International Journal of Soft Computing and Engineering (IJSCE)
ISSN: 2231-2307, Volume-3, Issue-2, May 2013

Tracking Multiple Moving Objects Using Gaussian Mixture Model

D. Hari Hara Santosh, P. Venkatesh, P. Poornesh, L. Narayana Rao, N. Arun Kumar

Abstract— The advance of technology makes video acquisition devices better and less costly, thereby increasing the number of applications that can effectively utilize digital video. Compared to still images, video sequences provide more information about how objects and scenarios change over time. For object recognition, navigation systems and surveillance systems, object tracking is an indispensable first-step. The conventional approach to object tracking is based on the difference between the current image and the background image. The algorithms based on the difference image are useful in extracting the moving objects from the image and track them in consecutive frames. The proposed algorithm, consisting of three stages i.e. color extraction, foreground detection using Gaussian Mixture Model and object tracking using Blob Analysis. Initially color extraction is done to extract the required color from a particular picture frame, after color extraction the moving objects present in the foreground are detected using Gaussian Mixture Model and Blob Analysis is applied on consecutive frames of video sequence, so as to observe the motion of the object, hence the moving object in the video sequences will be tracked.

Index Terms— gaussian mixture model, multiple object tracking blob analysis, background subtraction, foreground detection

I. INTRODUCTION

The moving object tracking in video pictures has attracted a great deal of interest in computer vision. Object tracking is the first step in surveillance systems, navigation systems and object recognition. There is a huge significance of object tracking in real time environment as it enables several important applications such as to provide better sense of security using visual information, Security and surveillance to recognize people, in Medical therapy to improve the quality of life for physical therapy patients and disabled people, to analyze shopping behavior of customers in retail space instrumentation to enhance building and environment design, video abstraction to obtain automatic annotation of videos, to generate object based summaries, traffic management to analyze flow, to detect accidents, video editing to eliminate cumbersome human operator interaction, to design futuristic video effects.

Manuscript received on May, 2013.

Mr. D. Hari Hara Santosh obtained his B. Tech. and M. Tech degrees from JNT University, Hyderabad in the year 2005 and 2010. Presently he is pursuing Ph.D, in Video Processing at JNTU, Hyderabad.

Mr. P. Venkatesh is currently pursuing his B. Tech. in MVGR College of Engg., JNT University, Kakinada, Andhra Pradesh.

Poornesh Patruni born in Visakhapatnam on 04/11/1991. He is currently pursuing B.Tech course in the Department of ECE, MVGR College of engineering, JNT University, Kakinada, Andhra Pradesh.

Mr. L. Narayana Rao is currently pursuing his B. Tech. in MVGR College of Engg., JNT University, Kakinada, Andhra Pradesh.

Mr. N. Arun Kumar is currently pursuing his B. Tech. in MVGR College of Engg., JNT University, Kakinada, Andhra Pradesh.

Tracking is a significant and difficult problem that arouses interest among computer vision researchers. The objective of tracking is to establish correspondence of objects and object parts between consecutive frames of video. It is a significant task in most of the surveillance applications since it provides cohesive temporal data about moving objects which are used both to enhance lower level processing such as motion segmentation and to enable higher level data extraction such as behavior recognition and activity analysis. Tracking has been a difficult task to apply in congested situations due to inaccurate segmentation of objects. Long shadows, full and partial occlusion of objects with each other and with stationary items in the scene are some of the common problems of erroneous segmentation. For robust tracking it is important to dealing with shadows at motion detection level and coping with occlusions both at segmentation level and at tracking level. Tracking in video can be categorized according to the needs of the applications. It is used in or according to the methods used for its solution.

There are two common approaches in tracking objects as a whole: one is based on correspondence matching and other one carries out explicit tracking by making use of position prediction or motion estimation. On the other hand, the methods that track parts of objects (generally humans) employ model-based schemes to locate and track body parts. Some example models are stick figure, Cardboard Model, 2D contour and 3D volumetric models combine motion estimation methods with correspondence matching to track objects. It is also able to track parts of people such as heads, hands, torso and feet by using the Cardboard Model which represents relative positions and sizes of body parts. It keeps appearance templates of individual objects to handle matching even in merge and split cases. In this paper the algorithm uses Gaussian Mixture Model, a background modeling method to extracting moving objects and for trajectory prediction.

Moving object tracking is the process of locating a moving object in time using a camera. An algorithm analyses the video frames and outputs the location of moving targets within the video frame. The main difficulty in video tracking is to associate target locations in consecutive video frames, especially when the objects are moving fast relative to the frame rate. Here, video tracking systems usually employ a motion model which describes how the image of the target might change for different possible motions of the object to track.

II. COLOR EXTRACTION

A color image is comprised of three basic colors which are Red, Green and Blue. So each pixel of a color image can be broken down into Red, Green and Blue values. As a result, for the entire image 3 matrices are obtained, each one representing color features. The three matrices are arranged in sequential order, next to each other creating a 3 dimensional m by n by 3 matrixes.

The three basic colors can be used to extract other color components from the image. They can be extracted using the following equations

$$R = image(:,:,1); \qquad G = image(:,:,2);$$
$$B = image(:,:,3);$$

The below figure shows the extraction of three basic components using the above equations.

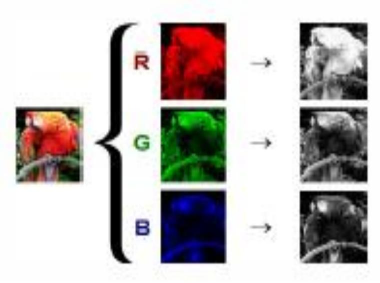


Figure 1. RGB color related with grayscale

This project mainly deals with tracking institutional buses which are yellow in color. Yellow is obtained by the combination of red and green colors. After extracting the basic components yellow color components of the image are extracted using R, G and B components using the following equation

$$Y = {(R+G)}/{2} - B$$

III. OBJECT DETECTION USING GAUSSIAN MIXTURE MODEL

A Gaussian Mixture Model (GMM) is a parametric probability density function represented as a weighted sum of Gaussian component densities. GMMs are commonly used as a parametric model of the probability distribution of continuous measurements or features in a biometric system, such as color based tracking of an object in video.

In many computer related vision technology, it is critical to identify moving objects from a sequence of videos frames. In order to achieve this, background subtraction is applied which mainly identifies moving objects from each portion of video frames. Background subtraction or segmentation is a widely used technique in video surveillance, target recognitions and banks. By using the Gaussian Mixture Model background model, frame pixels are deleted from the required video to achieve the desired results. The application of background subtraction involves various factors which involve developing an algorithm which is able to detect the required object robustly, it should also be able to react to various changes like illumination, starting and stopping of moving objects.

Surveillance is the monitoring of the behavior, activities or other changing information usually of people and often in a surreptitious manner. Video surveillance is commonly used for event detection and human identification. But it is not easy as think to detect the event or tracking the object. There are many techniques and papers introduced by many scientists for the backend process in the video surveillance. Different automated software's are used for the analysis of the video footage. It tracks large body movements and objects.

A. Background Subtraction

Background Subtraction is based on four important steps which are given below:

1) Preprocessing

Temporal or spatial smoothing is used in the early pre processing stage to eliminate device noise which can be a factor under different light intensity. Smoothing technique also includes removing various other elements like environment such as rain and snow. In real-time systems, frame size and frame rate are commonly adopted to reduce the data processing rate. Another key factor in preprocessing technique is the data format used by the background subtraction model. Most algorithms can handle luminance intensity which is one scalar value per each pixel

Figure 2 - Image on the left shows snowing and image on the right is a resultant of smoothing effect

In the figure 2, shown are two images which shows snow on the left and whereas with the application of spatial and temporal smoothing on right image results in the elimination of snow producing an more clear and effective image for background subtraction.

2) Background Modeling

This step uses the new video frame in order to calculate and update the background model. The main aim of developing a background model is that it should be robust against environmental changes in the background, but sensitive enough to identify all moving objects of interest.

3) Foreground Detection

In this step, it identifies the pixels in the frame. Foreground detection compares the video frame with the background model, and identify candidate foreground pixels from the frame. Commonly- used approach for foreground detection is to check whether the pixel is significantly different from the corresponding background estimate.

4) Data Validation

Finally, this step eliminated any pixels which are not connected to the image. It involves the process of improving the foreground mask based on the information obtained from the outside background model.

Most background models lack three main points:

1. Ignoring any correlation between neighboring pixels

2. The rate of adaption may not match the moving speed of the foreground object.
3. Non-stationary pixels, from moving leavers or shadow cast by moving objects are at times mistaken for true foreground objects.

B. Algorithm of Gaussian Mixture Model:

In order to give a better understanding of the algorithm used for background subtraction the following steps were adopted to achieve the desired results:

1. Firstly, we compare each input pixels to the mean 'µ' of the associated components. If the value of a pixel is close enough to a chosen component's mean, then that component is considered as the matched component. In order to be a matched component, the difference between the pixel and mean must be less than compared to the component's standard deviation scaled by factor D in the algorithm.

2. Secondly, update the Gaussian weight, mean and standard deviation (variance) to reflect the new obtained pixel value. In relation to non-matched components the weights 'w' decreases whereas the mean and standard deviation stay the same. It is dependent upon the learning component 'p' in relation to how fast they change.

3. Thirdly, here we identify which components are parts of the background model. To do this a threshold value is applied to the component weights 'w'.

4. Fourthly, in the final step we determine the foreground pixels. Here the pixels that are identified as foreground don’t match with any components determined to be the background.

C. General formula of Gaussian Mixture Model:

A Gaussian mixture model can be formulated in general as follows:

$$P(X_t) = \sum_{i=1}^{K} \omega_{i,t} \eta(X_t; \mu_{i,t}, \Sigma_{i,t})$$

Where, obviously,

$$\sum_{i=1}^{K} \omega_{i,t} = 1$$

The mean of such a mixture equals

$$\mu_t = \sum_{i=1}^{K} \omega_{i,t} \mu_{i,t}$$

that is, the weighted sum of the means of the component densities.

Where be the variable which represents the current pixel in frame , K is the number of distributions, and t represents time (i.e., the frame index), is an estimate of the weight of the ith Gaussian in the mixture at time t, is the mean value of the ith Gaussian in the mixture at time t, is the covariance matrix of the ith Gaussian in the mixture at time t.

D. Necessity of GMM:

The value of a pixel at time t in RGB or some other color space is denoted by $\vec{x}$. Pixel-based background subtraction involves decision if the pixel belongs to background (BG) or some foreground object (FG). Bayesian decision R is made by:

$$R = \frac{p\left(BG \middle| \vec{x}^{(t)}\right)}{p\left(FG \middle| \vec{x}^{(t)}\right)} = \frac{p\left(\vec{x}^{(t)} \middle| BG\right) p(BG)}{p\left(\vec{x}^{(t)} \middle| FG\right) p(FG)}$$

The results from the background subtraction are usually propagated to some higher level modules, for example the detected objects are often tracked. While tracking an object we could obtain some knowledge about the appearance of the tracked object and this knowledge could be used to improve background subtraction. In a general case we don't know anything about the foreground objects that can be seen or when and how often they will be present. Therefore we set p(FG) = p(BG) and assume uniform distribution for the foreground object appearance p($\vec{x}$|FG) = c. We decide then that the pixel belongs to the background if:

$$p\left(\vec{x}^{(t)} \middle| BG\right) > c_{thr} \left(= Rc_{FG}\right),$$

where c is a threshold value.We will refer to p($\vec{x}$|BG) as the background model. The background model is estimated from a training set denoted as X.

E. Methodology

The GMM is a mixture of k Gaussian distributions that point the change of state of the corresponding pixels from one frame to another. The algorithm developed applies gaussian mixtures to each frame and transforms images once colorful into binary images. For the corresponding pixels that undergo no state changes, the value 1 (black) is attributed and for pixels that undergo drastic changes in state, the value 0 (white) is attributed. Thus, it is possible generating the locations of all moving objects in the video.

This is how all the moving objects in the video are located using GMM.

Figure 3.a

Figure 3.b

As shown in the figure 3.a the pixels corresponding to the road which is background in the image do not undergo state changes, as a result the value 0 (black) is attributed and appears black as shown in figure 3.b. The pixels corresponding to cars undergo drastic changes in state, so the value 1 (white) is attributed and cars appears white as shown in figure 3.b

IV. BLOB ANALYSIS

For image processing, a blob is defined as a region of connected pixels. Blob analysis is the identification and study of these regions in an image. The algorithms discern pixels by their value and place them in one of two categories: the foreground (typically pixels with a non-zero value) or the background (pixels with a zero value).

In typical applications that use blob analysis, the blob features usually calculated are area and perimeter, Feret diameter, blob shape, and location. The versatility of blob analysis tools makes them suitable for a wide variety of applications such as pick-and-place, pharmaceutical, or inspection of

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

food for foreign matter.

Since a blob is a region of touching pixels, analysis tools typically consider touching foreground pixels to be part of the same blob. Consequently, what is easily identifiable by the human eye as several distinct but touching blobs may be interpreted by software as a single blob. Furthermore, any part of a blob that is in the background pixel state because of lighting or reflection is considered as background during analysis.

Blob analysis is used in finding blobs whose spatial characteristics satisfy certain criteria. In many applications where computation is time consuming, blob analysis is used to eliminate blobs that are of no interest based on their spatial characteristics, and keep only the relevant blobs for further analysis. It can also be used to find statistical information such as the size of the blobs or the number, location, and the presence of blob regions.

A. Blob Analysis Concepts

A typical blob analysis process scans through an entire image and detects all the particles, or blobs, in the image and builds a detailed report on each particle. This report usually contains approximately 50 pieces of information about the blob, including the blob's location in the image, size, shape, orientation to other blobs, longest segment, and moment of inertia.

Perimeter, angle, area, and center of mass are the parameters used to identify and classify these blobs. Using multiple parameters can be faster and more effective than pattern matching in many applications.

B. Blob analysis to calculate area:

The following area parameters are described below

Number of pixels – Area of a particle, without holes, in pixel units

Particle area – Area of a particle expressed in real units (based on image spatial calibration). This value is equal to **Number of pixels** when the spatial calibration is such that 1 pixel represents 1 square unit.

Scanned area – Area of the entire image expressed in real units. This value is equal to the product **(Resolution X×X Step) (Resolution Y×Y Step).**

Ratio – Ratio of the particle area to the entire image area. The percentage of the image occupied by all particles.

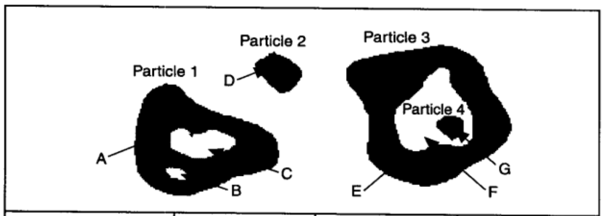
Ratio = $\frac{particle\ ar}{scanned\ ar}$

Number of holes – Number of holes inside a particle. The software detects the holes inside a particle as small as 1 pixel.

Holes' area – Total area of the holes within a particle

Total area – Area of a particle including the area of its holes. This value is equal to **Particle area + Holes' area**

Note: A particle located inside a hole of a bigger particle is identified as a separate particle. The area of a hole that contains a particle includes the area covered by the particle



Particle #	Particle Area	Holes' Area	Total Area
Particle 1	A	B + C	A + B + C
Particle 2	D	0	D
Particle 3	E	F + G	E + F + G
Particle 4	G	0	G

Fig 4: Different types of area indications

V. RESULTS

Figures and Initially R, G, B components of the images are extracted individually. Using these extracted components, yellow objects are detected while excluding all the remaining objects in the image. Later Gaussian Mixture Model is applied to extract the foreground containing yellow objects.

After extracting the foreground, yellow buses are separated from remaining yellow colored objects based on the area of blob. For calculating the area blob analysis is used. Besides area of blob, area of bounding box is also calculated and then ratio of blob area and bounding box area is found out. If the ratio is greater than the threshold set then it is classified as bus. Now bounding rectangles are drawn around the detected buses, with that the numbers of buses being tracked are displayed on the top of output video.

Figure 5.a

Figure 5.b

Figure 5.a shows the frame 1422 of the traffic video sequence which contains two yellow buses and two other non-yellow color vehicles, from this frame R, G, B components are extracted individually. Using these components yellow objects are detected while neglecting all the non-yellow color objects in the frame, the yellow objects appear brighter than the remaining objects. Now Gaussian Mixture Model is applied to extract the foreground containing only yellow. After extracting the foreground, in order to separate the yellow buses from remaining yellow color objects blob analysis is used. Using blob analysis areas of blob, bounding box are calculated and ratio of those areas is calculated. For the yellow bus, the calculated ratio exceeds the threshold set, as a result a bounding box is drawn around the bus and with that number of buses that have been tracked is displayed in black box on the top of output frame which can be

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

observed in figure 4.b. Though there are two yellow buses in the input frame bounding box is drawn around only one bus which is present in the ROI (Region Of Interest). ROI is the region present below the white line in the output frame as shown in figure 5.b.

Figure 6.a Figure 6.b

Figure 6.a shows the frame 3335 of the traffic video sequence which contains one auto and other non-yellow color vehicles, from this frame R, G, B components are extracted individually. Using these components yellow objects are detected while neglecting all the non-yellow color objects in the frame, the yellow objects appear brighter than the remaining objects. Now Gaussian Mixture Model is applied to extract the foreground containing only yellow. After extracting the foreground, using blob analysis areas of blob, bounding box are calculated and ratio of those areas is calculated. Though the auto is in yellow color, bounding box is not drawn around it as shown in figure 6.b because the ratio of areas does not exceed the threshold set.

Thus, this method is capable of discriminating the yellow buses from remaining yellow colored objects.

A. *Case study:*

TABLE 2: TRACKING EFFICIENCY OF THE PROPOSED ALGORITHM

Time slot	Buses passed	Buses detected	Efficiency in %
3:30 pm-3:45 pm	11	10	90.9
3:45 pm-4:00 pm	13	12	92.3
4:00 pm-4:15 pm	4	4	100
4:15 pm-4:30 pm	6	6	100
4:30 pm-4:45 pm	17	15	88.2
4:45 pm-5:00 pm	10	9	90
Overall Efficiency			91.8

In order to calculate the density of institutional buses and to compute the efficiency of proposed algorithm six time slots have been considered between 3:30 pm to 5 pm each of 15 minutes interval. In the table 7 it can be observed that the efficiency of the proposed algorithm varies between 88.2 to 100 percent. The causes for this variation observed are:

1. the average speed of the vehicles is below the threshold
2. occlusions caused while overtaking other vehicles

Fig 7: Failed case to detect due to occlusion

From the above figure it is observed that the second bus is not detected as the first bus occludes the second bus, as a result the second vehicle is not recognized as a bus.

From table it is observed that a total of 61 buses passed through, out of which 56 buses have been detected successfully. Thus the overall efficiency of proposed algorithm is 91.8.

VI. CONCLUSIONS

We have proposed an object tracking algorithm for video pictures, based on GMM and blob analysis on frames in a simple feature space. Simulation results for frame sequences with moving buses (especially yellow colored) video sequences verify the suitability of the algorithm for reliable moving object tracking. Initially R, G, B components of the images are extracted individually. Using these extracted components, yellow objects are detected while excluding all the remaining objects in the image. Later Gaussian Mixture Model is applied to extract the foreground containing yellow objects.

There may also be the concern that the linear motion estimation may fail for objects moving in a complicated nonlinear way. However, if the movement is not extremely fast, the deviation from estimated positions between successive frames is small, than correct tracking is reliably achieved. Furthermore, if mistracking occurred at some frame by reason of occlusion, newly appearing or disappearing objects, the proposed algorithm could recover correct tracking after a couple of frames. GMM makes the algorithm more robust, such that the tracker will recover tracking if occlusion takes place.

REFERENCES

[1] Takashi morimoto, Osama kiriyama, youmei harada,Object tracking in video images based on image segmentation and pattern matching IEEE conference proceedings, vol no 05, page no, 3215-3218, 2005
[2] Zehang, S., Bebis, G., Miller, R.: On-road vehicle detection: a review. IEEE Trans. Pattern Anal. Mach. Intell. 28(5), 694–711 (2006)
[3] Bin, D., Yajun, F., Tao,W.: A vehicle detectionmethod via symmetry in multi-scale windows. In: 2nd IEEE Conference on Industrial Electronics and Applications, 2007 (ICIEA 2007), pp. 1827–1831 (2007)
[4] Nedevschi, S., Vatavu, A., Oniga, F., Meinecke, M.M.: Forward collision detection using a stereo vision system. In: 4th International Conference on Intelligent Computer Communication and Processing, 2008 (ICCP 2008), pp. 115–122 (2008)
[5] Reynolds, D.A., Rose, R.C.: Robust Text-Independent Speaker Identification using Gaussian Mixture Speaker Models. IEEE Transactions on Acoustics, Speech, and Signal Processing 3(1) (1995) 72–83
[6] Boeing, A., Braunl, T.: ImprovCV: Open component based automotive vision. In: Intelligent Vehicles Symposium, 2008 IEEE, pp. 297–302 (2008)

[7] Wei, L., XueZhi, W., Bobo, D., Huai, Y., Nan, W.: Rear vehicle detection and tracking for lane change assist. In: Intelligent Vehicles Symposium, 2007 IEEE, pp. 252–257 (2007)
[8] Junqiu wang [Member IEEE], yasushi yagi [Member IEEE],‖Integrating colour and shape texture features for adaptive real time object tracking‖, IEEE transactions on image processing, vol no 17,no 2, page no 235-240,2007

Mr. D. Hari Hara Santosh obtained his B. Tech. and M. Tech degrees from JNT University, Hyderabad in the year 2005 and 2010. Presently he is pursuing Ph.D, in Video Processing at JNTU, Hyderabad. He is working as Assistant Professor in MVGR College of Engineering, Vizianagaram, AP. He has 5 publications in both International and National Journals and presented 19 papers at various International and National Conferences. His areas of interests are Image Processing, Speech processing and Video Processing.

Mr. P. Venkatesh is currently pursuing his B. Tech. in MVGR College of Engg., JNT University, Kakinada, Andhra Pradesh. His areas of interests are Image Processing, Speech processing.

Poornesh Patruni born in Visakhapatnam on 04/11/1991. He is currently pursuing B.Tech course in the Department of ECE, MVGR College of engineering, JNT University, Kakinada, Andhra Pradesh. He is a member of IEEE. His His areas of interest is image processing. His paper Video Fusion Using Pixel Averaging, Principal Component Analysis and Laplacian Pyramid – A Comparative Study was published in International Journal Of Advanced Computer Research Volume-2 Number-4 Issue -6 December-2012.

Mr. L. Narayana Rao is currently pursuing his B. Tech. in MVGR College of Engg., JNT University, Kakinada, Andhra Pradesh. His areas of interests are Image Processing, Speech processing.

Mr. N. Arun Kumar is currently pursuing his B. Tech. in MVGR College of Engg., JNT University, Kakinada, Andhra Pradesh. His areas of interests are Image Processing, Speech processing.