Dedication

I would like to dedicate my research work to my beloved parents,

All My Friends

&

Family Members

whom continuous encouragement, guidance and support remained a true source of motivation for me. I have always looked upon them, as a real source of Morale booster.

Also, to my knowledgeable Instructor Yasar Ayaz, who enacted like a beacon light.

Certificate of Originality

The substance of this thesis is the original work of the author and due reference and acknowledgement has been made, where necessary, to the work of others. No part of this thesis has been already accepted for any degree, and it is not being currently submitted in candidature of any degree.

Badar Ali

2011-MS-NUST-RIME-08

Thesis Scholar

Acknowledgement

All regards to Almighty Allah, the most beneficial, the most merciful and creator of everything what we know and what we don’t.

The current work is an absolute example of the team effort and support offered and rendered by the faculty and staff members of School of Mechanical and Manufacturing Engineering (SMME), NUST.

The contributions and contemporary technical support offered by Assistant Professor Dr. Yasar Ayaz (HOD Robotics & Intelligent Machine Engineering) and Assistant Professor Dr. Muhammad Naveed in development and execution of experimental work on Pioneer 3AT mobile robot using Stereo camera, deserves special recognition and acknowledgement.

An exemplary valuable guidance and support for the wholesome project has been provided by the Guidance Committee members, Assistant Professor Dr. Muhammad Naveed, Dr. Omer Gilani and Dr. Mohsin Jamil.

I would like to thank my family in general and my father, mother and my fiancé in particular for their consistence moral support and sparing their precious share of my time.

Lastly, I sincerely thank to Engr. Zeeshan Baig, Engr. M. Haseeb Hafeez and Engr. Umair Alam whom assistance and support throughout the tenure, made the transformation of an elusive project plan to current work reality.

Badar Ali

Table of Contents

Table of Figures

Abstract

Human-robot interaction is a very fundamental need for assistive and service robots. In order to interact with human beings, detection and tracking of people is a basic requirement. A great deal of work has been done in this area and various sensors have been utilized to reach the solution. Generally laser range finder and cameras have been the most common devices used for human detection but still there is a huge problem of human tracking in the presence of multiple human beings and at various illumination levels.

This research work presents a new approach of human detection and targeted person tracking by a mobile robot. The major contribution is the development of a novel Motion Detection based Tracking scheme which utilizes the differential images of three iterative frames of live video stream. Then the comparison is made with three other conventional techniques; cam shift, LK optical flow and particle filter. The robot is equipped with stereo camera and sonars. Human detection is done using Haar cascade classifier, followed by the implementation of a tracking method on the selected target person. Comparisons of these techniques are based on testing in different environmental conditions. Afterwards, Kalman filter based prediction and correction is done to accurately localize the target. Moreover stereo correspondence is done by template matching followed by linear triangulation, to get the 3D location of the target.

Thus, this research work presents the implementation of various techniques that make the robot to select a target person in the environment and then track and follow that target. Results of all these methods are presented and their evaluation is done in detail.

Chapter 1
INTRODUCTION

1.1 Introduction

Human detection and tracking has been a very popular field of research in last few years. The applications include surveillance, intelligent monitoring system, intelligent transportation, human-machine interaction, virtual reality and many more. Service and assistive robots are gaining importance due to their ability to serve and assist humans in homes, hospitals and offices. Human-robot interaction is therefore a very emerging field due to gaining involvement of robots in daily life acting as assistants, team-mates, care-takers and companions.

Presently, various methodologies (experimental approaches) are being employed for detection and following of human beings by mobile robots. Majorly used sensors for the completion of task are laser range finder and cameras. Few methods used the data fusion of laser range finder and omni-directional camera while others have only used stereo camera. Different methods have been implemented which can do target person tracking with in certain limitation of presence of other human beings and various vision related environment conditions.

In current research work, a new method of human detection and target person tracking is implemented which utilizes four different tracking techniques to tackle the problem of multiple human being's presence. Our major contribution is the development of a novel human tracking method which utilizes the computation of differential images to detect motion of target person around the image. Comparison of our novel method is also done with other conventional methods under various environmental conditions to evaluate the performance. Moreover, the method has been successfully implemented in the presence of multiple humans and under various illumination environments.

1.2 Aims

The aim of this work is to develop a novel approach of human detection and targeted person tracking by a mobile robot, majorly focused on development of novel tracking scheme and its comparison with existing conventional methods.

1.3 Objectives

The objectives of this research work are;

- Detection of all human beings in robot’s field of view and target person selection.
- Implementation of conventional tracking techniques for targeted person tracking in the left camera of stereo system.
- Implementation of stereo correspondence in the right camera of stereo system.
- Localization of target person in space with robot’s frame of reference.
- Steering of robot towards target person by maintaining a specific distance and angle from target.
- Development of novel differential images based tracking scheme.
- Comparison of our tracking method with other conventional methods.

1.4 Thesis Outline

In Chapter 2, literature related to various techniques / approaches has been reviewed. In Chapter 3, proposed method of human detection and target person tracking is presented and details of implementation of various tracking techniques are also explained. Chapter 4 presents the experimental results and analyzes the comparison of various techniques. Finally the conclusion and future prospect has been presented in Chapter 5.

Chapter 2

LITERATURE REVIEW

2.1 Fast Human Detector

Fast human detector is a human detection technique using cascade of classifiers based on oriented gradient histogram of images. This method is based on Histograms of Oriented Gradients (HoG) which describes the appearance of human body by the statistical distribution of orientation and magnitude of image gradient [1]. Overview of the method for visual object recognition with linear SVM based fast human detection is shown in Figure 2 - 1.

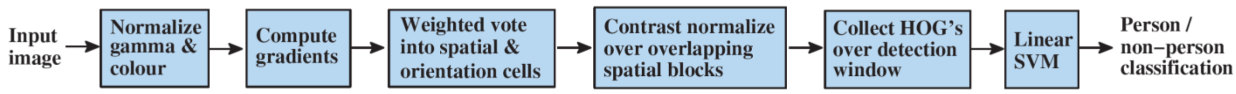


Figure 2 - 1: Fast Human Detection based Feature Extraction & Object Detection. [1]

2.2 Kinect Camera based Human Detection

Kinect is a motion sensing input device by Microsoft for Xbox 360 and Xbox One. In [2] a model based approach has been presented in which human detection is done by using a 2-D head contour model and a 3-D head surface model. A segmentation scheme has been presented for the detection of human beings from the environment. The overview of segmentation scheme presented in [2] is shown in Figure 2 - 2.

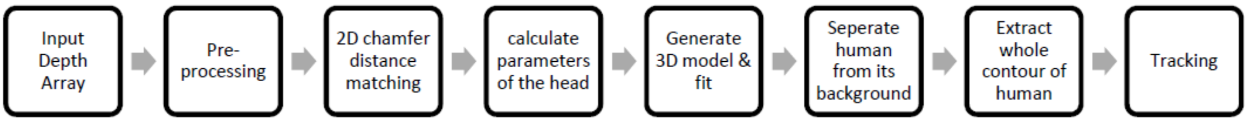


Figure 2 - 2: Human Detection using 2-D Head Contour & 3-D Head Surface Model. [2]

Another method of human gesture recognition using data mining classification methods by various human body joint positions captured by Kinect camera is presented in [3]. Three gesture patterns are stand, sit down and lie down and classification methods are back propagation neural network, support vector machine, decision tree, and naive Bayes. Furthermore they have also evaluated these classification methods. Overview of method presented in [3] is shown in Figure 2 - 3.

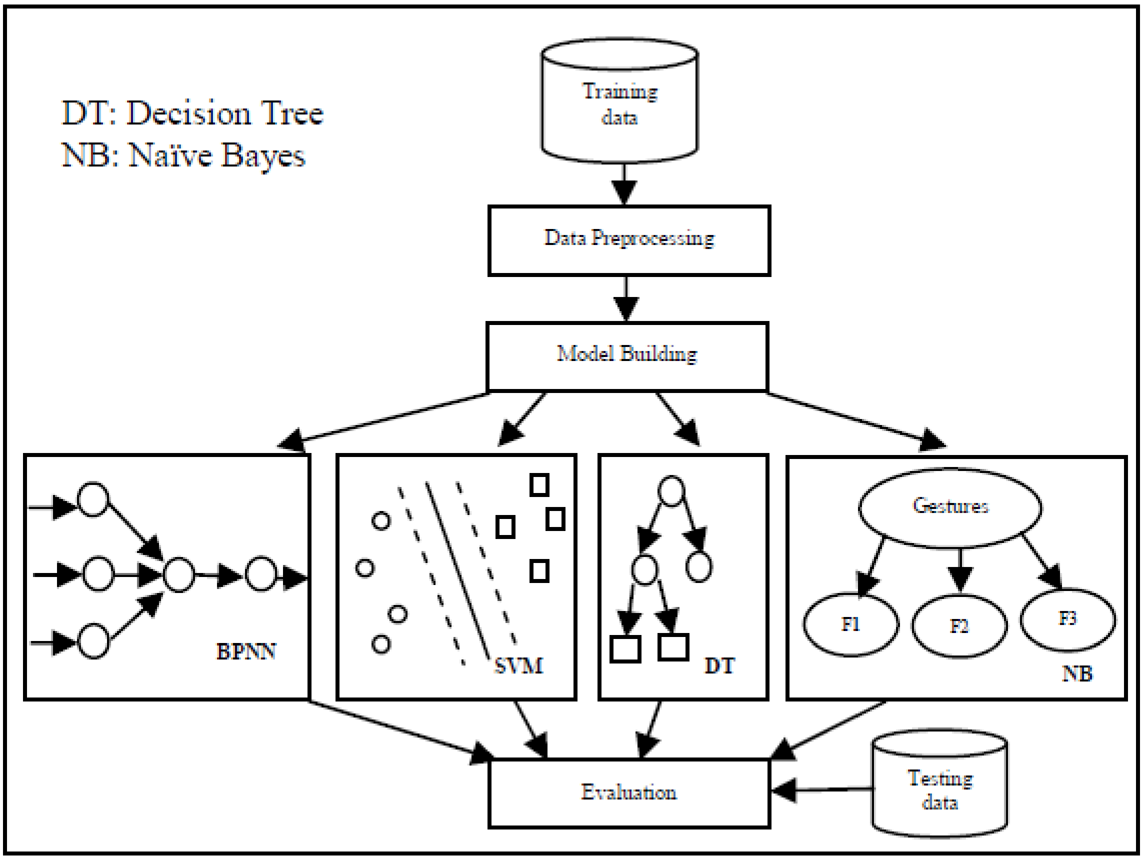


Figure 2 - 3: Human Gesture Recognition using Kinect Camera. [3]

2.3 Object Detection using Haar like Features

Object base detection technique presented uses a cascade of Haar like features to detect a particular object in an image. This method was proposed by Paul Viola et al., [4]. Haar like features are basically adjacent rectangular regions of pixel intensities at specific positions. The position of rectangles is used in a detection window to search target in image. Set of Haar like features [5] for edge, line, center round and diagonal line features is presented in Figure 2 - 4.

The method presented in [4] and [6] for Haar based object detection works primarily on three components. First is the generation of “Integral Image”, which is a representation of image that allows the features computation quickly. Second is the implementation of a learning algorithm, i.e. AdaBoost, which selects small amount of visual features from a larger set. Third is a method of cascading more complex features that allows quick discarding of background regions.

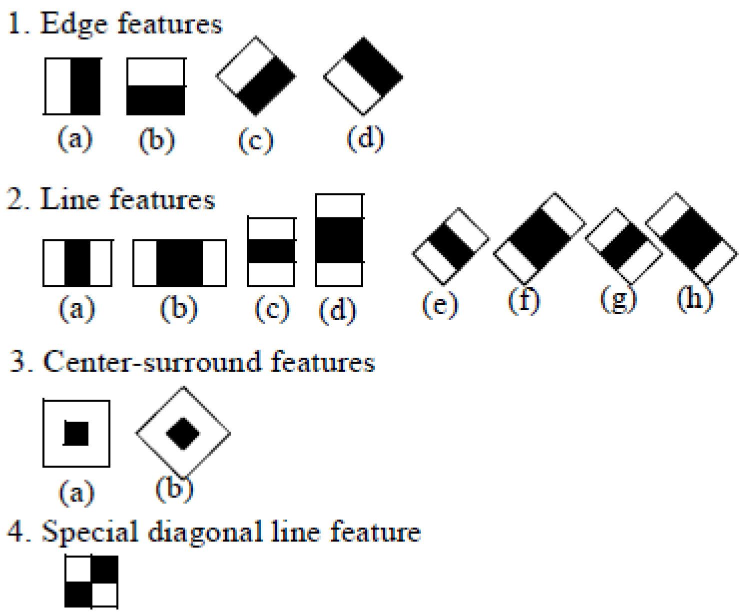


Figure 2 - 4: Set of Simple Haar like Features. [5]

In [7], evaluation of various Haar based detectors is presented. These public domain detectors are frontal faces, head, shoulder, left eye, right eye, eye pair, nose, mouth, upper body and lower body. All these detectors use Haar like rectangular features for their detection [7].

2.4 Mean Shift / Cam Shift

Mean Shift clustering technique is a tracking method used for tracking a particular object in a stream of images. Real time tracking of non-rigid objects using mean shift method is presented in [8]. The idea behind this technique is the iterative calculation of mean position of all the points (centroid) that lie within the kernel window. The next step is to translate the kernel window towards that centroid. The points are basically the detected features from a region of interest which needs to be tracked. In proposed work, it is actually the histogram of Hue component of HSV image. Figure 2 - 5 shows the mean shift vector of a region of interest.

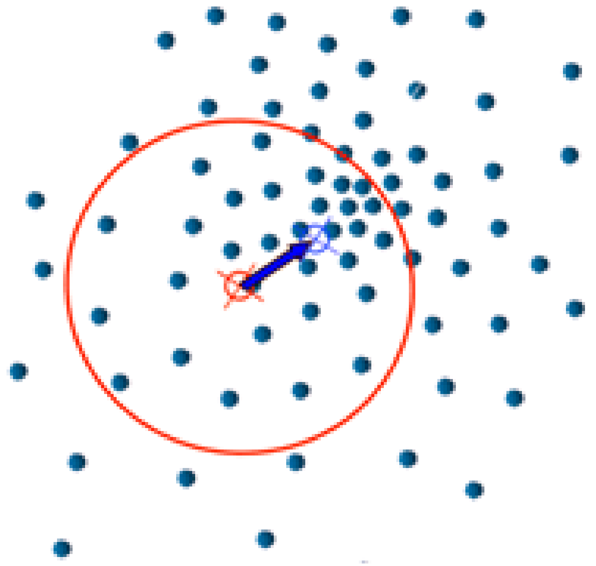


Figure 2 - 5: Mean Shift Clustering based Tracking.

Cam-Shift (Continuously Adaptive Mean-SHIFT) is an extension to mean shift which after finding the object center, adjusts the window size and find the optimal rotation. An improved object tracking method using Cam shift and Grey prediction model for good tracking in occlusions is presented in [9].

2.5 Lucas Kanade Optical Flow

Optical flow is the pattern of apparent motion of objects, in a video due to relative motion between camera and the scene. In actual, the method tries to compute the motion between two image frames taken one after the other. An Iterative Image Registration Technique based on optical flow is presented in [10]. Another Pyramidal Implementation of the Lucas Kanade Feature Tracker is presented in [11].

L-K optical flow method for tracking a template T(x) in an image from time t = n to time t = n+1, is the extraction of template T(x) (5 × 5 window) from the image at t = n in the image at t = n+1. This method takes an assumption that there is a constant flow in the neighborhood of the pixel. It then solves the optical flow equations of all the pixels of neighbor through least square method, followed by the combination of the information from neighboring pixels to resolve the ambiguity of optical flow equation.

The basic optical flow equation is given as:

$$I(x, y, t) = I(x + \Delta x, \quad y + \Delta y, \quad t + \Delta t)$$

Where, (x, y) are the points in the image at time t with the intensity I (x, y, t) and (x + Δx, y + Δy) are points at time t + Δt with the intensity I (x + Δx, y + Δy, t + Δt).

2.6 Particle Filter

Particle filter is the non-parametric sampling based Bayesian filter which estimates the position by a finite number of parameters. Theory and applications of particle filter has been discussed in [12]. Particle filter uses random sampling method to model the posterior probability. It uses N weighted discrete particles to estimate the probability by observing the prior data. Each particle has a state vector x along with its weight w. Then resampling and the posterior approximation are done.

Particle filter method, [13], can be divided into four steps: initialization, sampling, estimation, and selection. During sampling, each particle is taken from the distribution and then the weights are calculated based on probability in the distribution of each particle. Then, the posterior density of each particle is approximated based on its weight. Afterwards, particles are selected according to their approximated posterior density from previous step to get uniform weight distribution.

2.7 Existing Methods for Human Following Robot

In [14], Ren C. Luo et al. presented their developed scheme of a mobile assistive companion robot by linking a vision sensor and a laser range sensor to follow the target person. They have further used a speech and sound system detection method to target the issues of external environmental factors to attain sound source localization and speech interaction between target and robot. When the target gets lost during the tracking process, the robot informs the user, and waits for a clapping sound to re-localize target's location.

Songmin Jia et al., [15], presented a human detection and tracking approach using stereo vision and extended Kalman filter based methods. They used Hu moments of head and shoulder contour to recognize humans. Their contour filter processing is shown in Figure 2 - 6. They further used EKF to predict the location and orientation of target after contour processing. Further improvement in the robustness and real-time performance of their algorithm has been presented in [16] and complete algorithm is shown in Figure 2 - 7.

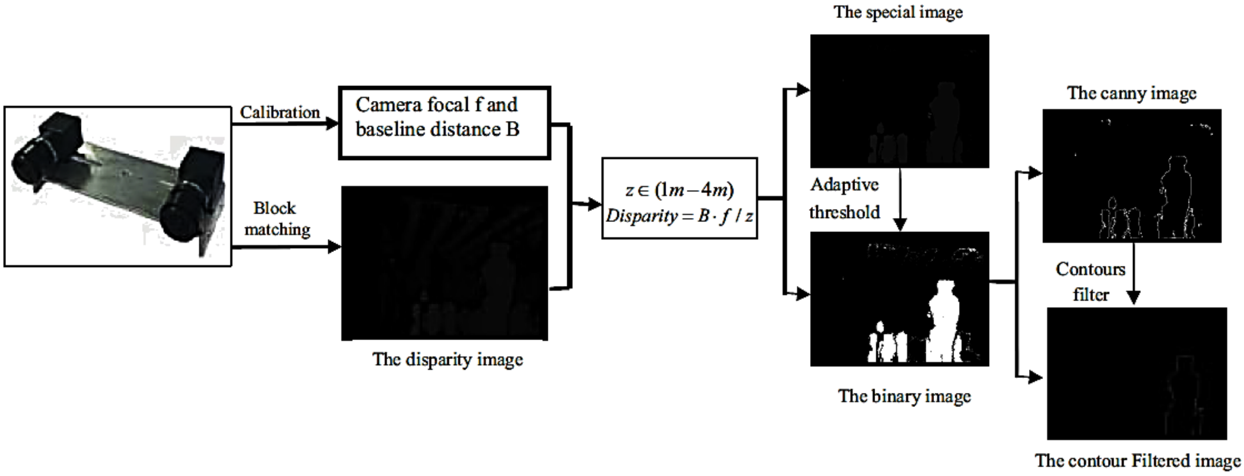


Figure 2 - 6: Human Contour Filter Process. [15]

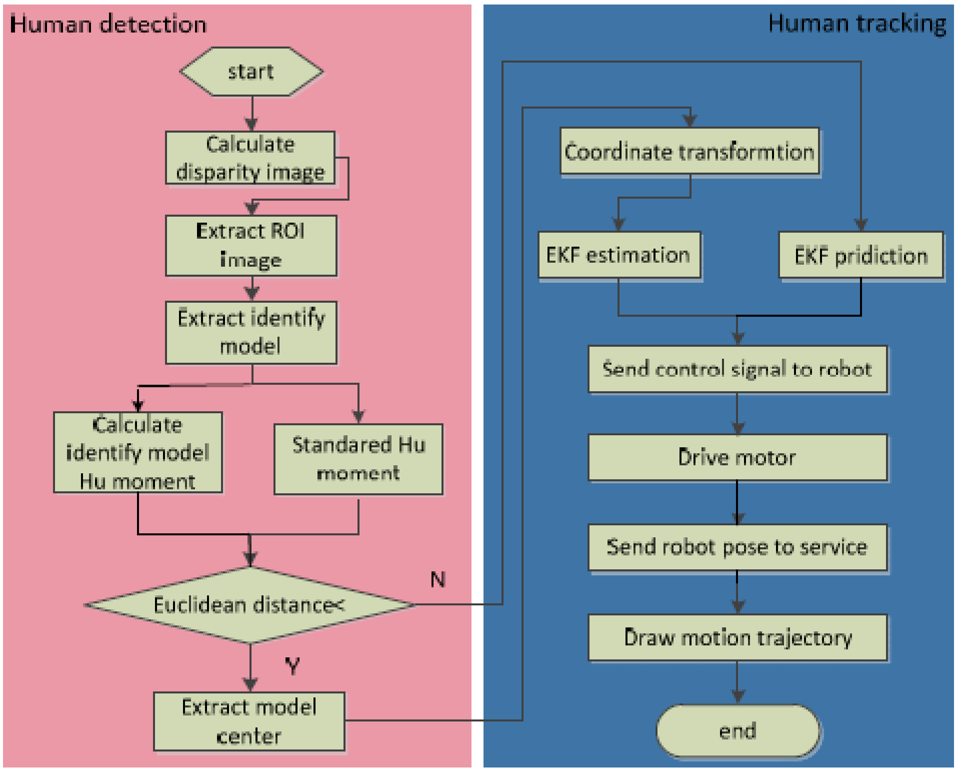


Figure 2 - 7: Human Detection & Tracking using Hu Moment and EKF. [16]

In [17], Mehrez Kristou et al. presented their approach for identifying and following the targeted person using laser rangefinder and omnidirectional camera. Their method fused the data of both these sensors by identifying the target using panoramic image and then tracking that targeted person using laser rangefinder. They initially register the human being by entering his data using camera and then performed localization using data fusion of LRF and Camera. Their method for registration and identification / localization is presented in Figure 2 - 8 and Figure 2 - 9 respectively. Further improvement in their approach is presented in [18].

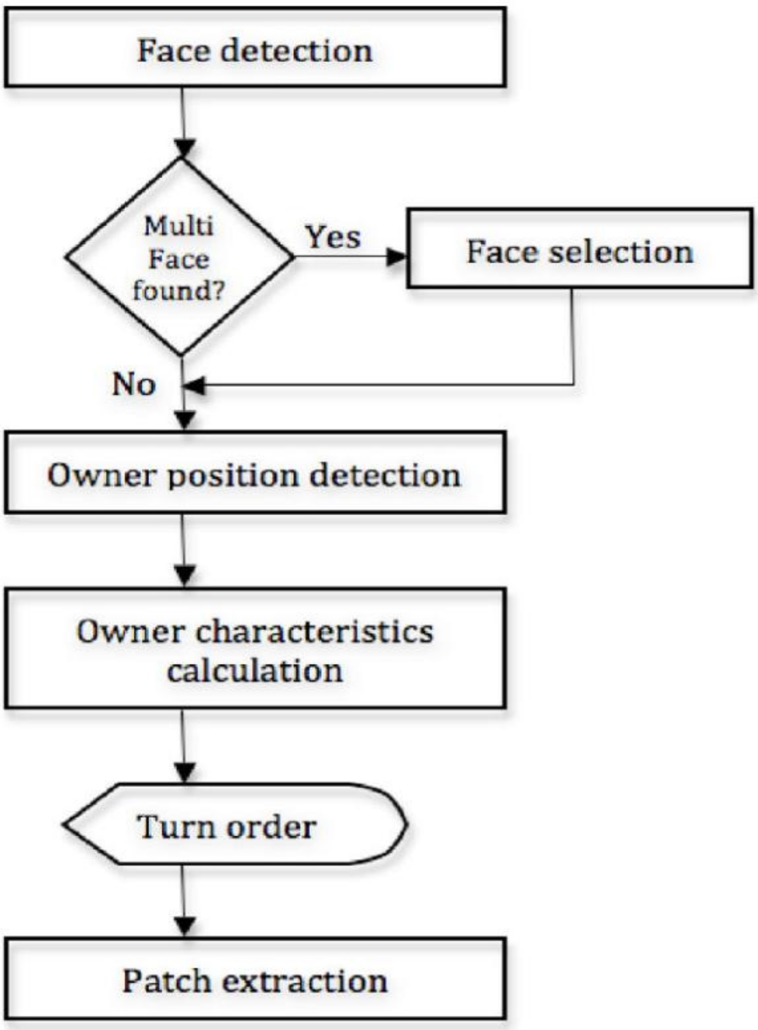


Figure 2 - 8: Human Registration Stage. [17]

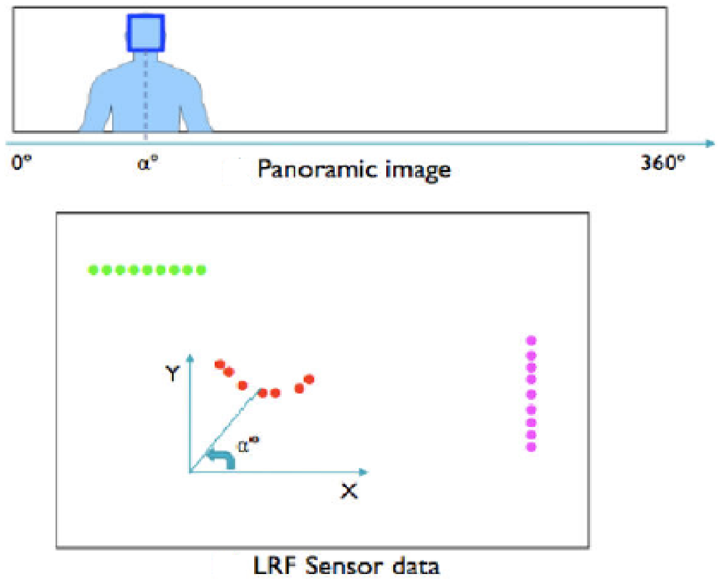


Figure 2 - 9: Face Localization in Panoramic Image & LRF Angular Position. [17]

Emina Petrović et al., [19], presented real-time vision based tracking method using modified Kalman filter. They used a stereo vision based detection method to get the features from 2D stereo images and then reconstructed them to into 3D object features to detect human beings. Block diagram for their human detection system is shown in Figure 2 - 10, whereas the block diagram for their human tracking system is presented in Figure 2 - 11.

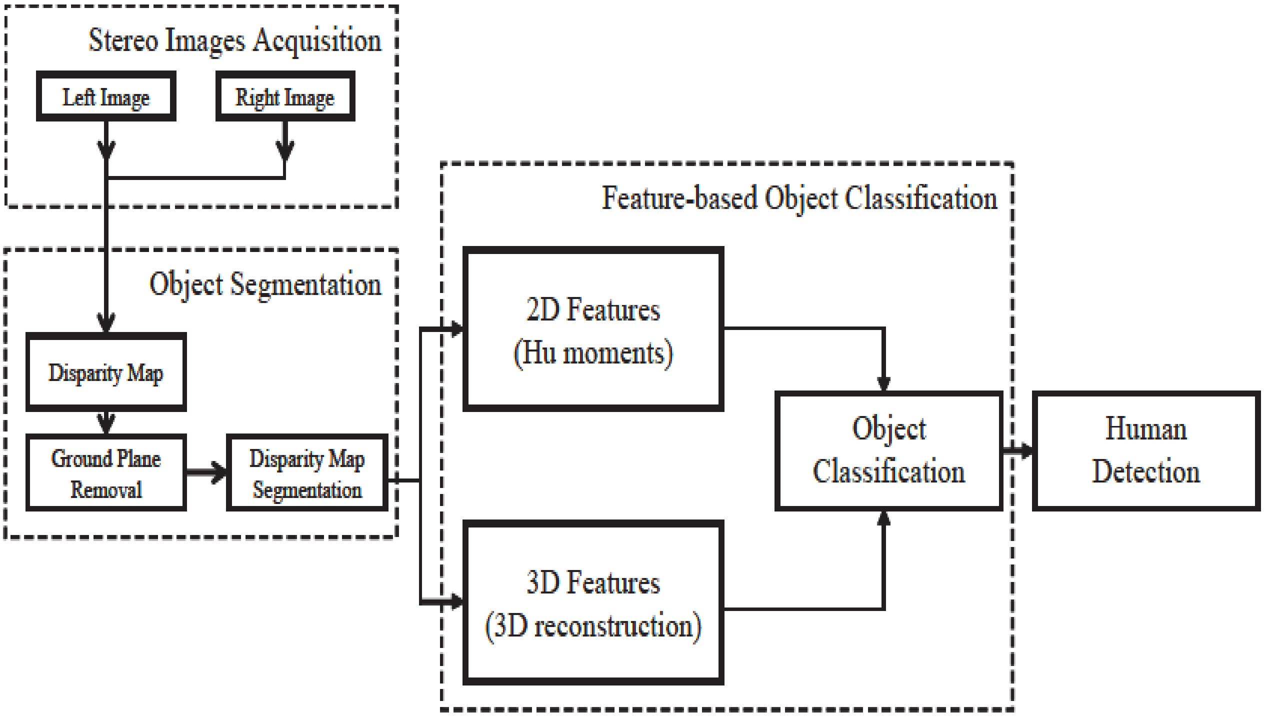


Figure 2 - 10: Block Diagram for Stereo Vision based Human Detection. [19]

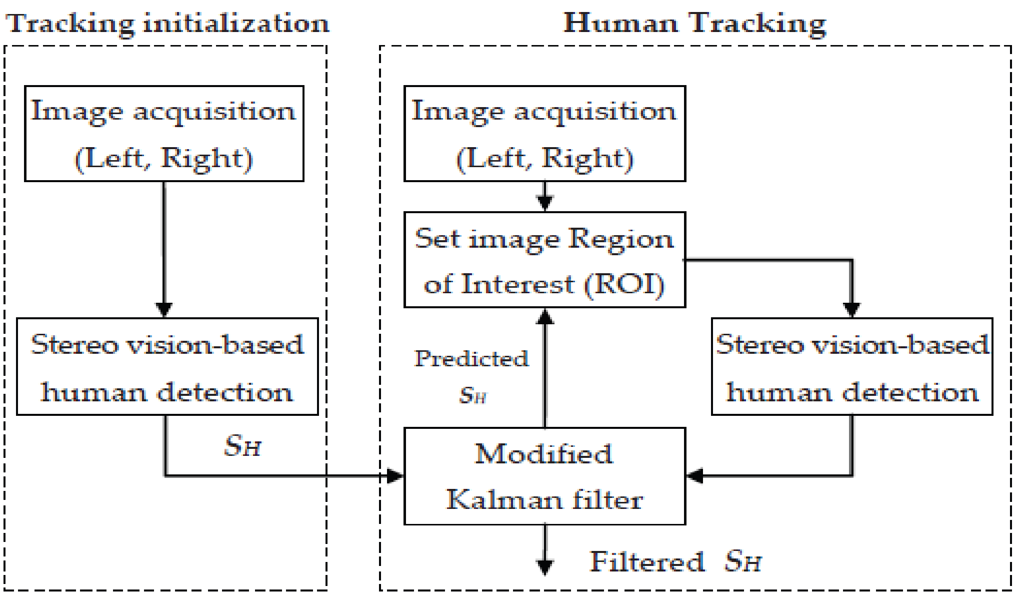


Figure 2 - 11: Block Diagram for Modified Kalman Filter based Human Tracking. [19]

In [20], Liyuan Li et al. proposed another system which integrates multiple vision models for multi person detection and tracking for social robots. Their method is maximum likelihood (ML)-based algorithm which combines the multi model detections in mean-shift tracking. First, a likelihood probability which integrates detections and similarity to local appearance is defined. Then, an expectation-maximization (EM)-like mean-shift algorithm is derived under the ML framework. In all iteration, the E-step estimates the associations to the detections, and the M-step locates the new position according to the ML criterion. To be robust to the complex crowded scenarios for multi person tracking, an improved sequential strategy to perform the mean-shift tracking is proposed. The methodology used by them is presented in flowchart of Figure 4 - 12.

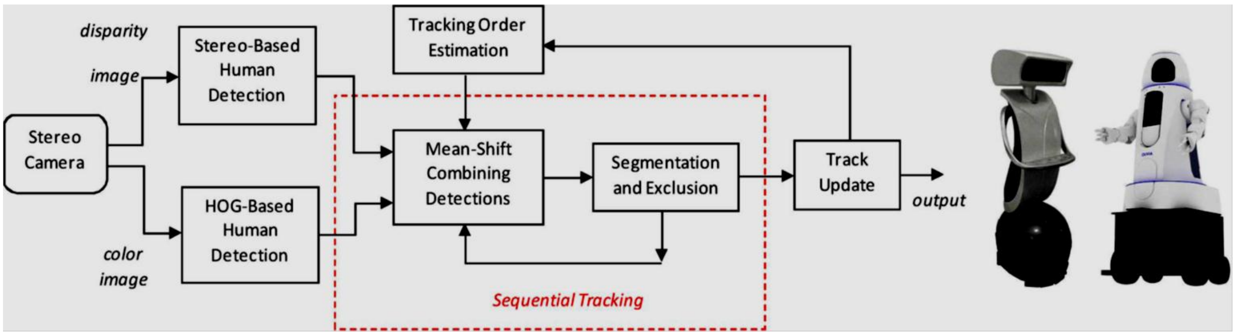


Figure 2 - 12: Block Diagram for Sound Source Localization based Tracking. [20]

Chapter 3

HUMAN DETECTION & TARGET PERSON TRACKING BY A MOBILE ROBOT

3.1 System Components

Our system for human detection and target person tracking is composed of hardware platform and software libraries which are described below;

3.1.1 Hardware Platform

Hardware platform of our system, as shown in Figure 3 - 1 (a), includes Pioneer 3AT mobile robot [21] and stereo camera [22]. In addition to this, the mobile robot is equipped with sixteen sonars (eight at front and eight at rear) which are used to avoid collision with the surrounding objects. Stereo camera is mounted at a height of 4.5 feet from ground to place it nearly at human chest level. Brief description of the hardware components are described below;

- ***Pioneer 3AT:*** Pioneer 3AT [21], as shown in Figure 3 - 1 (c), is a state of the art four wheel drive robotic platform from MobileRobots. P3-AT has an embedded computer; Ethernet based communications, laser, and other autonomous functions. P3-AT also has 8 forward and 8 rear sonars to sense obstacles from 15 cm to 7 m. P3-AT has the ability to reach at speed of 0.8 meters per second and can carry a payload of 12 kg. It also has the option of incorporating laser rangefinder for building a 2D map of the environment. P3-AT is compatible with various soft wares like ARIA, Mobile Sim, Mapper3, and Player/Stage.

  Stereo USB Camera: Minoru 3D webcam [22], as shown in Figure 3 - 1 (b), is a high speed web camera, fully compatible for stereoscopy. The two cameras are spaced at a distance of 55 mm from each other which is quite near to the distance between human eyes. If both the cameras are used together for video input, a resolution of 320 × 240 is required for each to work properly with a frame rate of 18 fps.

(a) (b) (c)

Figure 3 - 1: Hardware Platform of the System.
(a) Person Standing in front of Robot. (b) Stereo USB Camera.
(c) Pioneer 3AT Robot with Laptop.

- ***Laptop for Processing:*** For processing all software related libraries, Fujitsu AH 530 laptop is used. The laptop is having a Core i3 2.53 GHz-380M processor with 4 GB of Ram and 320 GB of Hard Disk. The laptop is provided with an operating system of Linux Ubuntu 12.04 to make a Server – Client communication with Pioneer 3AT mobile robot.

3.1.2 Software Libraries

In order to support the hardware, two software libraries of C++ are used: Player Cc [23] and Open CV [24]. Brief descriptions of these libraries are presented below;

- ***Player Cc:*** The Player project [23] is free software that enables research in robot and sensor systems. It is the most widely used robot control interface released under the GNU General Public License. Player Cc is a C++ platform used to establish server clien2t communication with various mobile robots. This library enables the robot's hardware to be controlled using standard C++ commands.
- ***Open CV:*** Open CV [24] is an open source computer vision library released under a BSD license. It has C++, C, Python and Java interfaces that supports Windows, Linux,

Mac OS, iOS and Android. Open CV is designed for computational efficiency and have many real-time applications. Open CV is used to perform all the computer vision related operations. This library has various built in functions that make it easy to implement various methods over the live stream of videos.

3.2 System Implementation on a Mobile Robot

The System Implementation section is classified into three sub sections. Firstly, the method adopted at initial level of research (Phase-I) is presented. Afterwards, the modifications are presented. Thirdly, the final method adopted in this research is presented.

3.2.1 Phase I

Flowchart of Figure 3 - 2 presents an overview of initially proposed method. The method has also been presented in [25] and [26]. As shown in the flow chart our system initially detected all the human beings found in front of the robot. It uses left camera to find the faces. This was done by using Haar Cascade Classifier. Then the user was asked to choose a target person from all the detected persons. When the user enters the target person's index, the robot instantly calculates the width and height of upper body and face by using stereo camera. It also calculated the legs width using laser data. Face / Upper body is detected in the left camera's image using the same Haar object detection approach whereas the right camera finds the detected face / upper body using the template matching method.

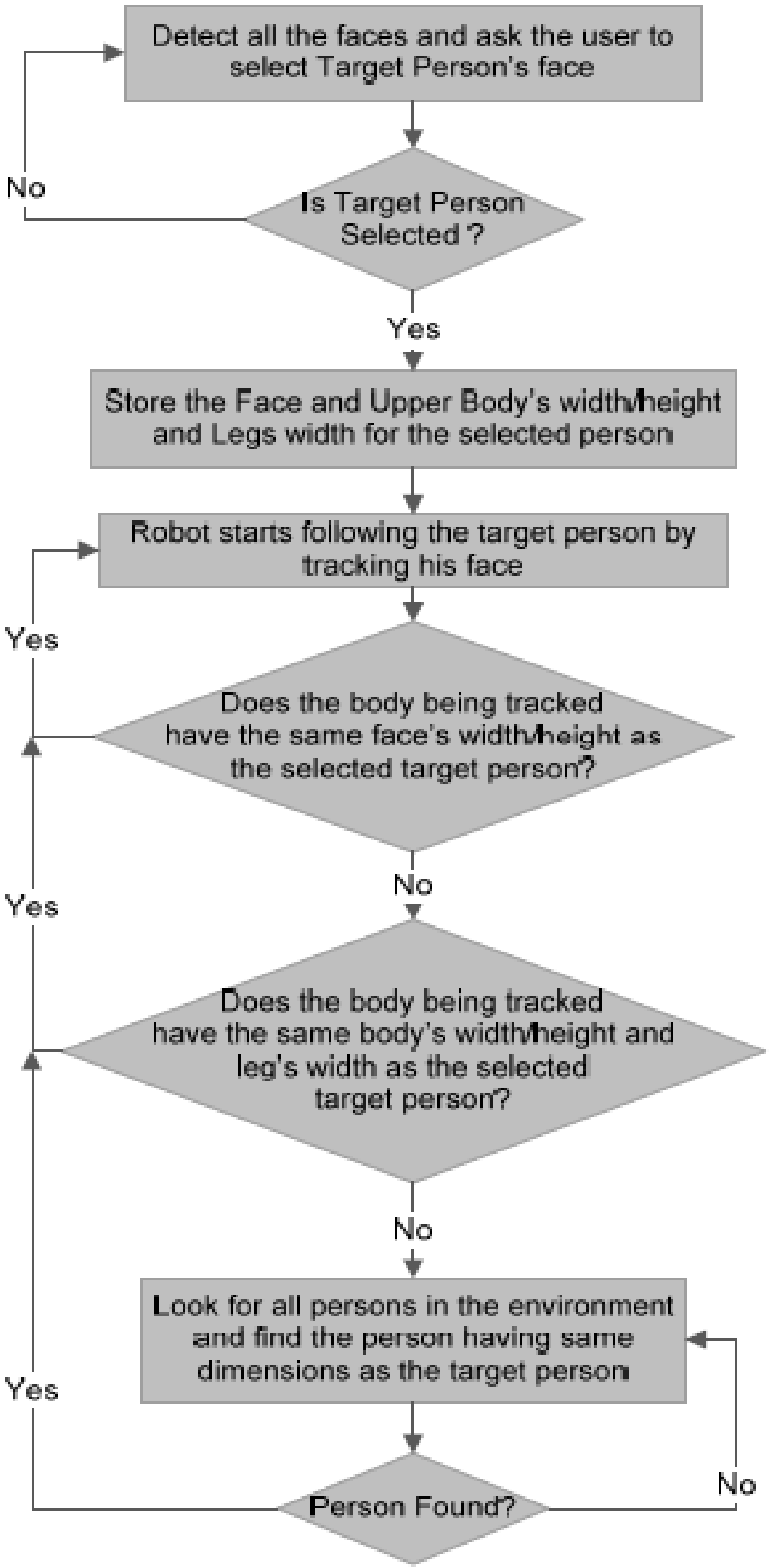


Figure 3 - 2: Block Diagram of Initially Proposed Method. [26]

On the other hand LRF was used to get the legs data based on the calculation of Euclidean distance between the detected legs. Thus when the target person is selected and his features are stored in the robot's memory, the next step of our algorithm is to follow the target person. For this we used Cam Shift tracking. So the left camera tracks the target person based on the detected face in the previous frame and then simultaneously Template of the tracking result from left camera's image is searched in the right camera's image to find the stereo correspondence between the two stereo cameras.

When the coordinates of tracked human being is found in both the camera's images, next step is to perform triangulation to find the distance and angle of the target person with respect to the camera. For this we have used Linear Triangulation technique. Triangulation will give the distance of robot from the camera. Thus the next step is to navigate the robot near to target person and maintain a distance and an angle from him. Thus our system has used Player Cc library to navigate the Pioneer 3AT mobile robot and the robot is programmed to maintain a distance of 1.5m±0.2m and an angle of $\pm 10^0$. Hence under normal circumstance the robot keep on following the target using Cam shift tracking and then template matching by maintaining the said distance and angle.

In case the robot has lost the target person, the robot will then make use of the stored data of target person and will start searching all the human beings in its proximity. Then the robot will find the person with the same dimensions as the previously stored legs, upper body and face data for the target person and will start following the target person. The complete approach for human detection, selection and tracking has been summarized in the flowchart of Figure 3 - 2.

3.2.2 Phase II

Flowchart of Figure 3 - 3 presents an overview of modified method that used two more tracking techniques. The proposed method initially looks for all the humans in front of the robot. This is done by using Haar feature based detection. When humans are detected, the control program asks the operator to select one target person. After the operator selected the target person the robot starts following the target by maintaining a distance of 1.5m and angle of $\pm 10^0$ from him.

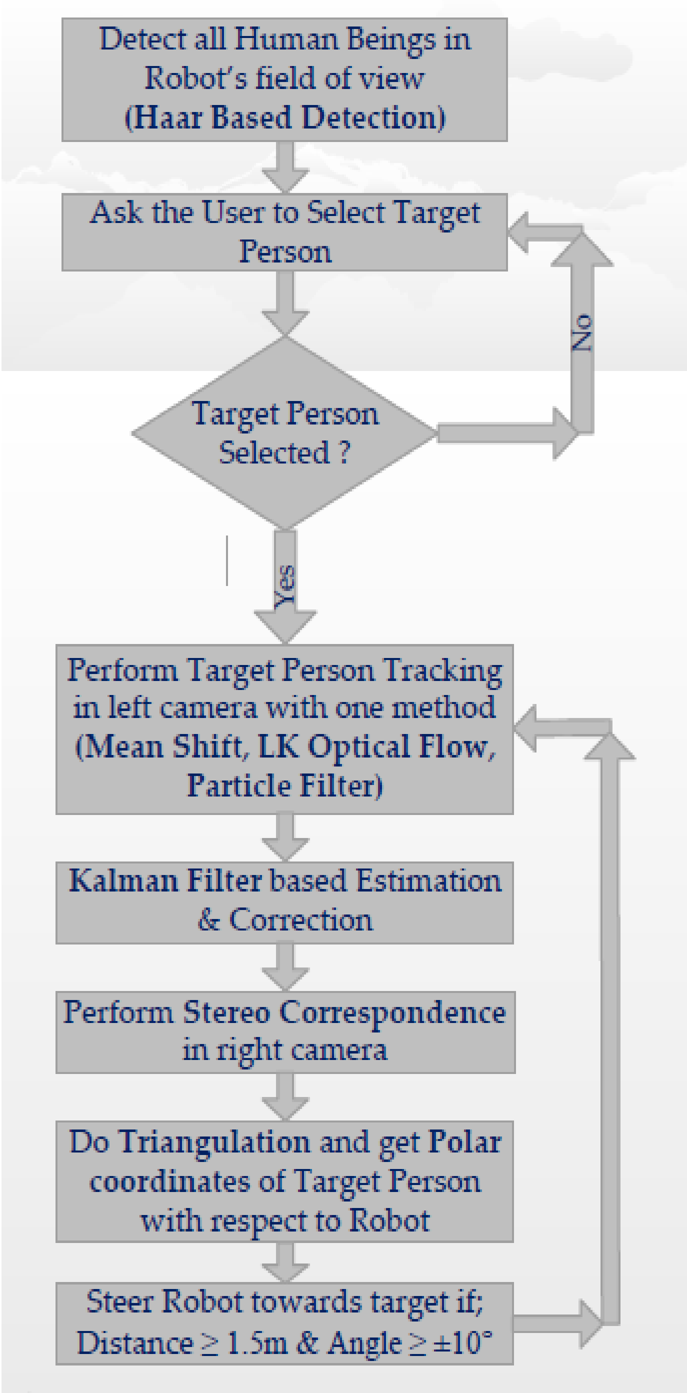


Figure 3 - 3: Block Diagram of Final Modified Method. [26]

The tracking is done using one of the three tracking techniques at a time: Mean shift, LK optical flow and Particle filter. These tracking techniques are also implemented independently to make their comparison with each other.

Furthermore, these techniques are facilitated with Kalman filter based estimation to assist more accurate localization of targeted person with respect to the robot. Afterwards, Stereo correspondence is done to get the position of target in right camera. Then using Linear Triangulation, the position of target in space is calculated which is used to steer robot towards target. Implemented methodology of all processes involved during target selection and tracking are explained below.

3.2.3 Phase III

Flowchart of Figure 4 - 12 presents an overview of our final modified method. At this level, we first implemented our novel method of motion detection based tracking which is best suited for illumination problem. Thus we incorporated this novel technique into our algorithm and implemented our adaptive target tracking algorithm to track the target person.

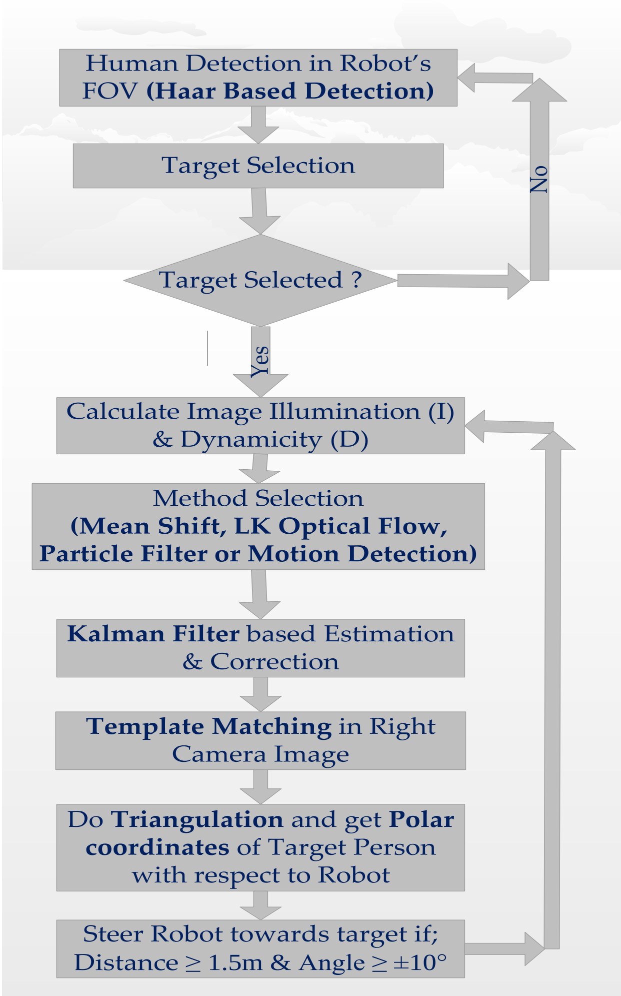


Figure 3 - 4: Block Diagram of Final Modified Method.

Furthermore, the comparison of these methods is also done under various illumination levels and dynamic conditions. Two parameters were defined i.e. Illumination (I) and Environment Dynamicity (D). Based on these parameters, the comparison of methods is done.

3.3 Human Detection & Target Person Selection

As mentioned earlier, the proposed scheme uses Haar like features for human detection. To implement Haar based object detection in Open CV, classifier is trained with a large number of samples of positive and negative examples of the target objects. Then it can be applied to any image to detect that target. This is done by moving classifier window to the whole image to detect particular object. Therefore, a trained set of classifier is employed for the detection of human upper body and face.

To perform people detection, only left camera is used. Initially, only upper body detection is done; and if detected, the faces are searched inside the detected bodies on the input stream of images. This check of face detection inside the detected upper body is done only to minimize the false detection of bodies. The algorithm of this scheme of detection is shown in Figure 3 - 5.

a. Classifiers Initialization

- i. *Load Body Classifier:* Cl_{body}
- ii. *Load Face Classifier:* Cl_{face}

b. Image Initialization

- i. *Get input image:* I_{in}
- ii. I_{gray} = *Convert* (I_{in}, grayscale)
- iii. I_{gray} = *hist_equalize* (I_{gray})

c. Body Detection

- i. *vec* det_body = *detect_multiscale*(I_{gray}, Cl_{body})

d. Face Detection inside bodies

- i. *vec* det_face = *detect_multiscale*(I_{gray}, Cl_{face})

e. Body-Face Check & Data Storing

```
i.  for all (det_body)
    if (det_face ► det_body)
        body [i] = det_body
        face   [i] = det_face
        next i
```

f. Target Person Selection

```
i.  if (selection_key == Pressed)
    input (k)
    target_id = k
  else
ii. goto (Step. c)
```

Figure 3 - 5: Implemented Algorithm for Human Detection and Target Person Selection.

After human detection, the next step is the selection of one target person out of all the detected persons. All the detected human beings are marked with a serial number which starts from the left most detected person. The control program then asks the operator to enter the serial number of one target person from all the detected human beings so that the tracking process could start. Therefore in this phase the target person selection is done.

3.4 Target Person Tracking with Left Camera

As mentioned earlier, three tracking techniques have been used in this work. These methods are implemented independently to compare them with each other. The implementation methodology of each of them is described below:

3.4.1 Mean Shift / Cam Shift based Tracking

Mean shift based tracking is first of the three tracking methods which was implemented in Open CV. This method requires the initial position of the object in order to track it. Therefore, input to mean shift is the location of selected target person during the target selection stage. The algorithm for mean shift is explained in Figure 3 - 6 (Step a-c).

a. ***Target Initialization***
   i. g = *get* target_ID
   ii. roi = *image* (target.rectangle)

b. ***Calculation of Probability Distribution***
   i. HSV = *rgb2hsv* (roi)
   ii. Hist = *imhist* (HSV:1)
   iii. Distribution = histogram_backproject (Hist)

c. ***Mean Shift Calculation***
   i. Mean_Calculation = Σ (Distribution) / n
   ii. *Move_Center* to *Mean_Calculation*

d. ***Cam Shift***
   i. *Recalculate* Area(Densed_Distribution)
   ii. *Adjust* Rectangle (Area)

e. ***Target Correction & Estimation***
   i. K = *Kalman_Predict* (pre_state, post_state)
   ii. Target = *Kalman_corect* (K, Move_Center)

f. ***Target Person Following***
   i. Calculate Position ‘P’ in 3D Space using *Target*
   ii. Robot Navigation towards ‘P’

Figure 3 - 6: Implemented Algorithm for Mean Shift based Tracking.

In the first step, the algorithm takes the back projection of the initial position of target person’s face / body. Then it finds the 2D color probability distribution of the target area. Afterwards, the target center is found by computing the mean of that 2D color probability distribution. After finding the target center, the search window is moved towards that mass center. This position is used to track the target iteratively in the stream of images. The search window size and orientation does not change in this method of tracking. The implementation and tracking results of mean shift based target tracking are shown in the next chapter.

3.4.2 Lucas Kanade Optical Flow based Tracking

In order to perform LK optical flow based tracking, firstly the image is converted into grayscale, and twelve points are selected which are mostly at the contours of the selected target person's upper body, face and shirt. After storing these points in a vector, iterative pyramid based optical flow is used to find corresponding features in next frame. Then a bounding box is generated which is a rectangle containing all the features having the flow.

This whole implementation is summed up in the flowchart of Figure 3 - 7 and the results of LK optical flow based tracking are discussed in the next chapter.

a. Tracking Initialization
   i. g = *get* target_ID

b. Points Specification
   i. I_{gray} = *Convert* (I_{in}, grayscale)
   ii. $P^{n}_{(1\sim12)}$ = Locations [body(g) & face(g)]

c. Optical Flow
   i. $P^{n+1}_{(1\sim12)}$ = *apply_lk_flow_equation* ($P^{n}_{(1\sim12)}$)
   ii. op_flow = *rectangle* [$P_{(top\_left)}$, $P_{(bottom\_right)}$]

d. Target Correction & Estimation
   i. K = *Kalman_Predict* (pre_state, post_state)
   ii. Target = *Kalman_corect* (K, op_flow)

e. Target Person Following
   i. Calculate Position *'P'* in 3D Space using *Target*
   ii. Robot Navigation towards *'P'*

Figure 3 - 7: Implemented Algorithm for LK Optical Flow based Tracking.

3.4.3 Particle Filter based Tracking

Particle filter is a recursive Bayesian filter based tracking method which is non-parametric. After the selection of target person, sampling of the selected area is done which is based on the motion model. These samples are drawn on the basis of proportionality of Gaussian weight at any location inside the region of interest. Weights are then calculated for each sample based on the similarity function of correlation of histograms for RGB channels of the selected target area. The next step is the resampling according to the calculated weights. Finally, the bounding window is moved towards the cumulative distribution of weights of the latest samples.

The implemented algorithm for particle filter based tracking is shown in Figure 3 - 8 and the implemented results are discussed in the next chapter.

a. Target Initialization

i. g = *get* target_ID

ii. Set *N = 200*

b. Sampling

i. Samples = *get_samples* [image(target_ID, N), gaussian]

ii. *for_each* samples

hist_sample(R, G, B) = *find_histogram* [image(rect. target_ID)]

Calc_Weight (n) = *Correlate* [hist(R,G,B), hist_sample(R,G,B)]

Weight(n) = *normalize* (Calc_Weight(n))

c. Resampling

i. Highest_weighted_sample = *find* [highest: Weight (1 → n)]

ii. Resamples = *get_samples* [Highest_weighted_sample, gaussian]

iii. *for_each* Resamples

hist_sample(R, G, B) = *find_histogram* [image(rect. target_ID)]

Calc_Weight (n) = *Correlate* [hist(R,G,B), hist_sample(R,G,B)]

Weight(n) = *normalize* (Calc_Weight(n))

d. Position Estimation

i. For all n

Largest = *find_largest* (Weights(1 → n))

ii. Particle = Resamples (largest)

f. Target Correction & Estimation

i. K = *Kalman_Predict* (pre_state, post_state)

ii. Target = *Kalman_corect* (K, Particle)

g. Target Person Following

i. Calculate Position *'P'* in 3D Space using *Target*

ii. Robot Navigation towards *'P'*

Figure 3 - 8: Implemented Algorithm Particle Filter based Tracking.

3.4.4 Motion Detection based Tracking

Along with above cited methods, we also implemented our novel method of Differential Images based Motion Detection of the target person. It actually uses three images termed as previous, current and next and calculated their differences. We first calculated the difference between previous and next image and then current and next image. Furthermore, a logical AND operations steered between both results and are threshold to detect the motion.

In order to neglect the false positives, two types of checks are incorporated: firstly the motion detected should be larger than 2 pixels, secondly calculation of standard deviation to get the distribution of motion in the image. This will eliminate the movement of objects other than the target person.

The implemented algorithm is presented in Figure 4 - 12. The movement of target person in three consecutive frames is searched to get track of the target and the implemented results are discussed in the next chapter.

a. Initialization

i. Get target location;

Rect Target = human(rectangle)

ii. Add a boundary of 10 pixels each side of Target;

Target.rectangle = target + 10

iii. Image Initialization;

prev = image (n – 1)
current = image (n)
next = image (n+1)

b. Differential Images

i. Calculate both Differential Images;

diff1 = next – prev
diff2 = next – current

ii. Apply Logical AND & Threshold the result

Threshold = *find_mean* [grayscale_intensity (image – image(rect.target_ID))]
Motion_image = diff1 AND diff2
Result = Motion_image [Target.rectangle]
Result = Result (all values > Threshold)

c. Optimization

i. Apply Erosion and Calculate Standard Deviation

Std_image= *calc_std_dev* (Motion_image)
Result = *imerode* (Result)
Std_result= *calc_std_dev* (Result)

ii. Detect Filtered Motion of Target

If (Std_result < Std_image)
Target.rectangle = Result. Motion

d. Target Correction & Estimation

i. K = *Kalman_Predict* (pre_state, post_state)

ii. Target = *Kalman_corect* (K, Result.Motion)

e. Target Person Following

iii. Calculate Position *'P'* in 3D Space using *Target*

iv. Robot Navigation towards *'P'*

Figure 3 - 9: Implemented Algorithm Motion Detection based Tracking.

3.4.5 Kalman Filter based Estimation & Correction

Kalman filter is an optimal estimator that minimizes the mean square error of the estimated parameters. It basically has two parts: Prediction and Correction. The X and Y coordinate of the tracked target from previous step is sent to the first step of Kalman filter i.e. prediction part. The second step is the measurement correction based on distance and velocity model. This step is implemented to more accurately localize the target person in the left camera. The results of Kalman filter based estimation and correction are discussed in the next chapter.

3.5 Stereo Correspondence with Right Camera

To find the location of tracked target person in right camera image, any stereo correspondence method is required to be implemented. In our implemented program, this is done by building a rectangular template of the tracked target in left camera. This patch is then searched in the right camera image by sliding it in the whole image to find the rectangular pixel coordinates. The matching technique we used for searching is normalized correlation. Its formulation is shown by Equation 2.

Rnorm_corr (x, y) = Rcorr (x, y) / Z (x, y) …….. (2)

Where,

Rcorr (x, y) = Σx, y [TL (x', y'). IR (x, y)]

Z (x, y) = Σx', y' TL (x', y'). Σx, y IR (x, y)

TL (x', y') = Patch of Template

IR (x, y) = Right Camera Image

Some other methods which can be deployed are squared difference, normalized squared difference, cross correlation, normalized cross correlation and correlation coefficient.

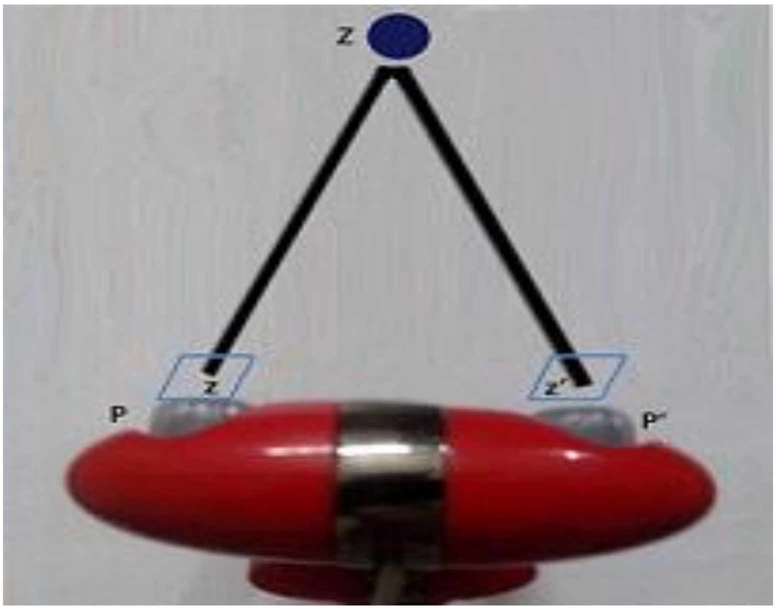


Figure 3 - 10: Stereo Camera based Triangulation.

3.6 Target Person Localization & Robot Steering

After finding the stereo correspondence of the target person, the next step is to get the exact location of target person with respect to the robot's origin. Linear Triangulation [27] is the method used for this purpose. It requires the camera projection matrix for both cameras, and the location of the corresponding point in both cameras. As shown in Figure 3 - 10, P and P' are the 3x4 camera projection matrix and z and z' are the points to be located in left and right camera images, then Z is the 3D location of the point with respect to the camera/robot's origin.

We have used the center point of target person in both cameras to get the exact location with respect to the robot. This 3D point is then used to get the angle and distance of target from the robot and then the robot is steered to maintain the distance of 1.5m and the angle of $\pm 10^0$ from the target. Furthermore, sonars are used to avoid the collision with other objects in the environment as it will stop robot motion and generate an error message. This can be further used to perform motion planning for the avoidance of static / dynamic obstacles in future.

Chapter 4

EXPERIMENTAL RESULTS & ANALYSIS

4.1 Introduction

In this chapter, the results of proposed human detection and tracking method are presented. The whole testing is done in more than ten different benchmarks. Here we are presenting the tracking results on three benchmarks. Human detection and tracking results are presented below;

4.2 Human Detection & Target Person Selection Results

Figure 4 - 1 shows the representation of laptop's screen (initialization window) with labeling of left and right image along with the terminal window. Figure 4 - 2 presents the four benchmarks whose results are presented.

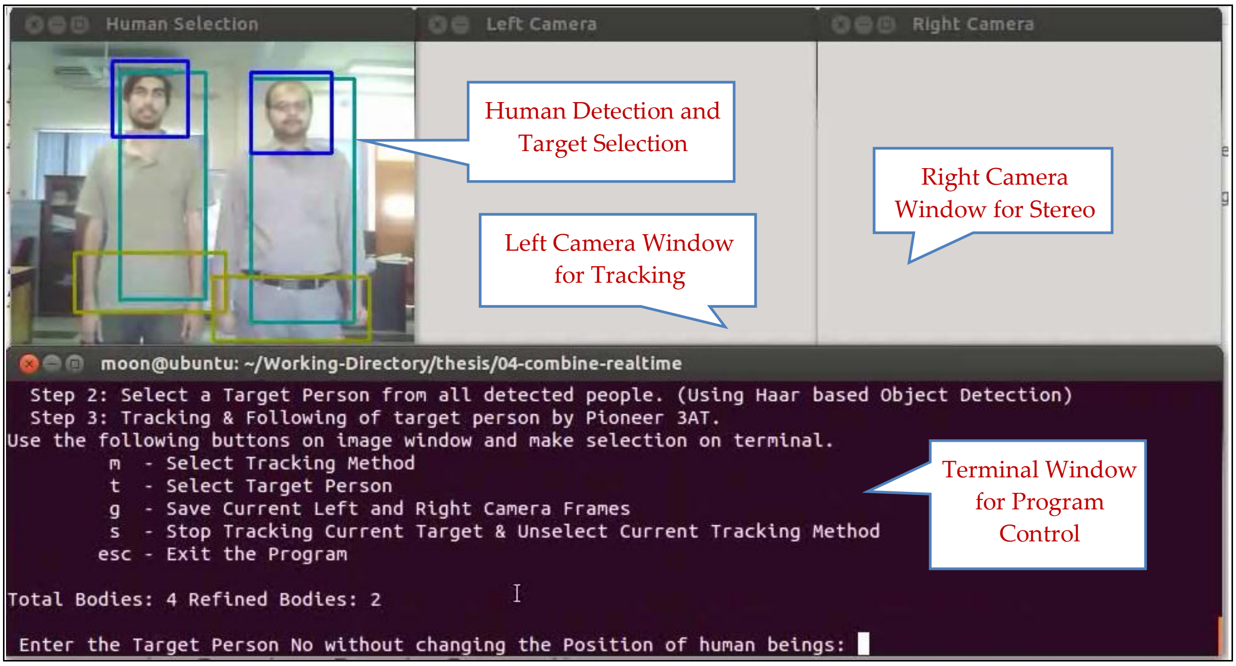


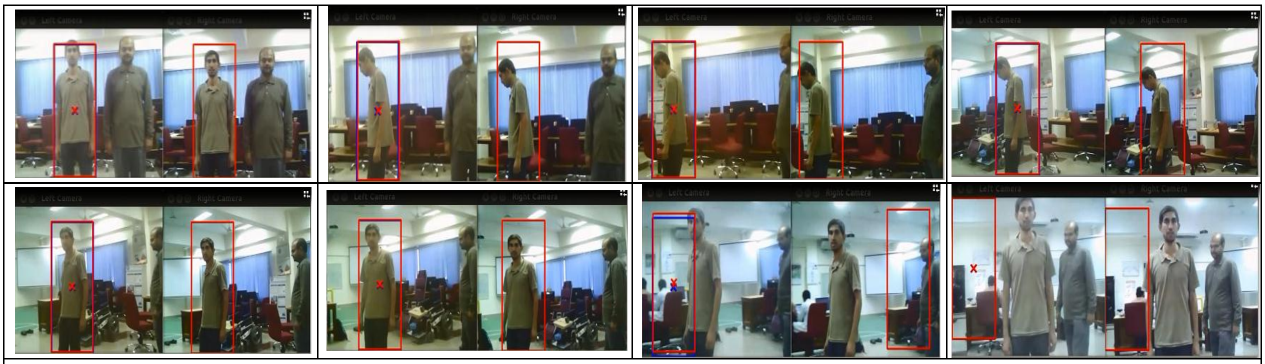
Figure 4 - 1: Initialization Window for Robot's Screen with Labeling

(a) Indoor Lab (b) Outdoor Corridor (c) Natural Background

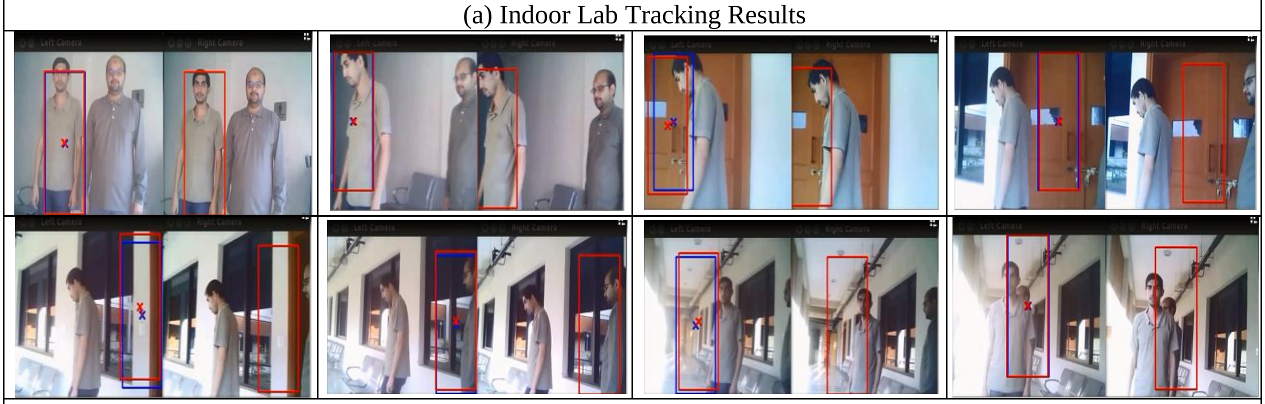
Figure 4 - 2: Benchmarks for Testing

4.3 Target Person Tracking & Stereo Correspondence Results

Figure 4 - 12 presents tracking results of Cam Shift based tracking for all three benchmarks. Sequential Images of Various Screenshots are presented. Cam Shift can prone to false tracking in case of illumination conditions when the background of environment is not much different from the target. Furthermore, the tracking results of mean shift method are not very good in images of high illumination i.e. I = 0.7.



(a) Indoor Lab Tracking Results



(b) Outdoor Corridor Tracking Results

(c) Natural Background Tracking Results

Figure 4 - 3: Testing Results for Cam Shift based Tracking.
(a), (b), (c): Sequential Frames of all Four Benchmarks

Figure 4 - 12 presents tracking results of LK Optical Flow based tracking for all three benchmarks. Sequential Images of Various Screenshots are presented. As mentioned earlier, twelve points are selected which are inside the contour of target person. Then optical flow of these points is used to track the person iteratively in the left camera image. It is observed that tracking efficiency reduces in case of large amount of motion in the environment near to the target person (D = 0.75). Also it undergoes loss of tracking points when the target has sufficiently moved from its initial position.

(a) Indoor Lab Tracking Results

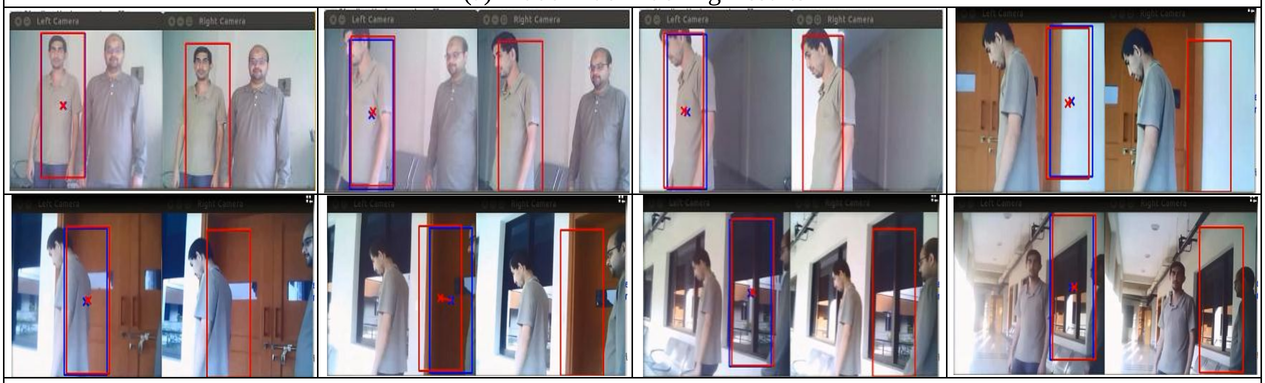
(b) Outdoor Corridor Tracking Results

(c) Natural Background Tracking Results

Figure 4 - 4: Testing Results for LK Optical Flow based Tracking.
(a), (b), (c): Sequential Frames of all Four Benchmarks

Figure 4 - 12 presents tracking results of Particle Filter based tracking for all three benchmarks. Sequential Images of Various Screenshots are presented. Particle Filter provided better results out of the above three, as it is a non-parametric method. Only weights are assigned to each sample based on the correlation of RGB histogram. The rest samples are drawn based on random number generation. But due to non- parametric approach, it is sample dependent. If there get some problem on runtime in sample selection by the algorithm, the tracking fails. Furthermore, the time consumption of Particle filter is highest of all four methods.

(a) Indoor Lab Tracking Results



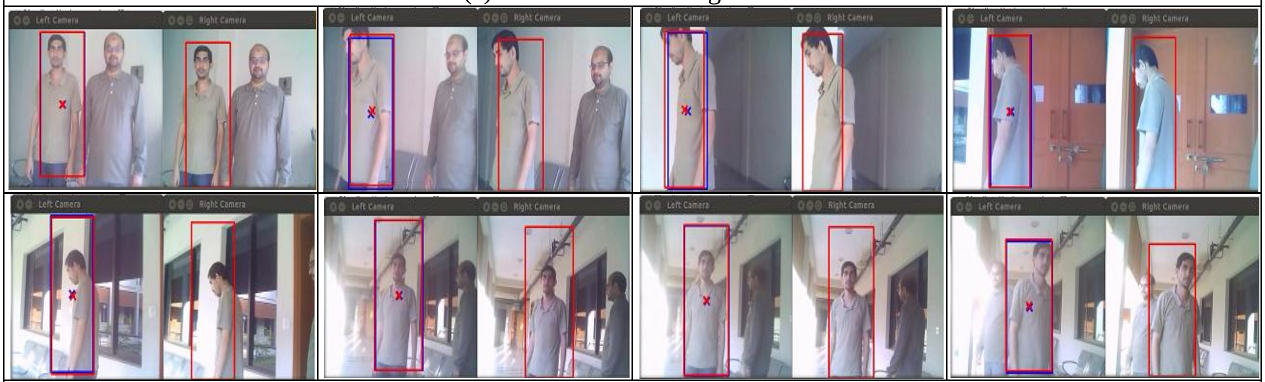
(b) Outdoor Corridor Tracking Results

(c) Natural Background Tracking Results

Figure 4 - 5: Testing Results for Particle Filter based Tracking.
(a), (b), (c): Sequential Frames of all Four Benchmarks

Figure 4 - 12 presents tracking results of Motion Detection based tracking for all three benchmarks. Sequential Images of Various Screenshots are presented. Motion Tracking is based on the motion vector generation of three iterative frames, and is therefore less affected by illumination. Furthermore, it provides better results of all when there is movement of other persons behind the target.

(a) Indoor Lab Tracking Results

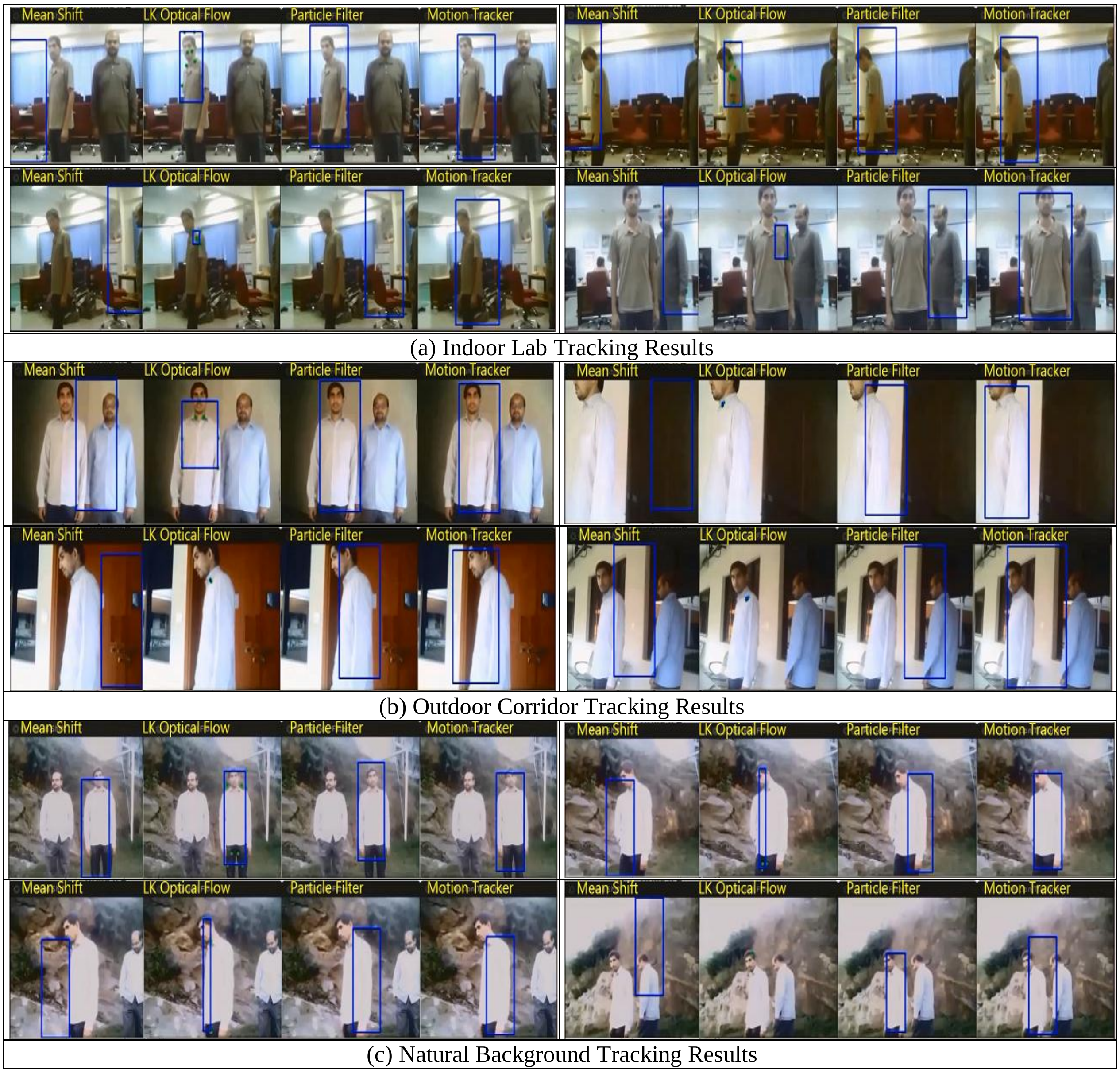


(b) Outdoor Corridor Tracking Results

(c) Natural Background Tracking Results

Figure 4 - 6: Testing Results for Motion Tracker based Tracking.
(a), (b), (c): Sequential Frames of all Four Benchmarks

Figure 4 - 12 presents the compared results of all four tracking methods with same input video frames of three benchmarks.



(a) Indoor Lab Tracking Results

(b) Outdoor Corridor Tracking Results

(c) Natural Background Tracking Results

Figure 4 - 7: Compared Tracking Results in Similar Frames.
(a), (b), (c): Sequential Frames of all Four Benchmarks

4.4 Robot Steering Results

Graphs for target person distance and angle from robot with respect to time are shown in Figure 4 - 8 and Figure 4 - 9. Graphs show that the robot is maintaining a distance of 1.5m and angle of $\pm 10^0$ from target.

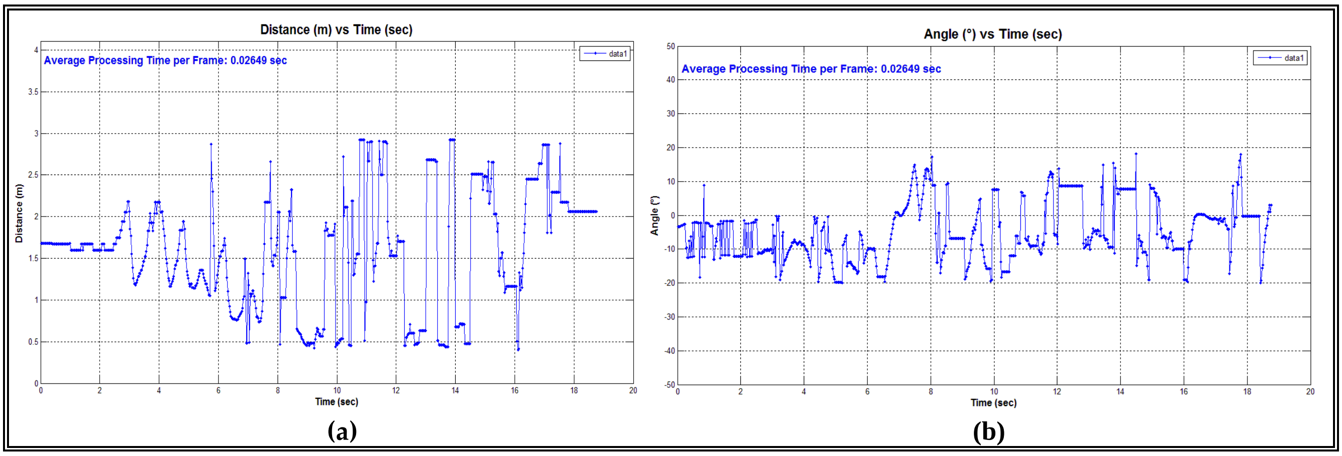


Figure 4 - 8: Target Position w.r.t Robot for Benchmark No. 1
(a) Distance Vs Time **(b)** Angle Vs Time

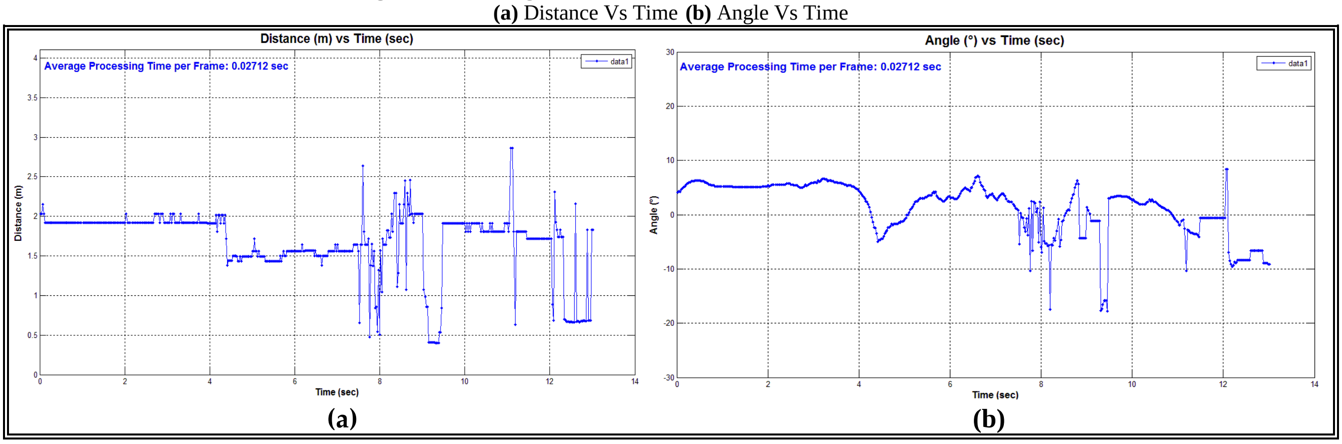


Figure 4 - 9: Target Position w.r.t Robot for Benchmark No. 2
(a) Distance Vs Time **(b)** Angle Vs Time

4.5 Comparison of Tracking Techniques

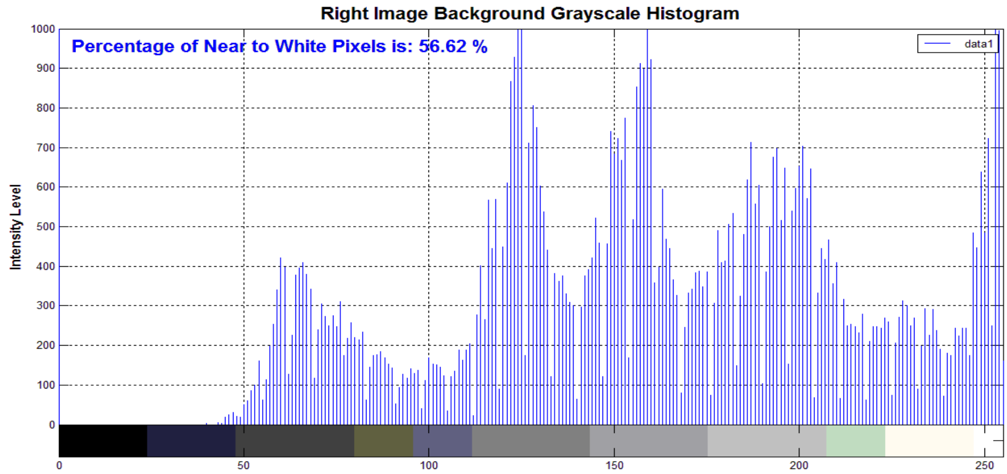
4.5.1 Methodology

Mean shift method uses the target's color histogram to generate probability distribution. LK optical flow uses the pyramid based flow of points computed with reference to previous frame. Whereas Particle filter is a sampling based probabilistic approach that select highest weighted sample. Samples are generated randomly and weights are assigned on the basis of correlation. Motion Detection is the technique which is based on the calculation of differential images to get the motion of target person.

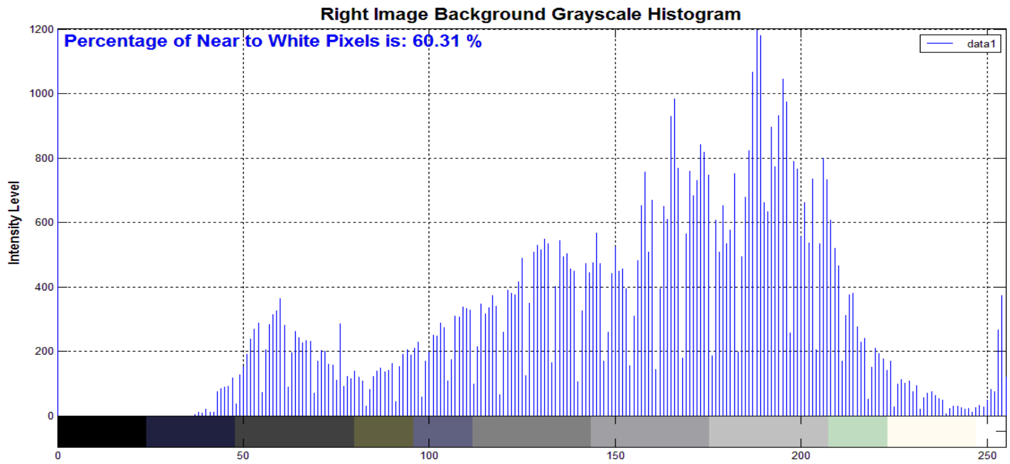
4.5.2 Illumination

Illumination is measured based on the percentage of near to White pixels in the grayscale histogram of image. Graph of Figure 4 - 10 present the grayscale histogram of all three benchmarks and also calculated the percentage of near to white pixels.

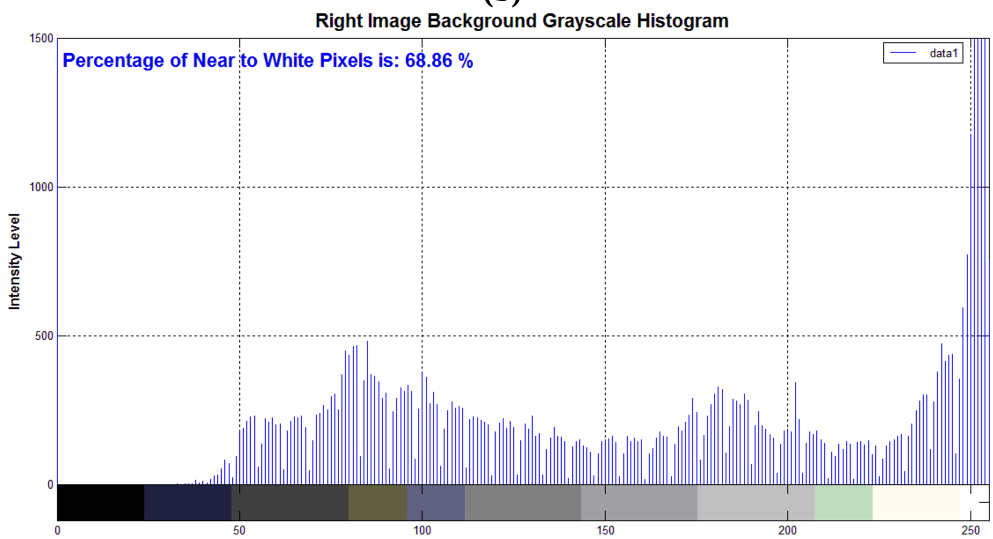
As the results of Figure 4 - 12 and Figure 4 - 12 shows, Mean shift method results in false tracking in all three benchmarks having illumination of 55%, 60% and 68% respectively. On the other hand LK Optical flow, Particle filter and Motion Detection based tracking has no influence of illumination in both benchmarks. (Figure 4 - 12 to Figure 4 - 12)



(a)



(b)

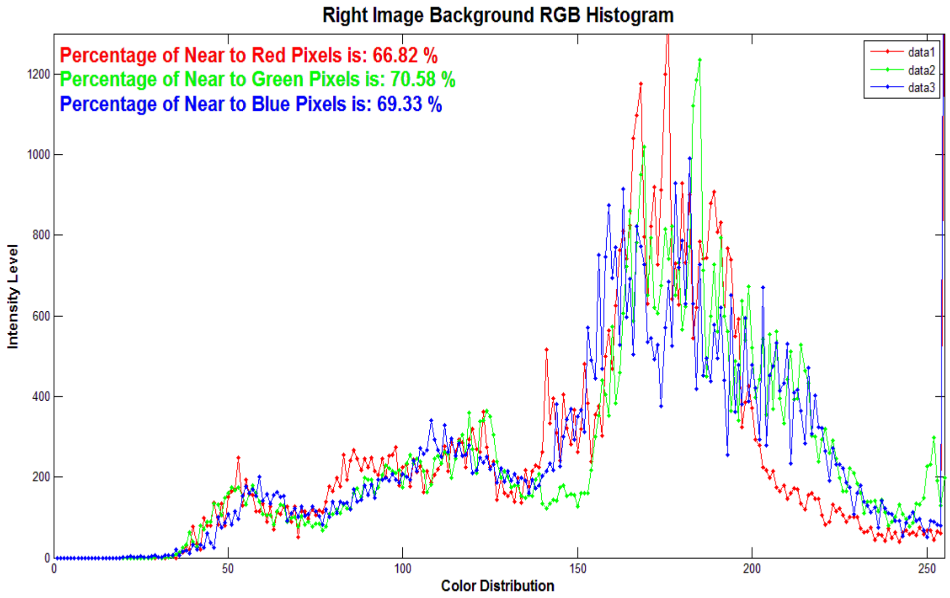


(c)

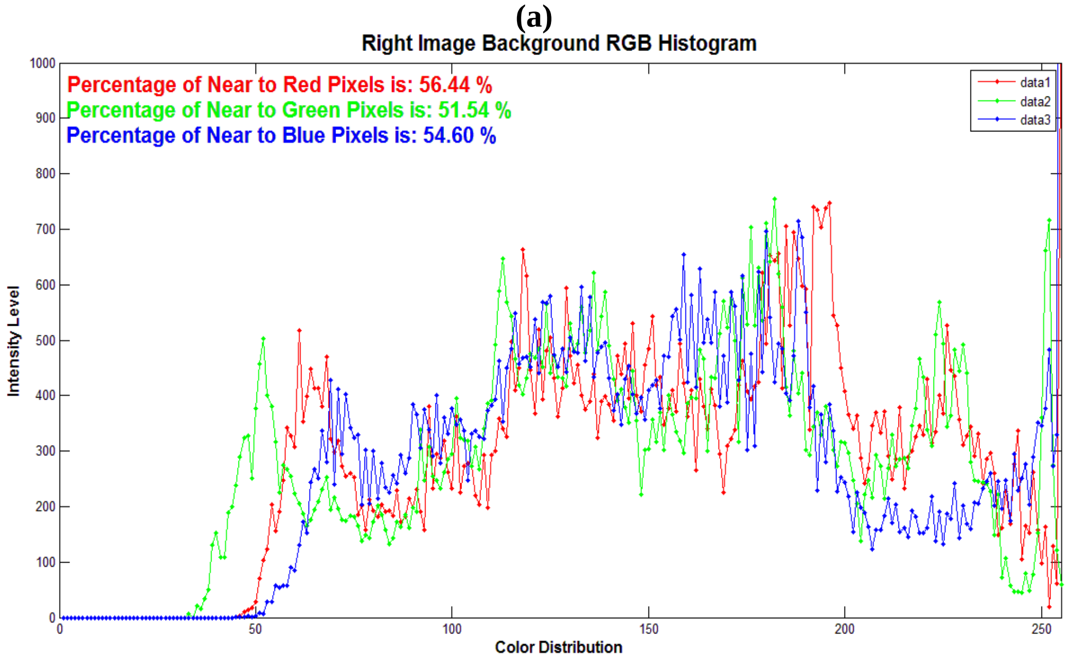
Figure 4 - 10: Grayscale Histograms for Near to white pixel Percentage calculation.
(a) Benchmark No 1 (b) Benchmark No 2 (c) Benchmark No 3

4.5.3 Background Colors

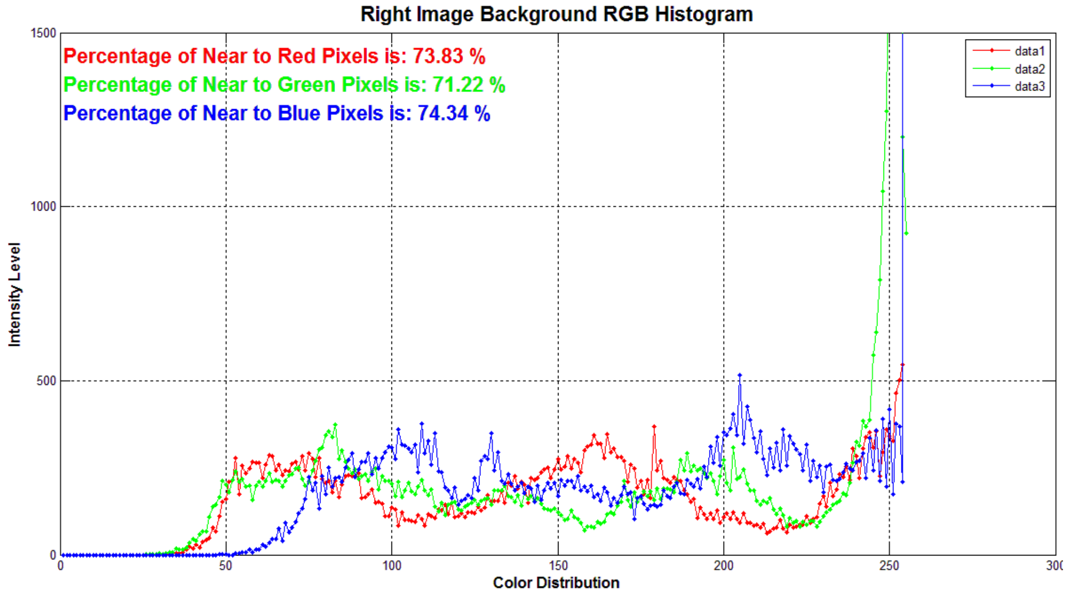
Background colors are measured based on the percentage of near to Red, Green and Blue pixels in the 3-Channel RGB histogram of image. Graph of Figure 4 - 11 present the 3-Channel RGB histogram of both benchmarks and also calculated the percentages of pixels of RGB colors in background.



(a)



(b)



(c)

Figure 4 - 11: 3-Channel RGB Histograms for Near to Red, Green & Blue pixel Percentage calculation. (a) Benchmark No 1 (b) Benchmark No 2 (c) Benchmark No 3

As the results of Figure 4 - 3 shows, Mean shift method results in false tracking in all three benchmarks having background colors of 66% Red, 70% Green, 69 % Blue; 58% Red, 51% Green, 54 % Blue and 73% Red, 71% Green, 44 % Blue respectively. On the other hand LK Optical flow, Particle filter and Motion Detection based Tracking has no influence of similar background colors in both benchmarks. (Figure 4 - 4 to Figure 4 - 12)

4.5.4 Movement in Background

Mean shift has no effect of movement in background. In Figure 4 - 3, another person is moving in the background of target person. Whereas LK Optical flow and Motion Detection based Tracking does false tracking when there is movement in the background (Figure 4 – 4 to 4 – 7). Particle filter like Mean shift has no influence of background movement.

4.5.5 Processing Time

Figure 4 - 12 presents the graph of processing time of four tracking methods in a similar video stream. Red Lines present the result of results of Cam Shift method, Green presents LK Optical Flow, Blue present Particle Filter and Yellow presents Motion Tracking results. All computation has been done on a core i-3 laptop with 2.4 GHz processor and 4GB of Ram. Average Processing time per cycle for Mean shift based tracking is 1.385 ms, whereas average Processing time per cycle for LK Optical flow based tracking is 0.726 ms.

Average Processing time per cycle for Particle filter based tracking is 24.088 ms. Average Processing time per cycle for Motion Detection based tracking is lowest i.e. 0.361ms.

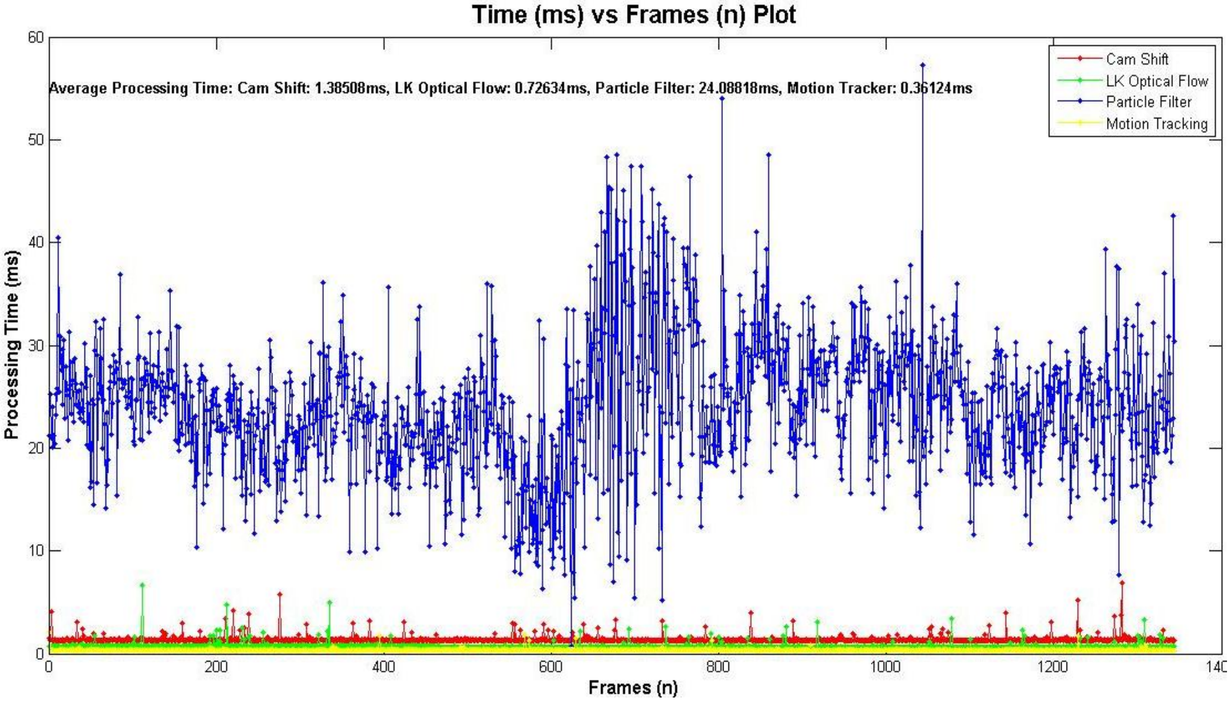


Figure 4 - 12: Processing Time Comparison for all Four Tracking Methods

Conclusively, based on the above comparison, it can be said that all these tracking techniques has its own merits and demerits. When there is no problem of illumination and background color differs hugely from target, Mean shift based method is a good choice. Whereas if there is no movement in background or the environment is not dynamic, Motion Detection based Tracking will be the best one to use. On the other hand, Particle filter is not majorly influenced by these above mentioned issues, as it is a non-parametric uniform random sampling method, but at the same time if random sampling goes wrong, it will lead to false detection. The whole comparison has been summed up in Table of Figure 4 - 13.

Parameters	Mean Shift	LK Optical Flow	Particle Filter	Motion Detection based Tracking
Tracking Methodology	Mean of Target's Color Histogram.	Flow of points in the contour of target.	Probability based weighted samples.	Calculation of Differential Images.
Illumination in Background (60% Illumination)	False Tracking under High Illumination	Have no influence of Background Illumination.	Have no influence of Background Illumination.	Have no influence of Background Illumination.
Effect of Background Colors (65% Red, 70% Green, 70% Blue)	False Tracking under High Illumination	Have no influence of Background Color similar to the target.	Have no influence of Background Color similar to the target.	Have no influence of Background Color similar to the target.
Effect of Movement in Background D > 0.56	No effect of movement in background.	False Tracking under movement in background.	No effect of movement in background.	False Tracking under movement in background
Processing Time	Average Processing time per cycle is **1.385 ms**.	Average Processing time per cycle is **0.726 ms**.	Average Processing time per cycle is **24.088 ms**.	Average Processing time per cycle is **0.361ms**.

Figure 4 - 13: Table for Comparison of Tracking Techniques.

Chapter 5
CONCLUSION & FUTURE WORK

5.1 Conclusion

In this research work, a novel approach for human detection and target person tracking scheme based on the implementation of four different vision based target tracking techniques is presented. The implementation is done on pioneer 3AT robot equipped with stereo camera and sonars. These tracking methods are implemented along with various other techniques for human detection, target person selection, stereo correspondence, triangulation, and robot navigation is utilized to finally make the robot track the target person.

Our main contribution in this research is the presentation of a new differential image based tracking method which utilizes the differential images of three iterative frames of live video stream. Then the comparison is made with three other conventional techniques; cam shift, LK optical flow and particle filter. The robot is equipped with stereo camera and sonars. Human detection is done using Haar cascade classifier, followed by the implementation of a tracking method on the selected target person. Comparisons of these techniques are based on testing in different environmental conditions. Afterwards, Kalman filter based prediction and correction is done to accurately localize the target. Moreover stereo correspondence is done by template matching followed by linear triangulation, to get the 3D location of the target.

Thus, this research work presents the implementation of various techniques that make the robot to select a target person in the environment and then track and follow that target. Tracking results of all tracking techniques along with stereo correspondence and target person selection are presented in three benchmarks and the comparison of techniques is done in detail.

5.2 Future Work

Future work will be focused on False Detection Models and its remedy in case the target is lost. This can be done by implementing more sensors in the system. Also if the target person gets out of sight from robot, some other perception mechanism will be required so that the target can be re-found by the robot. In one of the paper presented in Chapter 2, this was done using Sound Source Localization system. Furthermore, a motion planning technique will be required in future in order to tackle with the environments having highly dynamic obstacles. After full evaluation, this method can be further implemented in luggage cart carrier or in hospitals to become an application of service robot.

Publications

Journals (ISI Indexed)

Advancement in Human Tracking by a Mobile Robot

*Major Revision is recommended in Review & Revised Paper is submitted in **IEEJ Electrical and Electronics Engineering (TEEE-C), 2014 (Impact Factor: 0.327)***

Improved Method for Stereo Vision based Human Detection & Target Person Following by Mobile Robot

*Submitted & Review Awaited in **South African Journal of Industrial Engineering, 2014 (Impact Factor: 0.134)***

Conferences (Robotics Automation Society & EI Indexed)

Human Detection and Following by a Mobile Robot using 3D Features

*Published in **IEEE International Conference on Mechatronics and Automation (ICMA)**, pp. 1714-1719, Takamatsu, Japan, 2013*

Human Tracking by a Mobile Robot using 3D Features

*Published in **IEEE International Conference on Robotics and Biomimetic (ROBIO)**, pp. 2464 - 2469, Shenzhen, China, 2013*

Bibliography

[1] B. T. Navneet Dalal, "Histograms of Oriented Gradients for Human Detection," in *International Conference on Computer Vision & Pattern Recognition*, 2005.

[2] C.-C. C. J. K. A. Lu Xia, "Human Detection Using Depth Information by Kinect," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2011.

[3] C. N. B. W. Orasa Patsadu, "Human Gesture Recognition Using Kinect Camera," in *International Joint Conference on Computer Science and Software Engineering*, 2012.

[4] M. J. Paul Viola, "Rapid Object Detection using a Boosted Cascade of Simple Features," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001.

[5] A. K. V. P. Rainer Lienhart, "Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapid Object Detection," 2002.

[6] M. J. Paul Viola, "Robust Real-time Object Detection," *International Journal of Computer Vision*, 2001.

[7] O. D.-S. L. A.-C. J. L.-N. M. Castrill´on-Santana, "Face and Facial Feature Detection Evaluation," in *International Joint Conference on Computer Vision, Theory and Applications*, 2008.

[8] V. R. P. M. Dorin Comaniciu, "Real-Time Tracking of Non-Rigid Objects using Mean Shift," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2000.

[9] F. Y. Zhiwei Wang, "Object Tracking Algorithm Based on Camshift and Grey Prediction Model in Occlusions," in *International Conference on Computer Application and System Modeling*, 2012.

[10] T. K. Bruce D. Lucas, "An Iterative Image Registration Technique with an Application to Stereo Vision," in *Proceedings of Imaging Understanding Workshop*, 1981.

[11] B. Jean-Yves, "Pyramidal Implementation of the Lucas-Kanade Feature Tracker," *Intel Corporation, Microprocessor Research Labs*, 1999.

[12] F. Gustafsson, "Particle filter theory and practice with positioning applications," in *IEEE Aerospace and Electronic Systems Magazine*, 2010.

[13] P. ,. Z. T. ,. P. A. E. C. Li, "Visual Contour Tracking based on Particle Filters," *Image and Vision Computing*, 2003.

[14] C. H. H. T. T. L. Ren C. Luo, "Human Tracking and Following Using Sound Source Localization for Multisensor Based Mobile Assistive Companion Robot," in , 2010.

[15] L. Z. X. L. W. C. J. S. Songmin Jia, "Autonomous Robot Human Detecting and Tracking Based on Vision," in *IEEE International Conference on Mechatronics and Automation*, 2011.

[16] L. Z. X. L. Songmin Jia, "Robustness Improvement of Human Detecting and Tracking for Mobile Robot," in *IEEE International Conference on Mechatronics and Automation*, 2012.

[17] A. O. S. Y. Mehrez Kristou, "Panoramic Vision and LRF Sensor Fusion Based Human Identification and Tracking for Autonomous Luggage Cart," in *IEEE International Symposium on Robot and Human Interactive Communication*, 2009.

[18] A. O. S. Y. Mehrez Kristou, "Target Person Identification and Following based on Omnidirectional Camera and LRF Data Fusion," in *IEEE International Symposium on Robot and Human Interactive Communication*, 2011.

[19] A. L. D. e. R.-D. V. N. Emina Petrović, "Stereo Vision-Based Human Tracking for Robotic Follower," *International Journal of Advanced Robotic Systems*, 2013.

[20] S. Y. X. Y. Y. K. T. H. L. Liyuan Li, "Robust Multiperson Detection and Tracking for Mobile Service and Social Robots," in *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS*, 2012.

[21] Pioneer 3AT Mobile Robot. [Online]. http://www.mobilerobots.com/ResearchRobots/P3AT.aspx

[22] Minoru 3D Webcam. [Online]. http://www.robotshop.com/en/minoru-3d-webcam.html

[23] Player - Stage. [Online]. http://playerstage.sourceforge.net/

[24] Open CV. [Online]. http://opencv.org/

[25] K. F. I. Y. A. N. M. Badar Ali, "Human Detection and Following by a Mobile Robot using 3D Features," in *IEEE International Conference in Mechatronics and Automation*, 2013.

[26] A. H. Q. K. F. I. Y. A. S. O. G. M. J. N. M. F. A. M. S. M. W.-Y. K. M. R. Badar Ali, "Human Tracking by a Mobile Robot using 3D Features," in *IEEE International Conference on Robotics and Biomimetics*, 2013.

[27] P. F. S. Richard I. Hartley, "Triangulation," in *Computer Vision and Image Understanding*, 1997.