International Journal of Engineering Research and General Science Volume 3, Issue 1, January-February, 2015
ISSN 2091-2730

Object Tracking Using Background Subtraction Algorithm

Vinay D R, N Lohitesh Kumar

CIT, Gubbi, vinay.ise@gmail.com, lohitesh.kn@gmail.com

Abstract— Identifying moving objects from a video sequence is a fundamental and critical task in many computer-vision applications. An efficient algorithm for detecting a moving object using background elimination technique is proposed in this paper. In the post processing step, the morphological gradient operation with median filter is used to remove the noise and shadow regions which are present in the moving object. The experimental result shows that the clarity of the image obtained using background elimination technique is much better than using background registration technique. The experimental results show that the accuracy of detected counting vehicles is 94%.

Keywords— Frame Difference, Background Elimination, Background Registration, Background Subtraction Algorithm

INTRODUCTION

Identifying moving objects from a video sequence is a fundamental and critical task in many computer-vision applications. A common approach is to perform background subtraction, which identifies moving objects from the portion of a video frame that differs significantly from a background model. There are many challenges in developing a good background subtraction algorithm. First, it must be robust against changes in illumination. Second, it should avoid detecting non-stationary background objects such as moving leaves, rain, snow, and shadows cast by moving objects. Finally, its internal background model should react quickly to changes in background such as starting and stopping of vehicles.

This research began with a comparison of various background subtraction algorithms for detecting moving vehicles and pedestrians in urban traffic video sequences (Cheung and Kamath 2004). The approaches vary from simple techniques such as frame differencing and adaptive median filtering, to more sophisticated probabilistic modeling techniques is considered. While complicated techniques often produce superior performance, the proposed experiments show that simple techniques such as adaptive median filtering can produce good results with much lower computational complexity.

In addition, the pre-and post-processing of the video might be necessary to improve the detection of moving objects. For example, by spatial and temporal smoothing, the snow can be removed from a video. Small moving objects, such as moving leaves on a tree, can be removed by morphological processing of the frames after the identification of the moving objects.

The rate and weight of model updates greatly affect foreground results. Slow adapting background models cannot quickly overcome large changes in the image background (such as a cloud passing over a scene). These results in a period of time where many background pixels are incorrectly classified as foreground pixels. A slow update rate also tends to create a ghost mask which trails the actual object. Fast adapting background models can quickly deal with background changes, but they fail at low frame rates. They are also very susceptible to noise and the aperture problem. These observations indicate that a hybrid approach might help mitigate the drawbacks of each.

The new foreground validation technique have been created that can be applied to any slow-adapting background subtraction algorithm (Cheung and Kamath 2005). Slow adapting methods produce relatively stable masks and tend to be more inclusive than fast adapting methods. As a result, they can also have high false positive rate. Foreground validation further examines individual foreground pixels in an attempt to eliminate false positives. The proposed algorithm first obtains a foreground mask from a slow-adapting algorithm, and then validates foreground pixels by a simple moving object model built using foreground and background statistics as well as a fast-adapting algorithm.

Ground-truth experiments with urban traffic sequences have shown that the proposed algorithm produces performances that are comparable or better than other background subtraction techniques.

PRINCIPLE OF WORKING

If the subsequent frames can be subtracted which is clicked by cam, then the part of image which does not change (background) gets subtracted to give zero intensity (black). Only the part of image moved (moving object) don't get reduced to zero as intensity of pixels of two subsequent frames are different. So non-zero intensity is obtained for pixels which are corresponding moved object. Rest is simple. Just convert the image into binary and obtain the centroid of largest area of connected pixels!!

Background Subtraction

Background Subtraction is a process to detect a movement or significant differences inside of the video frame, when compared to a reference, and to remove all the non-significant components (background). Background subtraction is applied in many areas, such as surveillance system (to effectively segment the only moving object).

Steps to implement background subtraction

1. Learning Background – Capture the ten background frames and calculate the mean (μ) and the standard deviation (σ) with the below equations (1.1) & (1.2).
2. The assumption is that the value of the background was iid-normal distribution.

$$\mu = \frac{1}{n}\sum_{i-1}^{n} x_i \qquad \text{----------------} \quad (1.1)$$

$$\sigma = \sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2\right)} \quad \text{-----------} \quad (1.2)$$

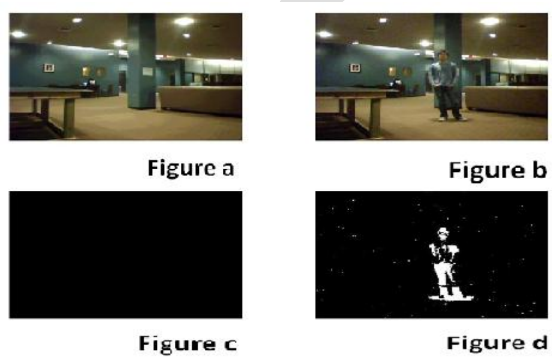


Fig. 1 a) Background Frame b) Non-background object is introduced d) Segmented Object

Algorithm

The idea of this algorithm is got from a paper "Implementation of an Automated Single Camera Object Tracking System Using Frame Differencing and Dynamic Template Matching" by guys from IIT-Kanpur and NIT-Nagpur. Although the paper is modified it to directly take and operate on rgb images. The algorithm works as follows:

1. Grab ith frame.
2. Grab subtract it from (i-3) th frame.
3. Convert the image into binary.
4. Fill small holes.
5. Label the connected pixels.
6. Run the loop to number of labels and find the label for maximum area.
7. Find centroid of the obtained area.
8. Mark the area if you just want to track or use the centroid information for other applications.
9. Go to step 1

Here subtraction is done with (i-3)th frame keeping in mind slow moving objects and implemented this algorithm completely in MATLAB.

ARCHITECTURE & MODELING

In many real-time applications like video conferencing, the camera is fixed. Some techniques proposed in paper [12] use global motion estimation and comparison to compensate the change in background due to camera motion. In the present algorithm, the assumption is that the background is stationary for the video clips considered.

The flow of the algorithm for background elimination is as follows: The first step is to read the video clip and it is converted to frames. In the first stage difference between frames are computed i.e. Fi and Fi+k.

In the next stage these differences are compared, and in the third stage pixels having the same values in the frame difference are eliminated. The fourth phase is the post processing stage executed on the image obtained in third stage and the final phase is the object detection.

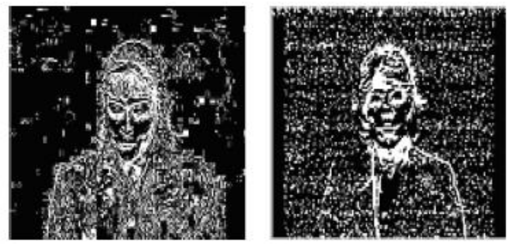
ISSN 2091-2730

Frame Difference

Frame differences are computed by finding the difference between consecutive frames but this will introduce computational complexity in case the video clips having slow-moving objects. Moreover this algorithm assumes a stationary background. Hence the difference between the frames at regular intervals (say, some integer k) is considered. If there are n frames, then (n/k) frame differences (FD) is obtained. The frame difference follows Gaussian distribution as indicated in equation (1.3).

$$p(FD) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(FD-\mu)^2}{2\sigma^2}) \qquad \text{----------- (1.3)}$$

Here, the mean and standard deviation of FD. The frame differences of some test sequences are as shown in Fig 2.



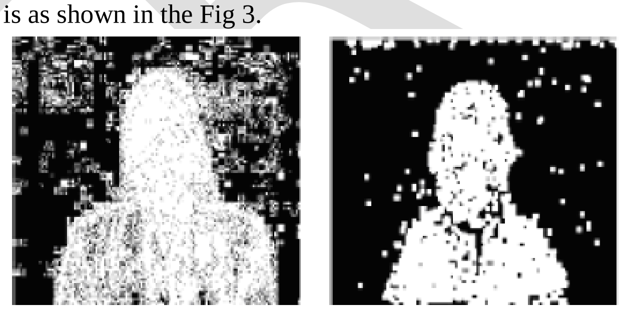
(a) Difference between 30th and 35th frames of Akhiyo (b) Difference between 12th and 15th frames of Claire

Fig 2: Frame Difference

Background Elimination

Once the frame differences are computed the pixels that belong to the background region will have a value almost equal to zero, as the background is assumed stationary. Many times because of camera noise, some of the pixels belonging to the background region may not tend to zero. These values are set to zero by comparing any two frame differences, say, FDi and FDj.

Thus, the background region is eliminated and only the moving object region will contain nonzero pixel values. The images obtained after background elimination is as shown in the Fig 3.



(a) Image of Moving Hands (b) Image of Claire

Fig 3: Background Elimination

Background Registration

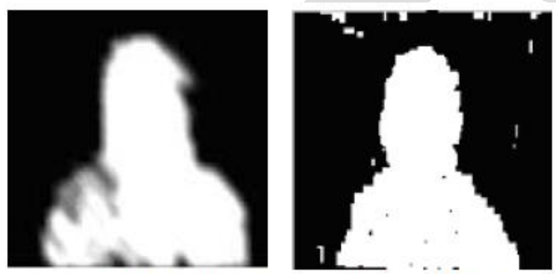
A general tracking approach is to extract salient regions from the given video clip using a learned background modeling technique. This involves subtracting every image from the background scene and thresholding the resultant difference image to determine the foreground image. Stationary pixels are identified and processed to construct the initial background registered image. Here the fact that vehicle is a group of pixels that move in a coherent manner, either as a lighter region over a darker background or vice versa. Often the vehicle may be of the same color as the background, or may be some portion of it may be camouflaged with the background, due to which tracking the object becomes difficult. This leads to an erroneous vehicle count.

Foreground Detection (Object Tracking)

Most vision based traffic monitoring system must be capable of tracking vehicles through the video sequence. Tracking helps in eliminating multiple counts in vehicle counting applications and it also helps in deriving useful information while computing vehicle velocities. Tracking information can be used to refine the vehicle type and also to correct errors caused due to occlusions. After registering the static objects the background image is subtracted from the video frames to obtain the foreground dynamic objects. Post processing is performed on the foreground dynamic objects to reduce the noise interference.

Post Processing

Many times due to camera noise and irregular object motion, there always exists some noise regions both in the object and background region. Most of the post processing techniques are applied on the image obtained after background elimination. Initially, order statistics filters are used, which are the spatial filters and whose response is based on ordering (ranking) the pixels contained in the image area encompassed by the filter. The response of the filter at any point is then determined by the ranking result. The current algorithm uses Median filter which is the best-known order-statistics filter. This filter replaces the value of a pixel by the median of the gray levels in the neighborhood of that pixel. The formula used is ˆ f (x, y) = median {g (s, t)}. After applying the median filter, the resulting image is converted into a binary image. The morphological opening technique is applied on this binary image. The opening of A by B is simply erosion of A by B followed by dilation of the result by B. This can be given as $A.B = (A \_ B) \oplus B$. Here, A is the image and B is a structuring element. After applying the above explained pre-processing techniques, the new image obtained is as shown in the Fig 4.



(a) Image of Moving Hands (b) Image of Claire

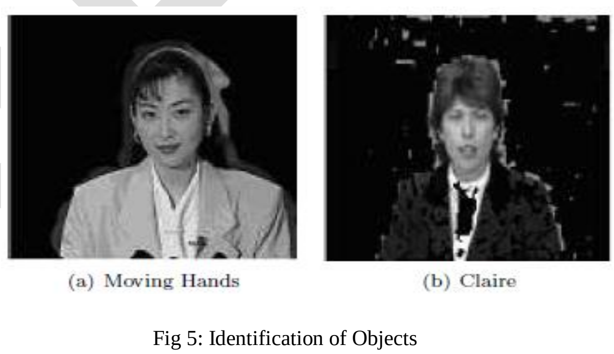
Fig 4: After Post Processing

Object Tunning

This is a post processing technique applied in some application. In the proposed algorithm median filter is used for noise elimination in both i.e. object and background. As the object boundaries are not very smooth, a post processing technique is required on the foreground image. The final output of the object tuning phase is a binary image of the objects detected termed mask1.

Object Identification

The image obtained after the pre-processing step has relatively less noise, so, the background area is completely eliminated. Now, if the pixel values of this image are greater than a certain threshold, then, those pixels are replaced by the pixels of the original frame. This process identifies the moving object as shown in Fig 5.



(a) Moving Hands (b) Claire

Fig 5: Identification of Objects

Object Counting

The tracked binary image mask1 forms the input image for counting. This image is scanned from top to bottom for detecting the presence of an object. Two variables are maintained i.e. count that keeps track of the number of vehicles and count register countreg, which contains the information of the registered object. When a new object is encountered, it is first checked to see whether it is already registered in the buffer, if the object is not registered then it is assumed to be a new object and count is incremented, else it is treated as a part of an already existing object and the presence of the object is neglected. This concept is applied for the entire image and the final count of objects is present in variable count. A fairly good accuracy of count is achieved. Sometimes due to occlusions two objects are merged together and treated as a single entity.

IMPLEMENTATION

Here two algorithms are proposed background elimination and background registration method which are implemented using Matlab. The performance analysis is done through the method of Least Squares. The least square method is normally used to find the bestfit, given two sets of data. According to the method of least squares, the best-fit must satisfy the rule given by equation (4).

Simulation is performed using Matlab Software. This is an interactive system whose basic data element is an array that does not require dimensioning. It is a tool used for formulating solutions to many technical computing problems, especially those involving matrix representation. This tool emphasis's a lot of importance on comprehensive prototyping environment in the solution.

The algorithm for Background Registration is as follows

```
ALGORITHM BGRegist ()
//Input: M Array
//Output: An Image with Registered Background in bg array
//Initialize array [b] to zeros
1. for i=1 to m
   for j=1 to n
   for k=1 to l-1
   if abs(double(T(i,j,l-k))-double(T(i,j,k)))¡10
      b(i,j)=T(i,j,k)
   end if
   end for
   end for
   end for
2. Convert b array values to unsigned integers and store it into array called background.
3. Fill the hole regions in image background and store it in bg array
4. Show the output images background, bg.
5. Declare two global variables m and n which stores the row and column values of video frames respectively.
```

The algorithm for counting is as follows

```
ALGORITHM Count()
//Input: d is specific video frame
//Output: An image with Foreground Objects is stored in c
//Initialize count=0 and count register buffer
//countveg=0
1. Traverse the mask1 image to detect an object
2. If object encountered then check for registration in countveg
3. If the object is not registered then increment count and register the object in countveg, labeled with the new count.
4. repeat steps 2-4 untill traversing not completed
```

A. Performance Analysis

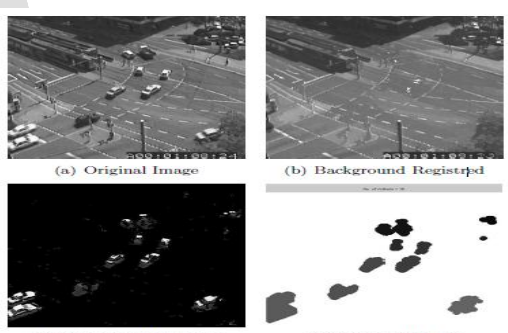


Fig 6 Video1

This video segmentation method was applied on three different video sequences two of which are depicted below. For the first video sequence Fig 6(a) depicts the original image, Fig 6(b) shows the background registered image, Fig 6 (c) the foreground detected objects obtained after background subtraction, and finally Fig 6(d) shows the count of the detected objects. The same is repeated for the next video sequence. The system is able to track and count most vehicles successfully. Although the accuracy of vehicle detection was 100%., the average accuracy of counting vehicles was 94%. This is due to noise which causes detected objects to become too large or too small to be considered as a vehicle. However, two vehicles will persist to exist as a single vehicle if relative motion between them is small and in such cases the count of vehicles becomes incorrect. An added advantage of this algorithm is, the segmentation logic is not intensity based, and hence vehicles whose intensities are similar to the road surface are not missed out. The results were successfully carried out on three videos; the accuracy of detecting the objects was 100%. The detected objects are then counted.

CONCLUSIONS

The proposed method introduces an efficient algorithm for detecting a moving object using background elimination technique. The experimental results obtained indicate that the clarity of the image obtained using background elimination technique is much better than using background registration technique. Good segmentation quality is achieved efficiently.

This paper also discusses an application system of traffic surveillance. Here an algorithm is developed to track and count dynamic objects efficiently. The tracking system is based on a combination of a temporal difference and correlation matching. The system effectively combines simple domain knowledge about object classes with time domain statistical measures to identify target objects in the presence of partial occlusions and ambiguous poses in which the vehicles are moving. The background clutter is effectively rejected. The experimental results show that the accuracy of counting vehicles reached 94%, although the vehicle detection was computational complexity of our algorithm is linear to the size of a video clip and the number of vehicles tracked.

As a future work a combination of higher dimensional features with some additional constraints may be tried so that adverse effects of some features can be compensated by contribution of others.

REFERENCES:

[1] Sikora, T., the MPEG-4 Video Standard Verification Model, IEEE Transactions, Circuits Systems, Video Technology, vol. 7, pp. 19-31, Feb. 1997.

[2] Nack F. and Lindsay A. T., Everything you wanted to know about MPEG-7: Part 2, IEEE Multimedia, vol. 6, pp. 64-73, Dec. 1999.

[3] Salembier P. and Marques F., Region-based Representations of Image and Video: Segmentation Tools for Multimedia Services, IEEE Transactions, Circuits Systems, Video Technology, vol. 9, pp. 1147-1169, Dec. 1999.

[4] Wang D., Unsupervised Video Segmentation Based onWatersheds and Temporal Tracking, IEEE Transactions, Circuits Systems, Video Technology, vol. 8, pp. 539-546, Sept. 1998.

[5] Y. Yokoyama, Y. Miyamoto and M. Ohta, Very Low Bit Rate Video Coding Using Arbitrarily ShapedRegion-Based Motion Compensation, IEEE Trans

[6] L.Wu, J. Benoise-Pineau, P. Delagnes and D. Barba, Spatio-temporal Segmentation of Image Sequences for Object-Oriented Low Bit-Rate Image Coding, Signal Processing: Image Communication., vol. 8, pp. 513-543, 1996.

[7] P. Salembier, Morphological Multiscale Segmentation for Image Coding, Signal Processing, vol. 38, pp. 359-386, 1994.

[8] N. T. Watsuji, H. Katata and T. Aono, Morpho- logical Segmentation with Motion Based Feature Extraction, In International Workshop on Coding Techniqus for Very Low Bit-Rate Video, Tokyo,Nov. 8-10, 1995.

[9] W. H. Hong, N. C. Kim and S. M. Lee, Video Segmentation Using Spatial Proximity, Color and Motion Information for Region-Based Coding, In Proceedings SPIE Visual Communications and Image Processing, vol. 2308, pp. 1627-1633, 1994.

[10] Choi J. C., Lee, S .W. and Kim, S. D., SpatioTemporal Video Segmentation Using a Joint Similarity Measure, IEEE Transactions, Circuits Systems, Video Technology, vol. 7, pp. 279-289, Apr. 1997.

[11] Vibha L., Venugopal K. R. and L. M. Patnaik, A Study of Breast Cancer Using Data Mining Techniques, Technical Report, University Visvesvaraya College of Engineering, Bangalore University, August 2003.

[12] Aach T., Kaup A. and Mester R., Statistical Model- Based Change Detection in Moving Video, Signal Processing, vol. 31, pp. 165-180,Mar. 1993.

[13] Neri A., Colonnese S., Russo G. and Talone P., Automatic Moving Object and Background Separation, Signal Processing, vol. 66, pp. 219-232, Apr. 1998.

[14] Stauder J., Mech R. and Ostermann J., Detection of Moving Cast Shadows for Object Segmentation, IEEE Transaction Multimedia, vol. 1, pp.65-76, Mar. 1999.

[15] D. Koller, J. Weber, T. Haung and J. Malik, Moving Object Recognition and Classification based on Recursive Shape Parameter Estimation, In Proceedings of the 12th Israeli Conference on Artificial Intelligence, Computer Vision and Neural Networks, pp. 359-368, Tel-Aviv, Israel, December 1993.

[16] D. Koller, J. Weber, T. Haung, J. Malik, G. Ogasawara, B. Rao and S. Russel, Towards Robust Automatic Traffic Scene Analysis in Real-Time, In Proceedings of the 12th International Conferenceon Pattern Recognition (ICPR-94), pp. 126-131, Jerusalem, Israel, October 9-13, 1994.

[17] Puzicha J., Hofmann T. and Buhmann J. M., Histogram Clustering for Unsupervised Image Segmentation, In IEEE Computer Society Conference Computer Vision and Pattern Recognition, Fort Collins, CO, pp. 602-608, June 1999.

[18] Chen S. C., Shyu M. L. and Zhang C., An Unsupervised Segmentation Framework for Texture Image Queries, In 25th IEEE Computer Society International Computer Software and Applications Conference (COMPSAC), Chicago, Illinois, USA, Oct.2000.

[19] Chen S. C., Shyu M. L. and Zhang C., An Intelligent Framework for Spatio-Temporal Vehicle Track-ing, 4th International IEEE Conference on Intelligent Transportation Systems, Oakland, California, USA, Aug. 2001.

[20] D. Beymer, P. McLauchlan, B. Coifman and J. Malik, A Real-time Computer Vision System for Measuring Traffic Parameters, In Proceedings of IEEE Conference Computer Vision and Pattern Recognition, Puerto Rico, pp. 496501,June, 1997,

[21] O. Masoud and N. P. Papanikolopoulos, Robust Pedestrian Tracking Using a Modelbased Approach, In Proceedings of IEEE Conference on Intelligent Transportation Systems, pp. 338343, Nov. 1997.

[22] J. Lipton, H. Fujiyoshi and R. S. Patil, Moving Target Classification and Tracking from Real-time Video, In Proceedings of IEEE Workshop Applications of Computer Vision, pp. 814, 1998.

[23] Gupte S., Masoud O., Martin R. F. K. and Papanikolopoulos N. P., Detection and Classification of Vehicles, In IEEE Transactions on Intelligent Transportation Systems, vol. 3, no.1, pp. 37 47, March 2002.

[24] Dailey D. J., Cathey F. and Pumrin S., An Algorithm to Estimate Mean Traffic Speed Using Uncalbrated Cameras, In IEEE Transactions on Intelligent Ttransportations Systems, vol. 1, No. 2, pp. 98-107, June, 2000.

[25] Grimson W. E. L., Stauffer C., Romano R. and Lee L.. Miyamoto and M. Ohta, Very Low Bif digital image processing