A PROJECT REPORT ON

"OBJECT DETECTION AND RECOGNITION USING MOBILE PHONE APPLICATION FOR VISUALLY IMPAIRED"

SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE
IN THE PARTIAL FULFILLMENT FOR THE AWARD OF THE DEGREE

OF

BACHELOR OF ENGINEERING
IN
INFORMATION TECHNOLOGY

BY

PRACHI V MUNSHETTIWAR	**B120388648**
KALYANI GOLEKAR	**B120388573**
SHREYAS MUNOT	**B120388626**
KALYANI RANE	**B120388655**

UNDER THE GUIDANCE OF
PROF. ADITI JAHAGIRDAR

DEPARTMENT OF INFORMATION TECHNOLOGY
MAEER's MIT COLLEGE OF ENGINEERING
PUNE – 411038
ACADEMIC YEAR 2015-2016

CERTIFICATE

This is to certify that the project report entitled

"OBJECT DETECTION AND RECOGNITION USING MOBILE PHONE APPLICATION FOR VISUALLY IMPAIRED"

Submitted By

Prachi V Mushettiwar	B120388648
Kalyani Golekar	B120388573
Shreyas Munot	B120388626
Kalyani Rane	B120388655

is bonafide work carried out by them under the supervision of **Prof. Aditi Jahagirdar** and it is approved for the partial fulfillment of the requirement of Savitribai Phule Pune university for the award of the Degree of Bachelor of Engineering (INFORMATION TECHNOLOGY).

This project report has not been earlier submitted to any other Institute or University for the award of any Degree or Diploma.

Prof. Aditi Jahagirdar
Internal Guide
Department of Information Technology

Prof. Dr. A. S. Hiwale
Head of Department
Department of Information Technology

External Examinar

Date:

Prof. Dr.M.S.NAGMODE
Principal

Place:
Date:

Acknowledgement

We take this opportunity to thank our project guide Prof. Aditi Jahagirdar and Head of the Department Prof. Anil Hiwale for their valuable guidance and for providing all the necessary facilities, which were indispensable in the completion of this project report. We are also thankful to all the staff members of the Department of Information Technology of MIT College of Engineering, Pune for their valuable time, support, comments, suggestions and persuasion. We would also like to thank the institute for providing the required facilities, Internet access and important books.

Prachi V Munshettiwar

Kalyani Golekar

Shreyas Munot

Kalyani Rane

LIST OF FIGURES

LIST OF ABBREVATIONS

ABBREVATIONS	MEANING
SIFT	Scale invariant Feature Transform
CRF	Corner Response Function
SURF	Speeded Up Robust Feature
PCA	Principal Component Analysis
RGB	Red Green Blue
HIS	Hue Saturation Intensity

LIST OF TABLES

LIST OF CONTENTS

CONTENTS

Abstract

The blind and the visually impaired people face diverse kinds of challenges. The independent mobility of these people is coming into danger due to scattered objects in house. Thus a significant help is offered by software applications for computers and touch-screen devices equipped with speech synthesizers that will help the blind people to perform their daily activities easily. In this system the solution is aimed at aiding the visually impaired people to recognize household objects by clicking a picture of it. The system is based on a dedicated object detection and recognition application running on Android system Smartphone. Image recognition results are communicated to the visually impaired user by means of pre-recorded verbal messages. The main goal is to design an application which would allow detecting and recognizing objects from images recorded by the camera of a mobile device and give back the information to the users. This application can also be further used in shopping malls to detect and recognize the objects in the malls.

CHAPTER 1

INTRODUCTION

The blind people face a lot of problems in their daily activities. An application is proposed that aims at aiding the visually impaired people in recognition of object. The system is based on a dedicated image recognition application running on an Android system Smartphone. Image recognition results are communicated to the blind user by means of prerecorded verbal messages.

1.1 Project idea

The application helps visually impaired people to perform their daily activities easily and more efficiently.

1.2 Motivation of project

Blind people: The blind and the visually impaired people face diverse kinds of life challenges. The safety and independent mobility of these people is coming into danger. We are motivated by them and have decided to do something for them.

1.3 Area of Project

Image Processing: It is a form of signal processing in which input is an image such as a photograph or a video frame and output of image processing either may be an image or characteristics or a set of parameters related to the image.

Mobile computing: It is a human computer interaction by which a computer is expected to be transported during normal uses. It involves mobile communication, computer hardware and software.

1.4 Statement of scope

The scope of this project is that it captures images of various household objects and compares the captured images with the already existing images in database and gives an output identifying the object in written as well as verbal form.

CHAPTER 2

LITERATURE SURVEY

2.1 D. Lowe[1] proposed

• Object recognition in cluttered real-world scenes requires local image features that are unaffected by nearby clutter or partial occlusion. The features must be at least partially invariant to illumination, 3D projective transforms, and common object variations.

• The difficulty of the object recognition problem is due in large part to the lack of success in finding such image features. However, recent research on the use of dense local features.

• This paper[1] describes an efficient method to identify stable key locations in scale space. This means that different scaling of an image will have no effect on the set of key locations selected.

• This paper[1] presents a new method for image feature generation called the Scale Invariant Feature Transform(SIFT). This approach transforms an image into a large collection of local feature vectors, each of which is invariant to image translation, scaling, and rotation, and partially invariant to illumination changes and affine or 3D projection.

2.2 D. Lowe[2] proposed

• Scale-space extrema detection: The first stage of computation searches over all scales and image locations. It is implemented efficiently by using a difference-of- Gaussian function to identify potential interest points that are invariant to scale and orientation.

• Keypoint localization: At each candidate location, a detailed model is fit to determine location and scale. Keypoints are selected based on measures of their stability.

• Orientation assignment: One or more orientations are assigned to each keypoint location based on local image gradient directions. All future operations are performed on image data that has been transformed relative to the assigned orientation, scale, and location for each feature, thereby providing invariance to these transformations.

2.3 Helal S., Mokhtari M. and Abdulrazak B.[3]

• The motivation of the project is the interdisciplinary research of a very complex topic of assistance of orientation and navigation of visually impaired people, in a known or unknown indoor environment.

• Assistive Technology is a generic term incorporating technology, equipment, devices, appliances, services, systems, processes and environmental change (Environmental Modifications) used by people with disabilities or older people to overcome social, infrastructural barriers, to actively participate in society and to perform activities easily and safely.

• Assistive systems for daily life can be classified as follows:

1. Personal care
2. Timekeeping, alarms and alerting
3. Food preparation and consumption
4. Environmental control/ household appliances
5. Money, finance and shopping.

2.4 K. Matusiak, P.Skulimowski1 and P. Strumiłło [4]

• This paper[4] proposed a new method of key point orientation. For each key point, a neighborhood is defined by the use of a Gaussian weighting function, which weights pixels gradient magnitudes and orientations that are used to form an orientation histogram. In the algorithm proposed here, for each pair of the descriptors of the key

points that were paired using a similarity measure, a difference of main orientations with precision to unit degree is calculated. On the basis of this data the application builds a histogram showing frequency of occurrence of all rotations. By using this histogram the most frequent difference is defined. As a result the most probable change in rotation of the object in a scene can be determined.

2.5 Rosten E., Drummond T.[5] proposed

- Corners play an important role in image processing and computer vision algorithms, because they are very distinctive features.

- For some of the approaches, the CRF (Corner-Response-Function) can be shown to be invariant in scale, rotation or even affine transformations. In general, the computational costs increase with the number of invariant corner features. Therefore, many traditional approaches focus either on the computation speed, or the distinctiveness of the detections.

- The feature vector of SURF is almost identical to that of SIFT. It creates a grid around the keypoint and divides each grid cell into sub-grids. At each sub-grid cell, the gradient is calculated and is binned by angle into a histogram whose counts are increased by the magnitude of the gradient, all weighted by a Gaussian. These grid histograms of gradients are concatenated into a 64-dimensional vector.

- The high dimensionality makes it difficult to use this in real time, so SURF can also use a 36-vector of principle components of the 64 vector (PCA analysis is performed on a large set of training images) for a speedup. SURF also improves on SIFT by using a box filter approximation to the convolution kernel of the Gaussian derivative operator. This convolution is sped up further using integral images to reduce the time spent.

CHAPTER 3

SYSTEM SPECIFICATION

3.1 Main Constraints

- The main constraint is the quality of camera which directly impact on the result of application. Good quality camera is needed.
- The resolution of image is another issue which impact on specification and also implementation of application.

3.2 Hardware Resources Required

- Processor: Pentium 4, 2 GHz and above
- RAM: 1 GB
- Disk: 40 GB
- Smart phone having good quality camera and computational qualities.

3.3 Software Resources Required

- Front End: JAVA
- Back End: MySQL
- IDE: Netbeans
- Operating System: Windows XP/Vista 7
- Documentation: MS- Office

CHAPTER 4

DESIGN

4.1 WHAT IS TO BE DEVELOPED

Firstly, the visually impaired person will click the image using the android application then that image is directly send to the server through the internet. Server contains the database about the all object details. Using that detail the server will send the corresponding text message to the user and at client side that text message is converted to the voice and in these way the object is recognized.

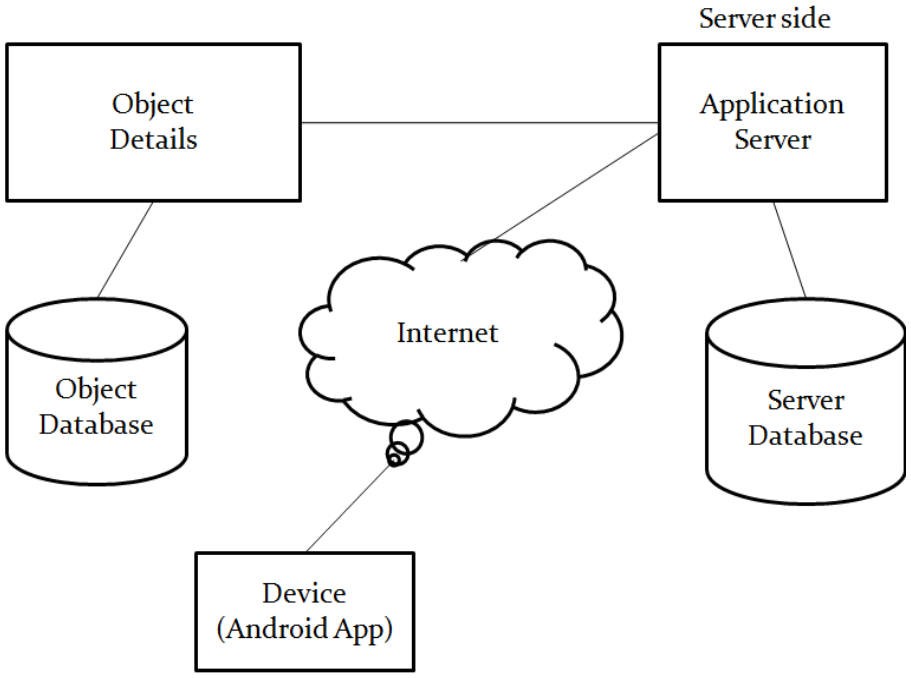

Figure 4.1: System Architecture

4.2 DATA FLOW DIAGRAM

A data flow diagram (DFD) is a graphical representation of the "flow" of data through an information system, modeling its process aspects. A DFD is often used as a preliminary

step to create an overview of the system, which can later be elaborated.A DFD shows what kind of information will be input to and output from the system, where the data will come from and go to, and where the data will be stored.

4.2.1 LEVEL 0 DFD

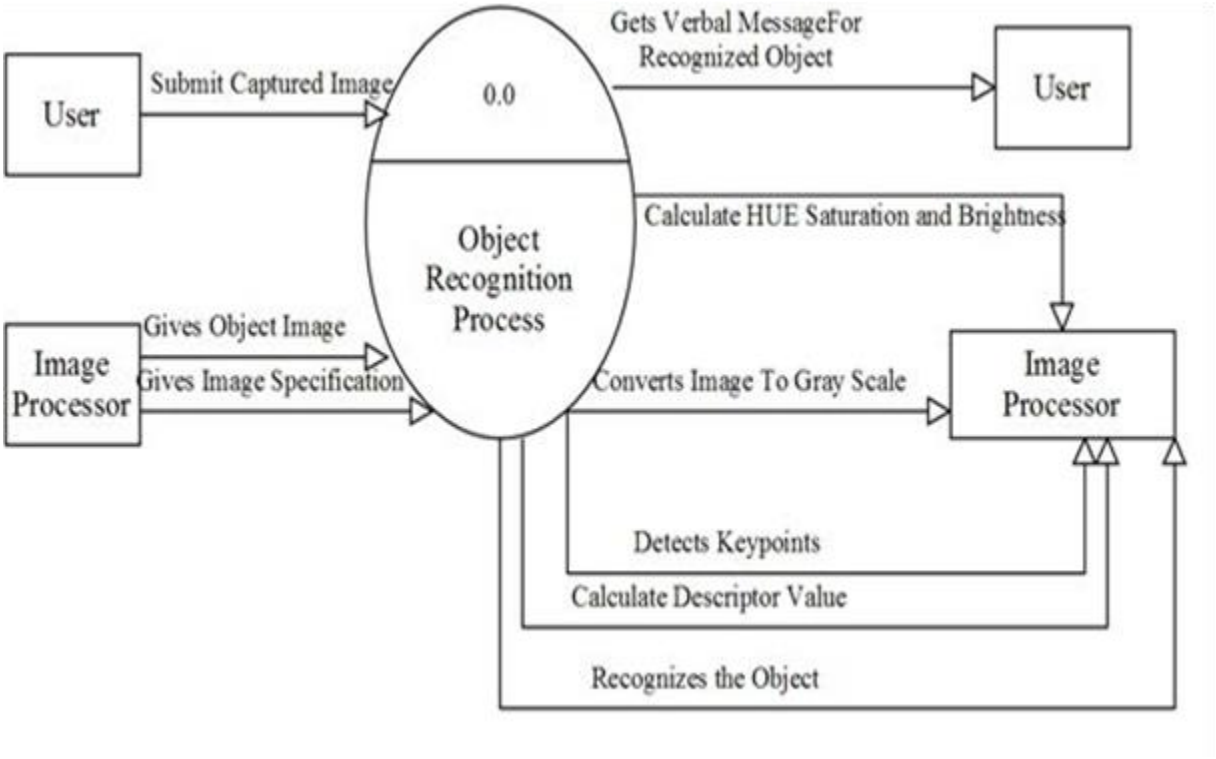

Figure 4.2.1: Level 0 DFD

4.2.2 LEVEL 1 DFD

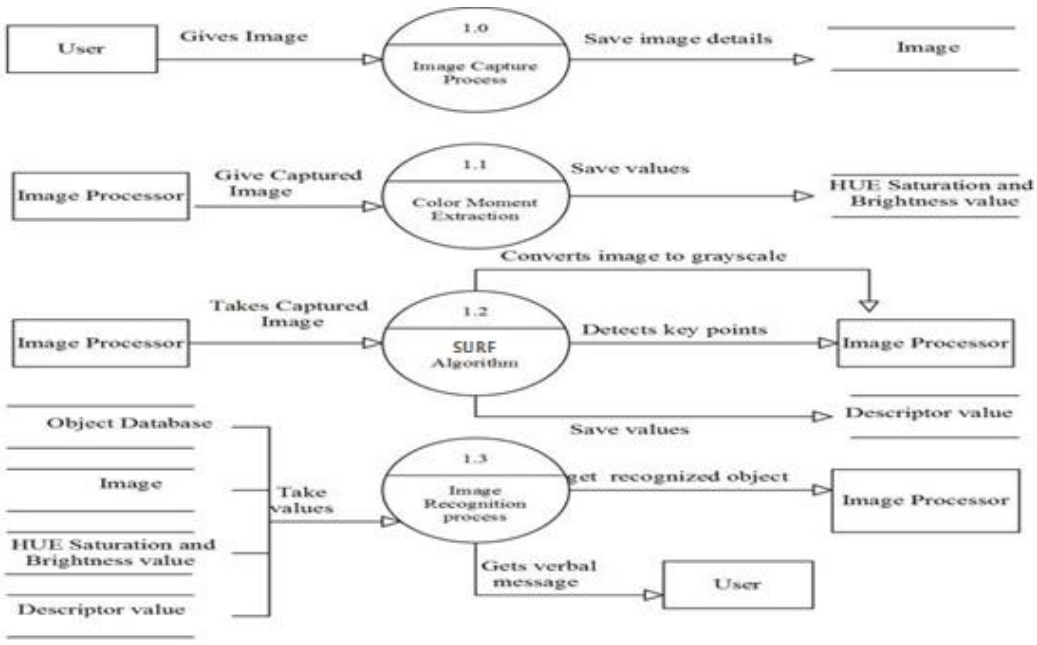

Figure 4.2.2: Level 1 DFD

4.3 UML DIAGRAMS

4.3.1 USE CASE DIAGRAM

In software and systems engineering, a use case is a list of action or event steps, typically defining the interactions between a role (known in the Unified Modeling Language as an actor) and a system, to achieve a goal. The actor can be a human, an external system, or time. Use case diagram depicts the interaction between the system and the end user. It describes the behavior of the system and helps us understand the role of actors in the system.

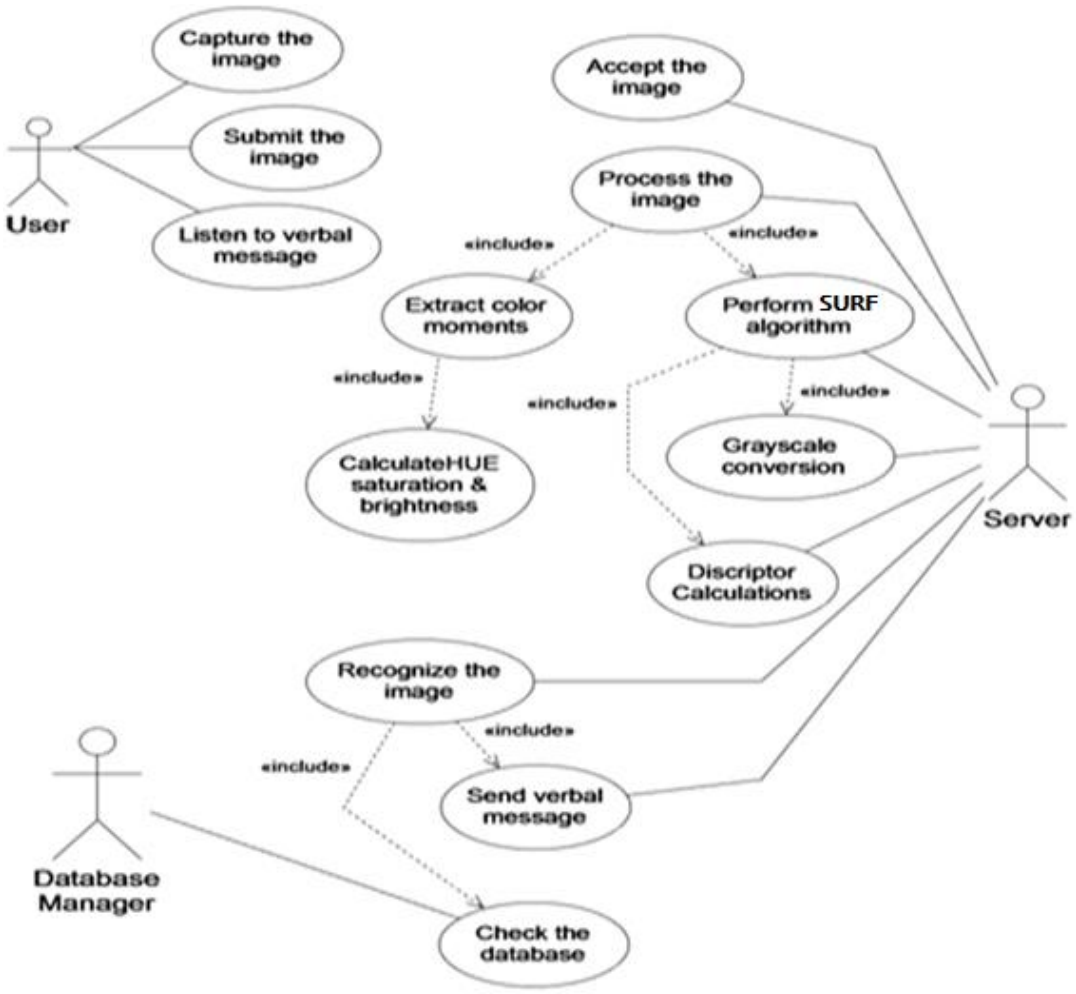

Figure 4.3.1: Use Case Diagram

4.3.2 ACTIVITY DIAGRAM

Activity diagram depicts a series of actions that defines a job or how work should be done. Activity diagram visualizes how tasks will flow between resources, whether they're machines or people and what conditions allow the sequence to move forward.

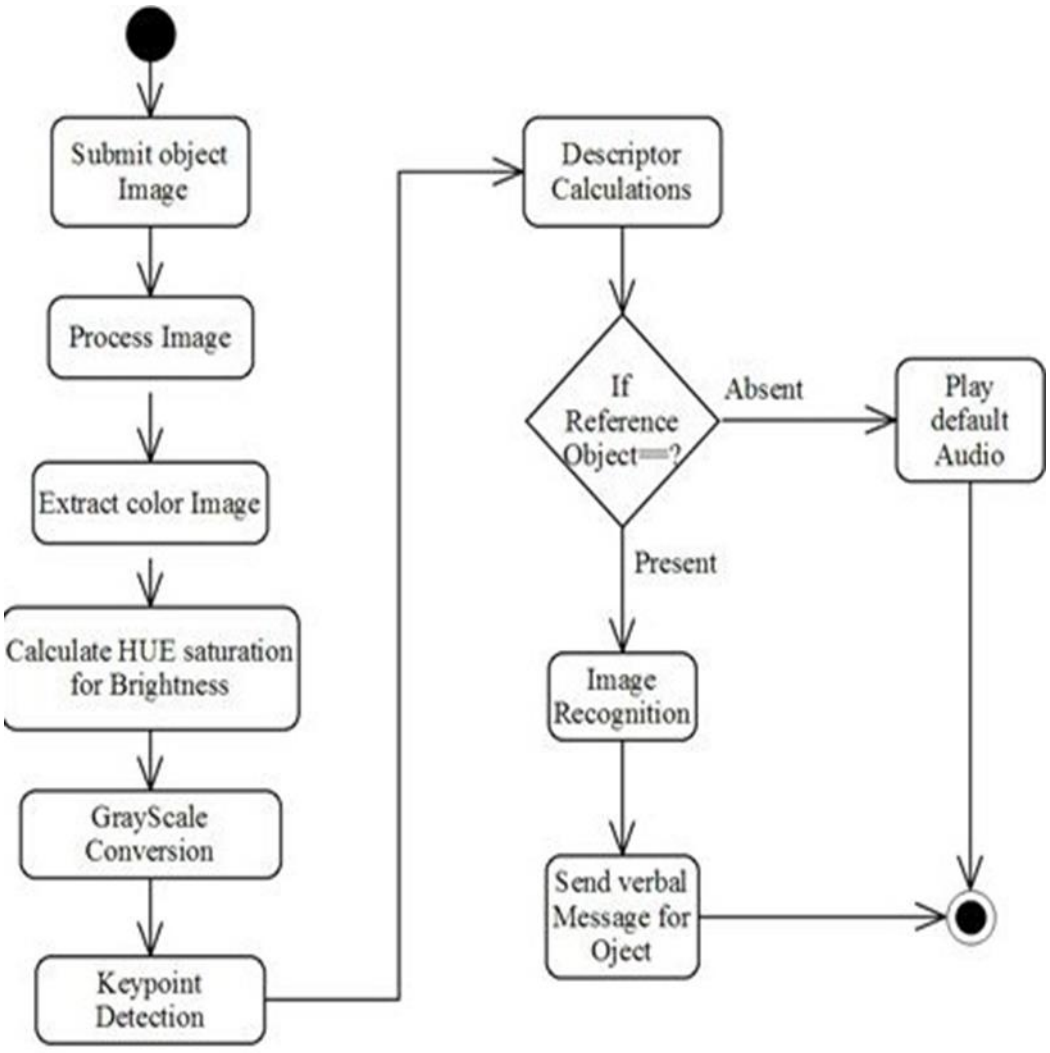

Figure 4.3.2: Activity Diagram

4.3.3 SEQUENCE DIAGRAM

A Sequence diagram is an interaction diagram that shows how processes operate with one another and in what order. The interaction is shown or arranged according to time sequence, so that we get an idea about what happens when and what comes next. It helps in understanding the flow of system with respect to time. The sequence diagram consists of objects, their lifelines, messages (call) to interact and activation bar for indicating which action is being performed at that particular instant.

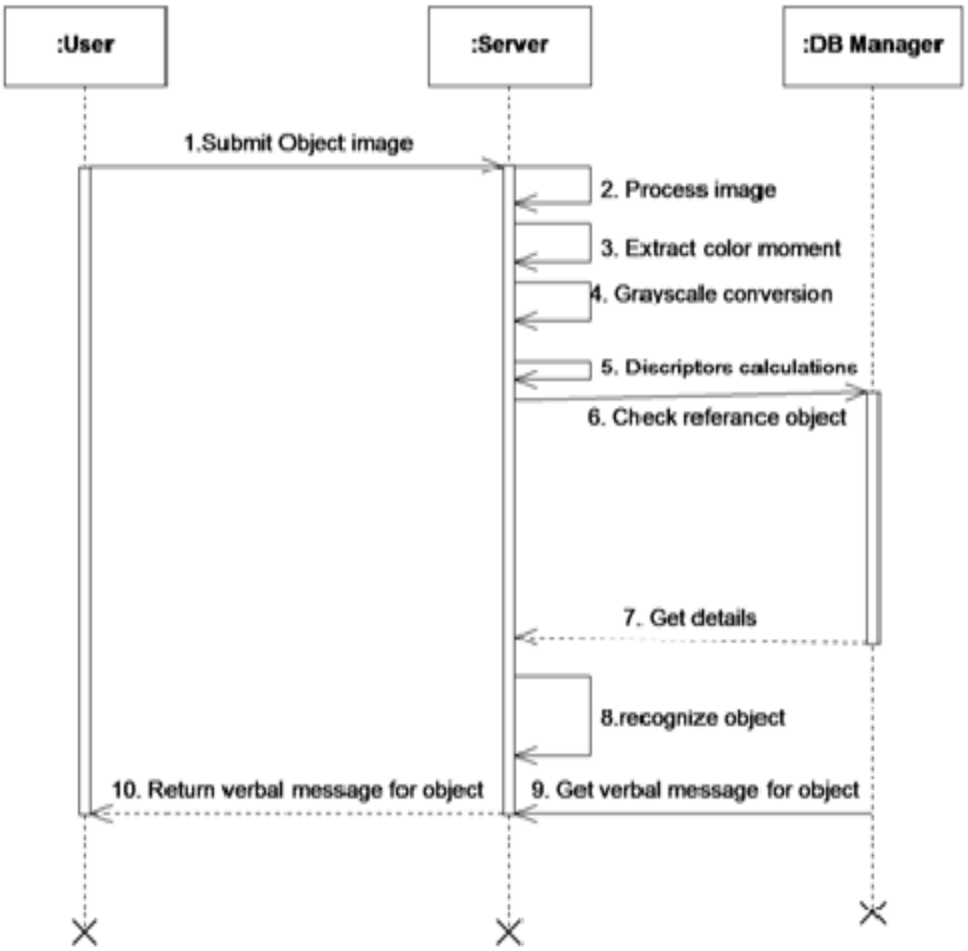

Figure 4.3.3: Sequence Diagram

CHAPTER 5
IMPLEMENTATION

5.1 METHEDOLOGY USED

In this application SURF algorithm is used to detect and recognize the objects. The clicked image will be processed using the SURF algorithm. Speeded Up Robust Features (SURF) is a local feature detector and descriptor that can be used for tasks such as object recognition or registration or classification or 3D reconstruction. It is partly inspired by the scale-invariant feature transform (SIFT) descriptor. The standard version of SURF is several times faster than SIFT and claimed by its authors to be more robust against different image transformations than SIFT.

To detect interest points, SURF uses an integer approximation of the determinant of Hessian blob detector, which can be computed with 3 integer operations using a pre-computed integral image. Its feature descriptor is based on the sum of the Haar wavelet response around the point of interest. These can also be computed with the aid of the integral image.

SURF descriptors can be used to locate and recognize objects, people or faces, to make 3D scenes, to track objects and to extract points of interest. In this algorithm the keypoints are calculated and these keypoints are matched with the keypoints of the images which are stored in the database. The steps of the algorithm are as follows:

1. Constructing a scale space. This is the initial preparation. You create internal representations of the original image to ensure scale invariance. This is done by generating a "scale space".

2. Then detect the interest points, use Hessian matrix approximation. Build the integral images and the scale space of image. Interest point description and matching, descriptor describes the distribution of the intensity content.

3. Based on sum of Haar wavelet responses, construct a square region centered around the interest point.

4. Get rid of bad key points. Edges and low contrast regions are bad keypoints. Eliminating these makes the algorithm efficient and robust. A technique similar to the Harris Corner Detector is used here.

5. Assigning an orientation to the keypoints. An orientation is calculated for each key point. Any further calculations are done relative to this orientation. This effectively cancels out the effect of orientation, making it rotation invariant.

6. Generate SURF features. Finally, with scale and rotation invariance in place, one more representation is generated. This helps uniquely identify features.

5.2 SUPPORTING MODULES

• Object recognition module: It would allow to recognize objects from images recorded by the camera of mobile device. This algorithm should be insensitive to image registration parameter i.e. scale rotation and lightning condition.

• Text to Voice Conversion Module: We convert the text into a verbal formats the actual and important conversion take place in this module i.e. verbal message. The Images are stored in the form of text then it converted into appropriate verbal form.

5.3 DATABASE CREATION

For database creation we have could not use databases available as open source as our project required real-time images. Therefore, we captured real-time images of objects and then inserted them into our database of object for further training. For each object we have approx 15-20 images in our database. Each object image is captured in different angles for greater accuracy purpose. As our algorithm searches for match serially

increasing number of images increases computation time. Hence, we have used minimal images for more accuracy and less computation time.

The following feature table is used to store the features/descriptors of the image. When the data is trained, the feature of the each image is stored in the database in feature table. When the image is clicked from the application that image is send to the database where the features of the clicked image is calculated and it is matched with the features stored in the feature table.

Name	**Type**	**Null**	**Extra**
file_id	int(11)	No	auto_increment
file_path	varchar(100)	Yes	NULL
color_feature	Blob	Yes	NULL
surf_feature	Blob	Yes	NULL
object_name	varchar(200)	Yes	NULL

Table 5.3: Feature Table

CHAPTER 6

TESTING

6.1 Introduction

The Testing phase forms an important part of the software development life cycle. Any software product has to be tested thoroughly before it is delivered to the end customer. Well tested software with limited features is certainly better than the one having many features with only a few of them working. This document provides a general overview of the testing strategy adopted for testing our product.

6.1.1 Goals and objectives

The software testing involves verification and validation of the Software produced. The objectives of testing are:

- Testing is a process of executing of program with the intent of finding an error.
- A good test case is the one where there is high probability of finding error.
- A successful test is the one which uncovers an as-yet-undiscovered error.

The objective of testing is systematically uncovering different classes of errors and to do so with minimum amount of time and effort. The data collected in the test provides a good indication of the software reliability and some indication of software quality as a whole. The results of testing will not only help to know which parts of the system are working below average but also helps to make the system more user friendly. Testing is considered as an unavoidable part of any responsible effort to develop a software system.

6.1.2 Statement of scope

The scope of the testing process is limited to the determination of the consistency and integrity of object recognition code generation.

6.2 Test Plan

This section describes the overall testing strategy and the project management issues that are required to properly execute effective tests**.**

6.2.1 Testing Strategy

A good testing strategy is the one using which a lot of errors can be easily found. The testing which is to be carried out is divided into a number of modules for a proper judgment of the quality of the software. The testing strategy mainly carried out was Module Testing. Test cases are plotted considering the above categories and correct functionality of various parts of the code is ensured.

1. Unit Testing:

This involves testing of individual modules. Here we have tested individual modules written for various operations.

- Module for SURF
- Module for Object Recognition
- Module for Text to Speech Conversion

2. Integration Testing:

The system as a whole is tested here. The system is said to be operating correctly if it passes these tests. After the different modules have been individually tested, we have to integrate them and tackle the issues during the integration.

3. Load Testing:

This includes testing of an application under heavy loads, such as testing of a web site on a specific configuration under a range of loads such as number of users, number of transactions to determine at what point the system's response time degrades or fails. We test the application under a heavy load such as providing multiple objects to it and then test it.

6.3 TEST PROCEDURE

6.3.1 USER SIDE TESTING

1	Test case id	1
2	Module to be tested	Camera mode is ON
3	Assumption	-
4	Test Data	Android phone camera/Emulator Camera
5	Test step	i) Go To Application. ii)Start application.
6	Expected Result	Camera must be ON
7	Result	Successful

1	Test case id	2
2	Module to be tested	Image Capturing
3	Assumption	Application must be started.
4	Test Data	Camera mode must be ON.
5	Test step	i)Capture Image
6	Expected Result	Image will be capture by camera.
7	Result	Successful

1	**Test case id**	**3**
2	Module to be tested	Text to speech conversion module
3	Assumption	-
4	Test Data	i)Response from server. ii)Object in text format.
5	Test step	i)Provide object in text format to text to speech conversion module. ii)Run it
6	Expected Result	Voice is generated which specify the object.
7	Result	Successful

1	**Test case id**	**4**
2	Module to be tested	Get Detail
3	Assumption	-
4	Test Data	Recently Captured Image
5	Test step	i)Capture any image ii)Wait for details
6	Expected Result	Object recognizes. Voice message generated.
7	Result	Successful

6.3.2 SERVER SIDE TESTING

1	**Test case id**	**1**
2	Module to be tested	SURF Algorithm
3	Assumption	-
4	Test Data	i)Take a recently captured Image. ii)Image in database
5	Test step	Provide both images to SURF algorithm for comparison.
6	Expected Result	Provide matching points of both images.
7	Result	Successful

1	**Test case id**	**2**
2	Module to be tested	Object Recognition Module.
3	Assumption	-
4	Test Data	Result of SURF algorithm.
5	Test step	i)Take result of SURF algorithm. i.e. matching points. ii)Run object recognition module.
6	Expected Result	Object recognition is done.
7	Result	Successful

CHAPTER 7

RESULT AND EVALUATION

7.1 SNAPSHOTS OF APPLICATION

Figure 7.1.1: Capturing image

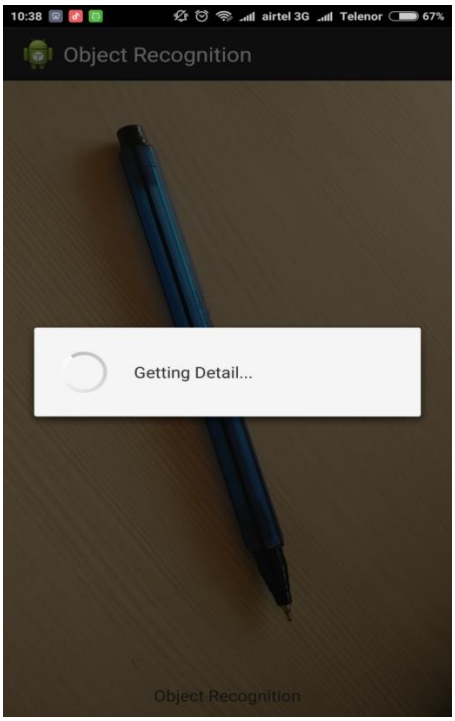

Figure 7.1.2: Getting details from database

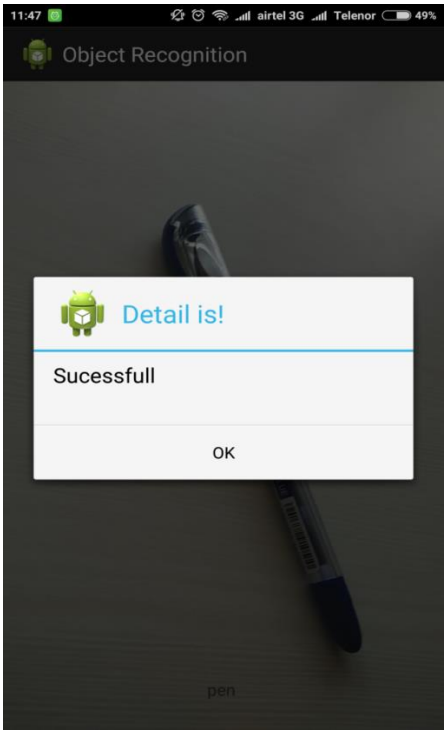

Fig 7.1.3: Successfully recognized object

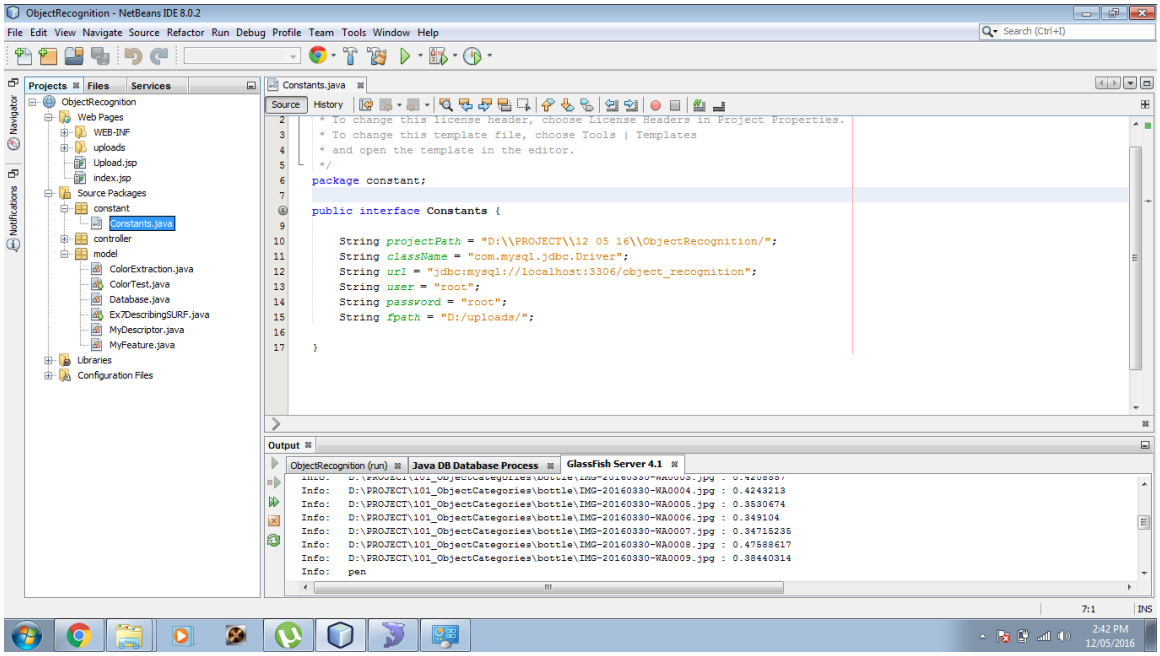

Figure 7.1.4 Server side result

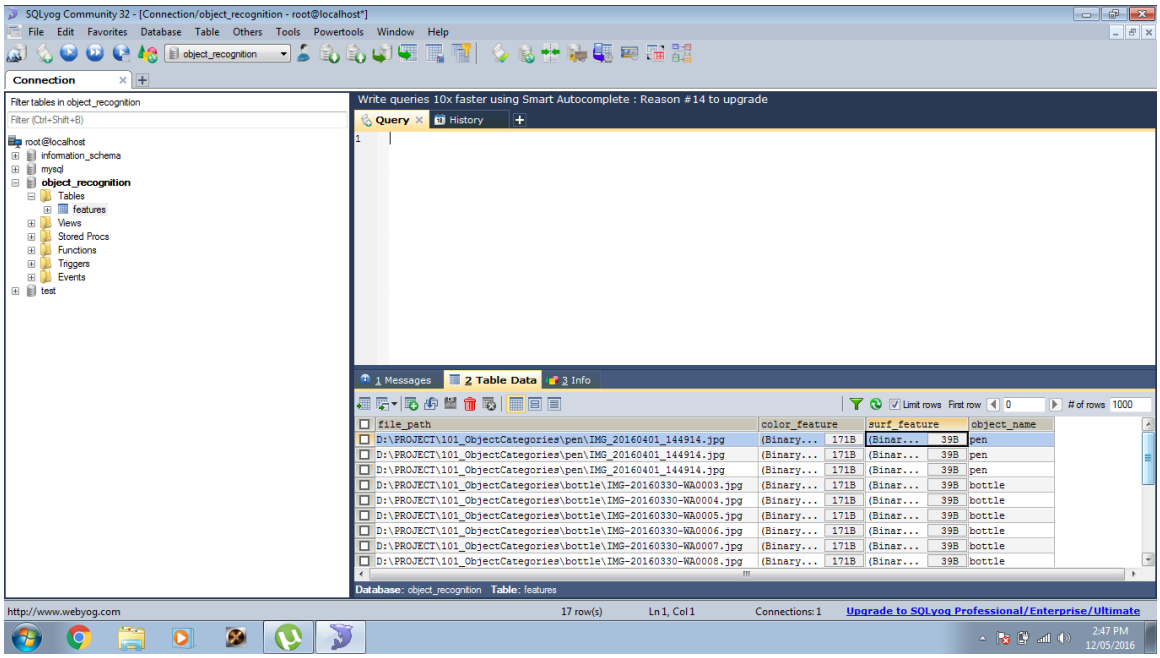

Figure 7.1.5: Database of objects

7.2 Evaluation

The following presents real world results. We have calculated the precision and accuracy of our project.

For accuracy and precision we have captured 25 different images of an object from different angles. Below table shows the outputs for the given inputs and their precision.

Input	**Output**
Pen	Pen
Pen	Pen
Pen	Pen
Pen	Pen
Pen	Pen
Pen	Pen
Pen	Pen
Pen	Pen
Pen	Pen
Pen	Pen
Pen	Pen
Pen	Bottle
Pen	No output
Pen	Pen
Pen	Pen
Pen	Pen
Pen	Pen
Pen	Pen
Pen	Pen
Pen	Pen
Pen	Pen

Table 7.2: Testing objects

Precision is defined as in equation,

$$Precision = \frac{Number\ of\ objects\ correctly\ detected}{Number\ of\ objects\ detected} = \frac{No - FA}{No}$$

where,

No is the Number of objects detected

FA is the number of false alarms i.e. the number of unmatched objects in output

$$\text{Precision} = \frac{24-1}{24} = \frac{23}{24} = 0.958$$

Precision of this application is 0.958

Accuracy rate is defined as in equation,

$$Accuracy\ rate = \frac{Number\ of\ objects\ correctly\ detected}{Number\ of\ times\ the\ algorithm\ was\ run} \times 100$$

$$Accuracy\ rate = \frac{24}{25} \times 100 = 96\%$$

Accuracy of this application to recognize the objects is 96 %

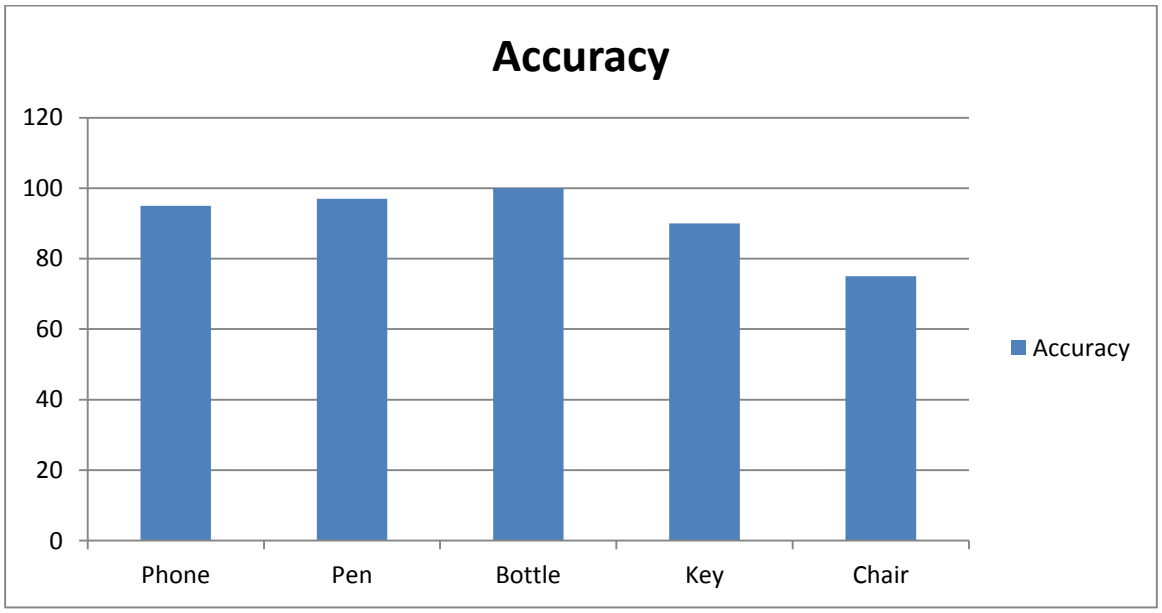

Figure 7.2(a): Accuracy graph with respect to particular objects

The accuracy for every object was checked during testing. The above graph shows the accuracy of few objects.

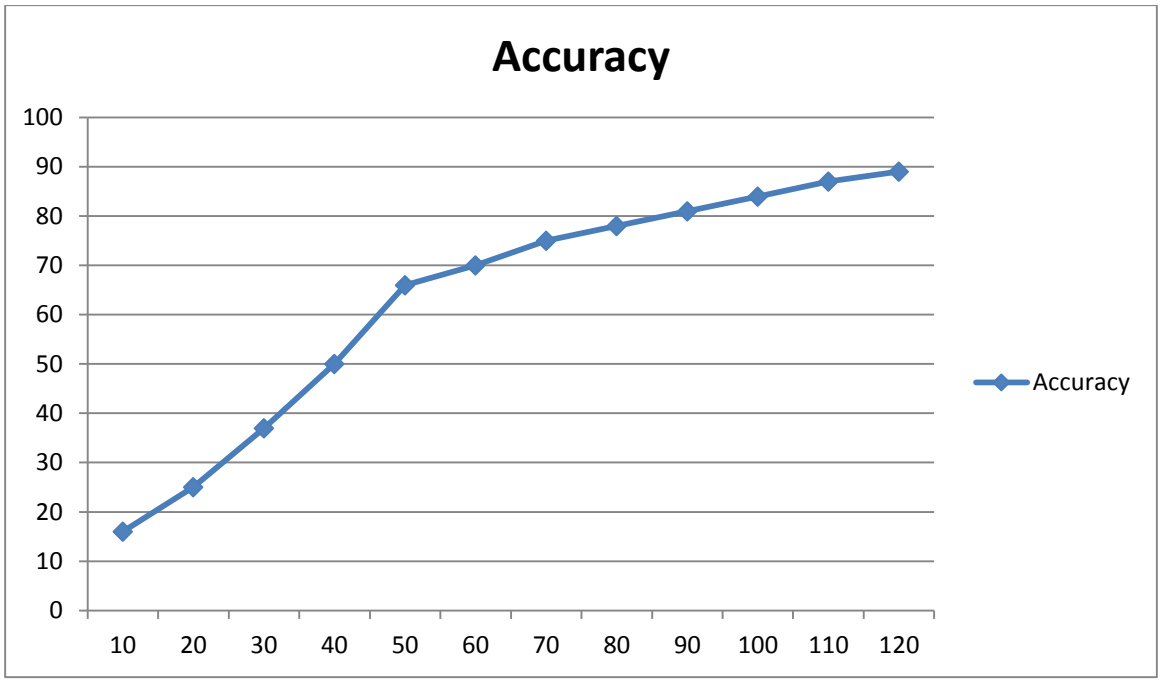

Figure 7.2(b): Accuracy graph with respect number of images in database

The above graph shows the accuracy of the application with respect to number of images in the database. As we increase the number of images for an object in the database the accuracy for recognition of an object increases.

CHAPTER 7

PLANNING & SCHEDULING

August: Requirement gathering phase, wherein we would be discussing about the idea of project selection, collection of related IEEE papers, analyze and study the existing system.

September: Planning phase. Here we would be doing discussion of topic with faculty and take suggestions, presentation of abstract and related papers, and collection of IEEE paper of proposed work.

October: Design phase. Here we would be doing the high level design, class diagrams, Sequence Diagram.

November-January: Implementation phase. We would be coding of our design. Code review would be done along with unit testing and validation.

February: Testing. Quality analysis testing would be done which includes sanity testing, integration testing and performance testing

Mid-March: Fully completed project.

CHAPTER 8

CONCLUSION AND FUTURE SCOPE

Now-a-days visually impaired people face diverse kinds of challenges in their daily activities. This system helps them in recognizing the household objects which are used in daily activities. The system communicates with the user by pre-recorded voice message which is given as an output to the recognized object. Thus this system helps the visually impaired people to perform their daily activities independently.

8.1 FUTURE SCOPE

Presently we have limited the scope of our project to recognize only household objects. Here we are providing the functionality that blind person can identify the objects which are generally used in their day to day life.

Hence future work will set up a system with same features but to include worldwide objects so that they can recognize any objects and also we will try to provide functionality to recognize all possible objects in same image.

REFRENCES

[1] D. Lowe (1999) Object Recognition from Local Scale-Invariant Features. International Conference on Computer Vision.

[2] D. Lowe (2004) Distinctive image features from scale-invariant key points. International Journal of Computer Vision.

[3] Gill J. (2008), Assistive devices for people with visual impairments by Helal S., Mokhtari M. and Abdulrazak B. (ed) The engineering handbook of smart technology for aging, disability, and independence, J Wiley and Sons, Inc, Hoboken, New Jersey, pp. 163-190.

[4] Matusiak K., Skulimowski P.(2012),Comparison of Key Point Detectors in SIFT Implementation for Mobile Devices, ICCVG Lecture Notes in Computer Science.

[5] Rosten E., Drummond T. (2006), Machine learning for high-speed corner detection. European Conference on Computer Vision.