Text Detection and Recognition for Semantic Mapping in Indoor Navigation

Submitted by

MUHAMMAD SAMI SIDDIQUI

MASTERS

IN

ROBOTICS AND INTELLIGENT MACHINE ENGINEERING (RIME)

YEAR

2015

THESIS SUPERVISOR

DR YASAR AYAZ

SCHOOL OF MECHANICAL AND MANUFACTURING ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

FORM TH-4

National University of Sciences & Technology

MASTER THESIS WORK

We hereby recommend that the dissertation prepared under our supervision by: Muhammad Sami Siddiqui . Registration No : NUST201260404MSMME62112F.

Titled: Text Detection and Recognition for Semantic Mapping in Indoor Navigation be accepted in partial fulfillment of the requirements for the award of Masters in Robotics and Intelligent Machines Engineering with grade .

Examination Committee Members

1. Name: Dr. Muhammad Naveed Signature:_____

2. Name: Dr. Mohsin Jamil Signature:_____

3. Name: Dr. Syed Omer Gilani Signature:_____

Supervisor's name: Dr. Yasar Ayaz Signature:_____

Date:_____

_____ _____

Head of Department Date

COUNTERSINGED

Date:_____ _____

Dean/Principal

DEDICATION

I thank to Allah Almighty the most gracious and the most beneficent for giving me everything in life. I dedicate this research work to my beloved family who support me in all crusts and troughs of life. All the gratitude to my professor supervisor who helped myself to obtain this milestone.

I am extremely grateful to the National University of Sciences and Technology, which facilitated me to complete my Master degree.

Humble thanks to my colleagues and other people who helped me with their best capabilities.

DECLARATION

We declare that gave any part of the work described in this thesis to support one application to another degree or condition that another university or other educational institution. In any act of plagiarism, if any, and we are fully responsible for all disciplinary actions taken against the United States based on the seriousness of the offense proved, even the abolition of the degree of our

COPY RIGHT DECLARATION:

• The copyright in the test this hypothesis on the student's book author. Copies (by any means) in its entirety, or extracts, may only be in accordance with instructions given by the writer and registered with the tiny library, NUST. The details can be obtained from the librarian. Should this page be part of all these copies produced. Additional copies (by any means) of copies produced in accordance with the provisions of these regulations can not be performed without permission (in writing) to the author.

• Proprietary intellectual property rights, which can be described in this thesis is dedicated to childhood, NUST, subject to any contrary prior agreement, and may not be available for use by third parties without the written permission of tiny, NUST describes the terms and conditions of any agreement.

• To learn more about the circumstances in which the detection and exploitation can happen is available from the tiny library, NUST and Islamabad.

ACKNOWLEDGMENTS

It is the only Allah Almighty who is the source of every knowledge. Thanks Allah Almighty that gave me power and determination to complete this research project. A special thanks to my beloved parents and respected teachers that it was a constant foundation of inspiration and enthusiasm for me. I am also grateful for my colleagues that have supported me whenever I needed their help.

Abstract

Detection and recognition of nameplates is used to increase the map of the robot built with locations of objects. This is very useful in service robots applications where many tasks will be the type to wear reading. The most common in the mobile robot service scenario where the problem is moving Android mobile autonomous in the domestic environment, and builds a map as it moves along, with headquarters at it, predictable the objects in their way and place them in the map.

In the research arena, utilizing the text has been favorite as a landmark in navigations and until now, different types of algorithms are being processed for indoor surroundings for initially detection and then recognition of objects like texts. The main objective of this project is to use a vision system in an autonomous robot to detect and identify signboards in an inside surrounding in such a way that it is able to recognize the desired nameplate in a given environment. Binarization the output image is brought to the recognition of OCR characters. a database of possible sets of signs that are created in the laboratory. From there, the algorithm more like comparing the strings match. We recorded the names of different signs that are in the native setting and target was accomplished of 80% accurateness in text recognition in an image. Robot Operating System is used for the simulations and a novel technique for detection of texts from the indoor signboards is developed.

Table of Contents

List of Table & Figures

List of Abbreviations

OCR	Optical Character Recognition
ROS	Robot Operating System

List of Publications

Research contributions of carried out research listed as follows.

- "Text Detection and Recognition for Semantic Mapping in Indoor Navigation ", International Conference on IT Convergence and Security 2015
- "Text Detection Algorithms for Robot's Navigation : A Survey " , Turkish Journal of Electrical Engineering & Computer Sciences (Submitted)

Research Profile of the Author:

- "Pose Recognition using Cross Correlation for Static Images of Urdu Sign Language." International Conference on Robotics and Emerging Allied Technologies in Engineering 2014.
- "Prefabrication Design of an Actuated Exoskeleton for Traumatized and Paralytic Hands." International Conference on Robotics and Artificial Intelligence 2012.
- " Selection of Suitable Control Techniques for Payload Anti-Swing and Trolley Position Problems of 3DOF Crane," 1st Applied Mechanical Engineering Conference 2014
- " Iterative Linear Quadratic Regulator (ILQR) Controller for Trolley Position Control of Quanser 3DOF Crane," International Conference on Green Computing and Engineering Technology 2015.

Chapter 1

Introduction

1.1 Introduction to Topic

Robots now are able to detect geometrical sign boards and text in a clustered environment. But still algorithms are not as robust as vision of human beings. Different types of sensors are available which can assist to move from a initial goal to an final target. But recently there has been a trend of researching on visual sensors and ways in which vision can help mobile robots to navigate autonomously in an environment. For this purpose, robot should have the ability of self localization to navigate in the environment without errors. It does this by building a map with matching of the perceived landmarks with the expected landmarks ,stored in its database. Following are its steps

- Get an input image
- Detection of landmarks
- Matching of detected landmarks with stored landmarks
- Position of robot in an environment is updated

In field of vision ,data have to be extracted from image to use it for other purposes. Given a image of the surrounding , two possible sets of information can be taken out :

a. Perceptual content and b) semantic content [1].

Perceptual data includes properties of intensity, shape, color, texture. While semantic content includes relation between objects like words, signs ,symbols and phrases. Special

focus will be on semantic content because text contained in the image is useful for us for following reasons:

1) It helps in describing data in the image

2) Text can extracted from image with relative ease

3) It has lots of applications like number plate recognition, understanding sign boards, meter readings etc

Optical character recognition (OCR) engines are found to be very successful in scanned texts but its performance is very low in real world images [2]. In scanned texts OCR separates texts from background by using technique of segmentation , which is not very effective for real world images. In real world texts exist in various forms of sizes, colors, fonts, orientations and it becomes difficult to learn and then recognize the characters. Compared to texts in scanned images, texts in real world are not so structured and region of appearance is not fixed. Because of above problems, accuracy and speed of OCR engines drops. For it the image has to be pre processed for it to be made readable for OCR algorithms. However, localization and detection of texts in real world is still very difficult and it is still an open problem and lot of research is going on it. Following are the main problems which arises in recognition of texts in natural environment [3, 4]:

• Even in the same scene of real world, texts exist in various forms of sizes, colors, fonts, orientations.

• Background in actual conditions is very complex and sometimes objects like grasses, bricks, signs are difficult to differentiate from texts.

• Extra noises, blurs , low resolution, different lightning conditions, occlusion also causes errors in the recognition stage.

1.2 Level of Research Already Carried Out on the Proposed Topic

No particular research has been carried out previously by the department in field of object recognition in mobile robotics. But throughout the last decade many researchers around the world have successfully extended SLAM from indoors to outdoors, to underground and underwater environments, to airspace, and from 2D to 3D. In the department of SMME also various projects have been completed and are going on in the research area of SLAM. The development of visually aided object recognition algorithm would be add on to enhance the performance of previous projects and open a vast arena of unexplored research.

1.3 Reason/Justification for the Selection of the Topic

Despite the recognition of the object it is one of the major research topics in the field of computer vision and robotics, and there is often a need for a system that can identify some of the objects in the environment - classified object detection capabilities. This a very exciting concept of aiding the mobile robots with human-like vision ability. It will enhance the functions of these systems significantly. Also it opens new ventures of research development in the field of vision and mobile robotics.

1.4 Objectives

The Objectives of the project can be broadly classified into two areas, the first one being detection and recognition of landmarks in an environment. And then integration of the algorithm in simultaneous localization and mapping.

1.5 Relevance to National Needs

An essential skill for the Android mobile system is completely independent of the ability to build a map of the environmental sensor data. This project is a direct application of cybernetics, a field that is currently an active frontier for research internationally. The development of it ensures establishment of research and development and awareness of this particular field in Pakistan. This work in rehabilitation can lead to development of new

industrial opportunities in this field. Pakistan will be having a head-start in the field of visually guided prosthetic hand, bringing country at the upfront of technology. Opportunities will be created to work in collaboration with the world's advanced centers of excellence in the field of robotics.

1.6 Advantages

The importance of the task of robotic devices, both in terms of economic and social perspective on the use of robots in the home and office environments, as well as help to the elderly and those with special needs. The visually aided wheelchairs (which is also a type of mobile robot) would allow for those who are physically disabled to maintain their lifestyle and continue to be able to fulfill daily tasks with more convenience.

The algorithms and technology developed can be applied in other fields of knowledge also, hence making it an ideal test bench for advancement in the technology. Moreover, the work shall be represented and published in international community, thereby promoting school as well as Pakistan in international robotics community. Currently there is not much work being done in Pakistan on this field.

The final product of this project will be a mobile robot capable of detecting and identifying objects in an environment and can also be implemented in wheelchair which could be sold commercially to rehabilitation centers.

1.7 Areas of Application

In industry, there are numerous applications in which identical autonomous agents are performing same tasks. This project is a unique blend of fields of vision and mobile robotics. This field has a tremendous application in all the industries having where individuals are required to perform daily activities work. All over the world a need of human-like vision system in a robot is now seen as a necessity, the system developed from this project will be beneficial not only to Pakistan but also to rest of the world.

Chapter 2

Literary Review

Semantic maps are used to make robots interact better with humans and to increase its ability to navigate and motion plan [5]. Work is also been done on using natural language with semantic maps to enable better socialization of robots with humans [6] . Semantics is study of landmarks and its relation to local objects. In it the robot collects selective information and stores it in database ,which can be used to interact with humans in more humanly-way. All these landmarks and information is organized by building a semantic map . One of the way is to build both semantic and metric maps simultaneously [7]. In the above level, semantic maps the landmarks of places and objects like human thinking while in the lower level metric maps stores geometrical properties useful to robot for smooth and safe navigation in an environment. Another way is to build metric map without building a metric map and vision helps in navigation of the robot [8,9].

There are many ways to detect texts and signs from an image in the scene. We would discuss it briefly to just give readers an idea how the semantic information can be extracted from real world. In [10] , vision system is used to detect signs on the road signs to help mobile robot with wheels. the theory of the algorithm is implemented in small bundles by Matlab .In edge information [11] Traffic signals are detected using different filters. In [12], the detection is carried out with the help of neural networks, and 96% will be trained by scaling conjugate gradient qualifier, the detection is carried out by means of the saturation intensity of the sound. based on the color recognition is with outdoor lighting problems. is in [13], the layers of gray images created and YCbCr format size is used to solve the problem is not smooth light level. The text in the real world, the image can be grouped together because of similar height texture. Parts contains texts and those who can not distinguish between support vector machines. In order to recognize the signs, uses remote

procedure excavation. In [14], recognized the posters of the local environment. The scenes in three different points of view and classified the coordinate x is well with him. The sample was then maps with maps and visibility of an object in the wall or on the wall in the direction perpendicular to the plane. In [15] plates and road signs removed by extraction area on the basis of the edge of the text algorithm. The algorithm detects the region on the edge detection, the text is finally extracted, and symbols. The extracted text area from the operator's expansion (using the features that are grouped).

Scale invariant feature transform (SIFT) [16] is very popular algorithm used by many systems for visual detection. In [17] ,an architecture is proposed with dual system. An improvement of the SIFT algorithm, Speeded Up Robust Features (SURF) is used for Simultaneous Localization and Mapping (SLAM) while laser detection is used for motion planning and obstacle avoidance. In [18], recognition of landmarks is done with two main supposition that the surface is planar and position of the object is fixed. A map is generated with locations of objects . In [19] , robot remembers the position and vertical edges of images of surroundings with help of omni directional image sensor. Later it is able to follow the same route to navigate to a goal point without a need for odometer system.

Autonomous navigation and obstacle avoidance can also be done by purely vision based as presented in [20] , here the algorithm takes points from 3D space and maps it to 2D map. Stereo cameras like Kinect can also be used for object detection to be in semantic mapping [21]. There is also work done on extracting information from outdoor environment [22] but its accuracy is very average. Sometimes landmarks are manually placed in the environment to help position detection for the robot, like in [23] color beacons are used as landmarks and are detected by single camera. Multiple cameras can also be used as in [24] . Algorithms are developed with ability to avoid hindrance from pedestrian movement with help of vision camera as done by [25] with help of Histogram of Oriented Gradients method

Texts are not the only property which can be used as labels for semantic mapping,in indoor navigation various objects can be used as a landmark. In [26] , door is extracted from real image to assist in navigating to a goal point. This approach is only useful if there are prominent line features and large colored segmented objects. Detection of doors by visual sensor is also adopted by [27] to reach a goal point, technique they relied on was dynamic window approach. In [28] , pictographs are used for aiding in navigation of a robot in an indoor environment. Camera acquires the image and SURF algorithm detects the sign board to estimate the position of robot in an environment. Distance is calculated from readings of odometer. Similar technique is used by [29] for navigation.

The robot recognizes the images by neural network technique and stores this information to create a database. If then the goal point is set from the available images in database, robot is able to navigate autonomously to that location. In [30], Convolutional neural network with six layers is used to detect and recognize different colored sign boards . The system detects blue or red color then extracts the contour and finally gives to algorithm to recognize the characters. In [31] a unique method of detecting obstacles is used by using variations in intensities of light ray directed at them. But its results are not as effective as using visual sensor for the same problem statement. In [32] , a new technique of using feature points to detect texts in real images was discussed. Process was divided into four stages with Frstner-Kthe corner detector find edge points and center point is found by sobel edge map. Tensor voting verifies the text detected region and non-text regions are extracted out.

• Only text

• Indoor navigation ,doors, colored objects, landmarks

• Outdoor navigation ,traffic signs

2.1 Perceptual content and OCR

Literature is classified on the basis of how a sign board is extracted from the surrounding. There are many way to recognize mages of sign boards found inside surrounding and recognize texts contained on the inner. They can be broadly categorized into three categories

– Color, Shape ,OCR

– Color, Shape , Neural

– Pictures

We will now discuss each in detail and what related work is done on it.

Input images sometimes have to be pre processed before texts from an environment being used for recognition. One of such techniques involves initially to extract color and focus on this region. Next shape will be detected and only those fulfilling the required criteria will be sent to the Optical character recognition engine for character recognition. There is some work done on using text recognition for robotic navigation. In [33] , defined landmarks and texts from room numbers are detected by the system and in [34] map of the indoor environment is already provided for robot's navigation . It reaches the goal point by initially comparing the model of door and then uses it to see text.

In [35] authors have developed a system which detects constant set of font, background and size with no focus on mapping. Constraints are part of every algorithm as no system is developed yet which has capability of text detection in surroundings like human beings. Two constraints in [36] are that the texts are placed few feet above the ground and are following defined guidelines. Also that the semantic mapping module is

independent of navigational module. Initially the robot roams in the indoor environment and take snaps of sign boards periodically.

Then character are recognized from OCR engine. Finally depth sensing is amalgamated with text recognition and location of it is calculated. Logistic regression classifier [37] is a very common technique used for text detection in a real world environment. Feature vectors like local variance and edge detection are extracted from every frame and calculations are then used for classifier. This module detects the possibility of texts occurring in an environment and then the next module recognizes the texts contained in it. Very convenient way to make an image readable for OCR engine is to threshold it to two levels [38]. Fluctuating lightning conditions varies the output of the threshold image, so a set of images with different of threshold levels are calculated. These output are then input to popular Tesseract engine and image with the highest score is stored [39]. In [40], colors are used to detect possibility of text occurring in an image and then detects edge points. Single block of color is usually recognized by connecting components method.

2.2 Perceptual content and Neurals

Using neural architectures for landmark detection is also a very common technique. In [41], an algorithm is developed which recognizes handwritten numbers and different traffic sign boards by using a filter trained with mixture of shifted filter responses to recognize patterns. Neural networks can be different techniques ,like in [42] fuzzy-neural is used to detect sign boards. It has advantage that changes in size and orientations will not alter the results of the algorithm.

Convolutional neural network and feed-forward neural networks is dependent on weights and connections on every layer. Steps for an algorithm to detect an object from camera feed:

– Direct camera to start obtaining live feed

– Grab a frame of fixed size from video for processing to be give for input layer

– Image processing is done and contours are passed on to neural architecture with multi level of layers. Different things are passed on to neural architecture

In different techniques, for example in some, shapes of the objects are segmented from the image.

– Descriptors are made from extracted objects to contain necessary information

– Recognition is done in different phases and levels. Output layer of Neural network recognizes different sign boards. Supervised and unsupervised training is conducted with to test similarities with dataset images.

In [43] ,color images are analyzed to detect texts in it by unsupervised method with multiple attributes of wavelet. Artificial neuron network method is used in [44] which learns from the training dataset and adapts to real time by varying weights of links. Linear discriminant analysis [45] is common method in statistics approaches. In it translation matrix is formed from objects extracted from an image and algorithm is able to differentiate each object from another.

2.3 Pictorial Detection

Pictures Detection and recognition of objects with just camera is difficult as depth perception is to be calculated for each object on the scene and variations of lightning conditions in an environment. If the landmark to detect is just an image without any text in it requires different techniques to be detected. Lot of work is done on this stream of vision too. A simple technique for object detection is template matching [46] , but it has limitations of detecting an object in variable lightning conditions. This is because it works by matching template image with input image. By using method of calculating changes in radiance value

[47] , limitations of lightning conditions can be encountered but its performance is compromised if scale of image is changed.

In [48], algorithm is proposed called as SIFT which extracts local features of an image and it has an ability to give desired results irrespective of changes in lightning conditions and scale. But the amount of time it requires for calculations make it unsuitable candidate in navigation for a robot. In [49] , SURF technique is proposed which is based on SIFT but is relatively faster in computations. It uses Haar wavelets to extract feature points from the object. This technique is also used by [50] for navigation of indoor service robot.

2.4 Semantic Content

Literature is full of images of workers extract the text. But most of them are in the primary process, a page of text on the screen. The next step is to define the text separately. Finally, the word from the individual shapes by processes such as machine learning occurs. In [51], and offers a new fast algorithm change line width (SVT). Dual Card with the end edge of the detector. Parallel lines to estimate the width of the stroke in the individual pixels. And classified those same width to one character. However, the effectiveness of this method depends on the levels of noise and blur. In [52], and the way the difference between the letters or characters based on engineering-related components. In [2], the author stable extremal regions the maximum amount of used properties, color and geometric characteristics of the text referred to output.

In [1], and MSER Almighty uses a neural network with the wrap. His best performance is good conditioning text in images with a simple Google Goggles figures while providing best performance there in the images of few words. The researchers studied the OCR performance with unlimited different point of view of texture and light. His goal was to improve the performance of the campus environment. Not carried out to detect

words. In [10], it uses the vision system to detect signs of road signs to help mobile robot with wheels. The algorithm is applied to the theory of demand in small packets of information. In the edges recognized [53] traffic signals c using different filters. In [12], the implementation of neural networks recognition and are getting 96% by expanding fi cations associated precision gradient coaches, is performed to detect the severity of tone saturation. Disclosure on the basis of Color had with outdoor lighting problems.

In [13], the layers are gray images and the use of forms of YCbCr size for solving the problem of the light level is not smooth. The text in the images of the real world can be classified due to a similar level of texture. Parts contains texts and those who are no different from the automated support beam. Character Recognition, is used this technique of drilling equipment remotely.

In [14], it was revealed the paintings of the local environment. It is supposed sections are classified in three diverse points of view and coordinate X is fine with it. It is then removed from the sample card map important article and health in the wall or on the perpendicular wall on the plane. In [15] removed the plates and traffic signs from the extraction zone algorithm is based on the edge of the text. The algorithm detects the region through the edge detection, then the text and the characters are deleted at the end. It is extracted from the text area expansion operator (using features that are assembled). And described in [36], was born SLAM map with text. System expects the text area seeded for logistic regression and read the characters in the OCR characters. In [39] the color portions are extracted. Divides combine to find the same blocks of color to display the text of the panel in a variety of colors and pixels.

In order to achieve effective results of current problems in the detection and identification of the text as it is displayed in the following order [54, 55]: • the text in the real world, there are many sizes, colors, textures, fonts, scales. • background overweight, significantly change the results, and the text is very difficult to distinguish. • beams of different colors, constipation, noise, sharpness, resolution does not help the cause

Chapter 3

TEXT DETECTION IN INDOOR ENVIRONMENT

In the field of robotics, and knows navigation and movement of the robot to find a safe passage from the current point to point away from the target. For this purpose, various types of visual sensors, sound, etc. They could be used. Optical sensors have Latham man advantage and to think for themselves to increase as the ability of robots. The ability to extract information from an image of great demand in the area of computer vision. There are two types of data that may be extracted from the image: first perceptual content, where we have some of the functions can, for example, to explain, for example, the shape of the object, color, texture

Secondly, there is a semantic content that deals with the issue of letters, words, symbols, etc., and how they relate to each other. Literature in image and text wide open. For example, in [6], the author would like the number of pixels without taking into account the need for a threshold. Thus, the text that extract most of these pixel blocks.

If the robot is to be understood that surrounds us, as human beings, then you should be able to recognize the information in the kitchen and in the bathroom, curtains or doors, etc. This is achieved through a semantic mapping, which transmits the information to the navigation ability. This will help to achieve better interaction between the human robot. [7] A concrete example of [5], which uses natural language semantic robot and makes better suited in the human environment. There are other ways to build a map of tags with metric cards and pumping semantic information in it. [51]

This helps cards (static and dynamic obstacles) and map semantic relationship is determined to move the detected objects. It will even place a primary and a form called an outdoor or go to the home environment ahead.

The following are the important points:

- Getting sample data
- Detection images Features
- If you observed the same signal, and then the link with the previous data
- Landmark is updated each time the position corresponds to a new reference.

If algorithms and optical character recognition letters of images in real-time use, then they are not effective for scanned documents. Thus, the main obstacle to the use of these processes is to excerpt the information in a way that will achieve the best results. A lot of work is in progress, research has been to do a better job in multiple lighting conditions, and the shadow color, background, and an orientation, etc. We discussed the possibility of the text to help map semantic and strong. The space required may have the image that was easily cut. And also it helps in the surf automatic robot.

To solution for our region to recognize the text of the clause in the inside surrounding , divided the scenario in three cases:

- First to find the color of the traffic lights in the environment
- Second be able to detect the shape of the billboard.
- Finally, when it was revealed the text to be recognized.

Figure 1-2 represents the sample pictures from both inside and outside environment.

Figure 1 Indoor Sign Board

Figure 2 Outdoor Sign Board

3.1 COLOR DETECT

We discovered that the required color red or green marker graph back projection. The advantage of being unable to meet varying illumination, in contrast to the clogging happen. This helps to find a region of interest in the image. We create an image from one channel to the image, the probability of an image similar to that required larger pixel size. Then, the input image taken into the schedule. In addition, this schedule with the schedule of projected back up, and is calculated by a choice of the corresponding pixel. The result after a threshold setting gives us the extracted image.

Originally, the graph calculating the color to find and where to find TNnew, TUnew. Find the relationship

$$Rnew = \frac{TNnew}{TUnew} \tag{1}$$

Then back projection of TEnew calculates the likelihood of necessary object

$$T\text{Vnew}\ (\text{TxN}, \text{TyN}) = \text{TEnew}[\text{Tgnew}(\text{TxN}, \text{TyN}), \text{Tanew}(\text{TxN}, \text{TyN})] \tag{2}$$

In this Tgnew called hue and Tanew called saturation point (TxN,TyN)

$$\text{TVnew}\ (\text{TxN},\text{TyN}) = \min\ [\text{TVnew}\ (\text{TxN},\text{TyN}),1] \tag{3}$$

Next step is to disc convolutional

$$\text{TVnew} = \text{TSnew} * \text{TVnew} \tag{4}$$

In which TSnew happens to be disc kernel.

Now the current location of the maximum points to the location of the object. If we wait threshold image area to an appropriate assessment provides a good result. Figure 3 displays the output graph revealed marker color and Figure 4 shows the pseudo code diagram of the color detection algorithm.

Figure 3 Color Detection of SignBoard

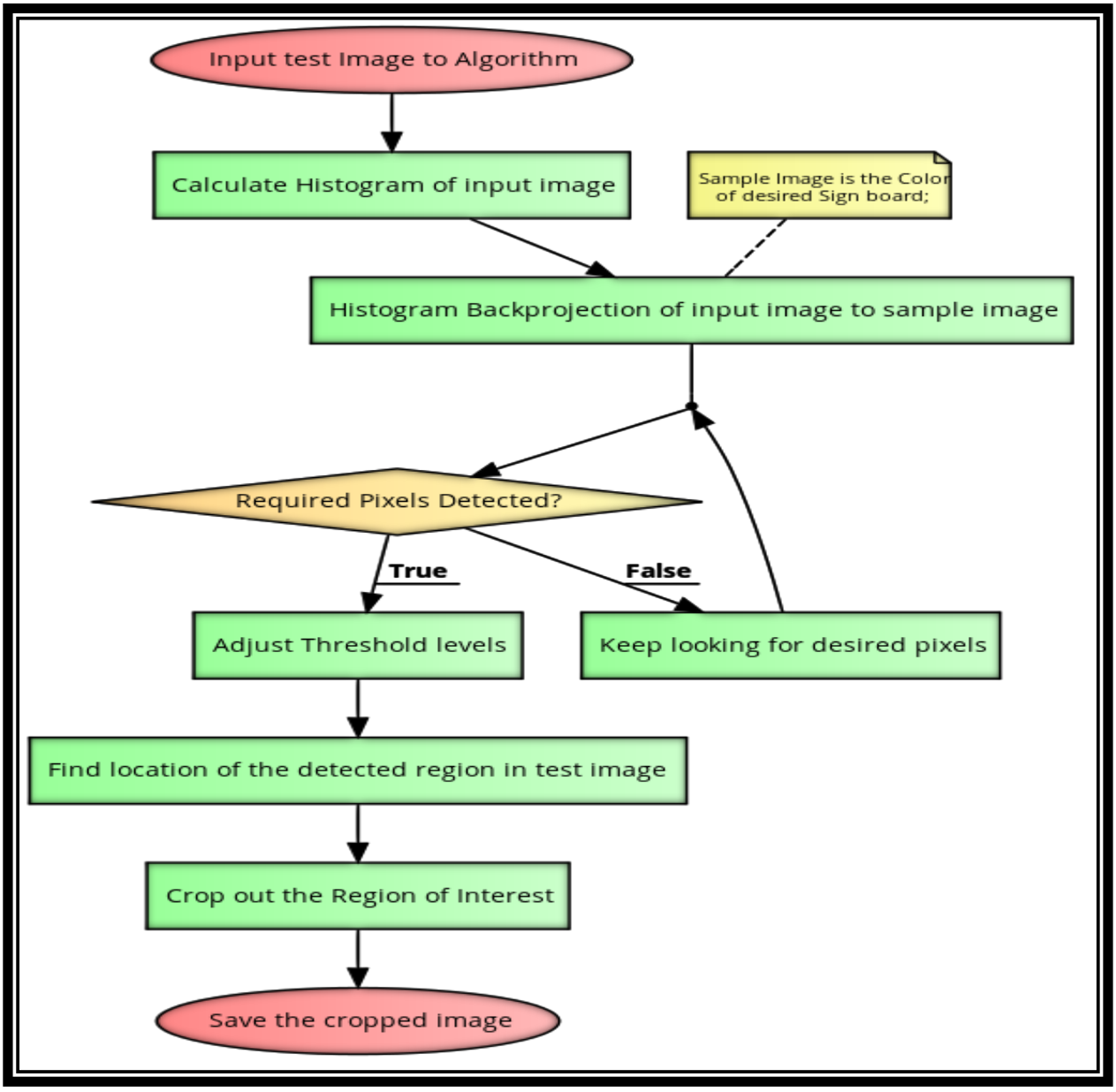


Figure 4 Pseudo Diagram of Algorithm for Color Detection

3.2 SHAPE DETECT

Having found the desired color in the environment, we need to find a way to make sure that the goal is for the panel. For this, it is the outline of the detected objects. Polygon curve is generated, which extracts a number of points of the object store. And to develop criteria to select an object in place, which are:

- Convex is what shape is.
- In it four things are found called vertices.
- In it the detected angles are all ninety degrees.
- Outline of a large area of relatively (even to remove this sound).

To change the image on a binary level solutions to gray, and manually set the threshold. Solve the problem of small noise, more we reduce the image first down, then rise again the same scale. Hakeem player has detected square shapes with different colors property.

To find out how wise edge detection system, the process is divided into the following steps:

1) It is known that size of the filter of gaussian disc is $2\mathrm{t}+1 * 2\mathrm{t}+1$ is:

$$H_{p\,o} = \frac{1}{2\,\pi\,\sigma^2} * exp\left(-\frac{(p-t-1)^2 + (o-t-1)^2}{2\,\sigma^2}\right)$$

2) Image points are high density gradient.

3) Removal of additional fine lines because of edge detection.

4) The threshold again to increase the chances of detection feature.

5) Removal of edges that have nothing to do solid edges.

It is expanded as a result of this operation to remove a certain image of the holes. Function below donations from all the features found in the squares and directed the picture. Figure 5 shows the output of the algorithm which displays exit sign as a prominient feature . Figure 6 and 7 show images of pseudo diagrams of the shape detect algorithms.

Figure 5 Detection of Exit Plate

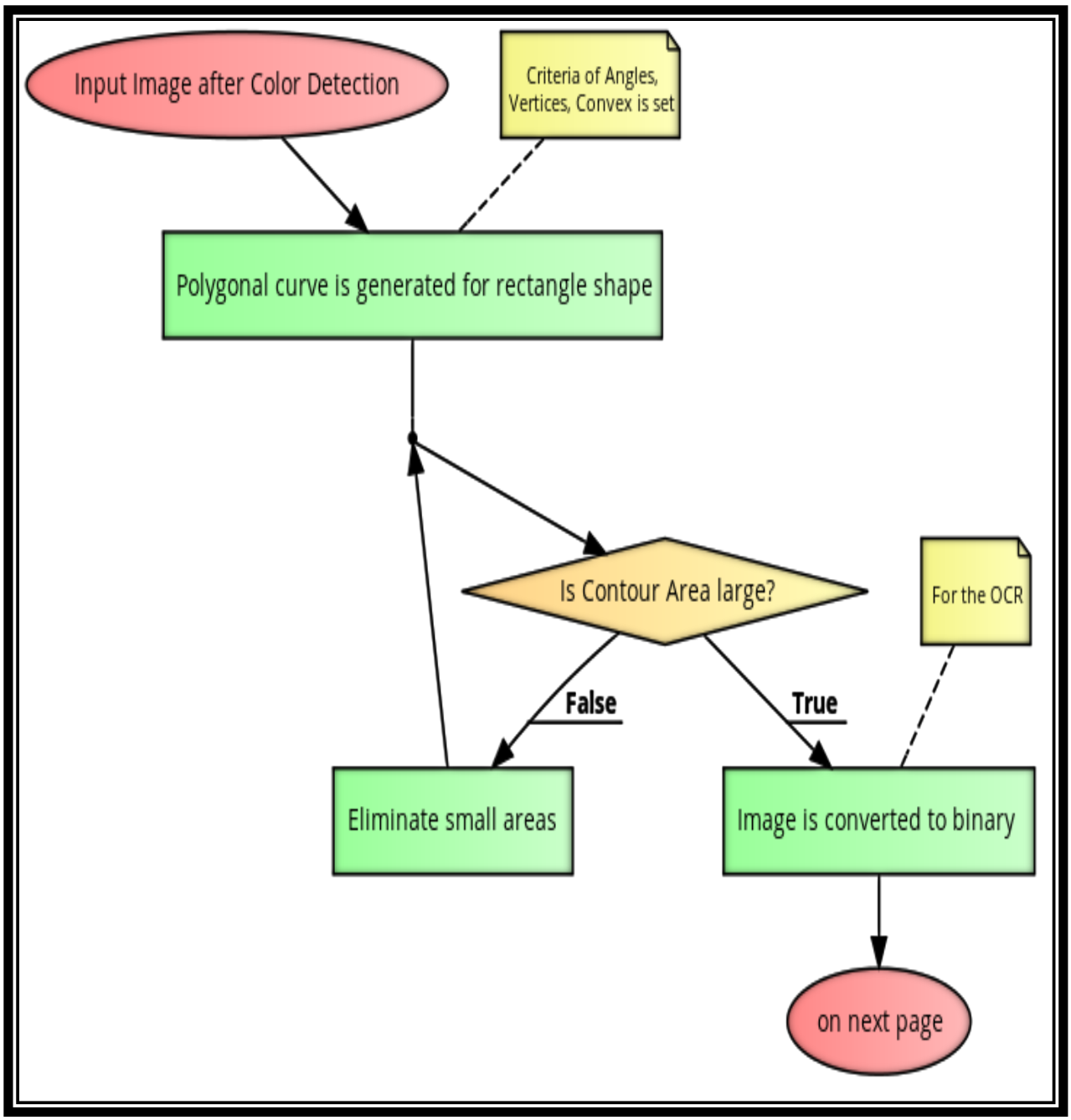


Figure 6 Pseudo diagrams of the Shape Detect Algorithms

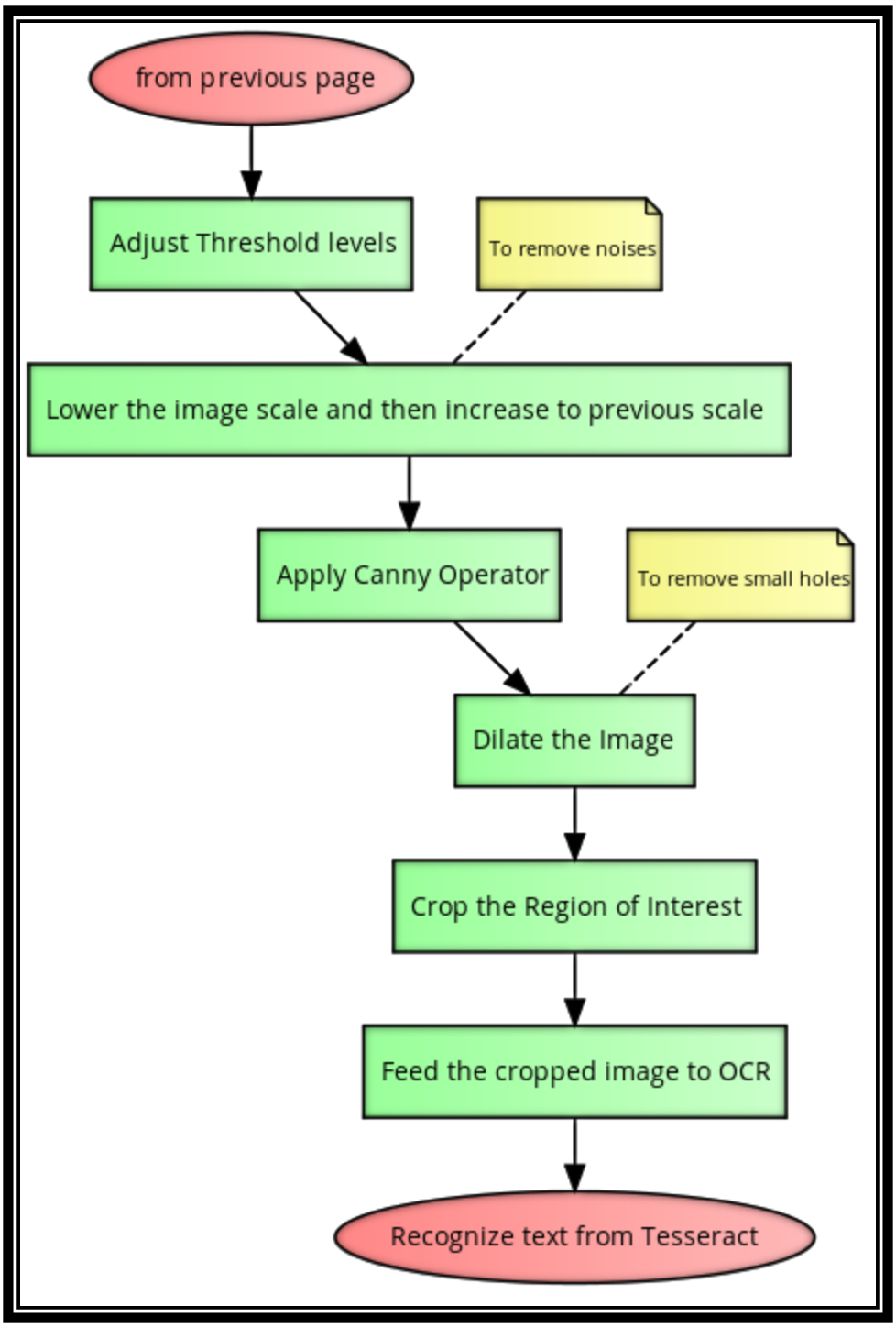


Figure 7 Pseudo diagrams of the Shape Detect Algorithms

3.3 TESSERACT

Optical Character Recognition is defined as a method of converting text in images to a computer printed text. Following are some of the applications on which OCR method can be used:

- To read documents like identity or business cards, brochures
- Scanning and maintaining database of receipts and invoices.
- Data entry system for passport or visa.
- Processing of bank statements and cheques.
- Read utility bills and legal bills.
- Car plate recognition in real time traffic.
- Scanning for required information in insurance papers.
- Finding phone numbers from a card and storing it.
- Conversion of printed books to an online edition.
- Search for images on soft copy of books.
- Making personal handwriting readable for computer system like tablets and smartphones.
- Assist blind and visually impaired persons.

After reading from above mentioned forms it can be processed and given as an input for machine to use it further like in machine learning, text mining and text to audio. Fields like computer vision, artificial intelligence ,image processing and pattern recognition research heavily in it.

OCR algorithms have evolved too much since its introduction. First they were capable of only recognizing texts in one image which had to be trained individually for each character. But now more robust algorithms are developed which have capacity to detect and recognize texts from more complex images like multiple columns and other expressions. .

Types of OCR

Text recognition algorithms normally an offline system, they process on static images. But some algorithms are also online , that is they learn the pattern on which the pen is moving, when it is lifted and then put on the paper and in which way the parts of hand writing was done. These online algorithms do not just depend on graphs and segments like offline algorithms. These dynamic recognition systems make algorithms more efficient.

There are mainly four types of Optical character recognition used in research arena :

- First one is Optical Character Recognition relying on using one character at a time.
- Second type reads words from the image by noticing spaces between them.
- Third type focuses on handwritten texts, they rely on techniques of machine learning. They are known as intelligent character recognition method.
- The fourth type known as Intelligent word recognition, it detects joined handwriting words from a sentence.

Pre-processing

Due to various reasons like noises, orientation, background clutter and object occlusions, optical character recognition systems tend to pre process the image to improve the accuracy of detection. Some of the techniques which are being followed in the research arena:

- One of the common technique is to adjust the orientation of texts in an image horizontally or vertically. Texts are rotated either clockwise or anti clockwise. This is known as Deskew.

- Another technique called as Despeckle is used to eradicate spots from image and adjust the edges to be more smooth.

- Binarization is the technique to make an image to just two levels of either zero (black) or one (white). Image is converted from grayscale to a binary image. It is a very simple technique to extract out the text from the background and efficient to an extent that commercial companies rely on heavily this technique for their algorithms. It depends on what type of input image is fed to the algorithm as various types of images need different binarization technique . Types of images may include like scanned documents, real time scene images , and documents which are very old dated.

- To remove objects other than glyph in shape of boxes and lines,technique known as line removal can used.

- Sometimes processing is done to distinguish different parts in text like columns , end and start of new paragraphs or recognizing captions of image. This process is known as layout analysis.

- In another technique a set of conditions are set and then the lines or different words and characters are recognized accordingly.

- Every language has different style of words and alphabets, so optical character recognition algorithms are modified to tune to specific script.

- If text in image like alphabets are not connected properly due to quality of the image then processing must be done to connect them properly so accuracy of the algorithm is increased.

- Scale of each sentence is kept to a constant ratio, so a process can be fixed to detect and recognize the text.

In occasions where spacing of fonts is equal ,then segmentation is done by drawing vertical lines in the image and finding the areas where text interacts with the lines. In scenarios of fonts being proportionate , complex methods are needed because spacing between the words are not equal and verticals lines may intercede more than one characters.

Character Recognition

Technique known as photocell OCR relies on the matching of matrixes of stored image with the input image on the pixel level. It is also called pattern matching or image correlation. Efficiency of this technique is highest in cases of texts that is typed rather than a new handwriting style being fed to the algorithm. The input section for the text algorithm should be excluded from rest of the image and the stored data should be as same font as the input data.

In order to recognize more robust hand written texts, techniques which fall under the section of intelligent optical character recognition are used. They extract the properties like lines and its orientation, number of its interaction and finding whether the loop is closed or not. These features are then correlated with vector defining each character stored in the data by algorithm. Classifiers are then utilized, like k-nearest, to match the input data with the stored data. There are many softwares working on optical character recognition, most famous of them is tesseract and cuneiform. The are based on technique to recognizes a character on two stages. The second stage is known as adaptive recognition and relies on the characters which are detected with good probability in the first stage to detect the other characters in second phase. Such technique is useful for texts which have low image standards like are not smooth or blur.

After you remove the sample card, Next, we need to recognize the text in the image. Optical character recognition is used to find a way and get the text in the images from the navigation of the robot around to read. This vision for the character recognition technology (OCR), which is mainly used for scanned images. This pattern of using access to the navigation of robots leads to errors and poorer recognition. Thus, we need to extract the image processing in the area of interest. This part will be cut and fed to the OCR recognition algorithm in the text. Our results prior to processing technology to improve the image of the text recognition.

Tesseract OCR is widely used to perform. The experimental results and found that, by the way, a unified bloc and one taken from the text works best. To enable them to detect paragraphs breaks, we automatically put all the pieces of the page. 8 is removed the picture from the panel after the process is complete appears. This picture is the entrance to the algorithm to identify the text. Figure 9 and 10 respectively shows the Pseudo code and its diagram for the algorithm of detecting text from an input image.

Figure 8 Extracted Region of Interest

Algorithm 1: Pseudo Code of the Entire Algorithm

```
Result: OCR from the sample image
Input test Image to Algorithm;
Calculate Histogram of input image;
// Sample Image is the Color of desired Sign board ;
Histogram Backprojection of input image to sample image;
while !Required Pixels Detected do
    Keep looking for desired pixels;
end
Adjust Threshold levels;
Find location of the detected region in test image;
Crop out the Region of Interest;
Save the cropped image;
Input Image after Color Detection;
while the color is within same range as required do
    if Contour Area large then
        //Criteria of Angles Vertices, Convex is set;
        Polygonal curve is generated for rectangle shape;
    else
        Discard small areas;
    end
end
//For the OCR ;
Image is converted to binary ;
Adjust Threshold levels;
//To remove noises;
Lower the image scale and then increase to previous scale;
Apply Canny Operator;
//To remove small holes;
Dilate the Image;
Crop the Region of Interest;
Feed the cropped image to OCR;
Recognize text from Tesseract;
```

Figure 9 Pseudo Code Algorithm

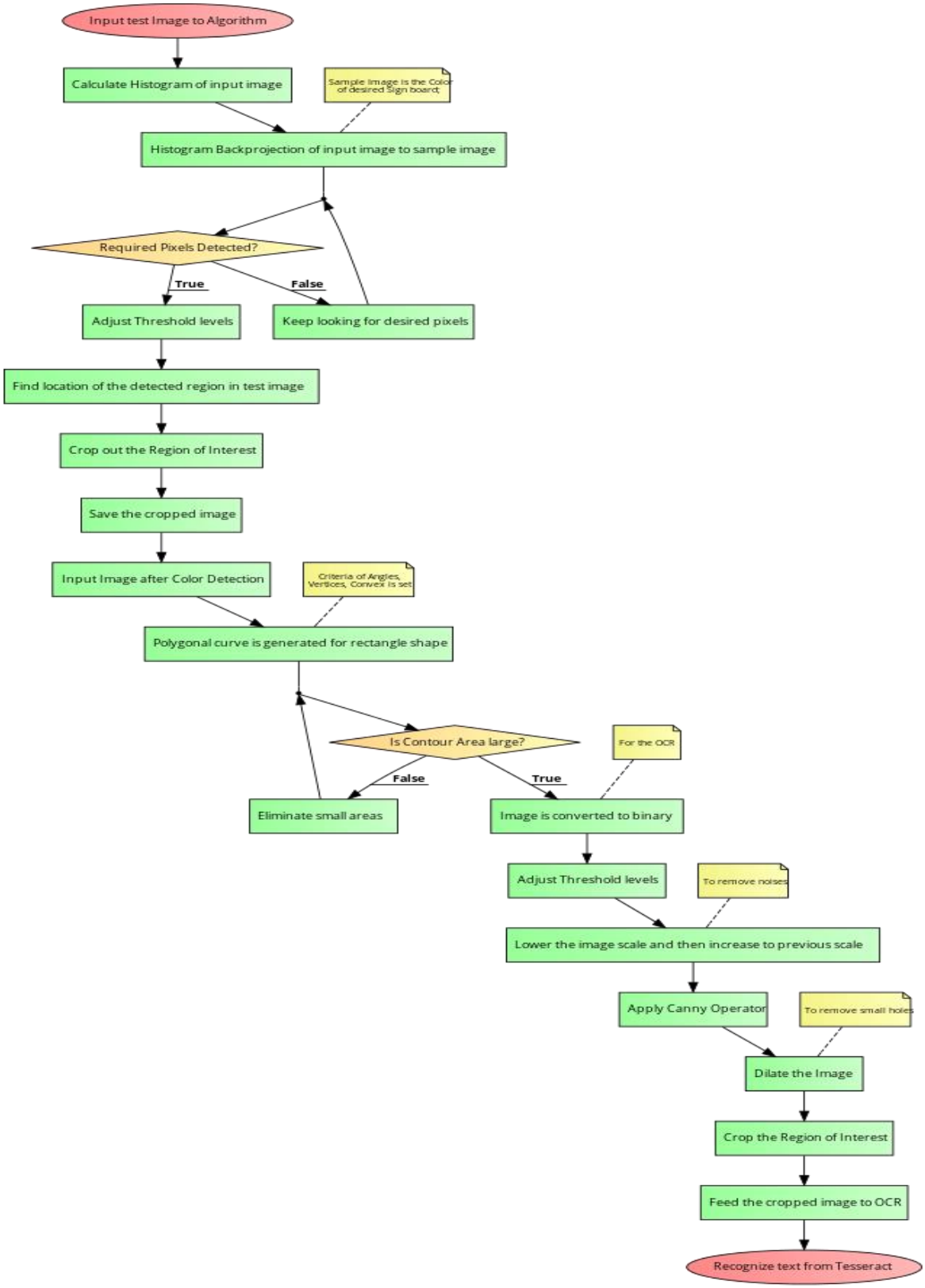


Figure 10 Pseudo Diagram for Algorithm

Chapter 4

SIMULATIONS

In this chapter , a brief overview of the simulations platforms which were used and how they can be related and useful in the big picture of the project. A table for the results achieved by performing the image recognition algorithm on test images is also presented.

4.1 ROS/ GAZEBO

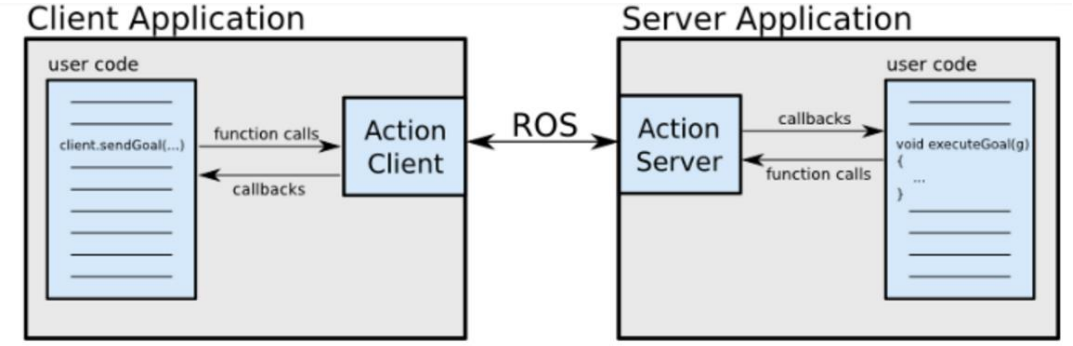
ROS

The Robot Operating System ,more commonly known as ROS, is created by Stanford Artificial Intelligence Laboratory as an framework which is open source and is used for the development of robotic softwares. It has applications in both academia and commercial arena. It has the advantage that the communications between two systems becomes easy by using the packages , called as stacks in ROS, which has usable executables. There are various kinds of robots which are compatible with ROS.

Master feature is very essential part of ROS. It functions as storing of information related to topics and services so that nodes are able to connect with each other. All communications between nodes are done through Master ,which gives the lookup information so that nodes can link themselves.

There are instances where a node needs to request another node for interaction, for this there is a feature in ROS called services. After a reply is sent from the requested node ,then only the communication takes place. In the condition when the reply from the node gets delay then the service can ask about the update of the request or altogether decline the request by

actionlib package. This type of protocol consists of three components and is based on ROS messages . The three components for these scheme are goal, feedback and result. The node requesting a service is called ActionServer and node which gets the request is called as ActionClient. A request send from ActionClient is called goal, which is received by ActionServer and it provides the required feedback to the ActionClient. Finally, at the end the ActionServer gives the result message to ActionClient that whether the goal was achieved or not. Below is the diagram of system:



After all the nodes are set, ros has a feature of ROS launch files. When these files are executed in ROS then all nodes written in it are executed and parameters for each node can also be set in this file. They can be launched from the local server or from the remote server.

GAZEBO:

Doing real world experiments can be costly and dangerous if it is done without any software simulations and testing. Gazebo is such kind of simulator in which various robots can be built and tested for mapping and navigation,checking efficiency of battery life, removing errors in grasping and handling etc. Gazebo is capable of building simulations for environment and robots in 3D, including features like lightning, inertia and gravity. Robot

Operating System (ROS) can used to establish a link between the robot and the simulation, and testing of the experiment can be done without any financial constraints.

To install Gazebo in the system, two major components can be used. One is gzclient, which shows the output graphical images for gazebo. And other is gzserver, where the system does all the computations and calculations. After this ,a connection has to be established between Gazebo and ROS (gazebo_ros_pkgs) so results of gazebo can be displayed through ROS.

GAZEBO_ROS Package :

- ROS node name : gazebo
- Plugins :

gazebo_ros_api_plugin

gazebo_ros_paths_plugin

GAZEBO_PLUGINS includes:

- Sensory
- Motory
- Dynamic Reconfigure

The graphical interface in Gazebo is good and has tools for simulation of lifelike events.

Below image is of the camera tool used for vision of the robot in an environment. Similarly ,many complex situations can be generated and visualised through gazebo and all the movements can be recorded for further use.

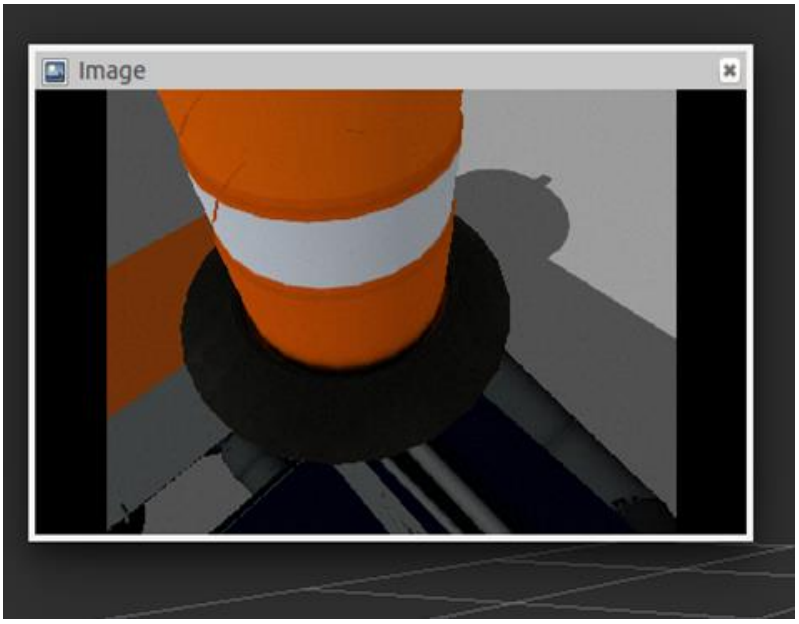


Figure 10. Image view in Gazebo of Simulated Camera

Experiment :

The algorithm for the analysis of performance under different conditions set lightning and harmonization, has been collecting pictures of local Testing indoor air. Experimental results show that the proposed method, we open the possibility of indoor environmental consulting signs and recognize text in it, despite the limited likelihood of unrest and changes in font size, alignment, text characters variable number and changing lighting conditions. The powerful algorithm in experiments in real time. The results showed to us that proposed was able to achieve its task and get out the required text from the image. We then use the tables is used as a reference signal and to assist in navigation of the robot. Model of wheelchair place in the arena took the chair as shown in Figure 6 and an image database was created.

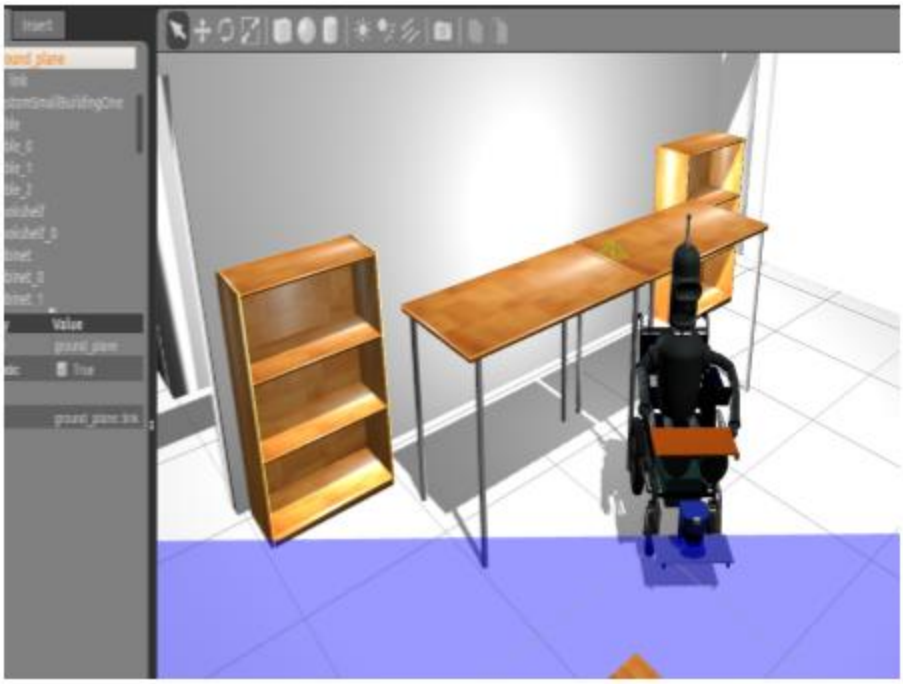


Fig. 6. Gazebo model of wheelchair

The difference in color density does not change the results as a threshold defined as algorithm. This method is strong in terms of font sizes, alignment, and the effects of lightning reflection of variables. Is brought binarize image output to the characters, OCR recognition. Data for possible groups of signs that have been created in the laboratory base. From there, NDS Fi algorithm like comparing strings match. For example, the output test card image for example inputs and outputs of the OCR is 0EXIT0. Comparing the method and it shows that the Chance of reaching much higher. In addition, we could see the names of various characters in the local environment. It is stated that in the end 80% accuracy in identifying text in an image.

Figure 7 . Exit nameplate detected

Figure 8 . FOAM region detected from image

Figure 9 . RED region detected from image

TEST SCENARIOS	WITHOUT PROCESSING, OCR OUTPUT	WITH ALGORITHM OCR OUTPUT	ACTUAL WORD
1	NIL	QEXIT	EXIT
2	NIL	ANiMAl FO-‘JSE	ANIMAL HOUSE
3	NIL	FOAE	FOAM

Table 1 : Output of words Detection from OCR

Slam Gmapping:

In ROS the map of the environment can either be generated or the older map can be retrieved as a ROS service. We have tested a saved map and played the bag file for the slam gmapping algorithm. The theory is that the robot is moved in the environment in a stable speed so that the algorithm saves the odometry and laser scan data for each orientation of the robot. In this way map is generated, below is the figure which we achieved during testing.

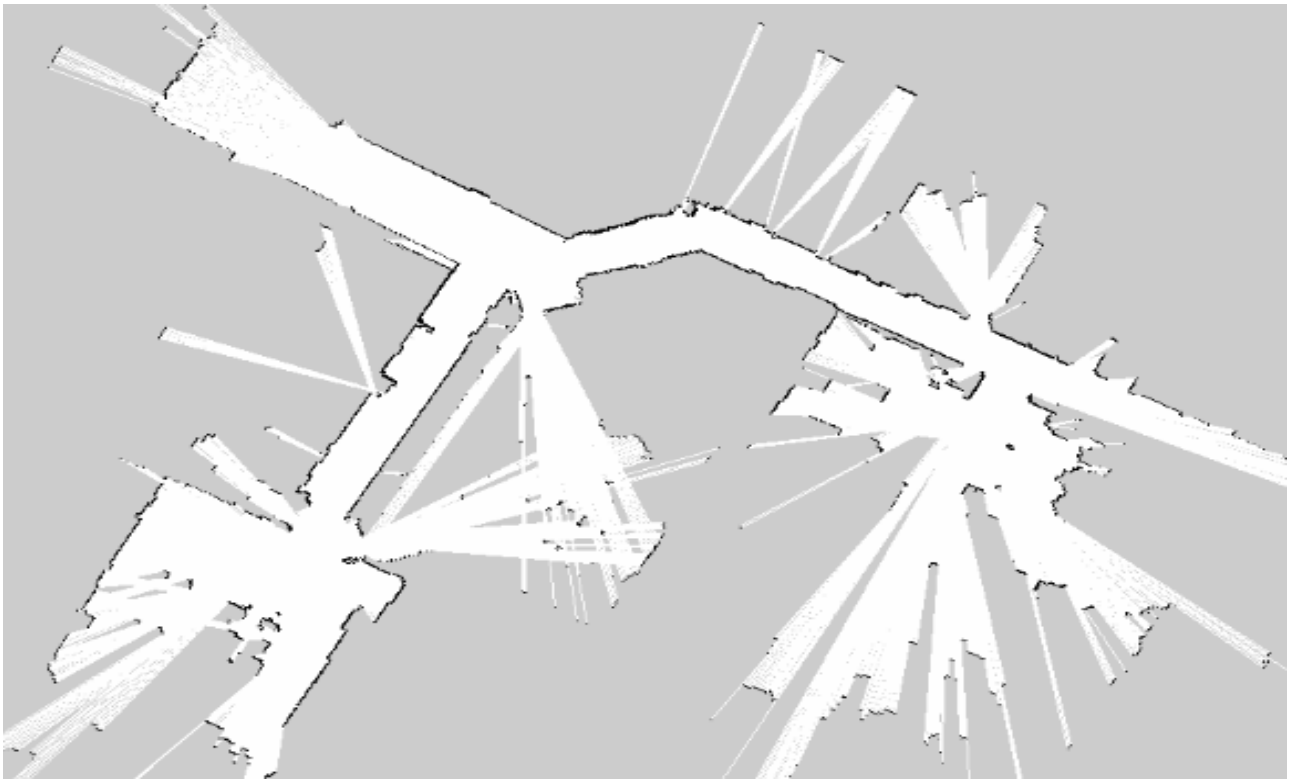


Figure 11. GMAPPING Algorithm in ROS

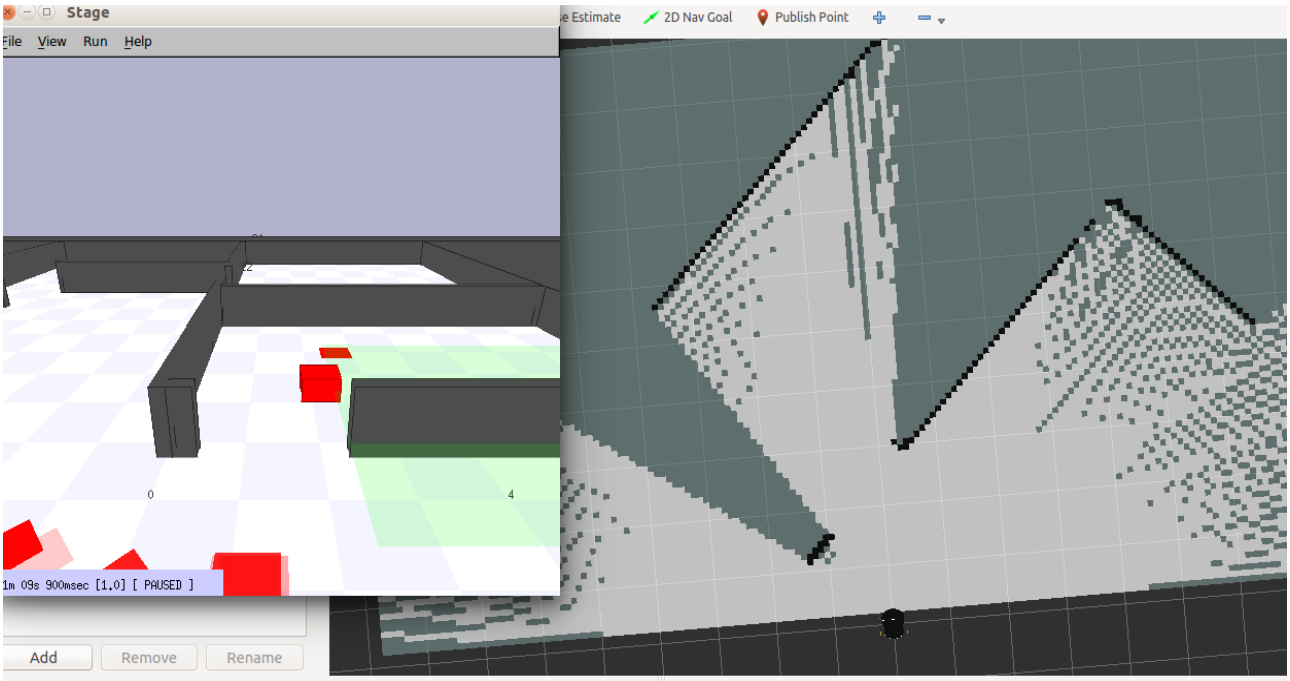


Figure 12. Laser Mapping Simulations

Chapter 5

CONCLUSION AND FUTURE RECOMMENDATIONS

The scope of this project is limitless and with the world rapidly towards the age in which robotics is becoming an essential part in the lives of the human beings, the technology developed can be utilized in various other projects too.

For future, we would continue the project and focus on:

- Using neural network architecture.
- Testing on Pioneer3AT robot.
- Marking landmarks in SLAM.
- Machine Learning in Navigation.

Below is the brief description of how the future work can be carried out . Portions of each module is already complete and is explained for the convenience of others working on it to catch up.

Object Detection :

A very common method for detection of objects is known as Haar like features, introduced by Paul Viola et al [56] . In this technique, there are set of rectangular regions which are aligned side by side. Intensities of pixels vary at different position for different set of rectangular box. Edge features, Line features, Center-surround features and Special Diagonal line features are the set of features in Haar like technique.
To detect the required object in the image, position of rectangular regions are utilized.
There are three stages (as mentioned in [57] and [58]) for the Haar like features to detect an object in a given image, they can be summarized as :

- Initially, to calculate the features in an image swiftly, an Integral image is produced.

- Then a learning Algorithm (e.g. AdaBoost) is executed to extract out a particular set of visual features from large amount of data.
- Finally, background regions are rapidly removed by using a technique comprising of cascading various complex features.

Object Tracking:

Object tracking is when we have to follow a targeted object in upcoming stream of images. One of the method to achieve this task for non-rigid bodies is by using the technique proposed in [59] called the mean shift method. The basic approach of this method is to create a kernel window in the image. Then mean positions of points are calculated iteratively which occur inside the kernel window. Then the kernel window is translated towards the object's centroid. Through all this process ,we obtain points which are actually the detected features for a particular region of interest. Continuously Adaptive Mean (Cam) Shift continues from the points obtained by mean Shift. Its main task is to find an appropriate window size and optimal rotation after the center of the object is found . In this paper [60], author has presented a technique for object tracking in occluded scenarios by using cam shift algorithm.

Optical Flow:

In optical flow , a relationship is established for the object moving in consecutive set of images and display the direction. There are many techniques for it, like the Iterative Image registration method shown in [61] and Lukas Kanade method discussed in [62]. The latter technique is a renowned one and it works on the principle that in the neighbouring pixel, optical flow remains constant.

Stereo Camera:

We can use both the right and left images from stereo camera to determine the location of the object to be tracked. One of the most common method is known as normalised correlation and in this technique we use the images of the tracked object from the left side of the camera be made as a template. Then this specific area be moved to the whole image of right side of the camera and obtain the pixel coordinates of the rectangular region.

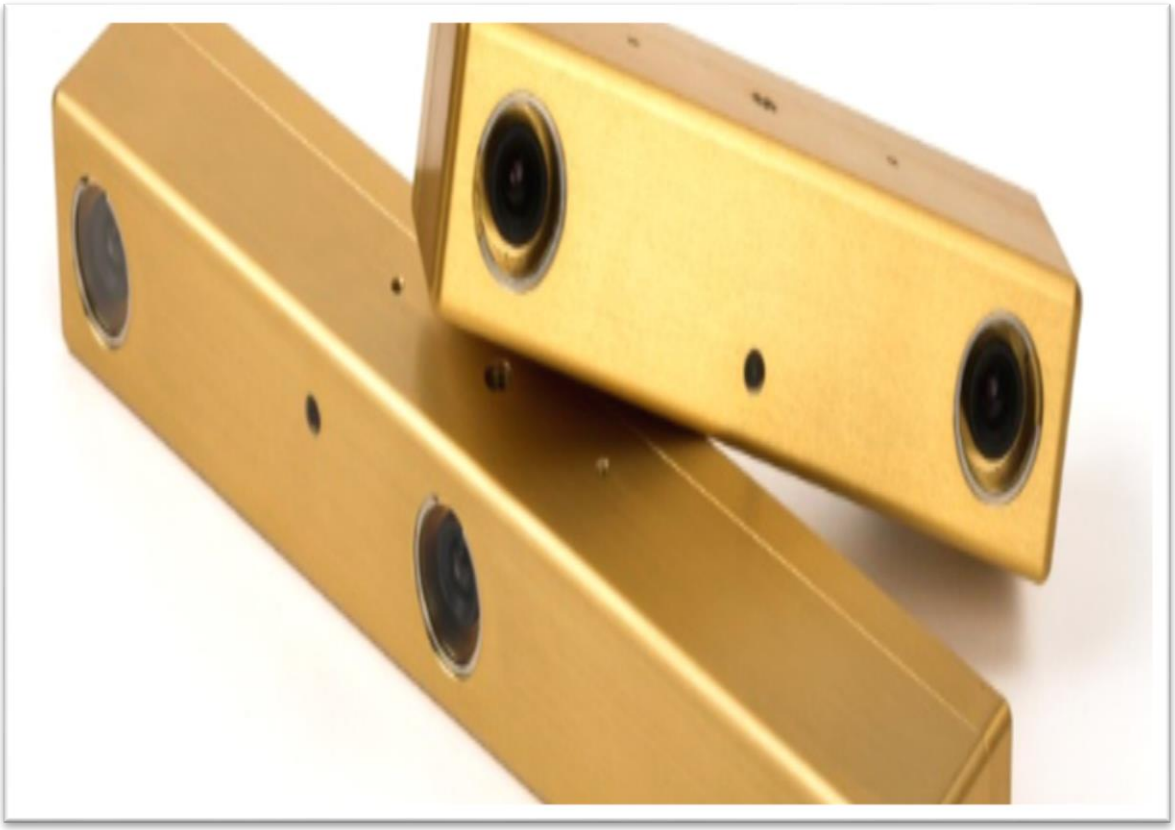


Figure 15. Integration of BUMBLEBEE CAMERA

Localization :

After this, the next step is very important in which the precise location of the rectangular region is calculated. One of the very popular technique for this is Linear Triangulation

method [76]. In this method we find the camera matrix of left and right cameras and locate the target region in both of them. After some further calculations we are able to find the location of the target point as a 3D location.

This 3D point can be utilized to calculate distance of the point from the origin of the robot's axis and the angle required to steer the robot towards the location. The P3at robot is equipped with sonar sensors, which are also installed in the wheelchair. Data from sonar sensors can be used to avoid static and dynamic obstacles and a better efficient motion planning algorithm can be designed.

Pioneer P3AT/ Wheelchair :

The pioneer P3AT is a state of the robot with many features such as four wheel skid-steer drive, DC motors, encoders and drive electronics. It is controlled via a server which is installed on top of it so that it commands can be executed without a need for connecting any other device on it. There are many extra features which can be included in it like arm, grippers, laser sensors etc. Because of all these it is a very useful laboratory equipment.

Below is the image of pioneer3At robot which is available in the laboratory. Same system of pioneer3AT is installed in the wheelchair project and by this way the wheelchair is made autonomous.

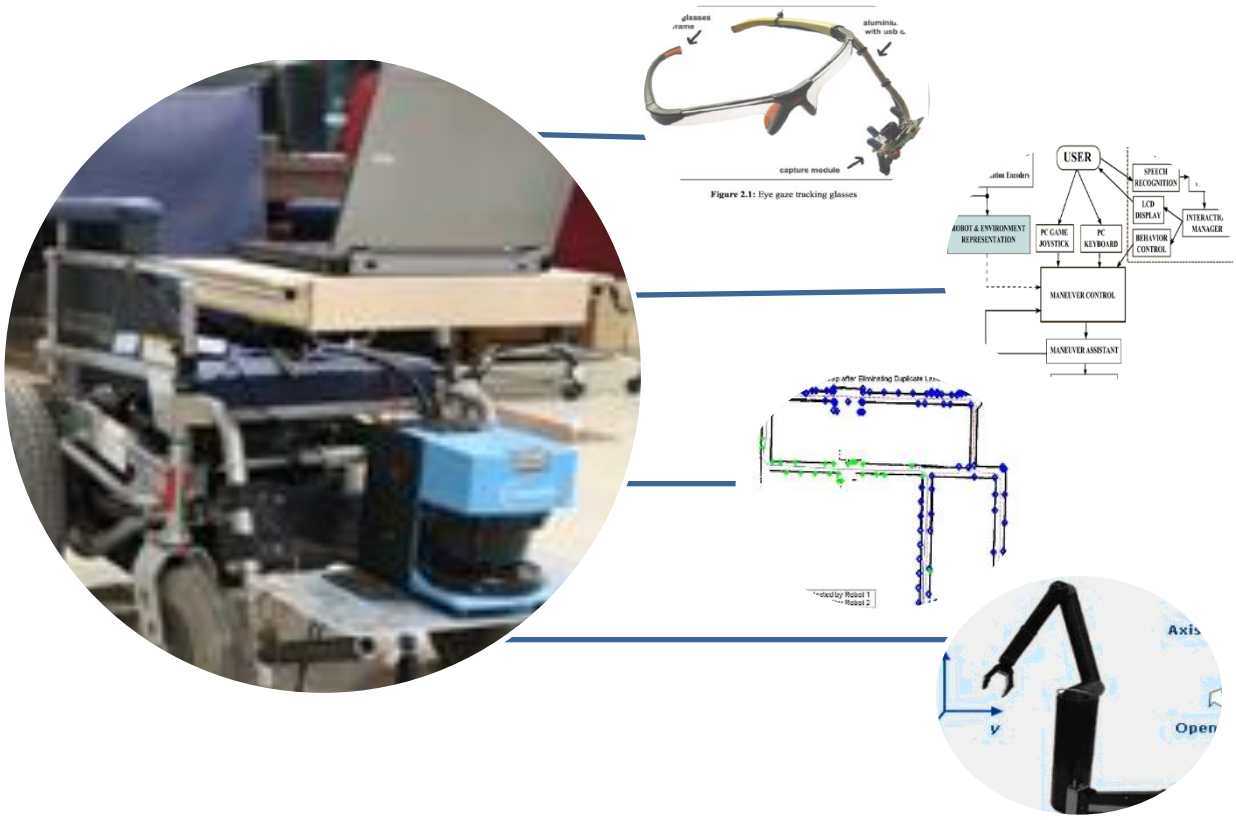


Figure 13. The Wheelchair Project

Figure 14. Testing of Pioneer3AT

Conclusion :

Vital are the landmarks element for mobile robotics, including a navigation system based on vision. This helps to determine the Android site, and develop a plan to go to the finish line. We suggested, a simple robust technology for the detection and text recognition on banners act as a reference point in the navigation system. As he developed algorithms to detect all the texts and internal environment, to determine very applicable in real life. The proposed method is an extension of the work and to assist in the ongoing investigation. Moving with the fashion of a larger project in a separate wheelchair on stage, and to plan a route in the room.

References

[1] P. Corke, O. Lam, F. Dayoub, and R. Schulz, “Text recognition approaches for indoor robotics a comparison” ,Australasian Conference on Robotics and Automation, 02 Dec. 2014.

[2] L. Neumann and J. Matas, “Real-time scene text localization and recognition”, in Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, 2012, pp. 3538.

[3] Qixiang, G. Wen, W.Weiqiang and Z.Wei, “ A Robust Text Detection Algorithm”, in Images and Video Frames, Information, Communications and Signal Processing , bari, 2003.

[4] Li and Y. Hu, “Salient traffic sign detection based on multiscale hypercomplex fourier transform”, in 4th International Congress on Image and Signal Processing (CISP) 2011, vol. 4, pp. 1963 .

[5] Pronobis, P. “Jensfelt, Large-scale semantic mapping and reasoning with heterogeneous modalities”, in IEEE International Conference on Robotics and Automation 2012, pp. 3515.

[6] Kostavelis, A. Gasteratos, “Learning spatially semantic representations for cognitive robot navigation”, Robotics Autonomous System 2013, vol. 61 (12), pp. 1460 .

[7] Zender, O. Martnez Mozos, P. Jensfelt, G.-J. Kruijff, W. Burgard, “Conceptual spatial representations for indoor mobile robots”, Robotics Autonomous System 2008 .

[8] Pronobis, B. Caputo, P. Jensfelt, H.I. Christensen, “A discriminative approach to robust visual place recognition”, in IEEE International Conference on Intelligent Robots and Systems 2006, pp. 3829.

[9] Ranganathan, F. Dellaert, “Semantic modeling of places using objects”, in Proceedings of the Robotics: Science and Systems Conference 2007 ,vol. 3, pp. 27.

[10] Soumya, V. Geethapriya, and S. Shankar Bharathi, “A novel edge detection and pattern recognition algorithm based on beamlet theory for a vision-based wheeled mobile robot” , in International Conference on Communications and Signal Processing (ICCSP) 2014, pp. 1197.

[11]Shopa, N. Sumitha, and P. S. K. Patra, “Traffic sign detection and recognition using OpenCV”, in International Conference on Information Communication and Embedded Systems (ICICES) 2014, pp. 1.

[12] Y. Nguwi and A. Z. Kouzani, “Automatic road sign recognition using neural networks”, in International Joint Conference on Neural Networks (IJCNN) 2006, pp. 3955.

[13] Saabni and M. Zwilling, “Text Detection and Recognition in Real World Images”, in International Conference on Frontiers in Handwriting Recognition (ICFHR) 2012, pp. 443,

[14] Kataoka, K. Sudo, and M. Morimoto, Region of Interest detection using indoor structure and saliency map, in Pattern Recognition (ICPR), 2012 21st International Conference on, 2012, pp. 3329 .

[15] Liu and J. Samarabandu, "An edge-based text region extraction algorithm for indoor mobile robot navigation", in IEEE International Conference on Mechatronics and Automation 2005, vol. 2, pp. 701.

[16] Stephen Se, "Vision-based global localization and mapping for mobile robots", in IEEE Transactions on Robotics 2003, vol 21, no. 3.

[17] L. Majdik, I. Szoke, L. Tamas, M. Popa, and G. Lazea, "Laser and vision based map building techniques for mobile robot navigation", in IEEE International Conference on Automation Quality and Testing Robotics (AQTR) 2010, vol. 1, pp. 1.

[18] Kim and R. Nevatia, "A method for recognition and localization of generic objects for indoor navigation", in Proceedings of the Second IEEE Workshop on Applications of Computer Vision 1994, pp. 280.

[19] Tang and S. Yuta, "Vision based navigation for mobile robots in indoor environment by teaching and playing-back scheme", in IEEE International Conference on Robotics and Automation (ICRA) 2001, vol. 3, pp. 3072 - 3077.

[20] Biswas and M. Veloso, "Depth camera based indoor mobile robot localization and navigation", in IEEE International Conference on Robotics and Automation (ICRA) 2012, pp. 1697.

[21] C. Astua, J. Crespo, and R. Barber, "Detecting Objects for Indoor Monitoring and Surveillance for Mobile Robots", Fifth International Conference on Emerging Security Technologies (EST) 2014, pp. 124.

[22] F. Wolf and C. N. S. Netto, "Vision-Based Outdoor Navigation Using Mobile Robots", IEEE Latin American Robotic Symposium (LARS) 2008, pp. 35.

[23] Tyukin, I. Lebedev, and A. Priorov, "The Development and Research of the Indoor Navigation System for a Mobile Robot with the Possibility of Obstacle Detection", in 16th Conference of Open Innovations Association (FRUCT) 2014, pp. 115.

[24]A. Arejolaleiba, F. Lerasle, M. Briot, C. Lemaire, and J. Hayet, "A smart sensor based visual landmarks detection for indoor robot navigation", in Proceedings of 16th International Conference on Pattern Recognition 2002 , vol. 2, pp. 848.

[25]Kim, J. Park, and Y. Do, "Monocular detection of pedestrians for the safe indoor navigation of a mobile robot", in IEEE International Workshop on Robot and Human Interactive Communication (ROMAN) 2013, pp. 270.

[26]Rous, H. Lupschen, and K.-F. Kraiss, "Vision-based indoor scene analysis for natural landmark detection", Proceedings of the IEEE International Conference on in Robotics and Automation (ICRA) 2005, pp. 4642.

[27]Zhang, L. Dou, H. Fang, and J. Chen, " Autonomous indoor exploration of mobile robots based on door guidance and improved dynamic window approach", IEEE International Conference on in Robotics and Biomimetics (ROBIO) 2009, pp. 408.

[28]Chugo, K. Hirose, K. Nakashima, S. Yokota, H.Kobayashi, and H. Hashimoto, "Camera-based navigation for service robots using pictographs on the crossing point", in 38th Annual Conference on IEEE Industrial Electronics Society (IECON) 2012, pp 14.

[29]Gopalakrishnan, S. Greene, and A. Sekmen, "Vision-based mobile robot learning and navigation" , in IEEE International Workshop on Robot and Human Interactive Communication (ROMAN) 2005,pp. 48.

[30]Purcaru, R.-E. Precup, D. Iercan, L.-O. Fedorovici, B.Dohangie, and F. Dragan, "Nrobotic mobile robot navigation using traffic signs in unknown indoor environments", in IEEE 8th International Symposium on Applied Computational Intelligence and Informatics (SACI) 2013, pp. 29.

[31]Boyu, G. Junyao, K. Li, Y. Fan, G. Xueshan, and B.Gao, "Indoor mobile robot obstacle detection based on linear structured light vision system", in IEEE International Conference on Robotics and Biomimetics (ROBIO) 2008, pp. 834.

[32]H. Nguyen, H. Choi, and G. Lee, " Text detection in scene images based on feature detection and tensor voting", in Journal of Information Processing Systems 2015, Vol.11, No.4, pp. 528-537.

[33]Mata, J. M. Armingol, A. D L. Escalera, and M. A.Salichs, "A visual landmark recognition system for topological navigation of mobile robots", in IEEE International Conference on Robotics and Automation (ICRA), 2001.

[34]Tomono and S. Yuta, "Mobile robot navigation in indoor environments using object and character recognition", in IEEE International Conference on Robotics and Automation (ICRA) 2000, vol. 1, pp.313.

[35]D. Letourneau, F. Michaud, and J. Valin, "Autonomous mobile robot that can read", EURASIP Journal on Applied Signal Processing 2004, pp. 2650-2662.

[36]Case, B. Suresh, A. Coates, and A. Y. Ng, "Autonomous sign reading for semantic mapping, in Robotics and Automation (ICRA), 2011 IEEE International Conference on, 2011.

[37]Lienhart, "Video ocr: A survey and practitioner's guide", In Video Mining, Kluwer Academic Publisher 2003 , pp. 155184.

[38] X. Chen and A. Yuille, "Detecting and reading text in natural scenes",in Computer Vision and Pattern Recognition 2004, vol. 2, pp.366.

[38]Smith, "An overview of the tesseract OCR engine", in ICDAR 07: Proceedings of the Ninth International Conference on Document Analysis and Recognition (ICDAR) 2007, Vol 2 , pp. 629.

[39]S. Fahn and P. Chang, "Text Plates Detection and Recognition Techniques Used for an Autonomous Robot Navigation in Indoor Environments", in 6th IEEE Conference on Robotics, Automation and Mechatronics (RAM) 2013.

[40]Azzopardi and N. Petkov, “Trainable COSFIRE filters for keypoint detection and pattern recognition”, in IEEE Transactions on Pattern Analysis and Machine Intelligence 2013 , vol. 35, pp. 490-503.

[41]Kantawong, “Road traffic signs guidance analysis for small navigation vehicle control system”, in Proceedings of IEEE International Conference on Vehicular Electronics and Safety (ICVES) 2007.

[42]Saoi, H. Goto and H.Kobayashi, “Text detection in color scene images based on unsupervised clustering of multi-channel wavelet features”, in Proceedings of 8th International Conference on Document Analysis and Recognition 2005, vol.1.

[43]Ganapathy and C.C.H.Lean, “Optical character Recognition Program for Images of Printed text using a Neural Network”, in Proceedings of the IEEE International Conference on Industrial technology 2006 , pp.1171-1176.

[44]Chen, J.Yang, J.Zhang, and A.Waibel, “Automatic detection and Recognition of Signs from Natural Scenes”, in IEEE Transactions on Image Processing, vol 13,no 1, pp.87-99,2004.

[45]Kanatani, and Y. Kanazawa, “Automatic Thresholding for Correspondence Detection Using Template Matching”, Proceedings of Statistical Methods in Video Processing Workshop 2002, Vol 12, pp.1502-1509.

[46]Kitazawa, Z. Li, and K. Yabuta, “Moving Object Extraction and Tracking Based on the Exclusive Block Matching”, Technical report of IEICE Image engineering 2008, Vol.108, pp.49-54.

[47]Nagahashi, H. Fujiyoshi, and T. Kaneda, “Object Type Classification Using Structure- Based Feature Representation”, in Proceedings of Technical Meeting on Systems and Control 2007, pp.39-44.

[48]Hirose, D. Chugo, S. Yokota, and K. Takase, “Service robots navigation using pictographs detection for indoor environment”, in 37th Annual Conference on IEEE Industrial Electronics Society (IECON) 2011, pp. 2170,

[49] V. Wu, R. Manmatha, and E. M. Riseman, Sr., “Textfinder: An automatic system to detect and recognize text in images” , in IEEE Transactions on Pattern Analysis and Machine Intelligence 1999 , vol. 21, pp.1224-1229

[50] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, “ROS: an open-source robot operating system”, in ICRA Workshop on Open Source Software, 2009.

[51] Boris Epshtein, Eyal Ofek, and Yonatan Wexler, “Detecting text in natural scenes with stroke width transform”, in IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2010, pages 2963-2970.

[52] Jin-Liang Yao, Yan-Qing Wang, Lu- Bin Weng, and Yi-Ping Yang, “Locating text based on connected component and svm”, in International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR) 2007, volume 3, pages 1418-1423.

[53] Shopa, N. Sumitha, and P. S. K. Patra, “Traffic sign detection and recognition using OpenCV ”, in International Conference on Information Communication and Embedded Systems (ICICES) 2014.

[54] C. Yao, X. Zhang, X. Bai, W. Liu, Y. Ma, and Z. Tu. “Rotation-invariant features for multi-oriented text detection in natural images”, in PLoS One 2013, 8(8) .

[55] Y.K. Lim, S.H. Choi, S.W. Lee, “Text extraction in MPEG compressed video for content-based indexing” , in Proceedings of International Conference on Pattern Recognition 2000, pp. 409-412.

[56] 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.

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

[58] 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.

[59] 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.

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

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

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

Appendix A:

Code for ROI select :

```
import numpy as np
import argparse
import cv2

#roiPts = [(42, 108), (47, 108), (48, 112), (45, 113)]
# initialize the current frame of the video, along with the list of
# ROI points along with whether or not this is input mode
frame = None
roiPts = []
inputMode = False

def selectROI(event, x, y, flags, param):
  # grab the reference to the current frame, list of ROI
  # points and whether or not it is ROI selection mode
  global frame, roiPts, inputMode

  # if we are in ROI selection mode, the mouse was clicked,
  # and we do not already have four points, then update the
  # list of ROI points with the (x, y) location of the click
  # and draw the circle
  if inputMode and event == cv2.EVENT_LBUTTONDOWN and len(roiPts) < 4:
     roiPts.append((x, y))
     cv2.circle(frame, (x, y), 4, (0, 255, 0), 2)
     cv2.imshow("frame", frame)

def main():
```

```
  # construct the argument parse and parse the arguments
  ap = argparse.ArgumentParser()
  ap.add_argument("-v", "--video",
     help = "path to the (optional) video file")
  args = vars(ap.parse_args())

  # grab the reference to the current frame, list of ROI
  # points and whether or not it is ROI selection mode
  global frame, roiPts, inputMode

  # if the video path was not supplied, grab the reference to the
  # camera
  if not args.get("video", False):
     camera = cv2.VideoCapture('/home/usama/Documents/Sami/eclipse/sample/codes/videos
nameplates/exit.mp4')

  # otherwise, load the video
  else:
     camera = cv2.VideoCapture(args["video"])

  # setup the mouse callback
  cv2.namedWindow("frame")
  cv2.setMouseCallback("frame", selectROI)

  # initialize the termination criteria for cam shift, indicating
  # a maximum of ten iterations or movement by a least one pixel
  # along with the bounding box of the ROI
  termination = (cv2.TERM_CRITERIA_EPS | cv2.TERM_CRITERIA_COUNT, 10, 1)
  roiBox = None

  # keep looping over the frames
  while True:
     # grab the current frame
```

```
(grabbed, frame) = camera.read()

# check to see if we have reached the end of the
# video
if not grabbed:
    break

# if the see if the ROI has been computed
if roiBox is not None:
    # convert the current frame to the HSV color space
    # and perform mean shift
    hsv = cv2.cvtColor(frame, cv2.COLOR_BGR2HSV)
    backProj = cv2.calcBackProject([hsv], [0], roiHist, [0, 180], 1)

    # apply cam shift to the back projection, convert the
    # points to a bounding box, and then draw them
    (r, roiBox) = cv2.CamShift(backProj, roiBox, termination)
    pts = np.int0(cv2.cv.BoxPoints(r))
    cv2.polylines(frame, [pts], True, (0, 255, 0), 2)

# show the frame and record if the user presses a key
cv2.imshow("frame", frame)
key = cv2.waitKey(100) & 0xFF

# handle if the 'i' key is pressed, then go into ROI
# selection mode
if key == ord("i") and len(roiPts) < 4:
    # indicate that we are in input mode and clone the
    # frame
    inputMode = True
    orig = frame.copy()

    # keep looping until 4 reference ROI points have
```

```
            # been selected; press any key to exit ROI selction
            # mode once 4 points have been selected
            while len(roiPts) < 4:
                cv2.imshow("frame", frame)
                cv2.waitKey(0)

            # determine the top-left and bottom-right points
            roiPts = np.array(roiPts)
            s = roiPts.sum(axis = 1)
            tl = roiPts[np.argmin(s)]
            br = roiPts[np.argmax(s)]

            # grab the ROI for the bounding box and convert it
            # to the HSV color space
            roi = orig[tl[1]:br[1], tl[0]:br[0]]
            roi = cv2.cvtColor(roi, cv2.COLOR_BGR2HSV)
            #roi = cv2.cvtColor(roi, cv2.COLOR_BGR2LAB)

            # compute a HSV histogram for the ROI and store the
            # bounding box
            roiHist = cv2.calcHist([roi], [0], None, [16], [0, 180])
            roiHist = cv2.normalize(roiHist, roiHist, 0, 255, cv2.NORM_MINMAX)
            roiBox = (tl[0], tl[1], br[0], br[1])

        # if the 'q' key is pressed, stop the loop
        elif key == ord("q"):
            break
    # cleanup the camera and close any open windows
    camera.release()
    cv2.destroyAllWindows()

if __name__ == "__main__":
    main()
```

Appendix B:

Code for EXIT crop :

```
import cv2
import numpy as np
import os

#mainimage = 'shoe.png'
mainimage = 'exit1.jpg'
erosion_size = 19

def angle_cos(p0, p1, p2):
  d1, d2 = (p0-p1).astype('float'), (p2-p1).astype('float')
  return abs( np.dot(d1, d2) / np.sqrt( np.dot(d1, d1)*np.dot(d2, d2) ) )

def find_squares(img , squares):


  bin = cv2.Canny(img, 0, 50, apertureSize=5)
  bin = cv2.dilate(bin, None)

  contours, hier = cv2.findContours(bin, cv2.RETR_LIST, cv2.CHAIN_APPROX_SIMPLE)

  for cnt in contours:
        cnt_len = cv2.arcLength(cnt, True)
            approx = cv2.approxPolyDP(cnt, 0.04*cnt_len, True)  # changed its value from 0.02
to 0.04
        cont_area = cv2.contourArea(approx)
            print str(cont_area) + ' : ' + str(len(approx))
```

```
                if len(approx) == 4 and cont_area >29000 and cv2.isContourConvex(approx):
            approx = approx.reshape(-1, 2)

                max_cos = np.max([angle_cos( approx[i], approx[(i+1) % 4], approx[(i+2) % 4] )
for i in xrange(4)])


            if max_cos < 0.3:
                squares.append(approx)

  return squares



def draw_squares(img , squares):


        for cont in squares:
                p = squares
                cv2.polylines( img, p, True, (0,255,0), 2,2)
        #cv2.imshow('crop', img)



if __name__ == '__main__':


        src = cv2.imread(mainimage)
        src = cv2.resize(src, (392,280))
        hsv = cv2.cvtColor(src,cv2.COLOR_BGR2GRAY)   # convert to grayscale (1 channel)
        #cv2.imshow('HSV', hsv)
        reti,binary = cv2.threshold(hsv,69,255,1)  #thresolding to binarize image)
        cv2.imshow('thresh', binary)
        #find contours in thresold image
```

```
	contours, hierarchy = cv2.findContours(binary,cv2.RETR_LIST,cv2.CHAIN_APPROX_SIMPLE)

	for idx,cnt in enumerate(contours):
		cv2.drawContours (binary, contours, idx, 255, -1,8,hierarchy,0)



	element = cv2.getStructuringElement(0 , (2 * erosion_size + 1, 2 * erosion_size + 1), (erosion_size, erosion_size) ) #MORPH_RECT -
	binary = cv2.erode(binary, element)
	binary = cv2.dilate(binary, element)
	#cv2.imshow('squares', binary)
	squares = []
	find_squares(binary,squares)
	#print squares

    #cv2.imshow('squares', binary)
	draw_squares(src,squares)
	#print squares
			#this is for cropping box
	if len(squares)==1:
		squares = np.array(squares)
		x,y,w,h = cv2.boundingRect(squares)
		roi =src[y:y+h, x:x+w]
		cv2.imshow('roi', roi)


	else:
		print 'More than one rectangle found'

	cv2.waitKey(0)
	os.chdir("/home/usama/Documents/Sami/python/python-tesseract/test-slim/")
	cv2.imwrite('crop.jpg',roi)
```

Appendix C:

Code for OCR :

```
#!/usr/bin/env python
# -*- coding: utf-8 -*-
#from __future__ import print_function
import tesseract
import gc
import pprint

api = tesseract.TessBaseAPI()
api.SetOutputName("outputName");
#api.Init(".","eng")
api.Init(".","eng",tesseract.OEM_DEFAULT)
api.SetPageSegMode(tesseract.PSM_AUTO)
#mImgFile = "eurotext.jpg"
mImgFile = "crop.jpg"

print("Method 1: Leptonica->pixRead")
pixImage=tesseract.pixRead(mImgFile)
print("Type of pixiamge=",type(pixImage))
print("repr(pixiamge)=",repr(pixImage))
api.SetImage(pixImage)
outText=api.GetUTF8Text()
print(("OCR output:\n%s"%outText));
api.End()
outText=None
tesseract.pixDestroy(pixImage)
```