Background Subtraction for Stationary and Moving Object Detection

Tilak Nanavati (201811001) & Charmy Patel (201811013)
Guided By: Prof. Pankaj Kumar
Dhirubhai Ambani Institute of Information and Communication Technology, Gandhinagar
Course: **Computer Vision - IT524**

Abstract— **Background subtraction is a technique for separating out the foreground elements from the background. Background subtraction is a major pre-processing step for moving object detection in many vision based applications such as video surveillance systems, human computer interaction, optical motion capture etc. Background subtraction segments moving objects by obtaining the difference between the background image and the current image. It is mainly used in many computer vision applications as a primary step for object recognition, classification and activity analysis tasks. Background subtraction can be applied for extracting information about particular vehicle with the help of images provided by traffic camera as well as content based video coding, real time motion gesture recognition or to keep a track of person,and further it can be used in autonomous vehicles, etc. Here as part of our course project we have investigated some of the key points for the foreground/background separation when the *depth* of the image is also provided. The dataset chosen to carry out the investigation is *SBM-RGBD* dataset as it contains inputs, ground-truth(GT) and depth. It sizes 1.11 GB.**

I. Introduction

Background modeling is a critical component for motion detection tasks. Here in this report we propose a background subtraction method based on frame differencing of the current image with the background image for stationary object detection. We also propose the frame averaging method for moving object detection. The prime objective of this project is to investigate whether adding one more dimension as depth to the given image can the accuracy for object detection be enhanced. The normal RGB images contain 3 channel information per pixel in the form of 3x1 vector. The novelty of RGBD dataset is that it not only provides the color details per pixel but also provides the depth information corresponding to that pixel. So for that dataset we can visualize the pixel as 4x1 dimensional pixel.[1]

To measure the object detection accuracy we take *Percentage Correct Classification* (PCC) and *Jaccard Coefficient* (JC). We can also calculate some more metrics as precision(P), recall(R) and F-score(F). The difference between PCC and JC metrics is that JC does not take into the consideration the True Negative(TN) values.

We have first analysed and investigated the stationary object detection task with the help of simple RGB image then we compared its results with the object detection ground truth(GT).[2] In the next step, with RGB+depth we have tried to detect the object and then again compared the results against the given ground truth(GT). For stationary object detection we are planning to use SBM-RGBD/ IntermittentMotion/ abandoned2 image sequence (size 138 MB). For this purpose we have used the *frame differencing* method for background subtraction.

Then we have analysed and investigated the moving object detection task with the help of *averaging filter* for simple RGB images and then we will compare its results to given ground truth(GT). After that we will take depth into the consideration and using RGB+depth we will detect the moving object. Then using procedure similar to above mentioned, we have compared its results to the ground truth(GT). For moving object detection we are using SBM-RGBD/IntermittentMotion/Sleeping-ds image sequence(size 222 MB). For this purpose we have used the average filtering (mean filtering) method for background subtraction.

We have also investigated the case when the inclusion of depth leads to the removal of shadow effects and produces finer details of the detected object. This is another aspect of this project's key findings that we have successfully removed the shadows with the help of depth in the moving/stationary object detection.

II. Problem Statement

A. *Object shadows mistakenly considered as the object parts.*

By using the conventional approach the shadows of the moving as well as stationary objects are many times misinterpreted as object parts which may lead to wrong object recognition, classification. The depth aspect can be used to mitigate this problem. Depth can be used to put a threshold on the extracted object region and then trimming it to make it particular detected object region.[1]

B. *Light switches or local gradual changes of illuminations*

Sometimes light switches or the local gradual changes of illumination may also result in improper object detection or the incorrect object classification [1]. In such cases the use of depth dimension can be helpful in a way that it is independent of the lightning effects and illuminations which in turn will result in the good discrimination factor for object detection.

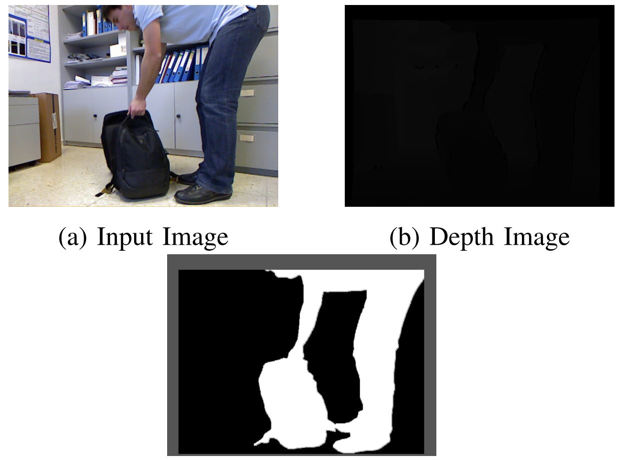
C. *Higher false positive rate in conventional approach.*

The shadow effects and the local gradual changes in the illuminations may result in incorrect object recognition as well as classification. Which eventually results in higher *false positive* (FP) rate. Using depth as the additional dimension has shown significant improvement in this direction. The depth as can be used independently of the pixel RGB intensity values, helps significantly in the correct detection of object regions.

III. Dataset

A. *Stationary Object Detection*

The dataset we used is SBM-RGBD /IntermittentMotion /abandoned2 which contains RGB images with its corresponding depth and ground truth images of size 640x480 pixels. It contains a video of 250 frames.



(a) Input Image (b) Depth Image

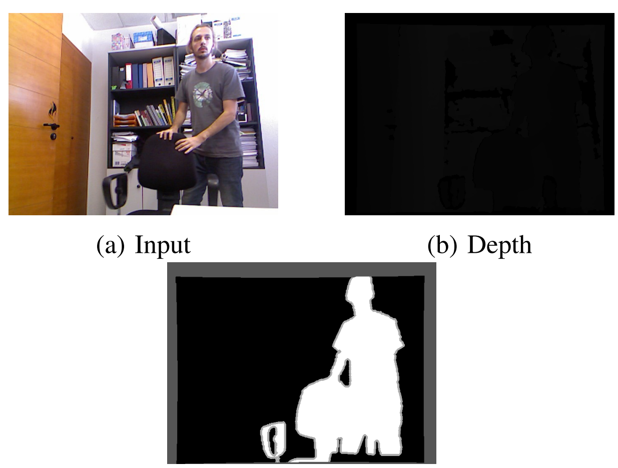
(c) Ground truth Image

Fig. 1: abandoned-2 dataset

Link to Dataset: http://rgbd2017.na.icar.cnr.it/SBM-RGBDdataset.html

B. *Moving Object Detection*

The dataset we used is SBM-RGBD /IntermittentMotion /Sleeping-ds which contains RGB images with its corresponding depth and ground truth images of size 640x480 pixels. It contains a video of 300 frames.



(a) Input (b) Depth

(c) Ground truth

Fig. 2: sleeping-ds dataset

IV. Proposed Approach

Depth data has opened new ways of dealing with the problems addressed above[1]. RGBD Data is collection of **RGB images** with the corresponding **depth images**. The proposed approach is to use the depth data as the fourth dimension of the pixel vector and use it to mitigate the above listed problems.

A. Stationary Object Detection

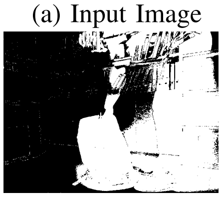
We have used frame differencing for stationary object

1) Without depth dimension [3]
In this we are converting our frames to gray-scale image. Then we subtract background frame from current frame which gives RGB differecnce image. After that we perform thresholding on RGB difference image.

2) With depth dimension [1]
In this again we are converting our frames to gray-scale image. Then we are subtracting background frame from current frame as well as subtracting background depth from current depth frame. After that we perform thresholding on RGB difference image as well as depth difference image.

(a) Input Image

(b) Ground Truth Image



(c) Result (Without depth)

(d) Result (With depth)

Fig. 3: Stationary Object background segmentation result - Frame Differencing

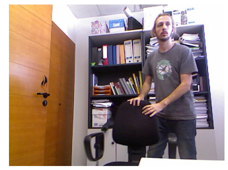
B. Moving Object Detection

We have used average filtering and gaussian method for background subtraction for this.

1) Average Filtering: In average filtering we take the average of the previous 5,10,15,20 frames depending upon the parameter tweaking and the results collectively. Then we subtract the averaged image from the current frame to get the moving object details.

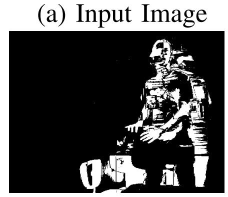
1) Without depth dimension
In this we are converting our frames to gray-scale image.Then we are averaging the previous 20 frames and subtracting it from current frame. After that we perform thresholding on RGB difference image.

2) With depth dimension [1]
In this again we are converting our frames to gray-scale image. Then we are subtracting average of previous 20 RGB frames from current RGB frame. We apply the same procedure for depth frames as well. After that we perform thresholding on RGB difference image as well as depth difference image.



(a) Input Image

(b) Ground truth Image

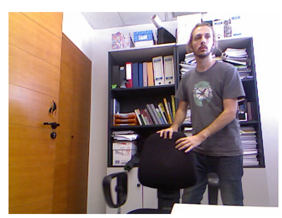


(c) Result (Without depth)

(d) Result (With depth)

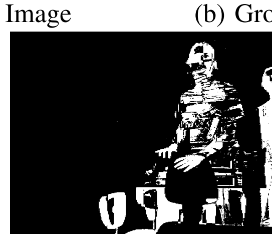
Fig. 4: Moving Object background segmentation result - Average filtering

2) Gaussian Filtering: In this we are using MOG inbuilt method. It uses a method to model each background pixel by a mixture of K Gaussian distributions (K = 3 to 5). The weights of the mixture represent the time proportions that those colours stay in the scene. The probable background colours are the ones which stay longer and more static.[4]



(a) Input Image

(b) Ground truth Image

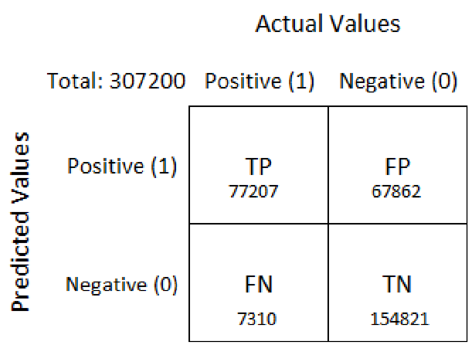


(c) Result

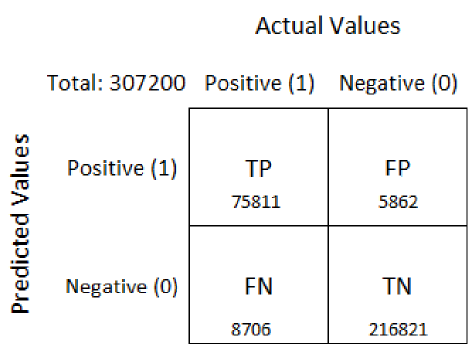
Fig. 5: Moving Object background segmentation result - Gaussian filtering

V. CONFUSION MATRIX

A confusion matrix is a table that is often used to describe the performance of a classification model (or ”classifier”) on a set of test data for which the true values are known. We have obtained following confusion matrices.

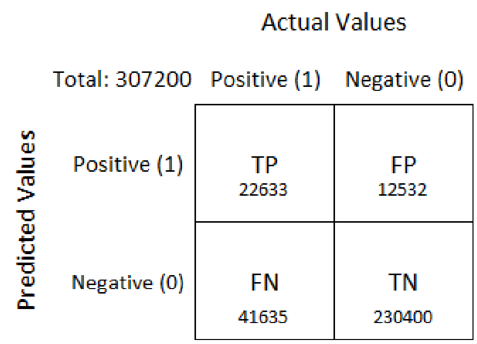


(a) Confusion Matrix (Without depth)

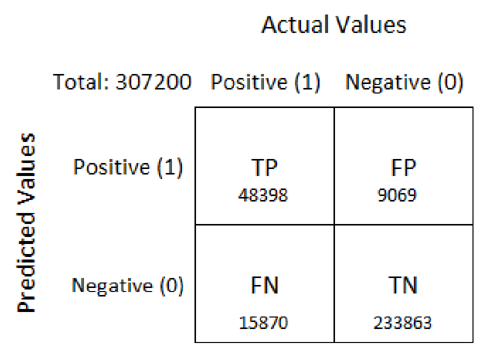


(b) Confusion Matrix (With depth)

Fig. 6: Confusion Matrix for Stationary Object Detection - Frame differencing

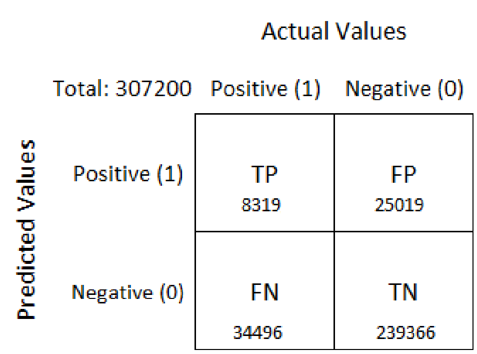


(a) Confusion Matrix (Without depth)



(b) Confusion Matrix (With depth)

Fig. 7: Confusion Matrix for Moving Object Detection - Averaging filter

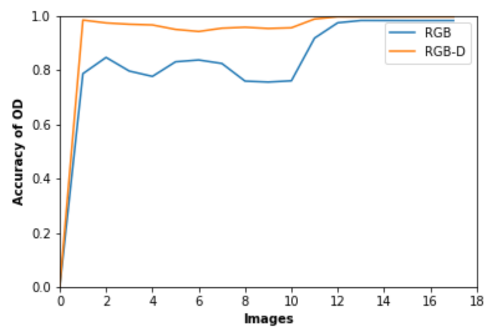


(a) Confusion matrix

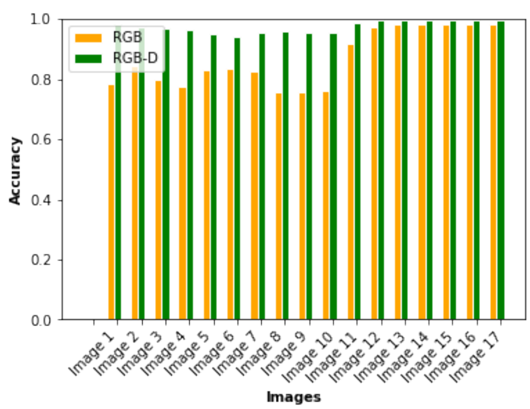
Fig. 8: Confusion Matrix for Moving Object Detection - Gaussian filter

VI. Comparison of RGB and RGBD data

In this section we present the results we obtained using RGB data first and then RGBD data. The plots showcase the head to head comparison when depth apsect is used and when it is not. The results show that when the depth is used as an additional dimension of an image the false positive rate decreases leading to good accuracy and F-score.[5]

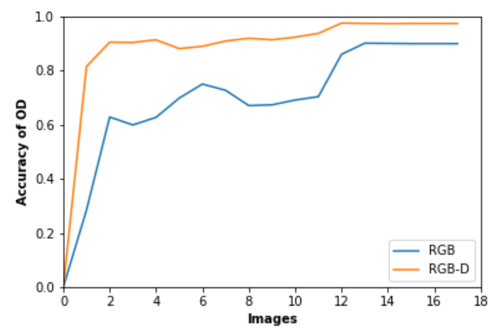


(a) Normal graph

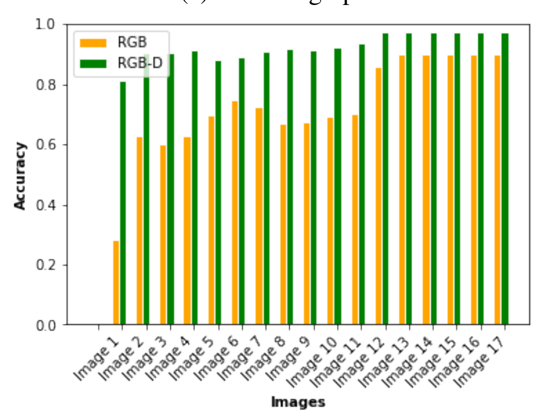


(b) Bar-Graph

Fig. 9: Accuracy graph for stationary Object Detection

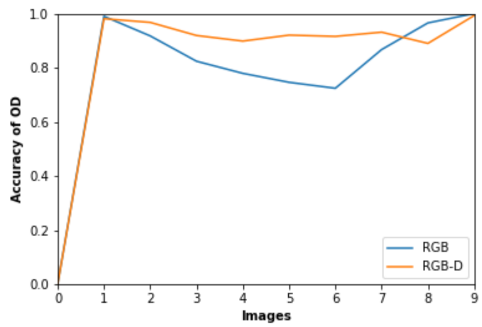


(a) Normal graph

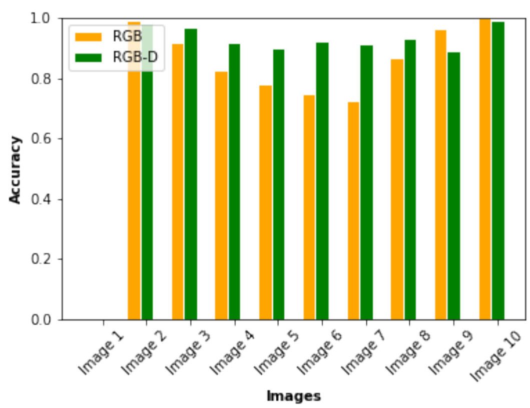


(b) Bar-Graph

Fig. 10: F-score graph for stationary Object Detection

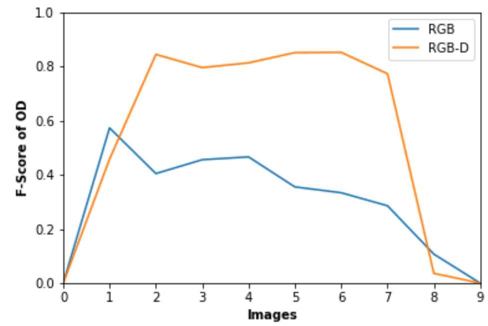


(a) Normal graph

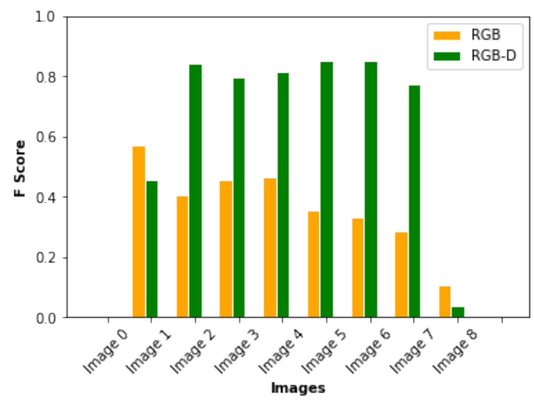


(b) Bar-Graph

Fig. 11: Accuracy graph for Moving Object Detection



(a) Normal graph



(b) Bar-Graph

Fig. 12: F-score graph for Moving Object Detection

VII. Future Work

As part of future work we would like to explore further techniques for background subtraction such as studying the Mixture of Gaussian considering depth parameter.[6], [7] we would also like to explore deep architectures for learning the background subtraction.[8] We would also like to further explore $3{*}\sigma$ rule discussed in class and given in paper which states, $|x_{i,j,k} - \mu_{i,j}| < 3 * \sigma_{i,j}$.[9]

VIII. CONCLUSION

After implementation of above discussed methods we can conclude from the obtained results that using depth we can more accurately detect the object compared to without depth implementations. The shadows are eliminated efficiently as well as the higher false positive rate is also significantly reduced using depth as fourth parameter.

The key take way from this project is utilization of recent hardware technology which can add up some more information and details to the image can help us get good results in terms of accuracy and precision.

References

[1] Lucia Maddalena and Alfredo Petrosino. Background subtraction for moving object detection in rgbd data: A survey. page 27, April.-May. 2018.

[2] Chung-Cheng Chiu, Sheng-Yi Chiu, and Sendren Sheng-Dong Xu. A background subtraction algorithm in complex environments based on category entropy analysis. page 21, April.-May. 2018.

[3] G. Rao and Ch.Satyanarayana. Object tracking system using approximate median filter, kalman filter and dynamic template matching. *International Journal of Intelligent Systems and Applications*, 6, 04 2014.

[4] Shuntaro Takahashi, Hiroyuki Furusawa, Takuya Ueda, and Yoshio Okahata. Translation enhancer improves the ribosome liberation from translation initiation. *Journal of the American Chemical Society*, 135(35):13096–13106, 2013.

[5] Pakorn KaewTrakulPong and R. Bowden. An improved adaptive background mixture model for real- time tracking with shadow detection. 2002.

[6] Y. Song, S. Noh, J. Yu, C. Park, and B. Lee. Background subtraction based on gaussian mixture models using color and depth information. In *The 2014 International Conference on Control, Automation and Information Sciences (ICCAIS 2014)*, pages 132–135, Dec 2014.

[7] Shireen Elhabian, Khaled El-Sayed, and Sumaya Ahmed. Moving object detection in spatial domain using background removal techniques-state-of-art. *Recent Patents on Computer Science*, 1:32–54, 01 2008.

[8] Massimo Camplani, Lucia Maddalena, Gabriel Moyá Alcover, Alfredo Petrosino, and Luis Salgado. A benchmarking framework for background subtraction in rgbd videos. In Sebastiano Battiato, Giovanni Maria Farinella, Marco Leo, and Giovanni Gallo, editors, *New Trends in Image Analysis and Processing – ICIAP 2017*, pages 219–229, Cham, 2017. Springer International Publishing.

[9] P. Kumar, S. Ranganath, Huang Weimin, and K. Sengupta. Framework for real-time behavior interpretation from traffic video. *IEEE Transactions on Intelligent Transportation Systems*, 6(1):43–53, March 2005.