ISSN 2278-3091

Volume 8, No.6, November – December 2019

International Journal of Advanced Trends in Computer Science and Engineering

Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse39862019.pdf

https://doi.org/10.30534/ijatcse/2019/39862019

Performance Analysis of Human Detection and Tracking in Changing Illumination

Mhalsakant Sardeshmukh[1], Vaishali Sardeshmukh

[1]JSPM NTC, Narhe, Pune, India, mmsardeshmukh2016@gmail.com

[2]Sinhgad Academy of Engineering Pune, India, vaishu.22.joshi@gmail.com

ABSTRACT

Human detection and tracking are useful in many applications like video surveillance, patent monitoring, scene understanding etc. The problem of human/ object detection becomes difficult in changing illumination and background. This paper proposed a comparative study of two methods of object detection and tracking in changing lighting. The first method suggested updates the background information continuously, and the other uses the depth information for segmentation. The performance of two ways compared based on the time required for segmentation and algorithmic complexity. It is observed that the use of depth information makes the system robust against varying illumination and less complex.

Key words : Human Activity Recognition, Varying illumination, Segmentation

1. INTRODUCTION

Over the past decade, our lives have been influenced by computers and worldwide tremendously. Computers extend our possibilities to communicate by performing repetitive and data-intensive computational tasks fundamentally. Because of recent technological advances, video data has become more and more accessible and now became very useful in our everyday life. Today, our commonly used consumer hardware, such as notebooks, mobile phones, and digital photo cameras are allowing us to create videos, while faster internet access and growing storage capacities enable us to publish directly or share videos with others. However, even though the importance of video data is increasing, the possibilities to analyze it in an automated fashion are still limited. In any surveillance application the first step is detection and tracking the human or interested object. This task becomes difficult in changing illumination. The background required to be updated continuously which increases time complexity. In this paper we proposed two approaches of human/object detection and tracking, one uses updating background continuously and other makes the use of depth information along with RGB for object detection and tracking. The comparison of these two

Approaches are presented in this paper which will help the researchers to select the appropriate method for their work.

2. BACKGROUND

The first step in any video surveillance application is Segmentation. For detecting the object knowledge of background, is essential. Any changes in illumination and viewpoint make this task critical. Differentiating the objects that are of similar color as of background turns more difficult. Depth information can be used along with color information to make the segmentation process robust and less complex. The traditional approaches for activity recognition in videos assume well-structured environments and they fail to operate in largely unattended way under adverse and uncertain conditions from those on which they have been trained. The other drawback of current methods is the fact that they focus on narrow domains by using specific concept detectors such as human faces, cars and buildings.

3.METHODOLOGY

3.1 Object detection and tracking in changing illumination and background

In many of the practical applications the illumination condition is not static whereas it keeps on changing. Detection and tracking of the object in this dynamic condition becomes difficult and challenging. This model detects and tracks the object in varying illumination. This method consists of continuously updating background for object detection and tracking. The foreground is extracted from the video sequence by learning background continuously by checking all pixel values from previous frames and subtracting it from test frame.

The system is executed in three steps

1. Continuously updating background modeling algorithm.
2. Object feature extraction.
3. Classification.

With background modeling algorithm, we can remove the constant pixel values means remove pixels whose values remain constant over some frames. Algorithm successfully removes noise, shadow effects and robust to the illumination change.

Real time features are extracted from the frames extracted from the video sequence. The features include the height, width, height width ratio, centroid. For a pedestrian, the periodic hand and leg movement is unique, and this is the feature that is used to classify the human from all other classes.

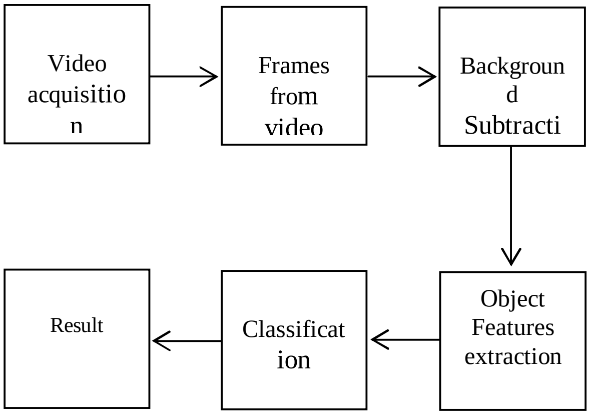


Figure 1: Segmentation using Background Subtraction

The real-time features that are acquired from the feature extraction stage are a base to classify the detected object into various categories such as pedestrian, group of pedestrian, vehicles, etc. Background subtraction is particularly a commonly used technique for motion segmentation in static scenes. It attempts to detect moving regions by subtracting the current image pixel-by-pixel from a reference background image that is created by averaging images over time in an initialization period. Some background subtraction techniques have many problems, especially when used outdoors. For instance, when the sun is covered by clouds, the change in illumination is considerable. In this framework, we have implemented continuously updating background modeling algorithm that ensures the removal of illumination change during the day, dynamic scene changes. The following figure shows the simple diagram for background subtraction.

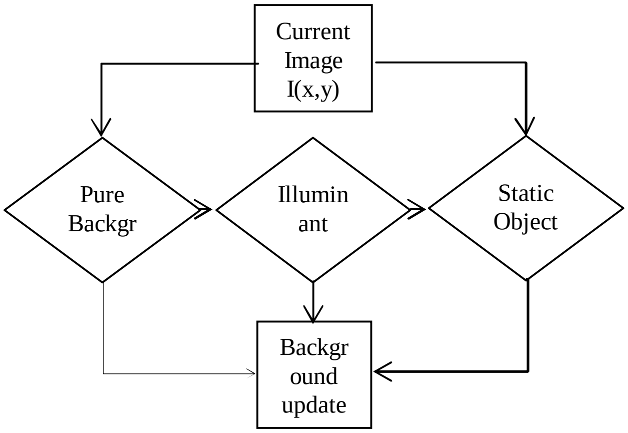


Figure 2: Background Modeling

The following equation gives the pure background.
$B(x,y) = I(x,y)$ if $| Bi(x,y) – Ii(x,y) | < Tbackground$, for all i ; else check illumination change.

b) Algorithm

1. *Start.*
2. *Read a video input.*
3. *Extract the frames from the video.*
4. *Convert each frame to grayscale.*
5. *Consider initial six frames as learning frames for threshold calculation.*
6. *Calculate threshold values for background.*
7. *Consider next ten frames for object detection.*
8. *Compare incoming frame pixels with threshold values.*
9. *If more than 40% of the pixels are changed then it is temporary illumination change, consider next frame else extract the foreground object from the entire frame.*
10. *Deduce boundary for foreground object.*

c) Parameters of the model
1. Video type : AVI consist of moving objects like humans and vehicles
2. Illumination conditions: Varying (outdoor scenes)
3. No of videos: 30
4. Frames per video 30 to 40

3.2 Human/ Object Detection using RGB –D Information

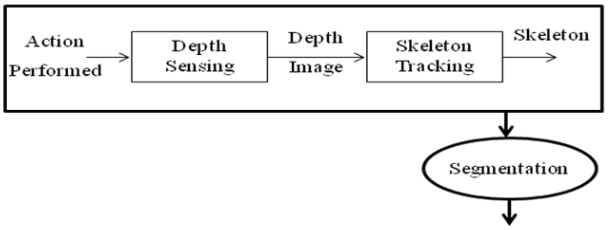


Figure 3: Segmentation using Depth Information

Proper segmentation is the first challenge faced in the activity recognition project. For background removal many methods require that a user have a known, uniformly textured, or colored background. For segmentation, we are considering the depth and color information of the image, into account. To recognize objects that have similar colors depth data can be segmented easier than color images. However, close objects with similar depths (e.g. two close people side by side) can not be easily identified at the same time. Having both depth and color data about the same frame is a quite common thing. Moreover, it is possible to recognize better the objects in the site by exploiting both geometries (depth) information and the color cues. By doing segmentation, we are generating a silhouette image. The steps included in segmentation are as follows:
1) We have done preprocessing on depth image that is used for reduction of noise that is coming from surround objects such as wall.

2) After preprocessing we have done binarization of a picture.

$$Bim(i,j)=\begin{cases}1 & if\ Pim(i,j)>1\\ 0 & if\ Pim(i,j)<1\end{cases}$$

Where, Pim = preprocessed image, Bim = binary image

3) We have obtained locations of standing persons using the sum of columns and rows.

$$Location_{Person\,Y\,axis}=i\ if\ Sum(i)_{horizontal}>T_H$$

We have selected first and last value of in Locationpersons y-axis as the y-axis limits for bounding box. Similar process is done for Locationpersons x-axis.

4) By using location on x and y-axis we have found the width and height of bounding box.

$$Width_{boundingbox1}=L_2-L_1$$

where L2= location(2) person x-axis and L1=location(1) person x-axis

$$Width_{boundingbox2}=L_4-L_3$$

where L4= location(4) person x-axis and L3=location(3) person x-axis

$$Width_{boundingbox2}=L_4-L_3$$

where L2= location(2) person y-axis and L1=location(1) person yx-axis

5) Later we have only selected the pixel locations that are present in bounding box boundaries decided by location of height and width-

$$I=\begin{cases}Bim(i,j) & if\ L_{|x}<i<L_{2x}\ \&\&\ L_{|y}<i<L_{2y}\\ 0 & elsewhere\end{cases}$$

Where, I = segmented image (i,j)
L1x = locationperson x-axis (1)
L2x = locationperson x-axis (2)
L1y = locationperson y-axis (1)
L2y = locationperson y-axis (2)

This gives us the segmented human/object present in the video that is moving. This proposed method is less complex and more accurate as we have not used the continuous updating of the background. Due to the use of depth information along with RGB obtained from Kinect sensor the segmentation becomes independent of illumination condition.

b) Parameters of the model
1. Dataset : MMS 3-D HAR database
2. Illumination conditions: varying
3. Background information: no
4. No of videos: 110
5. Frames per video: 100
6. Type of video: RGB and depth

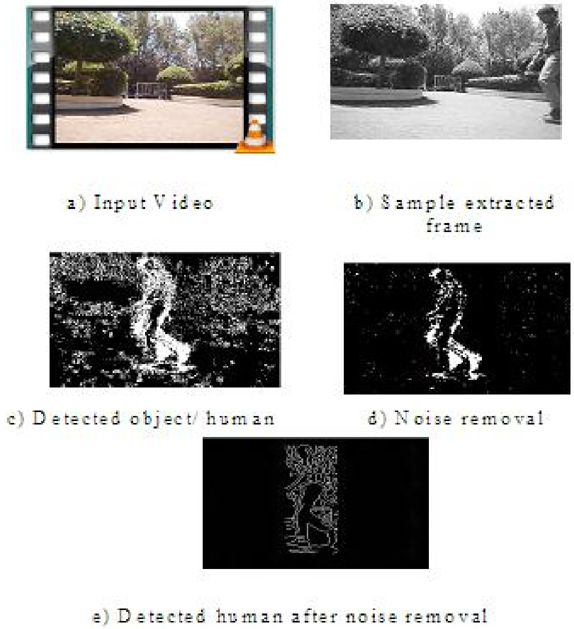
4. RESULTS

Experimentation carried out with model I i.e. object detection and tracking in varying illumination and background shows the results at different stages in detection and tracking of moving object in the video. This simulation is done on total thirty videos recorded by RGB camera in outdoor environment.

a) Result for video containing moving human

Table (1): Segmentation Result for Background Subtraction using Continuous Updation of Background

Table 1: Segmentation Result for model I



a) Input Video

b) Sample extracted frame

c) Detected object/ human

d) Noise removal

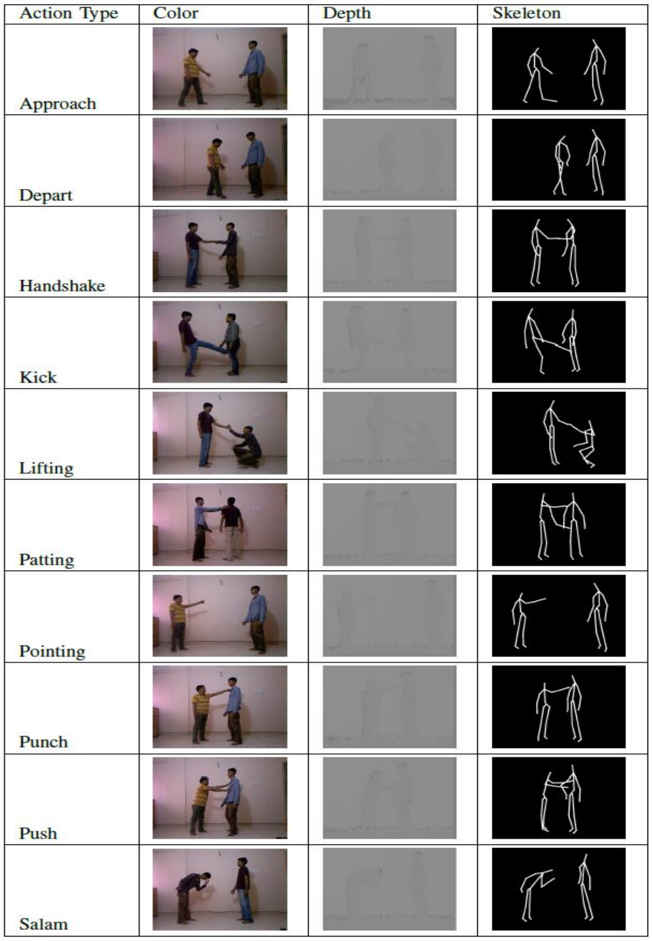
e) Detected human after noise removal

Experimentation on model II i.e. human activity recognition with RGB-D data is carried out on the MMS 3-D HAR dataset comprising total hundred and ten videos for ten activities performed by eleven individuals in four lighting conditions. Results obtained at various stages are given below.

a) Dataset

MMS 3-D HAR dataset consist of total hundred and ten videos of ten human-human activities recorded by eleven individuals in four lighting conditions. Table 4.2.1 shows all these activities and the recorded information

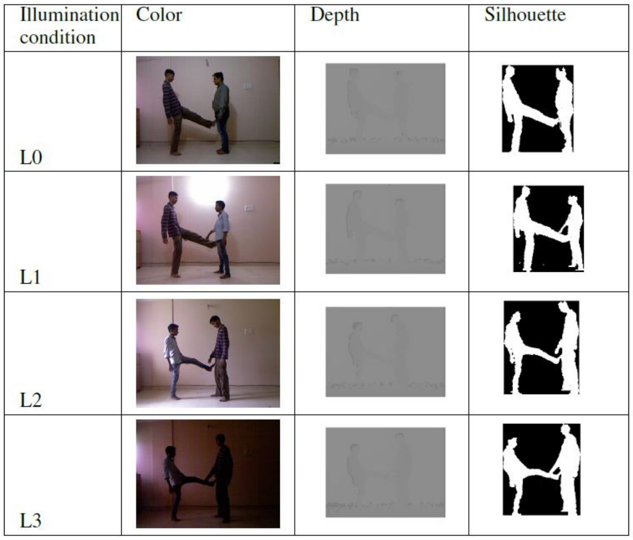
Table 2 : Dataset Categories with Depth and Skeleton Information



b) Segmentation

The first step in the activity recognition is segmentation. Table shows the detection of the silhouette in the different illumination condition.

Table 3 : Segmentation using RGB-D Data in Varying Illumination



From the results, it is observed that there is no effect of changing illumination on the extracted silhouette this is due to the use of depth information shown in the table along with RGB information. This makes the segmentation less complex, less time consuming and more accurate. Time required for segmentation in RGB-D case is around 3 sec whereas if background is updated continuously to tackle the problem of varying illumination it takes around 25 sec for segmentation. Therefore it is difficult to maintain the frame repetition rate required in video processing to make it real time.

Comparison of segmentation algorithm based on time complexity:

The computational and experimental models are compared for the time required for segmentation. Model 1 and model 2 tested with the videos recorded in varying illumination condition. For model 1 the videos are captured by sony pc camera, whereas for model 2 videos are captured by Kinect Microsoft sensor.

It is observed that the time required for model 1 is more as compared to model 2. The very important point which can be noted here is in case of model 2 even the illumination condition is varying the time required for execution is very less as compared to other models.

Table 4 : Comparison of segmentation algorithm based on time complexity

Sr. No.	Object Detection Algorithm	No. of Frames in video	Time Elapsed in sec	Lighting Condition
01	Background Subtraction using continuous updating	40	25.62	Changing
02	Background Subtraction using RGB-D data	100	3	Changing

5. CONCLUSION

In model II depth and color information is used in segmentation because of which segmentation is accurate and also independent of illumination condition. So, it is possible to detect human/ object from the recorded video with varying illumination conditions, which makes algorithm robust. Continuous updating the background is not required. Hence, algorithm becomes less time complex and the possibility of real time implementation increases.

REFERENCES

[1] Han, Sang Uk, et. al. “Empirical Assessment of a RGB-D Sensor on Motion Capture and Action Recognition for Construction Worker Monitoring.” Visualization in Engineering Springer Open Journal, Vol. 1, 2013, pp. 1-13. https://doi.org/10.1186/2213-7459-1-6

[2] Sardeshmukh M. M., M. T. Kolte, and D. S. Chaudahri. “Activity Recognition Using Multiple Features, Subspaces and Classifiers”, Swarm, Evolutionary, and Memetic Compu ting, Springer International Publishing, Vol.2, 2013. pp. 617-624.

[3] Victor Escorcia, Mara A. Dvila, Mani Golparvar-Fard and Juan Carlos Niebles, “Automated Vision-based Recognition of Construction Worker Actions for Building Interior Construction Operations using RGBD Cameras”, Construction Research Congress 2012, 2012 pp. 879-888.

[4] Fanellom Sean Ryan et al. “One-shot Learning for Real-time Action Recognition,” Pattern recognition and Image Analysis, Springer Berlin Heidelberg, Vol.2, 2013, pp. 31-40. https://doi.org/10.1007/978-3-642-38628-2_4

[5] Yamato J., Ohya J., Ishii, K., “Recognizing Human Action in Time-Sequential Images using hidden Markov Model,” Proceedings of IEEE Society Conference on Computer Vision and Pattern Recognition, 1992, pp. 379-385.

[6] Natarajan P., Nevatia R., “Coupled Hidden Semi Markov Models for Activity Recognition”, Workshop on Motion and Video Computing, 2007, pp.1-10.

[7] Li W., Zhang Z., Liu Z, “Action Recognition Based on a Bag of 3D Points, Proceeding of IEEE Computer Society Conference on Computer Vision and Pattern Recognition , Vol.2 , 2010, pp. 13-18. https://doi.org/10.1109/CVPRW.2010.5543273

[8] Poppe Ronald, “A Survey on Vision-Based Human Action Recognition”, Image and Vision Computing Journal IEEE, Vol. 28(6) , 2010, pp. 976-990.

[9] Thomas B. Moeslund, Adrian Hilton, Volker Kruger, “A Survey of Advances in Vision-based Human Motion Capture and Analysis”, Computer Vision and Image Understanding (CVIU) Elsevier Journal, Vol. 104 (2-3), 2006, pp. 90-126. https://doi.org/10.1016/j.cviu.2006.08.002

[10] Neil Robertson, Ian Reid, “A General Method for Human Activity Recognition in Video”, Computer Vision and Image Understanding (CVIU) Elsevier Journal, Vol. 104 (2), 2006, pp. 232-248.

[11] Sangho Park, Mohan M. Trivedi, “Understanding Human Interactions with Track and Body Synergies (TBS) Captured From Multiple Views”, Computer Vision and Image Under- standing (CVIU) Elsevier Journal, Vol. 11 (1), 2008, pp. 220-228. https://doi.org/10.1016/j.cviu.2007.10.005

[12] Michael S. Ryoo, Jake , K. Aggarwal, “Semantic Representation and Recognition of Continued and Recursive Human Activities”, International Journal of Computer Vision (IJCV), Vol. 82 (1), 2009, pp. 124-132. https://doi.org/10.1007/s11263-008-0181-1