GESTURE CONTROLLED INTERACTION WITH VIRTUAL OBJECTS

DISSERTATION-II

Submitted in partial fulfillment of the
requirement for the award of the Degree of

MASTER OF TECHNOLOGY IN

Electronics and Communication Engineering

by

Prashant Katiyar

(11309745)

Under the Guidance of

Dr. Kamalpreet Kaur

Assistant Professor,

Lovely Professional University

School of Electronics and Electrical Engineering

Lovely Professional University
Phagwara, Punjab

December, 2017

CERTIFICATE

This is to certify that Prashant Katiyar bearing Registration no. 11309745 have completed objective formulation/Base Paper implementation of the thesis titled, **"Gesture controlled interaction with virtual objects "** under my guidance and supervision. To the best of my knowledge, the present work is the result of his original investigation and study. No part of thesis has ever been submitted for any other degree at any University.

Dr. Kamalpreet Kaur

Assistant Professor

Division of Research and Development

Lovely Professional University

Phagwara, Punjab

Date:

Objective of the Thesis is Satisfactory/ Unsatisfactory

Examiner I **Examiner II**

ACKNOWLEDGEMENT

First and foremost, I would like to express my sincere gratitude and appreciation to my guide Dr. Kamalpreet Kaur, for his whole-hearted and invaluable guidance, inspiring discussions, encouragement, and support throughout my work. I found her always sincere in helping me even during her busiest hours of the day. Her ardor and earnestness for studies are respected and will never be forgotten. Without her sustained and sincere effort, this report would not have taken this shape.

I am indebted to all authors of the research papers and books referred to, which have helped me in carrying out the research work.

Prashant Katiyar

Reg. No: 11309745

DECLARATION

I, Prashant Katiyar, student of M. Tech in Department of Electronics and Communication of Lovely Professional University, Punjab, hereby declare that all the information furnished in this Dissertation-II report is based on my own intensive research and is genuine.

This report does not, to the best of our knowledge, contain part of my work which has been submitted for the award of my degree either of this University or any other University without proper citation.

Prashant Katiyar

11309745

ABSTRACT

Human computer interaction (HCI) devices interprets human's speech, gestures and other parameters and interface it to computer. Controlling and manipulating computer values and variables is not an easy task for every non-technical user's so as to provide an easy and user friendly environment, a dynamic algorithm is proposed which provides HCI for urban planning by manipulating virtual buildings and spaces by hand gestures. Proposed system tracks the index finger and thumb tips for interaction with virtual object like selection of single object between multiple objects, moving of objects around the platform. The objects are handled just like real world entity without any advance knowledge of 3-D object modelling, analysis in VRML (virtual reality modelling language).

LIST OF ABBREVIATIONS

HCI	…………	Human Computer Interaction
VRML	……….	Virtual Reality Modelling Language
VR	……….	Virtual Reality
AR	……….	Augmented Reality
IHRS	……….	Intelligent Hybrid Reality System
BLC	……….	Bright Light Condition
LLC	……….	Low Light Condition
RSS	……….	Reflective and Scratched Surface
SOFT	……….	Scene, Object, Face and Text
ROM	……….	Region of Manipulation

LIST OF FIGURES

LIST OF TABLES

Table of Contents

INTRODUCTION

Overview

virtual reality (VR) normally provides mainframe technologies that virtual reality headsets use to produce more unpretentious images, sounds and other ambiances that imitate a real environment or create an unreal situation. VR also simulates the physical presence of a user in this environment. VR is defined as "a realistic and immersive simulation of a three-dimensional 360-degree environment, created using interactive software and hardware, and experienced or controlled by body movement" or as a "compelling, interactive experience generated by a computer". A person who uses virtual reality equipment can "look around" in the artificial world, move with high quality VR, and interact with functions or items displayed in the headset. Virtual reality is displayed with a virtual reality headset. VR headsets are head-mounted safety glasses with a screen for the eyes. Programs can contain audio and sounds through speakers or headphones. The definition of virtual reality naturally comes from the definitions for both 'virtual' and 'reality'. The definition of 'virtual' is near and the reality is what we experience as human beings. So the term 'virtual reality' essentially means 'near-reality'. This can of course mean anything, but it usually refers to a specific type of reality emulation. We know the world through our senses and perception systems. At school we have all learned that we have five senses: taste, touch, and smell, see and hear. However, these are only our most obvious senses. The truth is that people have many more senses than this, such as a sense of balance. These other sensory inputs, plus a special processing of sensory information by our brains, ensure that we have a rich flow of information from the environment to our minds. Everything we know about our reality comes through our senses. In other words, our whole experience of reality is simply a combination of sensory information and our brains, observing mechanisms for that information. It goes without saying that if you can present your senses with fabricated information, your perception of reality would also change in response to it. You would get a version of reality that is not real, but from your perspective it would be experienced as real. Something we would call a virtual reality.

Gesture Controlled Virtual Urban Planning

With the immense influx of wearable computers [25], HCI has become the basic and important part to operate wearable computers. As computers are important part of human life. In this era of advance technology size of computer is getting small and ways to operate them are also varying. These wearable computers can be operated by various ways which can be roughly categorized into three categories; Touch Interface, Speech Interface and Gestures Interfaces.

Touch interface includes touch screen, ring-type pointing devices [24]. Which requires direct intervention of human with the computer in order to operate them. Touch interfaces provides fast and accurate response compared to other interfaces. Speech Interface involves speech and acoustic processing. Human voice is used to operate the devices in this scenario. As the advancement in speech processing and voice recognition makes HCI possible [26, 27]. Though Phonetic interaction is very natural way to make communications, expressing something for inputting every single command to the machine is not an appropriate way for wearable personal computer use in the public. At last gesture interface in which movements of specific body part is logged and interpreted. In this case hand gestures is considered as the data can be logged from the hand movement by two ways; Glove sensors [29, 30, 31] in which a glove is designed with sensors that sense the hand movement. This gives precise results but the human has to wear bulky devices which hinders the ease while interacting with the computer. This procreated the dynamic research towards vision based hand gesture techniques. Where hand movements are recognized by camera without any hindrance of bulky electromechanical sensors and wires.

Interacting with real world entity is sometimes difficult such as urban planning, military training, and medical training so virtual reality [28, 29] helps to overcome such issues. 3-D modelling of building, armaments, medical anomalies as well as analysis can be done easily. Manipulating these 3-D objects relays large amount of information and helps human for better understanding. VRML is the programming language through which these models are created, after successful creation they can be deployed to VR enabled devices.

Texture-Based Real-Time Character Extraction and Recognition in Natural Images

OCR is a unique tool used to recognize text from any image. It is basically developed by using the neural network and machine learning concepts, in which neural network is trained by making a database of different kind of images of same characters. Further, machine learning helps in making the database or the retrieving information from it for example k-nearest approach etc. There are many complex and tough languages in the world that are hard to understand or translate. To overcome these problems many techniques have been proposed which can help in translating the languages just by pointing the camera on it, for example, google glasses etc. In continuation of the research in this field, in the present work, a real-time character extraction and recognition algorithm using the texture based approach has been proposed to extract and recognize text or character from natural images. Most conventional algorithms uses canny edge detector to find out the edges for detection of characters but in this proposed algorithm convolution of Prewitt operator and image is used to calculate the edges of an image.

Social Media Access via Face Recognition

Face recognition is one the main task of this proposed system it is used to detect the no. of faces in the acquired image and then selected face will be recognized. Every face is linked with a unique ID which helps in making a discrete social media platform. Through this the user can access social information about any other person just by looking at person face. It can also help user to find any person in a crowded place. In current era of social media people like to share their views and idea on apps and website via mobile phone. Which involves a hazy process to check for every notification whenever their phone beeps, rather than just seeing the person. This process has another benefit that nowadays the youngster are always busy behind the phone screens. It will help those people to interact other people in real while accessing social without any hassles.

Object Detection and Recognition

Object detection is the major of all process. In this the proposed system uses very deep convolutional neural network for classification of objects such as GoogLeNet, ResNet, DenseNet for detection and recognition of objects in the acquired images. Object detection provides intelligence to this proposed system. For example let suppose the user wants the system to show "Red Sedan class car" then the system will find the cars in the image. After that categorizes them as sedan class, luxury class etc. then categorize them according to the colour of those detected cars. And finally highlight the sedan class cars having red colour in the acquired image. Now the next major contribution of this process to this proposed system is to locate the position of objects in the image. For the implementation of Hybrid reality, virtual objects must be positioned over real world objects or environment to provide an interactive and enhanced view to deal with the day to day tasks or any specific problem.

Scope of the Study

In this Proposed System, it is planned to design an intelligent hybrid reality system (IHRS). Which led to extract maximum information from a natural image or surroundings so as to help user to access information around him. For example translating any foreign language, using virtual maps or internet while walking and talking to a friend without any hesitation, study or learn complex 3-dimensional figures or object, and creating digital wireframe of any 3-dimensinal object just by taking images of it. To achieve this four processes are running in parallel, extracting out different information. Artificial intelligence concepts is used to make this system more smart and user friendly.

Objective of the Study

The main objective of this work is to extract maximum information from a natural image. If a natural image is dissected then four major things can be extracted and those are object (car, bus, human, furniture… etc.), Face, Text and Scene. Through their detection and recognition maximum information can be extracted out so as to implement IHRS (intelligent hybrid reality system). IHRS is combination of both virtual and augmented reality. Virtual reality helps in deploying virtual environment whereas augmented reality deploys virtual object over real environment. These above discussed parameter are termed as SOFT (Scene, Object, Face and Text).

- Translating any text into desired language inside image.
- Object Detection for implanting virtual objects in the scene.
- Face Detection for accessing social media interaction.

All these process are carried out to make the system intelligent by labelling the face, objects etc. so as to make the user task unbelievably fast and easy. All these computation were carried out in a parallel fashion to calculate every bit of information available at all times.

Proposed Work Plan with Timelines

The proposed work is 50% completed. There are four processes in which two processes are successfully completed. In the next coming semester it is planned to implement those two remaining processes. After implementing of all processes the focus of this thesis will turn it to deploy these implemented algorithm on hardware.

TABLE 8 PROPOSED SYSTEM PROCESSES TIME LINE

Process	Timeline	Aug	Sep	Oct	Nov	Dec	Jan	Feb	March	April	May
Literature Survey											
Phase 1: Problem Analysis											
Phase 2: Implementation											
Phase 3: Result Calibration											
Testing Phase											
Publication											

REVIEW OF LITERATURE

PAPER 1: “**APPLICATION OF MIXED REALITY**”

Sven Wursthorn, Alexandre Hering Coelho and Guido Staub [32]

Mixed or Augmented Reality (AR) systems support computer-generated virtual objects and objects from the real world to stimulate the vision of physical reality with the virtual objects in a way that allows a user to communicate with three-dimensional information in his natural atmosphere. Especially when it comes to natural calamities, the advantage of the enlarged representation of a user with additional supporting information is clear. Actions such as precautions for protection, actual response and refurbishment must be supported by a technology that promotes the productivity and performance of the user. This approach relates to a system that fulfills specific needs with regard to earth sciences and disaster management. We will present our work with the development of an AR system (ARS) and its components with examples of practical applications. The crucial point of this implementation is the use of laser scanning models with sufficient accuracy, reliability, topicality and completeness. VRML tiles produced using the laser scanning data are displayed in the operator's field of vision, along with additional models for water surfaces or simple functions such as points and lines. These geometric objects are visualized and controlled using the parent-child relationship of scene diagrams of Java3D. In this document, flood disasters are used to illustrate possible applications in a typical four-step disaster management process. In addition, some examples of the virtual water surface for the prediction of flood damage are discussed. The results of virtual water models are used by applying digital image processing techniques to laser scanning data to provide tactical information for disaster management.

PAPER 2: “**MIXED REALITY IN VIRTUAL WORLD TELECONFERENCING**”

Tuomas Kantonen, Charles Woodward, Neil Katz VTT Technical Research Centre of Finland, IBM Corporation [33]

In this article, the author presents a teleconferencing application for Mixed Reality (MR) based on Second Life (SL) and the virtual world of OpenSim. Augmented Reality (AR) techniques are used to display virtual avatars of participants in external meetings in real physical spaces, while Augmented Virtuality (AV), in the form of video-based gesture detection, captures human expressions for driving avatars and for manipulating virtual objects in virtual worlds. The use of Second Life to create a shared enlarged space to represent different physical locations enables us to incorporate the application into the existing infrastructure. The application is implemented with open source Second Life viewer, ARToolKit and OpenCV libraries.

PAPER 3: "RAPID CREATION OF PHOTOREALISTIC VIRTUAL REALITY CONTENT WITH CONSUMER DEPTH CAMERAS"

Chih-hun Chen, Mark Bolas, Evan Suma Rosenberg [34]

Virtual objects are essential for building environments in virtual reality (VR) applications. Making photorealistic 3D models is not easy, however, and manufacturing the detailed 3D model of a real object can be time-consuming and labor-intensive. An alternative way is to build a structured camera series such as a light beam to reconstruct the model of a real object. However, these technologies are very expensive and not practical for most users. In this work, we demonstrate a complete end-to-end pipeline for capturing, processing and displaying image-dependent 3D models in virtual reality from a single RGB-D camera for consumers. The geometry model and the camera trajectories are automatically reconstructed from an RGB-D image sequence that has been recorded offline. Based on the HMD position, selected images are used for real-time model rendering. The result of this pipeline is a 3D network with image-dependent structures that are suitable for real-time display in virtual reality. Specular reflections and light burst effects are especially noticeable when users view the objects from different perspectives in an environment with a main route.

PAPER 4: "COORDINATING ATTENTION AND COOPERATION IN MULTI-USER VIRTUAL REALITY NARRATIVES"

Cullen Brown, Ghanshayam Bhutra, Mohamed Suhail, Qinghong Xu, Eric D. Ragan [35]

There has been limited research into attempts to address environments with multiple user stories in virtual reality. Therefore, a number of questions need to be answered about dealing with story progression and maintaining user presence in a virtual environment with multiple users. We created a multi-user virtual reality story experience in which we want to study a series of guided camera techniques and a number of can-see techniques to determine how different users can best be exposed to the same story. In addition, we describe our preparatory work and we plan to study the effectiveness of these techniques, their effect on the presence of users and in general how multiple users believe that their actions influence the outcome of a story.

PAPER 5: “**RAPID OBJECT DETECTION USING A BOOSTED CASCADE OF SIMPLE FEATURES**”

Paul viola and Michael Jones [36]

This article describes a machine learning approach for visual object detection that is capable of processing images extremely quickly and achieving high detection speeds. This work is distinguished by three important contributions. The first is the introduction of a new image representation, the "Integral linage", with which the functions used by our detector can be calculated very quickly. The second is a learning algorithm, based on AdaBoost, which selects a small number of critical visual features from a larger series and produces highly efficient classifiers. The third contribution is a method of combining increasingly complex classifiers in a "cascade", allowing background areas of the image to be quickly discarded, while spending more on promising object-like regions. The cascade can be seen as an object-specific focus-of-attention mechanism that, unlike previous approaches, provides statistical guarantees that end-of-life regions are unlikely to contain the object of interest, in the field of face detection, the system provides detection rates comparable to those of the best previous systems. Used in real-time applications, the detector operates at 15 frames per second without using image differences or skin color detection.

PAPER 6: “**ROBUST TEXT DETECTION IN NATURAL IMAGES WITH EDGE-ENHANCED MAXIMALLY STABLE EXTREMAL REGIONS**”

Huizhong Chen, Sam S. Tsai, George Schroth, David M. Chen, Radek Grzeszczuk and Bernd Girod [37]

Detecting text in natural images is an important condition. In this article we propose a new text detection algorithm that uses edge-enhanced Maximally Stable Extremal Regions as basic letter candidates. These candidates are then filtered using geometric and area width data to exclude non-text objects. Letters are paired to identify text lines, which are then separated into words. We evaluate our system using the ICDAR competition data set and our mobile document database. The experimental results show the excellent performance of the proposed method.

PAPER 7: “**EXPLORATION OF THE CHANGING STRUCTURE OF CITIES: CHALLENGES FOR TEMPORAL CITY MODELS**”

Clementine Perinaud, Georges Gay and Gilles Gesquiere [38]

The ALARIC project (project for research into incremental urban changes) is dedicated to the production of geohistoric information about two former industrial cities in the Lyon-Saint-Etienne region (19th and 20th century). The exploration of the increasing nature of urban change implies that it is determined when certain historical processes took place, such as the shift from one-off to systematic construction and changes in city planning strategies. In concrete terms, the case study examines the emergence of local urban projects to compare the transformation processes of the urban fabric and to specify the pace of urban change. Historical records will be accessible through a virtual mapping environment based on the temporal reconstruction of cities in two dimensions, and sometimes for special cases in three dimensions. This document presents methodological principles for reconstructing former cities to explore urban changes and requirements for the effective sharing of hypotheses about the different states of the past urban landscape and associated transformation processes.

RESEARCH METHODOLOGY

To achieve Intelligent Hybrid Reality System four major processes are running in parallel. Detailed working with architecture is discussed under this section. This section deals with the working and architecture of the two implemented processes. IHRS is working on Computer Vision, Machine Learning and Virtual Reality concepts and techniques.

GESTURE CONTROLLED VIRTUAL URBAN PLANNING

This proposed technique is divide mainly into three parts viz. hand gesture modelling, analysis and recognition of gesture. It is subdivided into various parts as shown in (fig. 1).

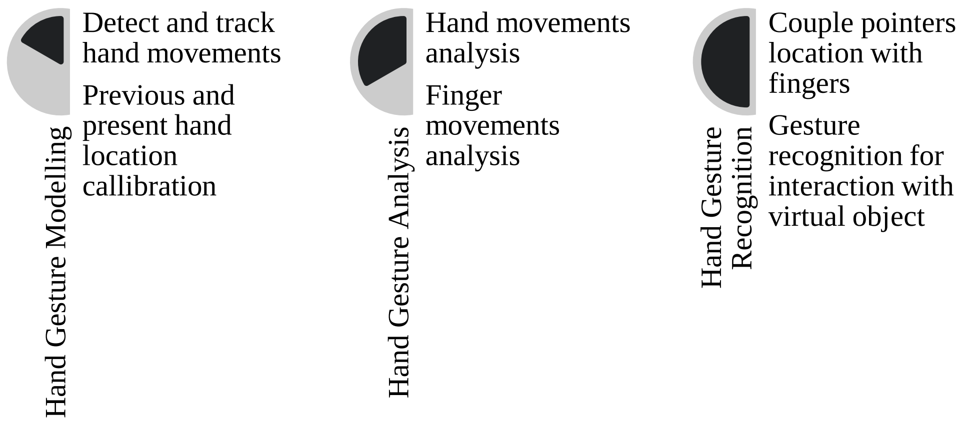


Figure 1 The Basic Functioning of Hand Gesture Recognition System

Hand Gesture Modelling

In hand modelling the basic objective is to detect and track the movement and calibrate the previous and present hand location with the pointer. Image will be captured by camera which is used for the extraction of information from the hand posture. In this proposed system one finger and one thumb is used as pointer in 3D workspace. Red and Blue markers are attached to finger and thumb respectively.

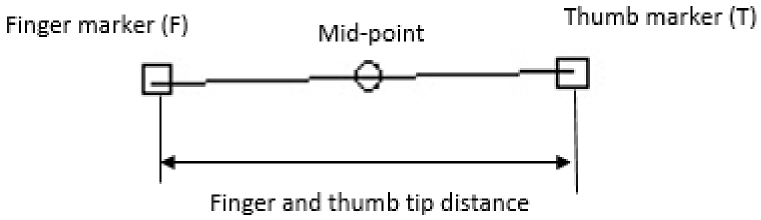


Figure 2 Finger coupled pointers

The figure (2) shows the coupling of index finger and thumb with the pointer. Finger maker (red colour) and thumb marker (blue colour) are denoted by F and T. there is two more parameters which are midpoint and distance between finger and thumb. The whole process sums up in three steps as shown in figure (3)

Figure 3 Modelling of hand gesture

Hand Gesture Analysis

After modelling of hand gesture valuable information can be extracted that can be used drive or operate virtual components or object of the designed system. In the analysis section the calculated distance between the finger and thumb and mid-point coordinates are normalized to match up the image resolution with the region of manipulation (ROM). ROM is the area allocated with in which interaction with the virtual world can be done. If the pointers coordinates move out of the ROM then it will stop at edge location of the ROM allowing real life experience. Now, the mid-point coordinate is used to selection of particular object multiple object. The distance between the finger and thumb tip is used to grabbing the object.

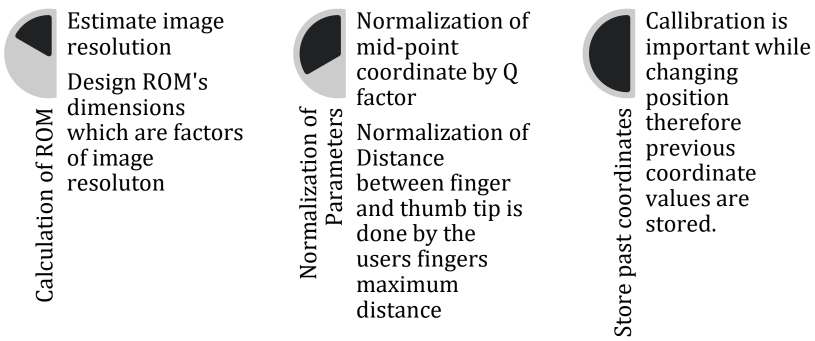


Figure 4 Analysis of hand gestures

Calculation of ROM, normalization of parameters is calculated in the second phase of this proposed system:

Area of Image Frame, Ai = W*H (1)

ROM = Nw * Nh (2)

ROM = Q * Ai (3)

Q is the factor depends on the area selected for urban planning. As shown in figure 5.1.5 there is difference between the dimensions of the image frame and ROM which is Q times. As the geographic areas are always in distorted shapes therefore a rectangle area is deployed to combine it with coordinates extracted and normalized from image frame. Normalization of parameters such as mid-point coordinate and distance between the finger and thumb tip.

Mid-Point Coordinate In Rom Work Space = Q * Mid-Point Coordinate In Image (4)

Maximum distance between users finger and thumb be D, the normalized distance be N and Euclidean distance between the two marker be E

N = E / D (5)

In mathematics, the Euclidean distance is the straight-line distance between two points in Euclidean space. With this distance, Euclidean space becomes a metric space. The Euclidean distance between points **A** and **B** is the length of the line segment connecting them $\overline{AB}$ **.**

$$\overline{AB} = \sqrt{(X_a - X_b)^2 + (Y_a - Y_b)^2 + (Z_a - Z_b)^2} \quad (6)$$

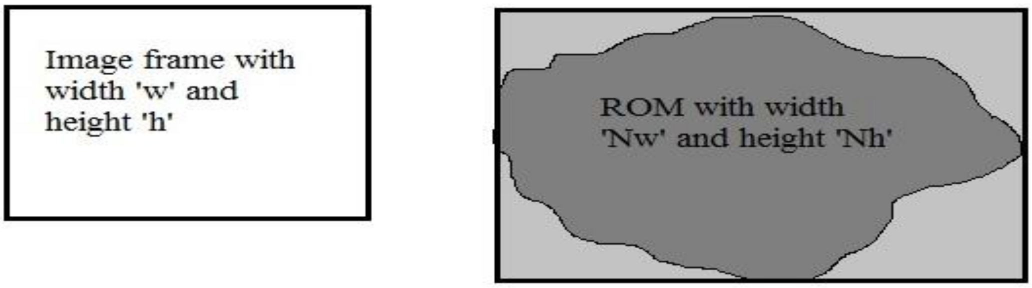
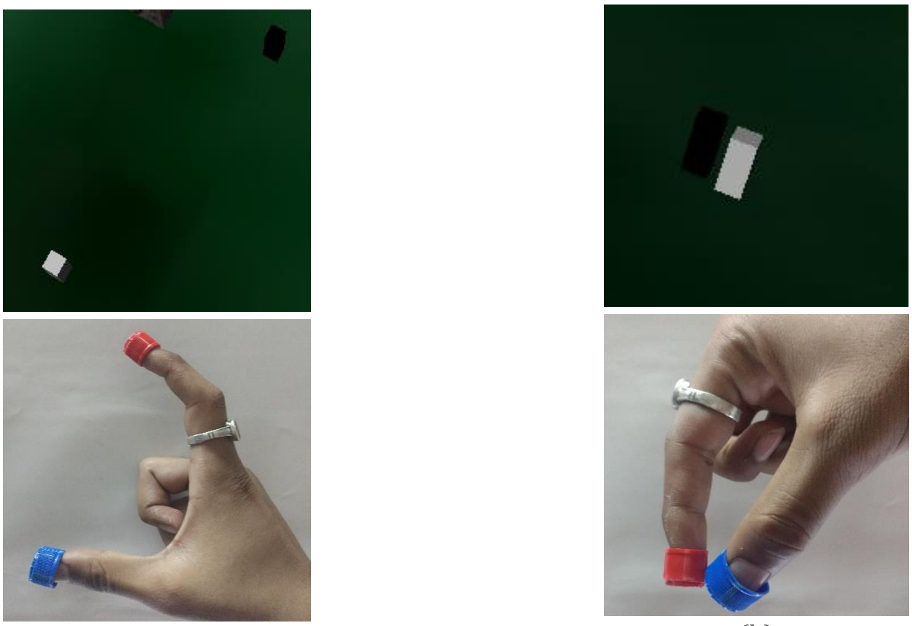


Figure 5 image frame and ROM where dark grey shows area for urban planning where building will be placed whereas light grey shows restricted area

Hand Gesture Analysis

Recognition consists of two modes selection and grabbing of object. Firstly selection of particular is done by moving the mid-point close to the centroid of object and while selecting the object the distance between the pointers should be greater than the normalized value shown in figure (6.a). Then, grabbing of object takes place by bringing finger and thumb tip closer by this the distance between the pointers get reduced when this distance is approximately a little less than to width of the object, it is ready for movements according to pointers movements as shown in figure(6.b).



(a) (b)

Figure 6 shows the pointer in selection mode when they are far away from each other figure b shows the pointer in grabbing mode they are close to each other

TEXTURE-BASED REAL-TIME CHARACTER EXTRACTION AND RECOGNITION IN NATURAL IMAGES

The proposed algorithm is mainly divided into four major sections. Firstly, pre-processing of the images is carried out to remove physical anomalies after which filter and operators are applied for detection of edges. Further, noise and artifacts are removed so that extraction and recognition of character can be done.

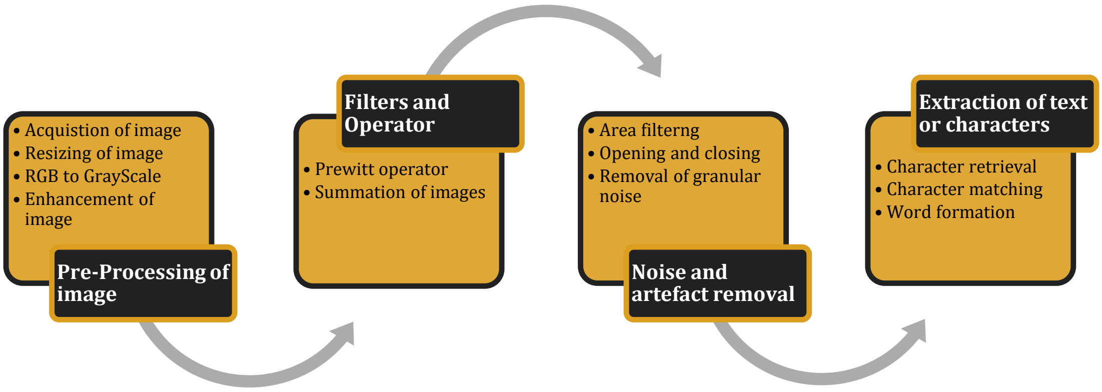


Figure 7 Block diagram of the proposed algorithm

Pre-Prcosseing of Images

Natural images are divided into three categories for pre-processing first is bright light condition images (BLC) second is low light condition images (LLC) and the last is reflective and scratched surfaces images (RSS). During pre-processing, the image is first, acquired and resized, then converted into two-dimensional matrix data from three-dimensional matrix data. The image acquired is in the form of a matrix of the pixel that is needed to be adjusted or resized as per requirement. It is necessary because the image acquired may be from different means like mobile phones, low sized screenshot, VGA and web camera which produce different size of images and may take very long time to acquire the text area or it will not produce the appropriate output. There is a need to increase or decrease the size hence providing a platform where computation is reduced to an optimum level. This task is done by resizing the captured image. After resizing the image so obtained needs to be converted into a gray image. The conversion of RGB image is too narrow down the matrix elements range and reduce down the computations. As the RGB image obtained is a three-dimensional matrix. Hence it has high complexity in manipulating or processing it. An image is consist of three major components, brightness (intensity), Hue and Saturation. Grayscale images are composed of shades of gray, varying from black to white (0 to 255 levels). Whereas, an RGB image contains the mix of three colors, and require all the three properties to assimilate an image. So, there is a need to convert an RGB image into the grayscale image to eliminating the hue and saturation component from the former image. Now after conversion of the color image to gray image enhancement of the image is necessary to extract out valid and accurate information from the image shown in figure (8). So the image is categorized into three bright light condition are directly moved to next phase of application of filter and operators. Low light image's contrast is enhanced with the help of histogram equalization. And in last reflective and scratched images are enhanced with the help of High-frequency emphasis function [6] as the main objective is to get a final image from which text extraction is easier compared to the captured image.

$$\mathbf{H_E\,(U, V) = A + B\; H_{HP}\,(U, V)} \qquad (7)$$

Where a and b will be positive and a value will lie in the range of 0.25 to 0.5 and b>a therefore its value will lie in the range of 1.5 to 2.5.

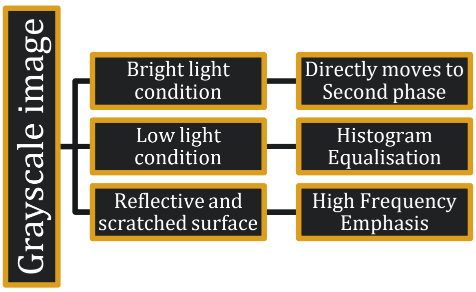


Figure 8 Enhancement of image after conversion of captured image into grayscale image

Application of Filters and Operators

After pre-processing of the image the edges of the gray image are evaluated it will differentiate between the noise, artefacts and area of interest. The edge detection is done with the help of Prewitt operator. This operator is used in image processing, majorly with edge detection algorithms. The Prewitt operator is used for convoluting the image with a small, separable, and integer valued filter matrix, in horizontal and vertical directions. The operator calculates the gradient of the image intensity at each point, giving the direction of the largest possible increase from light to dark and the rate of change in that direction. The result shows the changes in the pixel values and direction which are magnified after the application of filters and operators. The Prewitt operator is generally a (3x3) matrix contain three values i.e. 0, 1,-1. The portion that has the value of positive one gets more contrasted as of other. These are basically of three types of Prewitt operator [8].

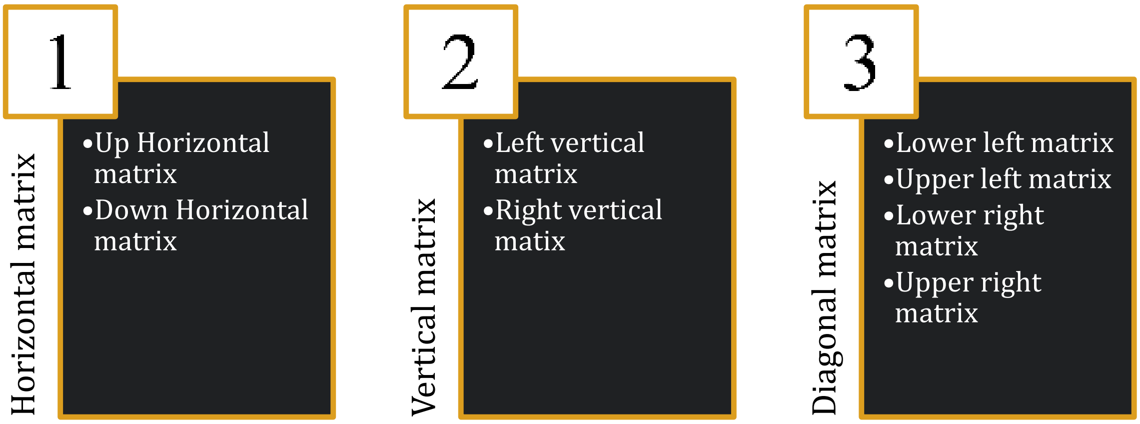


Figure 9 Types of Prewitt Operators

Prewitt Operator

The operator uses two 3×3 kernels which are convolved with the original image to calculate approximations of the derivatives - one for horizontal changes, and one for vertical. **A** is defined as the source image, and **Gx** and **Gy** are two images which at each point contain the horizontal and vertical derivative approximations, which can be computed as:

$$\mathbf{GX} = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} * \mathbf{A} \tag{8}$$

$$\mathbf{Gy} = \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} * \mathbf{A} \tag{9}$$

Where “*" here denotes the 1-dimensional convolution operation.

The *x*-coordinate is defined here as increasing in the right direction, and the *y*-coordinate is defined as increasing in the down direction. At each point in the image, the resulting gradient approximations can be combined to give the gradient magnitude, using:

$$G = \sqrt{Gx^2 + Gy^2} \tag{10}$$

Gradient’s direction is calculated by

$$\Theta = \text{ATAN2 (GX , GY)} \tag{11}$$

Θ is 0 for a vertical edge which is darker on the right side.

The detection of edges in the image is carried out with the help of Prewitt matrix. All the eight images are combined together to form a true edge detected image with all possible edges and have enhanced in its contrast

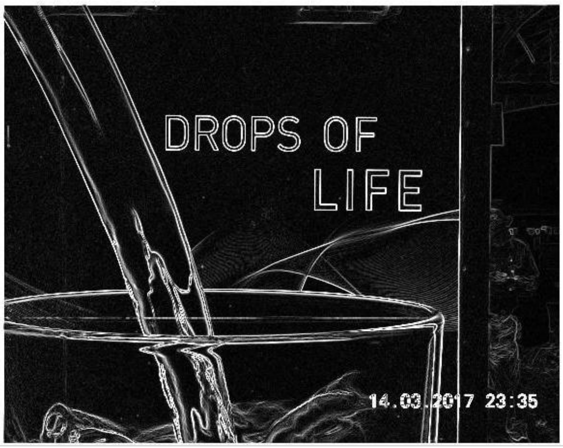


Figure 10 Resultant image “I”, summation of 8 convoluted image with Prewitt operator

Noise and Artefacts Removal

A histogram is developed after the application of filters and operator which is used to analyze the values of the pixel present in the image ‘I’. By analyzing figure (12.a) it is observed that most of the pixels present are toward zero only which means that portion that it has more dark part than the white part. As observed in the figure (10) the text has all white portion and its value is towards 256. Afterwards, the thresholding is carried out to differentiate pixels in two category white (high value) and black (low value). Figure (12.d) shows that a lot of pixels are removed from the binary image is obtained with the edges of the text detected, there is an advantage that the obtained image consists of only two-pixel values that reduce down the computation time significantly.

An area filter is used to remove all the other component and keeps only the required region of interest. Granular noise is removed by it. A specific range of areas is allowed under which the filter allows only connected component or connected area of that area range rest other pixels are rejected just like a bandpass filter. Morphological operation are applied after area filtering such as opening by using structuring element as diamond 3x3 matrix. After this the background turns to black whereas text appears in white.

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$$

up horizontal prewitt matrix

convolved image A1

$$\begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

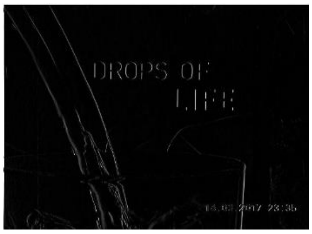
down horizontal prewitt matrix

convolved image A2

$$\begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}$$

left vertical prewitt matrix

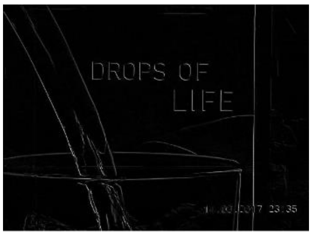
convolved image A3



$$\begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}$$

right vertical prewitt matrix

convolved image A4



$$\begin{bmatrix} 0 & -1 & -1 \\ 1 & 0 & -1 \\ 1 & 1 & -1 \end{bmatrix}$$

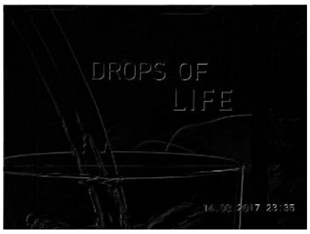
lower left diagonal prewitt matrix

convolved image A5

$$\begin{bmatrix} 0 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & -1 & 0 \end{bmatrix}$$

upper left diagonal prewitt matrix

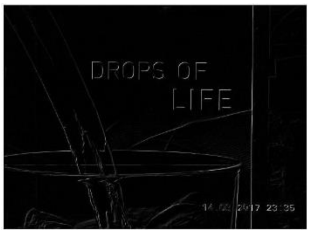
convolved image A6



$$\begin{bmatrix} -1 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

lower right diagonal prewitt matrix

convolved image A7

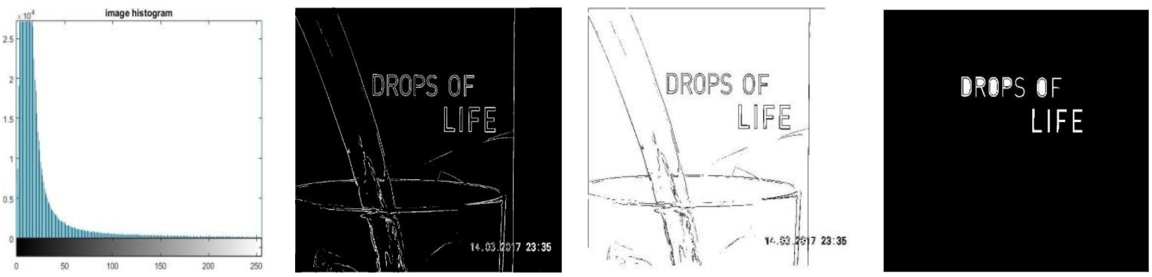


$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & -1 \end{bmatrix}$$

upper right diagonal prewitt matrix

convolved image A8

Figure 11 types of Prewitt operator and the convolved images with their respective matrix



(a) (b) (c) (d)

Figure 12 Noise and artefacts removal (a) histogram of resultant image "I" (b) Binary image obtained after thresholding (c) negative of image (b) for application of area filter (d) image obtained after noise and artefacts removal

Extraction of Text or Character

Afterward, cropping of the image is done to extract out the text and characters only, but it is an unstructured scenario, therefore, location is unknown so cropping of the binary data matrix is done row-wise i.e., first rows are checked one by one whether any white pixel present or not. On getting a white pixel that location is stored [4]. Through this procedure, horizontal location is estimated shown in figure (13). Now same procedure is carried out for columns the red color is used to represent the cropped part of the image and the results are shown in figure (13). Then to extract each letter a matrix is created which stores all the values of columns where the value changes in combination i.e., in the first column the value stored are 1 which is non-zero column so the next column's value it stores retrieves from the column with all values zero then again non zero and so on. With this, every part get extracted with some extra components i.e. the space between two characters and it has all values zero in it just like the character space is very much smaller in size. Therefore matrix with all element zero and less in size is removed, thus protecting words space as shown in figure (14.a) and (14.b).

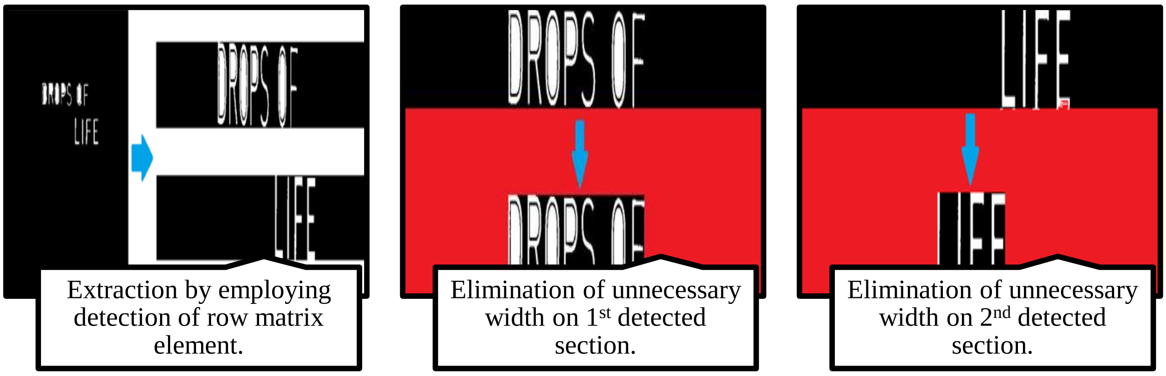


Figure 13 Extraction of words from different lines

Figure 14 Extraction of characters from words (a) Words segmentation for 1st section (b) Words segmentation for 2nd section.

PROPOSED SITE PLAN

As two process are already done. Result of one process are accepted by Springer CCIS series. The other paper in under review stage.

The other two remaining process are currently under process. One of them is going through literature survey and the other one is yet to be started. It is estimated that four processes result will be accepted for publication in the coming January end. After successful completion of all process it is planned to further improve those results and implement proposed in biomedical application.

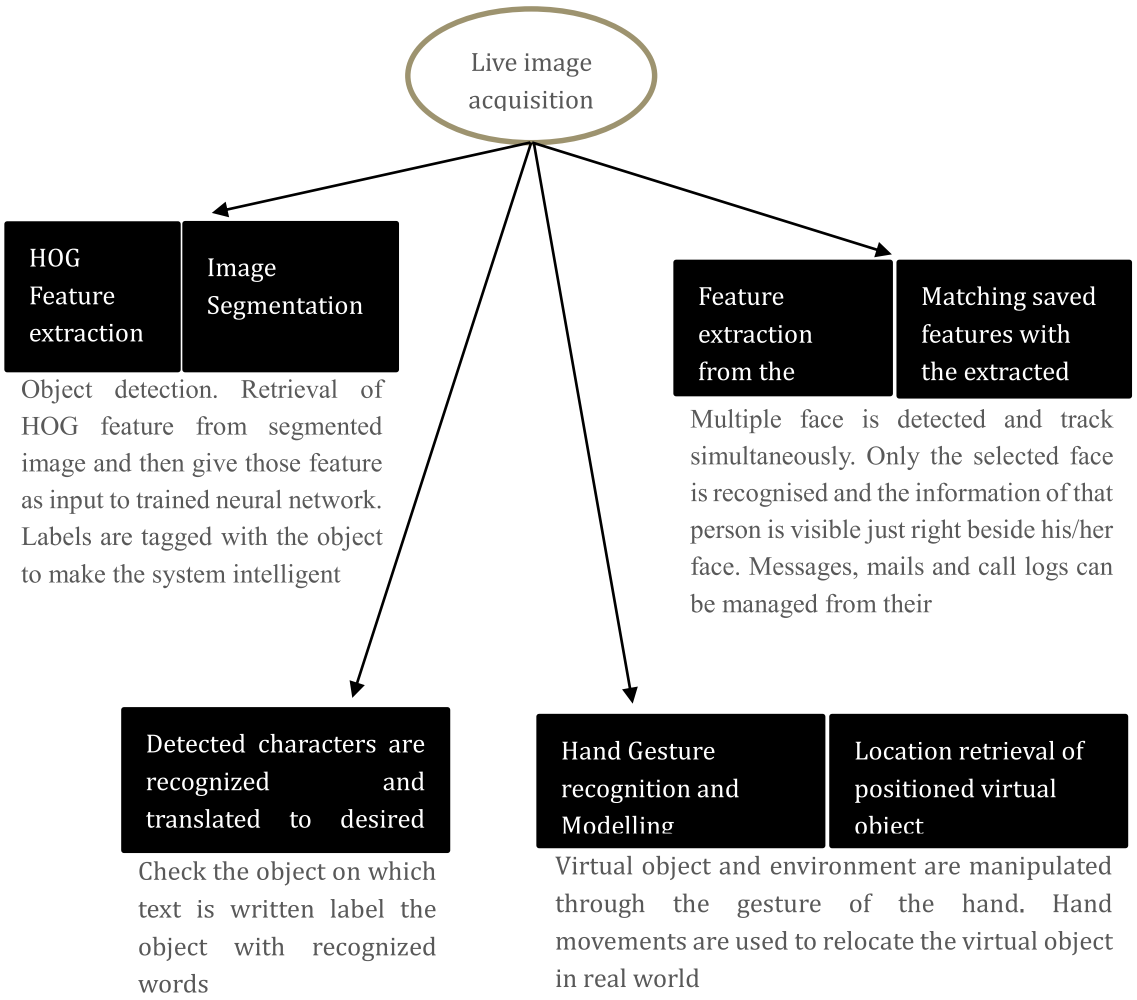


Figure 15 Architecture of IHRS

RESULTS AND DISCUSSION

Complete results of the partial proposed system are discussed under this section and the remaining part is under implementation.

Gesture Controlled Virtual Urban Planning

Here Q factor is 0.1

Resolution of image is 640x480.

Dimensions calculation of ROM: $N_w = 0.1*640$

$= 64$ units

$N_h = 0.1* 480$

$= 48$ units

Size of ROM= 64x48.

Value of Q depends on geographical data and the user requirement

TABLE 1 SHOWING NORMALIZATION ON MID-POINT COORDINATES

COORDINATES OF MID-POINT IN AN IMAGE FRAME	COORDINATES OF MID-POINT IN ROM WORKSPACE
(151,160)	(15.1,16)
(10,30)	(1,3)
(640,480)	(64,48)
(0,28)	(0,2.8)

Now buildings coordinates are calculated that are shifted or moved in ROM after selection and grabbing of buildings with the help of pointers. After grabbing of building the building will follows the normalized mid-point coordinates of the finger and thumb tip.

TABLE 2 MOVEMENTS AND PARAMETER OF CYLINDRICAL BUILDINGS CORRESPONDING TO HAND MOVEMENTS

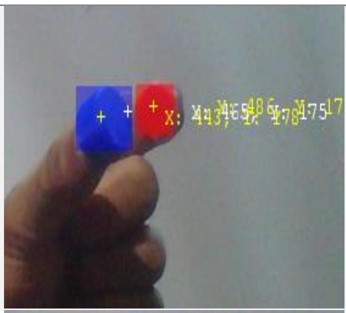
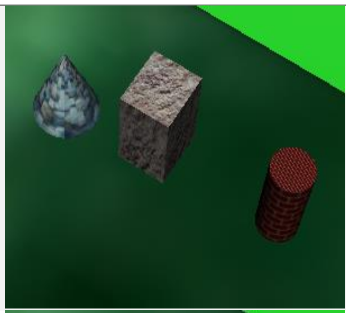
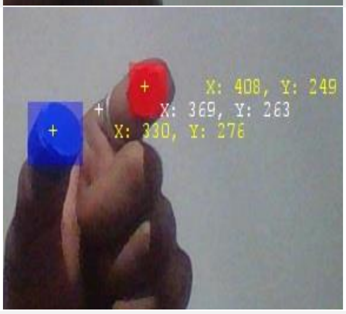
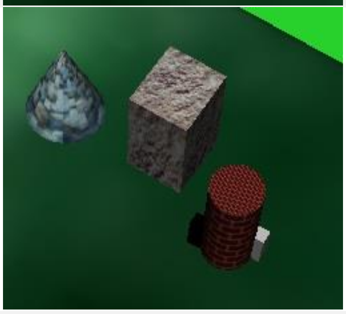
	Image frame coordinates	*Building location*	*Parameters of building*
Initial state			Blue pointer • X-Coordinate=44.3 • Y-Coordinate=17.8 Red pointer • X-Coordinate=48.6 • Y-Coordinate=17.2 Mid-point coordinate = [46.5,17.5] Distance between the two pointers = 4.3 units
Final state			Blue pointer • X-Coordinate=33.0 • Y-Coordinate=27.6 Red pointer • X-Coordinate=40.8 • Y-Coordinate=24.9 Mid-point coordinate = [36.9,26.3] Distance between the two pointers = 8.3 units

TABLE 3 MOVEMENTS AND PARAMETER OF CUBOIDAL BUILDINGS CORRESPONDING TO HAND MOVEMENTS

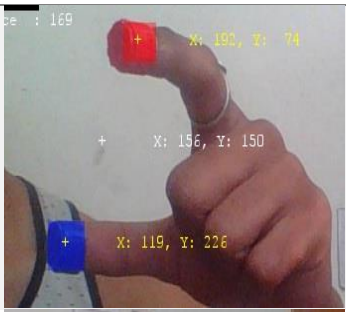
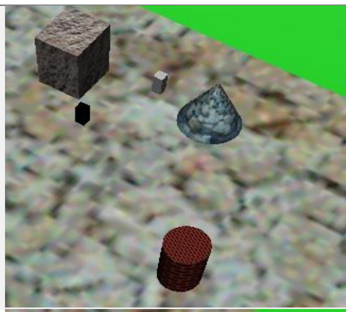
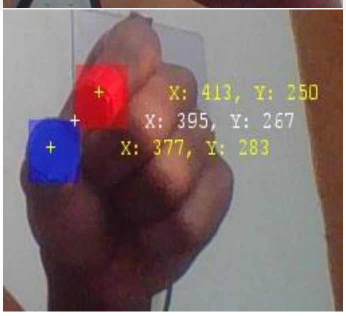
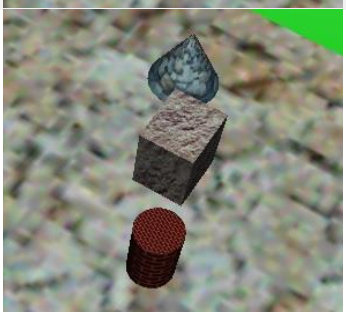
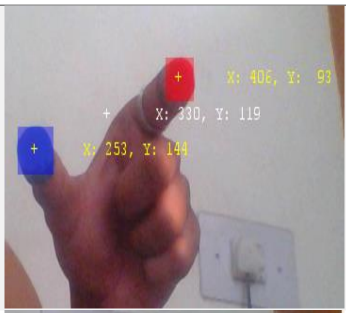
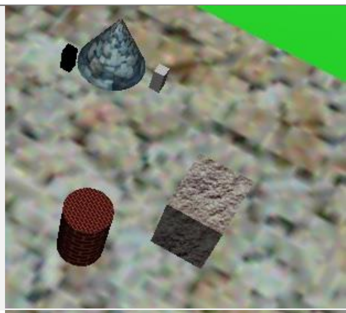
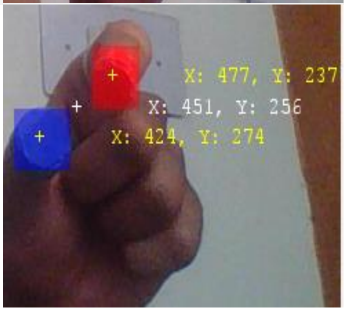
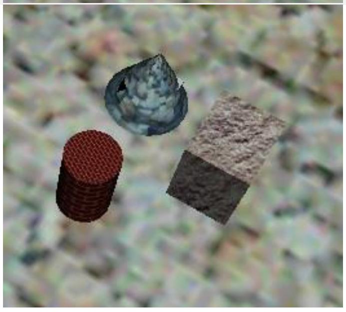
	Image frame coordinates	*Building location*	*Parameters of building*
Initial state			Blue pointer • X-Coordinate=11.9 • Y-Coordinate=22.6 Red pointer • X-Coordinate=19.2 • Y-Coordinate=7.4 Mid-point coordinate = [15.6,15.0] Distance between the two pointers = 16.9 units
Final state			Blue pointer • X-Coordinate=37.7 • Y-Coordinate=28.3 Red pointer • X-Coordinate=41.3 • Y-Coordinate=25.0 Mid-point coordinate = [39.5,26.7] Distance between the two pointers = 4.9 units

TABLE 4 MOVEMENTS AND PARAMETER OF CONICAL BUILDINGS CORRESPONDING TO HAND MOVEMENTS

	Image frame coordinates	*Building location*	*Parameters of building*
Initial state			Blue pointer • X-Coordinate=25.3 • Y-Coordinate=14.4 Red pointer • X-Coordinate=40.6 • Y-Coordinate=9.3 Mid-point coordinate = [33.0,11.9] Distance between the two pointers = 16.1 units
Final state			Blue pointer • X-Coordinate=42.4 • Y-Coordinate=27.4 Red pointer • X-Coordinate=47.7 • Y-Coordinate=23.7 Mid-point coordinate = [45.1,25.6] Distance between the two pointers = 6.5 units

Texture-Based Real-Time Character Extraction and Recognition in Natural Images

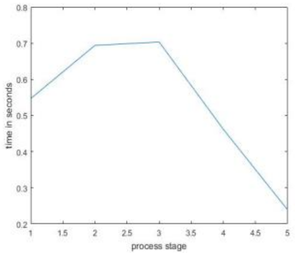
A technique has been developed for detection as well as extraction of text and character. Robust classes and functions of image processing techniques are employed in this proposed algorithm. Captured image is converted into binary image then row-column detection is done after that letters are extracted from word. Extracted letters are matched with the templates to recognize the character. Figure 8 shows the results obtained from this proposed algorithm.

TABLE 6 DETAILED ANALYSIS OF DIFFERENT NATURAL IMAGES

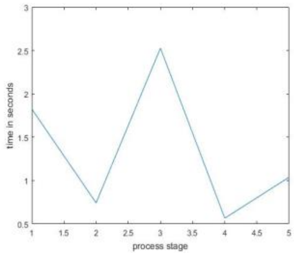
Image type	Original Image size	Resize Factor	No. of edges	Total Time (in sec)	AHE	HFE
BLC image (fig. 16.1.c)	4.11 MB 3456x4608	0.3	131392	10.530	No	No
LLC image (fig. 16.2.c)	3.90 MB 4608x3456	0.1	136383	7.4773	Yes	No
RSS image (fig. 16.3.c)	1.71 MB 1605x2185	0.1	117186	2.6444	No	Yes

Table 6 shows is the detailed analysis of three different natural image in first case it is observed that large size image total time required for output is 10.5301 sec which is considerably high compared to other two cases. In last two cases adaptive histogram and high frequency emphasis is used respectively for enhancement of image to get desired results.

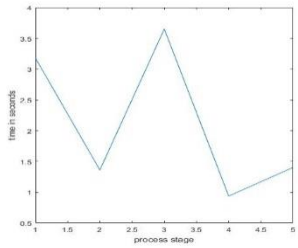
TABLE 7 GRAPHICAL REPRESENTATION OF TIME TAKEN IN CORRESPONDING STAGES



BLC image (figure 16.1.c)



LLC image (figure 16.2.c)



RSS image (figure 16.3.c)

Now the time is different for different category of images as it is observed from table 7 the LLC image of figure 16.2.c and RSS image of figure 16.3.c is time curve is similar. It is clear from the time curve that noise and artefacts removal took the most time of the proposed system.

TABLE 8 ACCURACY AND ERROR ANALYSIS OF PROPOSED SYSTEM

Image type	**Image no.**	**Total no. of character**	**Correctly matched character**	**Mismatched character**	**Missed character**	**Accuracy (in %)**	**Error (in %)**
BLC image	Fig(16.1.a)	8	8	0	0	100	0
	Fig(16.1.b)	11	11	0	0	100	0
	Fig(16.1.c)	16	8	0	8	50	0
LLC image	Fig(16.2.a)	14	11	3	0	78.57	27.27
	Fig(16.2.b)	11	11	0	0	100	0
	Fig(16.2.c)	7	7	4	0	100	57.14
RSS image	Fig(16.3.a)	20	14	2	4	70	10
	Fig(16.3.b)	15	15	0	0	100	0
	Fig(16.3.c)	63	50	0	13	79.36	0

Accuracy and error analysis of every image used in this proposed system is done in table 8 in this three parameters are calculated to estimate accuracy and error of proposed system. Matched character parameter is used to calculate accuracy and mismatched character parameter is used to calculate error in the system. From the tables (1, 2 and 3) it is concluded that number of missing character is more than the number of missed character, so enhance the results proposed system has to be more sensitive towards low font size character.

Figure 16 (1) BLC images (2) LLC Images (3) RSS Images

SUMMARY AND CONCLUSION

Virtual reality is one of the emerging and dynamic technologies used for entertainment, education, space and military training etc. Virtual objects and environments are simple and easy to design and sure are cost effective. The virtual objects can be handled like real world entity. HCI by hand gestures allows the person to interact with virtual objects and environment without any advance knowledge of programming in VRML, 3-D modelling etc. In this paper two blocks (as pointers) are developed for interactions. These pointers are coupled with the index finger and thumb which are responsible for management of multiple virtual objects across the virtual world. Virtual reality has countless and diverse applications which will lead boom in the current scenarios of the market and research industry. There are numerous ways to further enhance this approach.

1. High resolution satellite 3-D terrain and elevation data can be used for planning of city, town and metropolitan, etc. as it gives a realistic view
2. In this proposed project only two pointers are designed and coupled, one with index finger and other with thumb. For complex gestures other fingers can be used for partial manipulation of object such as segmentation, resizing, rotation etc.

As virtual and augmented reality have immense number of application like virtual maps, internet surfing, video games etc. The target of this proposed system is to combine all those process in which information can be extracted from a natural image and create a powerful virtual environment in which user will easily interact with the real environment. After that intelligence has to deployed on this system so as provide an user friendly experience and make it easier for user to handle the extracted information by the system

REFERENCES

1. Chen Huizhong, Sam.S.Tsai, Georg Schroth, David M.Chen Radek Grzeszczuk, Bernd Girod. "Robust Text Detection in Natural Images with Edge-Enhanced Maximally Stable Extremal Regions." Image Processing (ICIP), 2011 18th IEEE International Conference on. IEEE, 2011. doi:10.1109/icip.2011.6116200
2. Gonzalez Alvaro, Luis M. Bergara, J Javier Yebes, Sebastian Bronte. "Text location in complex images." Pattern Recognition (ICPR), 2012 21st International Conference on. IEEE, 2012.
3. Li, Yao, and Huchuan Lu. "Scene text detection via stroke width." Pattern Recognition (ICPR), 2012 21st International Conference on. IEEE, 2012.
4. Neumann, Lukas, and Jiri Matas. "Real-time scene text localization and recognition." Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012. doi:10.1109/CVPR.2012.6248097.
5. Wilhelm Burger and Mark J. Burge (2008). Digital Image Processing: An Algorithmic Approach Using Java. Springer-Verlag. (truecolor) extensions". Archived from the original on 16 March 2012. Retrieved 23 March 2012
6. Gonzalez, Rafael, C; Woods, Richard E (2008). Digital Image Processing, 3rd Edition. Pearson Prentice Hall. p. 577,188.
7. Philip Howard. "True-Color GIF Example". Archived from the original on 22 February 2015. Retrieved 23 March 2012.
8. J.M.S. Prewitt "Object Enhancement and Extraction" in "Picture processing and Psychopictorics", Academic Press,1970
9. Dim, Jules R.; Takamura, Tamio (2013-12-11). "Alternative Approach for Satellite Cloud Classification: Edge Gradient Application". Advances in Meteorology. 2013. doi:10.1155/2013/584816
10. Michael Freeman (2005). The Digital SLR Handbook. Ilex.
11. V.Kasmat and S. Ganesan "An efficient implementation of the Hough transform for detecting vehicle license plates using DSP's " IEEE International Conference on Real-Time Technology and Application Symposium Chicago USA pp. 58-59 2005. doi:10.1109/RTTAS.1995.516201
12. S.H. Park K.I. kim K. Jung and H.J. Kim "Locating car license plate using Neural Network " Electronic Letters Vol.35 No.17 pp. 1474 - 1477 1999. doi:10.1049/el:19990977
13. K.K. Kim K.I. Kim J.B. Kim and H.J. Kim "Learning-Based Approach for License Plate Recognition" Proceeding of IEEE Signal Processing Society Workshop Vol.2 pp.614-623 2000. doi:. 10.1109/NNSP.2000.890140.
14. Stephen Rawls, Hauigo Cao, Ekraam Sabir and Prem Natarajan “Combining deep learning and language modeling for segmentation-free OCR from raw pixels” IEEE International conference on Arabic script Recognition and Analysis (ASAR), 2017. doi: 10.1109/ASAR.2017.8067772
15. Guillaume Chiron, Antoine Doucet, Mickael Coustaty, Muriel Visani, Jean-Philippe Moreux. “Impact of OCR Errors on the Use of Digital Libraries: Towars a better Access to information”

ACM/IEEE Joint Conference on Digital Libraries (JCDL), 2017. doi: 10.1109/JCDL.2017.7991582
16. Mohammad Azim Ul Ekram, Anjani Chaudhary, Ashutosh Yadav, Jagadish Khanal, Semith Asian.“Book organization checking algorithm using image segmentation and OCR”.IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS), 2017. doi: 10.1109/MWSCAS.2017.8052894
17. Vladimir I. Pavlovic, Rajeev Sharma, Thomas S. Huang “Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review”, IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, July 1997
18. Koichi Hirota, Kazuyoshi Tagawa “Interaction with Virtual Object using Deformable Hand”, IEEE Virtual Reality Conference 2016
19. Jan Jacobs, Bernd Froehlich, “A Soft Hand Model for Physically-based Manipulation of Virtual Objects”, IEEE Virtual Reality 2011
20. Yoshiyuki Kojima, Yoshihiro Yasumuro, Hiroshi Sasaki, Ichiroh Kanaya, Osamu Oshiro, Tomohiro Kurodo, Yoshitsugu Manabe, Kunihiro Chihara Naist, “Hand Manipulation of Virtual Objects in Wearable Augmented Reality”, Proceedings of the seventh International Conference on Virtual Systems and Multimedia(VSMM01), 2001 IEEE
21. Jun Rekimoto and Katashi Nagano: The World through the Computer: ‘’Computer Augmented Interaction with Real World Environment”, In Proceedings of UIST’95, pp.29-36, 1995
22. Takeshi Kurata, Takashi Okuma, Masakatsu Kourogi and Katsuhiko Sakae, ” The Hand-Mouse: A Human Interface Suitable for Augmented Reality Environment Enabled by Visual Wearbles”, Technical Report of IEICE (PRMU)2000-156, pp. 69-76
23. Kouichi Ishibuchi, Keisuke Iwasaki, Haruo Takemura and Fumio Kishino: “Real-Time Vision-Based Hand Gesture Estimation for Human-Computer Interfaces”, The Journal of The Institute of Electronic Information and Communication Engineers, D-2, Vol. J79-D-2 No.7, pp.1218-1229(in Japanese)(1996)
24. Masaki Fukumoto and Yoshinobu Tonoma; “Body Coupled Finger Rings”, Wireless Wearable Keyboard, http://acm.arg/sigchi/chi97/proceeding/paper/fkm.htm
25. PattThad Starner, Joshua Weaver, and Alex Pentland Proceedings of the IEEE International Symposium on Wearable Computers, Cambridge, MA, 1997
26. L.R. Rabiner and B.Juang, Fundamentals of Speech Recognition. Englewood Cliffs, N.J.: Prentice Hall, 1993.Waibel and K.F. Lee, Readings in Speech Recognition. Morgan Kaufmann, 1990
27. J.A. Adam, “Virtual Reality,” IEEE Spectrum, Vol.30. n0. 10, pp. 22-29, 1993.
28. H. Rheingold, Virtual reality. Summit Books, 1991.
29. S.S. Fels and G.E. Hinton, “Glove-Talk: A Neural Network Interface between a Data-Glove and a Speech Synthesizer,” IEEE Trans. Neural Networks, vol. 22, no. 12, pp. 79-89, 1991.
30. D.J. Sturman and D. Zeltzer, “A Survey of Glove-Based Input,” IEEE Computer Graphics and Applications vol. 14, pp. 30-39, Jan. 1994
31. D.L. Quam, “Gesture Recognition with a DataGlove,” Proc. 1990 IEEE National Aerospace and Electronics Conf., vol. 2, 1990.

32. Sven Wursthorn, Alexandre Hering, Coelho and Guido Staub, “Application of Mixed Reality” 10th IEEE International Symposium on Mixed and Augmented Reality, 2011
33. Tuomas Kantonen, Charles Woodward, Neil Katz VTT Technical Research Centre of Finland, IBM Corporation,”Mixed Reality in virtual World Teleconferencing” IEEE Virtual Reality Conference (VR), 2010, pp. 179-182.
34. Chin-fen chen, Mark Bolas, Evan Suma Rosenberg, “Rapid creation of photorealistic virtual reality content with consumer depth cameras”, IEEE Virtual Reality 2017, pp. 473-474.
35. Cullen Brown, Ghanshayam Bhutra, Mohamed Suhail, Qinghong Xu, Eric D. Ragan, “Coordinating attention and cooperation in multi-user virtual reality narratives”, IEEE Virtual Reality 2017, pp. 377-378.
36. Paul Viola, Michael Jones, “Rapid Object Detection using a Boosted Cascade of Simple Features”, Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, Vol. 1, pp. 511-518.
37. Huizhong Chen, Sam S. Tsai, George Schroth, David M. Chen, Radek Grzeszczuk and Bernd Girod, “Robust Text Detection in Natural Images With Edge-Enhanced Maximally Stable Extremal Regions”, 18th IEEE International Conference on Image Processing, 2011, pp. 2609-2612.
38. Clementine Perinaud, Georges Gay and Gilles Gesquiere, “Exploration of the Changing Structure of Cities: Challenges for Temporal City Models”, Digital Heritage, 2015