify the two sets of blobs. Blobs with larger area size are been taken as human object and smaller blobs are discarded for any further processing.

2.3 Coordinate of center of blob:

Center of Gravity (CG) can be determined easily of a given blob, as the blob always maintains a regular geometrical shape.

2.4 Angular Deviation of blob:

Blob is segment through connected component labeling. Global threshold is used to transform the gray scale image into a binary image [15]. The regions are identified more accurately from binary image. But human body parts may not be identified as a integrated body due to factor such as, color of the dress may be identical with the background, which in terms may lead to differentiation problem of identifying whole body part of human being. For e.g., head may look like separated from main body, if the person is wearing some clothes in the neck area, which is exactly matching with the background color of the image frame. We propose a methodology, explaining how the deviation of individual regions with respect to CG can be used to merge the blobs, assuming CG of the blob will not change its location. There may be possibility that all the three blobs, we considered may not be related to same human being. This similarity feature will be used to merge the blobs.

In figure 3, ABCD, EFGH, IJKL describes three blob namely blob1, blob2 and blob3 at certain instance of time of certain image frame. These three blobs individually identifying three individual image segments in the image frame. In the video sequences the blobs position may get changed due to the body movement, due to the movement of any other moving object in the frame. The blob position changes to $A'B'C'D'$, $E'F'G'H'$ and $I'J'K'L'$. These position changes of the blobs occur due to the movement of the object. 'O', 'R' and 'U' are the CG of the blob1, blob2 and blob3 respectively, which remain invariable independent of blob position. A vertical line PQ is drawn, perpendicular to horizontal line AB or CD of blob1, which crosses through the CG 'O' for blob1 (fig 3(a)).

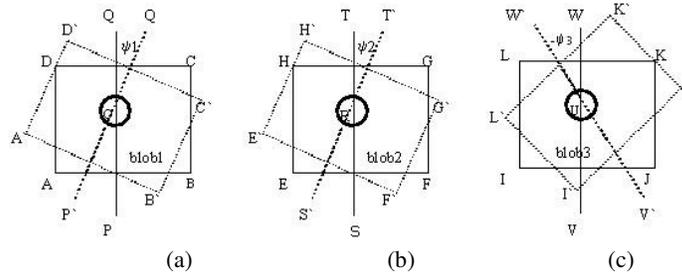


Fig 3: Angular deviation of CG

After the movement of Human or Object, blob1 position changes to A`B`C`D`. Another vertical line P`Q`, which is perpendicular to A`B` or C`D` is drawn through CG of ‘O’ of blob1, The angle between the lines PQ and P`Q` is calculated say ψ_1 and that will be is the Angular Deviation of CG for blob1.

For a particular image frame, blob1, blob2 and blob3 are having angle of deviation with respect to CG’s are ψ_1 , ψ_2 and $-\psi_3$ respectively. For blob3 the angle of deviation is ‘-’ve ($-\psi_3$), as the motion is opposite to the motion of previous two blobs.

BLOB	Angular Deviation1	Angular Deviation2	Angular Deviation3
Blob1	ψ_1	τ_1	φ_1
Blob2	ψ_2	τ_1	φ_1
Blob3	$-\psi_3$	τ_3	$-\varphi_3$

Table 1: blobs having angular deviation

Table1 shows individual blob’s angular deviation with respect to CG due to the movement of the object and its corresponding blob for three iteration. The difference between blob1 and blob2 angular deviation, termed as $(\zeta_1)= \psi_1 -\psi_2 \cong 0$. Angular deviation of blob3 with respect to blob1 and blob2 are $(\zeta_2)= \psi_1 - (-\psi_3)$ and $(\zeta_3)= \psi_2 - (-\psi_3)$. It clearly shows that $\zeta_1 < \{\zeta_2, \zeta_3\}$ because according to our assumption ψ_1 and ψ_2 are having equal or nearly equal value. There is a high probability that, blob1 and blob2 are part of same human being as they are having similar body movement, which result to the similar movement to their corresponding blob. Based on the feature, similar Angular Deviation of CG of blob3 is to be discarded as part of blob1 and blob2.

All features of a particular blob are stored in their respective feature vectors. Considering significant features, the blobs are tacked using euclidean distance. The distance between CG’s of blob of two consecutive frame is considered. Tracking is performed by matching features of blobs in current frame with the features of the blobs in previous frame. The difference between the feature vectors of each blob in current frame with each of previous frame is calculated. We do an exhaustive matching among N blobs in the current frame with M blobs in the previous frames, so a total of $N \times M$ matching is required. As, small numbers of objects are considered in the scene, this exhaustive matching is not time consuming. This difference is obtained by using Euclidian distance given by

$$\text{Dist}=\sqrt{\sum(f1-f2)} \text{ -----(1)}$$

Minimum distance between two blobs is selected and remaining are discarded. This process is continued for complete video and thus tracking of multiple people is achieved.

3. COMPUTER ALGORITHM:

Steps:

- 1 A background image with no moving object is taken
- 2 Background image is modeled to cope with noisy environment
- 3 Median filter is applied remove noise from image
- 4 Background subtraction is done pixel by pixel in current frame with background frame to get foreground object as

$$\text{IMG}=\text{current_img}(x,y)-\text{background_img}(x,y)$$
- 5 For calculation of features, image blob is obtained on detected object with the help of matlab
- 6 For each blob a feature vector is calculated which consists of
 - 1 size of blob
 - 2 center coordinates of blob
 - 3 average color of blob
- 7 Now matching of blob is done in sequence of frames for tracking by calculating Euclidian distance between each pair of blob's feature as

$$\text{Dist}=\sqrt{\sum(F(I,k)-F(j,k))}$$
- 8 The blob pair with minimum distance is considered as tracked pair and other pairs are discarded for that corresponding blob
- 9 Finally, trajectory of tracked blob is plotted in a graph

4. EXPERIMENTAL RESULT

In this case the experiment is performed in the indoor environment. Some random frames are taken from video, which having no moving object in frames (Fig 4a)). Then median filter is applied to eliminate the noise in image. Moving object or foreground object is calculated by subtracting background image from current frame (Fig 4b)) where moving object is present. Image subtraction is done by pixel wise. For the ease of processing,

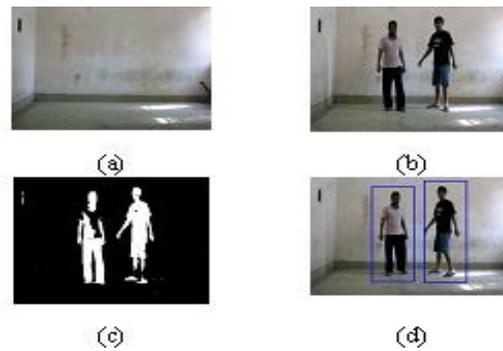


Fig 4: (a) Background image, (b) Current frame, (c) Binary image and (d) Detected blobs

subtracted image is converted to binary image (Fig 4c)). After finding the moving regions, the noise is removed by morphological operations such as erosion and dilation applied.

5. CONCLUSION

In this paper, we have presented methods of segmentation of foreground object by background subtraction and tracking of multiple people in indoor environment. The background subtraction method is selected, because it gives maximum number of moving pixels. Angular Deviation of CG gives the related blobs, which is used as single object for tracking. We used feature based tracking, as it is faster than other methods. Future work can also be done on finding good threshold value during matching for creating new object hypothesis and minimum size of blobs. Tracking can be done on individual body parts like head, hands, legs etc for higher-level analysis of human activity.

REFERENCES

- [1] Hu, W.; Tan, T.; Wang, L.; Maybank, S.; "A survey on visual surveillance of object motion and behaviors", *Systems, Man and Cybernetics, Part C, Volume 34, Issue 3, Aug. 2004.*
- [2] Xu, L.; Landabaso, J. L.; Lei, B.; "Segmentation and tracking of multiple moving objects for intelligent video analysis", *BT Technology Journal, Vol 22, No 3, July 2004*
- [3] Toyama, K.; Krumm, J.; Brumitt, B.; Meyers, B.; "Wallflower: principles and practice of background maintenance", *7th IEEE International Conference on Computer Vision, Volume 1, 20-27 Sept. 1999 Page(s):255 - 261*
- [4] Giovani Garibotto, Carla Cibeil, "3D Scene analysis by Real-Time Stereovision", *IEEE, 2005.*
- [5] N.R.Pal and S.K.Pal. A review on image segmentation techniques. *Pattern Recognition, 26(9):1277-1294, 1993.*
- [6] Elgamal A.; Duraiswami R.; Harwood D. and Davis L.; "Background and foreground modelling using nonparametric kernel density estimation for visual surveillance", *Proc of the IEEE, 90, No 7 (July 2002).*
- [7] McKenna, S. J.; Jabri, S.; Duric, Z.; Rosenfeld, A.; Wechsler, H.; "Tracking groups of people", *Computer Vision and Image Understanding, 80, pp 42—56 (2000)*

- [8] Stauffer, C.; Grimson, W. E. L.; “Adaptive background mixture models for real-time tracking”, Proceedings of CVPR, Jun 1999, pp. 246-252.
- [9] Zhou, Q.; Aggarwal, J. K.; “Tracking and classifying moving objects from video”, Proc of 2nd IEEE Intl Workshop on Performance Evaluation of Tracking and Surveillance (PETS’2001), Kauai, Hawaii, USA (December 2001).
- [10] Stauffer C.; “Estimating tracking sources and sinks”, Proc of 2nd IEEE Workshop on Event Mining (in conjunction with CVPR’2003), 4, Madison, Wisconsin (June 2003).
- [11] Comaniciu, D.; Ramesh, V.; Meer, P.; “Real-time tracking of non-rigid objects using mean shift”, Computer Vision and Pattern Recognition, 2000
- [12] Adams, R.; Bischof, L.; “Seeded region growing”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 16, Issue 6, June 1994 Page(s):641 – 647
- [13] Xu, M.; Ellis, T. J.; “Partial observation vs. blind tracking through occlusion”, In Proc of BMVC’2002, Cardiff, pp 777—786 (September 2002).
- [14] Dorin Comaniciu, Peter Meer, “Mean Shift: A Robust Approach Toward Feature Space Analysis” IEEE transaction(2002).
- [15] Bizhong Wei, Ning Ouyang, YueLin Chen, Xiaodong Cai, “Automatic Color Blob Segmentation and Fast Arbitrary Shape Tracking”, Institution of Engineering and Technology (2008)

ABOUT THE AUTHORS



Debmalya Sinha received his B.Tech degree in Information Technology from Haldia Institute of Technology, Haldia, India and ME in Software Engineering from Jadavpur University, Kolkata. He has more than 3 years of Industrial experience in the area of software development and research. Presently, he is pursuing his PhD as an Institute Research Scholar at the National Institute of Technology, Durgapur, India.



Gautam Sanyal has received his B.E and M.Tech degree National Institute of Technology (NIT), Durgapur, India. He has received Ph.D (Engg.) from Jadavpur University, Kolkata, India, in the area of Robot Vision. He possesses an experience of more than 25 years in the field of teaching and research. He has published nearly 50 papers in International and National Journals / Conferences. Two Ph.Ds (Engg) have already been awarded under his guidance. At present he is guiding six Ph.Ds scholars in the field of Steganography, Cellular Network, High Performance Computing and Computer Vision. He has guided over 10 PG and 100 UG thesis. His research interests include Natural Language Processing, Stochastic modeling of network traffic, High Performance Computing, Computer Vision. He is presently working as a Professor in the department of Computer Science and Engineering and also holding the post of Dean (Students’ Welfare) at National Institute of Technology, Durgapur, India.