CAMERA TRACKING HUMAN MOVEMENT USING IMAGE PROCESSING

OMER MOHAMED OMER BASLEEM

A report submitted in partial fulfillment of the requirements for the degree of Mechatronics engineering

FACULTY OF ELECTRICAL ENGINEERING

UNIVERSITY TEKNIKAL MALAYSIA MELAKA

2016

“I hereby declare that I have read through this report entitle camera tracking human movement using image processing and found that it has comply the partial fulfillment for awarding the degree of Bachelor of Electrical Engineering (Mechatronics)”

Signature: …………………………………………..

Supervisor’s Name : Miss Nur Maisarah binti Mohd Sobran

Date : 6 / 1 / 2016

I declare that this report entitle camera tracking human movement using image processing is the result of my own research except as cited in the references. The report has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature: ....

Name: Omer Mohamed Omer Ba-saleem

Date : 6 / 1 / 2016

Dedicated to my beloved parents

ACKNOWLEDGEMENT

I would like to express my special thanks of gratitude to my supervisor Miss Nur Maisarah binti Mohd Sobran for her sincere supervision along the project.Also I would like to thank my panels, lectures and friends for their assistances into my project by giving comments and advice to improve it.

Lastly. I want to thank my beloved parents that support and motivate along my study period in UTeM. I wouldn’t be able to make my way without them

ABSTRACT

Nowadays, personal assistive robot is an active area of research in order to make robots see the person and make adjustment based on the human behaviour. However, some problems facing in this project such as the direction of the target object and the changing of the image size that occur. The target position, the mutlti-person tracking and the changing of the background can affect the result of a vision tracking system. In this project, the objectives are developing a vision system to track the movement of the human body, analyse the performance of the human tracking camera and build a graphical user interface (GUI) for the programme. For the hardware component in the methodology, there are some important component chosen for the tracking human system such as 5 Megapixel webcam, Arduino Uno as a microcontroller, and a servo motor. The method of tracking the human body consist of two stages, a detecting process and tracking process. Detecting the upper body of the human by extracted it features and tracking the feature point on the detected body. Based on the deviation angle between the centre of the camera and the centre of the detected body, the servo will rotate to track the human body. The experiments are to test the accuracy of the camera during tracking. When the body having different distance from the camera and with different angels which are 45° and 135° from while the target facing the camera at 90°. As the result, the accuracy of the camera tested at 1.5 meters and 2 meters, the camera during tracking got an accuracy reach to 93.6%.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDIX

CHAPTER 1

INTRODUCTION

1.1 Project background:

This project is about tracking human movement using image processing approach. Robot is the invention that the human future depends on it. Especially those robots that depends on a vision system because it make the robots more interact with the surrounding environment in real life. So many researchs had been done to find ways to communicate and interact with robots using vision techniques. Human body tracking is an active area of reserch in order to make robots see the person and adjust based on the human behavior.

For the tracking system using image processing, it consists of a camera mounted horizontally on the top of the motor. The camera will track the person in real time and adjusted based on the direction of the target. Moreover, human body tracking depends on the detection as a frist stage by extracting the features of the part of the body that need to be tracked. Many researches to get the aim of the target and track the person by using images that come from the camera which have been reported based on human characteristic such as face, human body shape, background subtraction, etc.

This report is organized as follows: Chapter 2 introduce the literature review and related works of image processing techniques. Furthermore, Chapter 3 will presents the project methodology. Then Chapter 4 shows the result and the discussion. Lastly, Chapter 5 concludes this report and lists some recommendations for further research on this particular field.

1.2 Motivation:

The ability to detect and track human motion is a beneficial tool for higher-level applications. Robots based on vision can be used to detect and track human motion for some application that requires tracking process. One of the applications that could be realized with reliable human motion detection and tracking is a smart shopping cart. Nowadays, customers still pull or push the shopping cart during the shopping trip. However, to make the shopping trip more comfortable a smart shopping cart based on vision will be used to track the customer during shopping. The smart shopping cart has the ability to follow the customer inside malls and supermarkets [1]. Vision system is a suitable way to interact between the smart shopping cart and the customer. Using vision techniques for tracking the person will be more comfortable and convenient for customers at the future.

1.3 Problem statement:

Robots based on vision system interacting with the environment surrounding it by extracting the features from it. But, how can make the robot communicate and interact with human movment by tracking it. Vision system is helping to make robots more effective interacting when it comes to detecting recognizing and tracking. The camera of the robot starts tracking the target person, some problems that might occur because of the movement of the target. Direction of the target is one of the problems that occur during the tracking process. In addition the position and orientation of the human in front of the camera might be a problem due to the changing of the size of the image. Moreover, during tracking the target camera more than one target might make the camera not be able to track the target person if somebody enters the frame of tracking. Last the changing of the background of the picture and that changing because of the camera moving.

1.4 Objectives of the Project:

i- To develop a vision system that can track the movement of the human body.
ii- To analyze the performance of the human tracking camera system.
iii- To build a graphical user interface (GUI) for the system.

1.5 Project scope:

The tracking algorithm will be developed using MATLAB software. The Hardware of this project will be a camera mounted on a servo motor moving along x axis to ensure keep traking the human body. Moreover, This project will be developed for tracking human in indoor environment with good illumination condition and the testing it will by tracking single body infront of the camera with different distance between the camera and the human body.

CHAPTER 2

LITRETURE REVIEW

2.1 Related works:

Recently, there has been a great interest in tracking system in the modern following robots. In [2] automatic camera tracking system had developed in order to use for TV program production. The camera can detect and tracking the performer's facial position using image processing. To ensure continuous real-time detection, individual algorithms are applied such as the head shape and the frontal face detection skin color and identification process.

One of the following robots applications is a robot can carry a luggage and follow the customer in [3] a robot quipped omnidirectional camera to take a panoramic image of the target and then tracked using Laser Range Finder (LRF). The picture that comes from the omnidirectional camera depends on the color histogram matching algorithm with the down sample a logarithm which both will be implemented using OpenCV to accelerate the identification and tracking process.

ApriAttenda is one of the humanoid robots that's depends on the vision based on target detection. The idea of detecting and tracking that the body region appears in the input image so a group of feature points on the target region when the target region is moving the feature points will be detected and the region will be followed [4].

In [5] following robot using kinetic sensor. Kinetic is a sensing device can observe human and object in 3D. Kinetic sensor has two parts the first one is the RGP camera the second is 3D depth sensor. The IR emits infrared light and the projector will record the reflective spots and then collecting the reflected spot producing a depth image and according to the depth data the counterpoint of the body will be obtained.

In [7] Histogtam of Oriented Gradients used with mean shift logarithm to track a moving human walking inform of a moving camera. The paper shown if the tracking the full body of the human in a big area is not suitable for mean shift logarithm because of the variance of histogram of target, especially when the target is moving which mean taking time for calculating the histogram of the large area. However the paper showed that also using a head of the human as a target and the color of the hair a robust way to track the human target in front of the camera.

In [8] Using human segmentation beased on shape to track mutiple people in difficult situation the method use 2 stages in order to track the human which will start with the segmented the human body shape to detect the target based on the human hypothesis which produced by the human shape model and then tracking it by using the kalman filter.In[11] use using morphological method and used edge detection to get the outline shape of the human then make a tamblet using object skeleton which one of the morphological method to define human expression and then track it using the kalman tracking method.

Table 2.1 : Summarizes the past related work articles:

Article	Author	Method	Feature
Automatic tracking camera system utilizing the position of faces in the shot image [2]	T.Tsuda, M,Okuda, K. Mutou, Y.Nishida.	Viola Jones algorithm	Using shape, face and skin color as a crossfire at the same time.
Target person identification and following based on omnidirectional camera and LRF data fusion [3]	M. Kristou, A. Ohya, S. Yuta.	HOG & Laser range finder	Using the HOG for detecting the target and robust tracking using LRF
People-following System Design for Mobile Robots Using Kinect Sensor [5]	G. Xing, S. Tian, H. Sun, W. Liu, H. Liu.	Depth images	Collecting the reflected spot producing a depth image in order to detect the human body
Human Body Detection and Tracking from Moving Cameras [7]	G. Li, Y.Xu, X. Shi, S.Wu.	HOG & meanshift	Detect the upper body of the human using Histogram. Tracking in a small area using meanshift to make the tracking robust.
Segmentation and tracking of multiple humans in complex situations [8]	T.Zhao, R.Nevatia, F.Lv.	Segmentation	Detecting the human moving pixel intensities and can also detect multiple humans moving at outdoor environment.

2.2 Detection and Tracking Techniques:

There are several techniques to detect and track an object, the aim is to find a proper way to track a person such as:

2.2.1 Viola Jones algorithm:

The Haar like features is a detection method that uses to detect an object with real time rates. The basic of this method is given bounding box applied on the target that needs to track called it as a detection window.

Haar like feature consider adjacent rectangular window in the specific location inside the detection window. This method depends on the difference between the sums because to make the detection it need to sum up the pixel intestines in each region. To get better detection Viola and Jones present a four horizontal and vertical feature for better coverage area.What makes Haar like feature is a wide usage, it is because it needs a short time to detect instead of the other methods which is due to integral images.

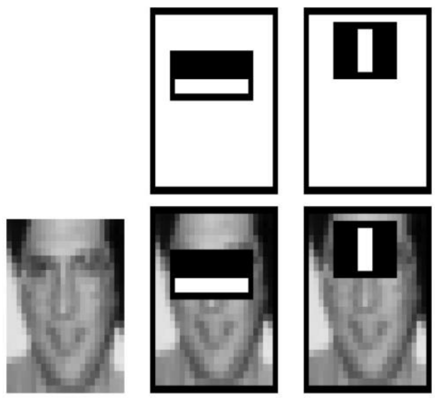

Figure 2.1: Haar like features [16]

In [16] Viola Jones algorithm has been used to detect a human parts such as frontal face, mouth, nose, eye pair and upper body. The cascade object detector has been used the algorithm to detect human upper body by indicating the head and shoulders area.

To get better detection Viola and Jones present a four horizontal and vertical feature for better coverage area. The principle of the rectangular features the values from the different from of the summation. The sum between the pixels in white rectangular subtracted from the pixels in the dark rectangular.

$$F_i = \sum(r_{i,white}) - \sum(r_{y,dark}) \qquad (2.1)$$

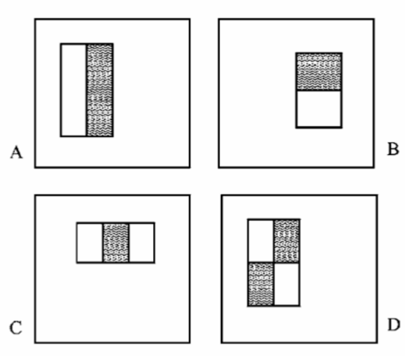

Figure 2.2: Haar-like features detectors

To find these rectangular features a weak learning algorithm is designed to select the single rectangular features which separate the positive from the negative examples. The learner will determine the best threshold for each feature.

$$h_j(x) = \begin{cases} 1 & if\ f_i(x) < \theta_i \\ 0 & otherwise \end{cases} \qquad (2.2)$$

When:

h_j= classifier

f_i=feature

θ_i=threshold

Cascade object detector using viola-Jones algorithm to detect parts of human body such as upper body.The algorithm increase the detection performance with reducing the computing time. More efficient classifiers can be constructed while rejecting as much as negative sub windows though detecting most of the positive example.

The detection process will generate a generation decision tree called a “cascade”. The cascade applies increasingly more complex binary classifiers, which allows the algorithm to rapidly reject regions that do not contain the target. If the target has not been founded in any stage of the cascade classifier, the detector immediately rejects the region and processing is terminated.

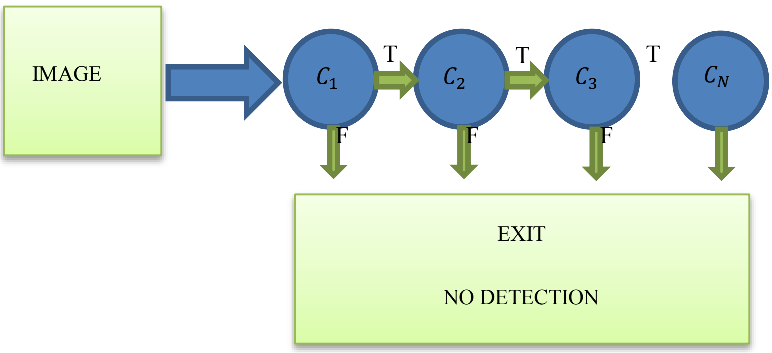

Figure 2.3: Cascade algorithm process

2.2.2 Histogram of Oriented Gradient (HOG):

Histogram of Oriented Gradient [6] is an image processing feature technique used to detect the object. The idea is to detect the object appearance and shape based on the distribution of the local image intensities gradient or edge orientation. To implement this feature the object need to divide the object that needed to detect into cells. Cells is a small region of the image that needs to be detected which will be calculated in term of the intensities gradient or edge orientation.

The advantages of the histogram of oriented gradients is its support translation and orientation of the object such as during moving which make HOG technique is the best way to detect human.

Histogram of Oriented Gradients logarithm is a powerful technique for good performance. But it has a disadvantage which it needs strong processor to run the algorithm. Moreover, even in strong processor and a low resolution image the processor will achieve a little bit frames per second. In the end we can conclude that by this method is not a suitable way to detect since the raspberry pi processor 700 MHz. Although It needs a fast and powerful processor to make the detection process done, but Histogram of Oriented Gradients consider the best way of detecting human and many vision applications had been done using histogram such as detecting pedestrian.

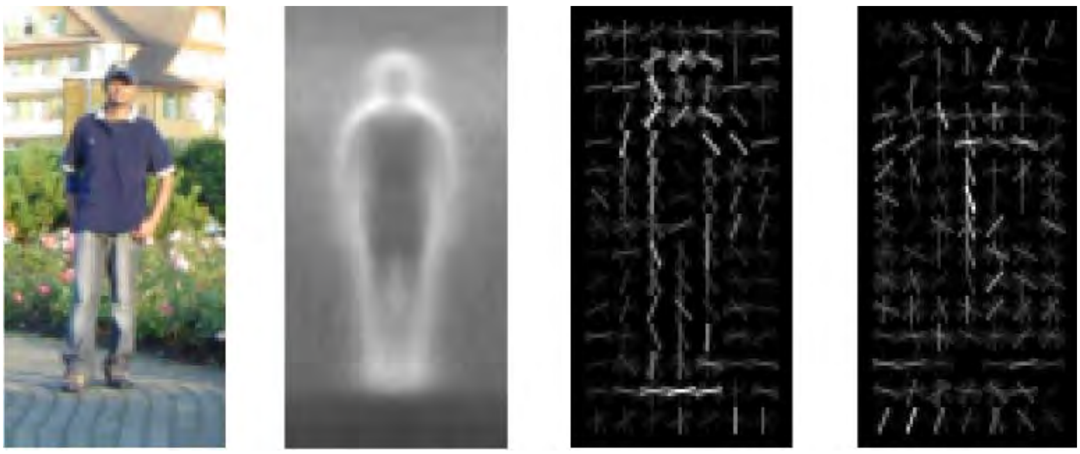

Figure 2.4: Gradient image after filtering [6]

2.2.3 Kanade–Lucas–Tomasi Tracking (KLT tracker):

In [12, 13, 14] when the camera moves, the patterns of the image intensities are changing in a complex way. However, Images are usually related to each other when they are taken in short time and that because it refers to the some scene taken from with slightly difference in the viewpoints.

By assuming the intensities of an image region unchanged when it's displaced. The image sequence satisfies. As the image motion model is not perfect and because of image noise. The image sequence formula can be written as:

$$I(x,t) = I(\partial(x), t + \Delta t) + n(x) \tag{2.3}$$

Consider an image sequence I(x, t).Where δ(x) is the motion field, specifying the warping that is applied to image points between time instant t and t+Δt and n is the noise function.The translation of motion can be approximate by the fast-sampling hypothesis. Later image at time t + Δt can be obtained by moving every point in the current image, taken at time t, by a suitable amount of the displacement vector. The tracker need to be computes the displacement d for a number of selected points For each pair of successive frames in the sequence.

$$\varepsilon = \sum_W [(I(x+d), t + \Delta t) - I(x,t)]^2 \tag{2.4}$$

Where W is a small image window cantered on the point for which d is computed.The control on the new position must be done when the displacement has been found and the new position of the point has been determined. The controlling of the new position given by a template window around the point in frame n and a slave window around the matched point in frame n+1, a cross-correlation coefficient ρ is computed. The corresponding feature in frame n+1 is accepted if the computed ρ. is bigger than a user-defined threshold value ρ0. Tomasi and Kanade in [13] improved the technique by tracking features that are suitable for the tracking algorithm thought using the same basic method for finding the registration due to the translation. The features can be selected if the value of the eigenvalues of the gradient matrix larger than the value of the threshold.

2.2.4 Background subtraction:

The background subtraction is a technique used for detecting an object through extracting the foreground image this method also called foreground detection. This method used to detect moving object in real time using static camera. This approach detects moving objects from the difference between the current frame and a reference frame called “background model” [10].

2.2.4.1 Frame Differencing:

The method of the Frame differencing technique is by subtracting the current frame (t+1) and the background model which is the previous frame t.

$$D(t+1) = |v(x,y,t+1) - v(x,y,t)| \tag{2.5}$$

Threshold (Th) is used to improve subtraction between frames and remove the image noise.

$$|v(x,y,t+1) - v(x,y,t)| > Th \tag{2.6}$$

The advantage of frame differencing technique is its very quick to adapt changes in lightening or camera motion. If the detected object stopped it will not be detected any longer due to comparing between the frames.