Smartphones as Visual Prosthesis

Alireza Shoaei Shirehjini

S320783

Abstract

Smartphone have become far more widely used during last decade. Ubiquitous daily computation power along with a rich set of sensors have added value to our life significantly. It has provided the opportunity to improve the quality of life for visually impaired people. The camera embedded in these mobile devices has enabled researchers to explore different ways to augment the vision sense as a solution for visually impaired people. Indoor and outdoor navigation systems that provide tactile or audio feedback and magnification applications have used the camera in some extent. People with more moderate vision loss tend to avoid technologies that use tactile and audio feedback. Therefore, those techniques that enhance vision can highly benefit them.

This study is research and development of a magnifier prototype developed on smartphone platform for visually impaired people. This magnifier is operated by hand gestures. Our prototype offers a way to easily scroll and pan through the magnified content. The algorithm will be technically evaluated. In the last phase an experiment will be conducted to compare the prototype with normal magnifier software that only uses the device's camera.

Contents

1. Introduction

1.1: Research Questions

The following list includes the research questions:

- How can smartphones meet the needs of visually impaired users?
- What is the required image quality, and how to ensure it?
- How the user is going to interact with the prototype system?

1.2: Project Objectives

The following are the project objectives:

- Finding a solution to enhance the vision of those who belong to the low vision part of the vision continuum.
- Improve the quality of the field of view in magnification application running on smartphones
- Providing easy horizontal and vertical scroll experience
- Allowing for efficient and effective enlargement of objects located in distance.

1.3: Problem Statement

World Health Organization (WHO) estimated that 253 million people live with vision impairment: 36 million are blind and 217 million have moderate to severe vision impairment. Vision impairment is scattered across a continuum from complete vision loss to more moderate impairments. The International Classification of Diseases represented by WHO defines the levels of vision as following:

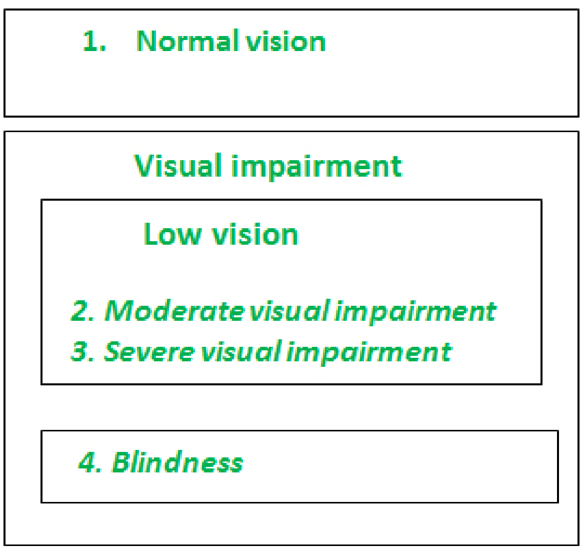

Figure 1.1 Low vision levels (WHO, 2015)

Different solutions have been proposed to address the issues of visually impaired. Assistive technologies that provide auditory or haptic feedback can significantly improve the lives of blind people, whereas magnifiers, font and contrast adjustments are more convenient for partially sighted people. Digital magnifiers and optical devices are the most commonly used tools among low vision users. The current rehabilitation services tend to use a multidisciplinary approach that combines the optical and digital aids. Inevitably, this makes these assistive technologies expensive, acting as a hindrance. Moreover, many of these vision aids often stigmatize the visually impaired and encourage the fear of having a stereotype applied to them.

Smartphones can considerably reduce the dismay of being stand out as a visually impaired in a public setting, letting their low-vision define them. Smartphones offer a rich potential to act as a visual prosthesis for visually impaired users. Actually, more importantly is that smartphones are used by everyone. They provide a much more suitable platform because they have become pervasive at an affordable cost and include lots of features as well as enough computational power. Features included in these phones allow for magnification, photo capture, GPS with voice navigation, illumination, and more.

Crossland et al. (2014) described the use of consumer devices such as smartphones, tablets, computers and e-readers by people with vision impairment by running an online survey. They experienced that accessing books, internet and applications are the primary uses of these devices by visually impaired. In addition, they have observed that people may use these electronic devices as vision aids. Using camera LED flash to see in dark environments or taking photographs to further enlarge the image are among such uses. 107 of the respondents have used smartphones and almost more than half of the smartphone users indicated that they have used the camera to see more easily.

While a lot of research have been done to cover the blindness level of vision continuum shown in figure 1, like object detection and recognition, navigation assistants or obstacle detection, only few studies try to address the problems of low vision. This study tries to find a solution for a common issue among low vision by utilizing smartphone's camera as a magnifier device. This will provide the opportunity of having a portable magnifier with no extra cost. The study involves the research and development of a smartphone based magnifier for visually impaired people which is operated by hand gestures. Document navigation (scrolling and panning) and selective magnification of distant objects in outdoor places like public transportation are the objectives of this project.

Low vision users rely on magnified views to navigate through contents effectively and efficiently. But, they face a serious loss of content. It affects the user experience and performance directly. Finding the important parts of the text such as beginnings of each paragraph is not an easy task in magnified views (Gowases et al, 2011).

Printed text and documents are the primary source of knowledge in real world which only people with clear vision are able to fully benefit from. Visually impaired users are using magnifiers to be able to read through these resources. There are various types of magnifiers categorized by size or technology. Unfortunately, each one may impose some restrictions on users. In the following section an alternative solution is proposed.

1.4: Proposed Solution

Smartphones seem to be a good option for a low cost and portable magnifier. Actually, a smartphone can be quite expensive, but the fact that the users have already invested in and own smartphones makes this an attractive option. The embedded power camera in smartphone has a great potential to help visually impaired people. However image jitter of hand held phone camera can prevent efficiency either for reading text or clear vision of distant objects. A stabilized video stream, captured by the camera, can be a reliable source for efficient magnification. Users may have problems while scrolling or panning the enlarged text as it is necessary to find the new portion of the text after reading the current lines. Usually, users have to move the device either horizontally or vertically to continue reading. It may cause confusion and fatigue. The solution is to update the words or lines being read in both directions by tracking the user's fingertip.

In addition, selective magnification can be another application of a magnifier. It is difficult to properly aim at the desired object while magnifying. It requires to aim towards the object with an accurate angel or otherwise, it entails keep moving camera to find the object. Hence, detecting the pinch gesture performed in front of the camera can enable the user to instantly enlarge the targeted object.

2. Background

2.1 : Magnifier Technology

Reduced visual acuity and contrast sensitivity associated with visual field loss are the most important issues that visually impaired users have to deal with in daily life. Reduced mobility, difficulty in reading and writing not being able to recognize faces are different disabilities caused by these impairments. Elderly low visions prioritize obtaining help for reading and completing daily activities as their primary concerns. Whereas, their hobbies such as watching television, mobility and independent traveling are considered as secondary objectives. Conventional Low Vision Aids (LAVs) such as optical device can help low vision users to compensate their reduced visual acuity. However, these optical devices are often task specific, expensive and may require the visually impaired users to have multiple aids for different requirements. The need for integrated solutions that address a wide range of user requirements can be well met with electronic magnification devices.

Magnifiers should equip their users to maintain their independence by enabling them to read and write. Technology is advancing at a rapid pace bringing more complexity into our lives. The needs and requirements of people with vision impairment have also been changing. It mandates that in designing visual aids different aspects such as performance, comfort rating and user preferences should be considered carefully.

There are a variety of different magnifiers based on the tasks they perform. The desktop magnifier, also known as closed circuit television systems (CCTVs) consist of a camera that displays the magnified image on a monitor. myReader (Harrison, 2004) is an enhanced CCTV type of magnifier in an attempt to eventually replace the earlier versions of CCTV magnifiers. It is designed in such a way to improve the usability. The redesigned machine surpasses the previous CCTVs by rearranging the text into different display modes to allow for different visual impairments and it provides a choice for different text layouts. Other functionalities like photo storage and increased digital zoom capability are some it distinctive features.

Hand-held portable magnification devices can be used for reading labels or restaurant menus. Some of these devices also include optical character recognition and text output to read the enlarged text aloud while it is being displayed on the screen. Depending upon user's expectations each type may have advantages or disadvantages. CCTVs are expensive and not portable, but they are more accurate and allow user a faster use with less fatigue and strain. On the other hand, portable handheld devices are less expensive and provide enhanced contrast and improved brightness which are important while on the go. They provide smaller field of view compared to CCTVs. Also, it may cause fatigue while reading a document which requires the user to hold the device steady for quite a while.

Technologies such as smartphones or wearables enable mobility. The attempts to provide mobility for magnifier users goes back to a time long before the creation of handheld portable devices. The earliest head mounted low vision imaging system (LVIS) was developed at Johns Hopkins University. It is a battery operated system which is equipped with two auto-focus cameras. It provides contrast enhancement and variable magnification. The magnification happens by means of aspheric lenses that project the magnified image on the mirrors located in front of the eyes. The device is also equipped with two monochrome cameras which provide unmagnified view field. User can switch between the magnified and unmagnified views by controls on the control unit which is worn as a belt. User has control over magnification, contrast and brightness. In addition it is possible to reverse any changes that are made by the user, for example image contrast can be reversed.

Figure 2.1 Head mounted LVIS device (Harper et al, 1999)

Figure 2.1 shows a user who has worn the head mounted LVIS proposed by Harper et al. (1999). The two monochrome cameras are depicted along with the lenses that provide magnification. By using the control unit, user can toggle between two views. It was one the earliest attempt that was made to solve the problems encountered by visually impaired. It was proceeded by further attempts to quickly build next generations. Technology has been evolving and that made it possible to create more operable and compact systems. With the advent of CCTVs and potable magnification devices, all the functionalities of that LVIS was available through a much more convenient platform.

Currently, smartphones are being considered as an alternative for both CCTVs and portable magnifiers as they provide much better resolution than before. Many third party applications have been developed to turn phone's camera into portable magnifier such as TapTap See, Optelec Magnifier and a large number of other magnification apps available on google play store and apple app store.

A technology which low vision persons can benefit from is finger reader. It is wearable device that can be mounted on user's finger to either project the text being scanned on a screen or provide audio as well as haptic feedback. Research has been done to use a finger mounted camera combined with Microsoft HoloLens technology to create an augmented reality magnifier for visually impaired. It is not fully wearable, especially in public places and besides HoloLens is a quite expensive technology (Stearns et al, 2017).

Figure 2.2 Reading text by HoloLens (Stearns et al, 2017)

MIT media laboratory also developed a prototype magnifier which utilizes a finger mounted camera to provide audio feedback. The system will read aloud the text at user's fingertip. It also warns the users, whenever their fingers starts drifting away from the line. People with more moderate vision loss tend to use their vision instead of audio or tactile feedback and may find themselves reluctant to use this solution.

Figure 2.3 Finger-mounted reading device for the blind (MIT News Office, 2015)

Pucihar&Coulton. (2013) proposed a system to enhance the virtual transparency by designing a digital magnifying glass implemented on commercially available mobile phones. The quality of the transparency is affected by image blur, unfocused camera, lightning condition and camera resolution. Image blur happens for objects which are at a distance smaller than hyper-focal distance. Hyper-focal distance is the distance between a camera lens and the closest object which is in focus when the lens is focused at infinity. This a handheld augmented reality system that replaces parts of the image with digital content.

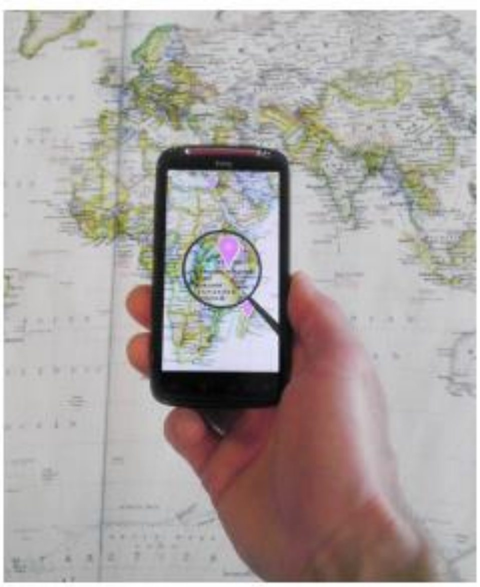

Figure 2.4 Augmented reality magnifying glass (Pucihar& Coulton 2013)

Figure 2.4 shows the magnifying glass prototype augmenting the pin icon to mark the capital cities on a printed map.

The limitation of this prototype is that user needs to change between the active area of the display which indeed is the magnified view field and the global real-state of screen. Aguilar& Castet. (2017) proposed a system that defines a region of interest (RIO) based on gaze direction. User can alternate between above mentioned views or even magnify part of the screen that he/she wants.

2.2: Image Stabilization

Magnification is happening by projecting the video stream captured by camera on a screen. The quality of captured scene may be affected by factors such as unfocused camera, poor lightening condition and handshake effect, leading to reduced quality in magnified view. Image jitter is inevitable in an enlarged image or video stream due to handshake. This can decrease the quality of the magnified text or distant object to high degree and make the magnification task cumbersome. Handshaking problem can further result in a reduced field of view letting the targeted object frequently slip out of the screen even with slight hand movements. Image stabilization techniques can eliminate the unlikely results of hand shake event and provide a steady and clear view. Image stabilization are divided into two categories. In Optical image stabilization (OIS) a part of the optical system is moved in response to the unintentional vibrations or some extra hardware is used to remove the jitter. These techniques are usually applicable for photography and filming equipment. OIS techniques increase both the size and cost of the system. For small cameras software based image stabilization techniques are more appropriate. The main difference between optical and digital image stabilization is the place that compensation occurs. In digital methods changes will be applied after the frame is recorded while in optical methods the modification of optical system such as prism or lenses actually enables the stabilization.

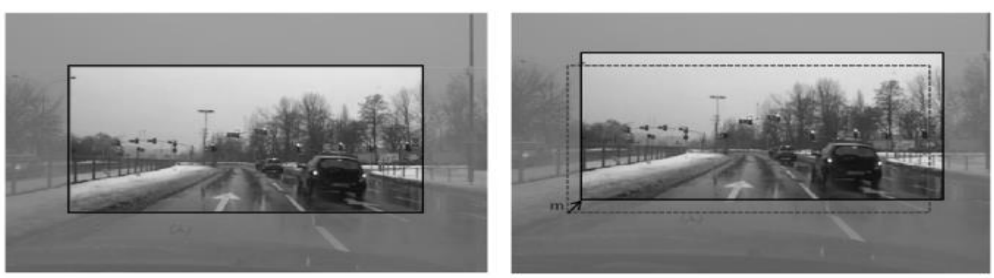

Figure 2.5 The basic idea of digital image stabilization (Lamza& Wrobel, 2012)

Figure 2.5 is a simple illustration of digital image stabilization by moving a window in inside the captured video. The margin of the movement is determined according to the maximum amplitude of the vibrations.

Digital image stabilization (DIS) techniques follow two main approaches: 2D and 3D methods (Li, & Zhao, 2014). 3D methods require heavy calculation, hence, they are not practical for real-time application. Motion estimation and motion compensation are the main steps in 2D DIS algorithms. Motion estimation basically means extracting global motion vector of neighbor frames. Motion compensation is responsible for removing jitter from the scene. Block matching algorithms are widely used in motion estimation phase, however they are computationally expensive and researchers have been trying to optimize these algorithms for real-time and mobile platform. Optical flow method and feature point matching are also used for motion estimation but not suitable for smartphones because they are computationally expensive (Xie et al, 2016).

Smartphone Internal sensors are the other option for implementing image stabilization. One application of using internal sensors in image stabilization is removing rolling-shutter distortion. It is done by analyzing the data that is received from accelerometer and gyroscope sensors.

Figure 2.6 Distortion effect and rectified frame using accelerometer and gyroscope (Hanning, et al, 2011)

Figure 2.6 is an example of this method. Unlike global shutter cameras where frames are captured at once, rolling shutter cameras record the frame line by line starting from top to bottom. As a result vertical lines will appear with unlikely gradient. This distortion can be compensated by sensing position and/ or orientation of the device (Hanning et al, 2011). Sensor based techniques are not effective and efficient in mobile usage scenarios where the frame is not only being affected by device's orientation.

Dynamic enlargement of the text size and magnifying specific parts of the content were studied as potential solutions to deal with handshake event on mobile device's screen. An anti-shake mechanism is proposed by Hsiao& Wei. (2017). the primary idea of this methods is motion compensation which is a technique for predicting a frame in a video based on previous frames. Figure 6 describes how the elements in this method are configured:

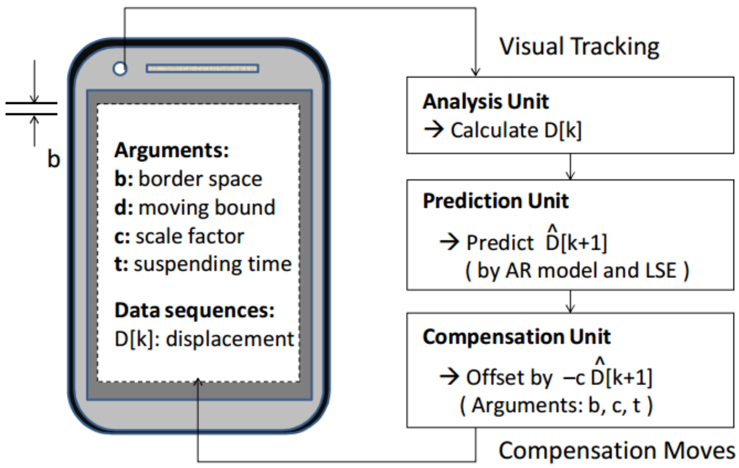

Figure 2.7 Appearance and infrastructure of the anti-shake mechanism (Hsiao& Wei, 2017)

This method shrinks the display area to create a buffer space, so the active display area can be adjusted to compensate the vibrations. This is possible by moving the active display area in opposite direction when the amount of displacement is calculated.

One technique for real-time video stabilization is computing the motion in frames captured by the camera (Li et al, 2013). It is done by motion estimation between captured frames and a frame as reference. This method was tested on an iPhone smartphone equipped with a telescope for achieving deeper zoom levels. The visually impaired performance in detecting letters in distance improved and this method was more effective than the one implemented using the phone's gyro-sensor.

2.3: Hand gestures control for mobile device

The advances in technology enabled powerful displays. Currently, mobiles rely on touch as an input method. The touchscreen method is generally faster compared to indirect methods such as physical buttons or knobs due to the reduction of the time needed for translation of physical distant moved to the virtual distant moved on the screen (Lee& Sanford, 2012). However, touchscreen may pose some limitation preventing efficient interactions with smartphones. Particularly, mobile phones and wearables such as smartwatches or fitness trackers continue to become more compact which makes the touch area increasingly limited. Evidently, occlusion and fat finger problem are likely to happen only to reduce the actual touchscreen real-estate. In addition consumers expect multi-touch functionality that makes interaction with the device even more challenging. Both the industry and researchers have been trying to incorporate new techniques to extend the input area of these devices. HCI researchers have attempted to explore the deferent possibilities to enhance the interaction modalities.

Several studies investigated the finger gestures for back of device interaction. CamTrackPoint uses a 3D printed ring that is mounted on camera's bezel, then by tracking the lights passed through the finger, it is able to sense the finger gesture (Yamada et al, 2017). LensGesture is another back-of-device input method which is able to detect the full or partial occlusion as well as swiping of finger on the camera. It is done by analyzing the image sequence captured by camera and measuring the redness level of skin caused by light passing through the skin (Xiao et al, 2013). They try to address common problems which may occur while using smartphones on the go. Users need to hold the phone with one hand, leaving the thumb finger as the only interacting channel with the device. Fat finger problem, occlusion problem and reachability problem are likely to happen. These studies show that the rear camera can be used as an input channel which is able to capture hand gestures and translate them into meaningful actions on the device.

SideSight (Butler et al, 2008) supports virtual multi-touch hand gestures by applying optical sensors implanted around the body of small mobile device such as cellphones, media players and PDAs. These are infra-red (IR) proximity sensor embedded along each edge of the device which are able to sense the presence and position of fingers. However, it is required to place the device onto a flat surface. Thus the whole area surrounding the device will become a multi-touch track pad.

Figure 2.8 SideSigth Prototype (Butler et al, 2008)

Figure 2.8 shows using the prototype of SideSight for rotating a picture (Kratz& Rohs, 2009).

HoverFlow explores the feasibility of around-device interaction by using IR sensors. This interaction method is suitable for small wearable devise such as wrist watches and wireless headsets. Figure 8 shows how one can interact with small mobile devices by performing coarse hand gestures around a devices which is enabled by HoverFlow (Kratz& Rohs, 2009).

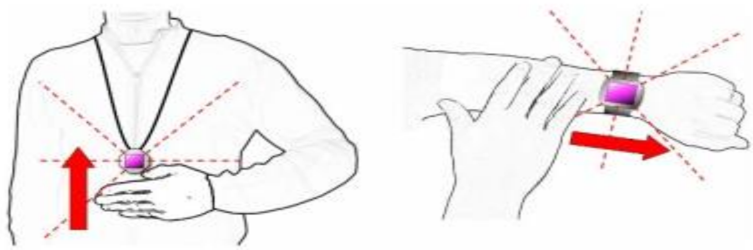

Figure 2.9 Interacting with small mobile devices by performing coarse hand gestures (Ketabdar et al, 2010)

MagiTac (Ketabdar et al, 2010) is another effort made to efficiently use the 3D around-device area to interact with mobile devices without occlusion. It utilizes the internal compass sensor of smartphones. A properly shaped magnet is worn, then, the changes in magnetic field around the device when gestures are being made, will be captured by compass sensor. The temporal pattern of these gestures will be the basis for calling different commands on the mobile device.

SideSwipe (Zhao et al, 2014) detects hand gestures in areas around and above cellphones by leveraging GSM signals. It does need external signal source or a costume transmitter, however, some modifications to the mobile device are required.

Wearable devices such as smartwatches and fitness trackers, also, have been a target for HCI researchers. Most of these devices include a built-in inertial measurement unit (IMU). IMU is an electronic device that measures linear and angular motions usually with triad of gyroscopes and triad of accelerometers. Through, analyzing the data acquired via IMU (Yeo et al, 2016) could determine the actions performed by the user around the device.

Watch Commander (Kubo et al, 2016), a bezel to bezel swipe gesture, is another proposed solution to provide barrier free interaction with smartwatches. This method is only introduced for rectangular smartwatches. It is a single-finger gesture that starts and ends at a bezel. Users can invoke different commands such as starting timer or messenger.

All the techniques described above impose some limitations on users. They require to equip either the mobiles, the environment or the user. Users need to place the device on a flat surface, wear a ring or modify the device in order to perform the hand gestures. In addition, it requires software development efforts to enable a phone to leverage the IMU functionality. Hence, other approaches which take advantage of the embedded camera of the device seem more efficient and effective. Herein, the device is being held with one hand while the non-involved hand makes the gesture in front of the camera. Afterwards, the smartphone detects the hand movements by using computer vision techniques, translating them into operable commands.

Due to the advances in graphic hardware for mobile devices, rotation of 3D scenes are now possible. This capability is especially useful in gaming and allows CAD tools implementation for smartphones. This is usually done via a virtual track ball that maps 2D input gestures to 3D objects rotation. PalmSpace is a technique designed specifically for addressing the task of 3D object rotation by phone's camera. The camera detects hand motions made in proximity of the device (Kratz et al, 2012).

Figure 2.10 Using the pose of the flat hand behind the device (Kratz et al, 2012).

Figure 2.10 shows how the changes in flat hand posture triggers the 3D object rotation.

Reflective glasses or visors such as ski goggles, helmet visors, sun glasses or diving masks can enable gesture communication with smartphones. In this scenario, the field of view of mobile's front facing camera is extend by the reflection on appearing on the reflective surfaces.

Figure 2.11 GlassHands extends the input space around mobile devices (Grubert et al, 2016).

Figure 2.11 depicts the idea of GlassHand. The camera's field view is expanded to enable the area around the device as an input field. Both the device's screen and the objects around it are reflected on the sunglass. Now, user is able to simultaneously make gestures on the screen as well as the area beside the phone (Grubert et al, 2016). Panning and scrolling of printed text are potential use cases of this approach. However, this method confines the user to wear some apparels which may not be comfortable for some them.

Song et al. (2014) use a machine learning algorithm to outstretch the interaction space around the smartphones by using the built-in camera of smartphones. Their algorithm is optimized for smartphones and it is able to classify hand states and salient features like fingertips.

Figure 2.12 Usage scenarios for in-air hand gesture machine learning based algorithm (Song et al, 2014)

Figure 2.12 shows how this algorithm complements the touchscreen. There are some usage scenarios that are depicted. The part A shows panning and zooming a document. This can be further expanded to perform the same task on printed texts and document. B shows the bimanual map browsing. User can use the screen to scroll the map while magnifying it with in-air pinch gesture. D displays the tool parametrization in a drawing app by rotate gesture. Finally, in part D while the player uses the screen to scroll the game scene, shooting is being done by the free hand in front of the camera.

Similar to the previously mentioned study, Behand is another in-air gesture interface for mobile phones introduced by Caballero et al (2010). It tries to facilitate the interaction with mobile devices by allowing them to manipulate three-dimensional virtual objects in their phones by hand gestures. Furthermore, during a user study participants descried the proposed concept useful, and found it innovative, fun and acceptable. The interaction model of Behand originates from the idea that real objects have in-hand 3D interaction modality. In their video that shows how the prototype interacts with real objects, they have depicted usage scenarios in which user can augment holding different objects and trigger a function on the real object. Turning on an electric kettle is one the examples that has been shown in their video. It is called direct manipulation of objects which actually changes the state of an object and is a key feature of their proposed system. Auditory feedback is also included to enhance user perception.

Recognizing sign language gestures has also been an interesting topic among researchers. Hartanto& Kartikasari (2016) developed a prototype to detect static Indonesian sign language for android smartphone. They segmented the image based on skin color detection and a specific color range to remove the background from image. This allowed them to detect the hands. They outlined their proposed system as following:

- Skin segmentation and contour detection
- Hand feature extraction

- Classification using back Propagation Neural Network

Features of the hand are some characteristics that will shape the hand gesture such as fingertips, distance from fingertips to the palm center or angle between two adjacent fingers. These features will be classified and recognized by neural network classifier. Because resources are limited in mobile platform, the classification algorithm should be fast and as light weight as possible. Back propagation neural network was chosen and they were able to reduce the time for training this classifier significantly by weighting the hand features extracted from images.

Lahiani& Elleuch (2015) proposed a system for hand gesture recognition. Their proposed method consists of four steps:

- Hand segmentation
- Smoothing
- Feature extraction
- Classification

The primary idea is to enable the smartphone to handle the all the steps for performing the training process without the need for connecting to an external computer over an internet connection. For hand segmentation they have used color-based segmentation. Furthermore, to remove the noise, some filters are applied on the image to create a new image by modifying and replacing the pixel values of the original picture. This is done to actually smooth the image while preserving the edges. This image filter is called median filter. For classification of hand features an optimized support vector machine technique (SVM) for mobile devices is used.

In a follow up work (Lahiani et al, 2016), they tried to improve the hand recognition step in their first research paper mentioned above. Viola-Jones is an object detection classifier that uses the Haar-like features of the objects for classification. This algorithm is computationally expensive for mobile platforms. Hence, they used a variant that uses Local Binary Pattern (LBP) features.

2.4: Scrolling and panning

This section deals with the necessity of reducing the efforts required for scrolling and panning. These operation are sometimes referred to as vertical and horizontal scrolling in the literature. To provide a better usability and enhanced user experience is often advised by web design professionals to avoid horizontal scrolling specially for small screen and handheld devices. It is achieved by responsive designing of web pages or other solutions proposed in the literature.

In our system users need to adjust the view field whenever they reach the boundaries of screen either vertically or horizontally. In CCTVs scrolling and panning happens automatically by electronic arms, while for handheld magnifiers and magnifier applications that run on smartphones users need to move the device over the text to update the view. It will result in fatigue and distress. Resembling the pan action, horizontal scroll bars are not considered a good practice in designing web pages. Some users find them unusual, or the designers express that a valuable space is taken and hence become useless by horizontal bars at the bottom of the page. WCAG 2.0 success criteria 1.4.8 deals with visual presentation to ensure that the layout of visually rendered text does not interfere with its readability. The ability of adjusting the text color, size and the background contrast will improve the comprehension of low vision users in reading text. It is a good practice to present the text to people with vision disabilities in narrow blocks of text as long lines can create a huge barrier. Also line spacing allows them to track the text and find the end of each paragraph.

This success criteria also includes provisions for provisions for text resizing to ensure that users are able to adjust the scale of the rendered text without any need for vertical or horizontal scrolling. The text should be rendered in such a way that does not require user to scroll horizontally back and forth to read a line of text. To properly render the text and relieve the user of adjusting the view, it should be mapped to a virtual layout on the screen. It can be achieved by magic lens paradigm. Magic lens filters are new interface tools. They create an arbitrary shaped region over an object and combine it with an operator that updates this view. The advantage of magic lens is that they only affect the part of the object presented in the arbitrary shaped view and the remaining space of the screen will remain intact. One their intuitive applications is augmented reality showcasing of products in exhibitions and other points of sale.

The advent of smartphones and tablet requires developers to create robust views for different screen sizes. Responsive design has been a promising solution that matches the design to the device specific configuration. It reduces the effort needed for resizing, scrolling and panning while viewing a web page.

The Horizontal scrolling problem stems from the fact that small physical screens represent a much larger virtual space. It is only possible to see only a small portion of a virtual workspace at different time instances.

Several studies have attempted to simplify the panning task. Virtual scroll ring (Moscovich& Hughes, 2004) is a software simulation technique that maps the circular motions of user’s finger into vertical scroll movements. This method is applicable for portable devices like laptops, which lack the physical mouse hardware. The users makes the circular motions on the touchpad trough their fingertip. It makes the scrolling task simple because people prefer a continuous motion over the method of hardware scroll ring.

Alli& Feng (2013) propose a mechanism for auto scrolling the search results in mobile devices. They argue that the web pages are mostly designed for desktop environments, hence, viewing them on smartphones with much more limited screen real-state requires constant and repeated horizontal and vertical scrolling which results in fatigue subsequently. One possible solution is to automate this cumbersome task. They have implemented their solution based on manipulating the term frequency and inverse document frequency of the document object model of web pages. The results of their study indicate that users are able to significantly save time during searching and filtering the needed information

Map navigation with mobile devices necessitates the users to execute a vast number of panning. The performance of navigating maps is negatively affected by small screen sizes and detailed information such as ATM locations or a specific grocery store are only visible by high levels of zooming. Users are required to scroll and pan extensively to be able to browse the map. Rohs& Essi (2006) explored methods for map navigation and compared their performance. Their results demonstrate that moving a handheld device over a static map performs better than having a virtual map in a static device.

Similar to our work, Rohs et al (2007) attempt to use the 3D space to externalize user interface of handheld mobile devices. In this way smartphone will act as a window to interact with objects which have fixed physical position and orientation. In this work, they tested halo, halo&zomm and pan navigation methods on spatially-aware handheld display. The task was reading a printed text. After their user study, participants described the pan method as being the slowest one among the three methods.

A technique for viewing large documents on small screen mobile devices which presents three degree-of-freedom is proposed by Mooser et al (2008). However, they use the device motion to allow user to scroll through large portions of the document, it is their primary concerns to minimize the hand movements for scrolling.

3. Method

For implementing the prototype, The OpenCV library for real time image processing will be used. In the following section, different components of our prototype are discussed. A brief overview of evaluation approach is also presented. Furthermore, it will be concluded by exploring the areas in which visually impaired people's life can be empowered highly from image processing.

The prototype magnifier should provide the user with a steady video frame as an input. User interacts with the software by hand gestures. Our prototype magnifier will consist of the following modules:

1. Image stabilization module
2. Hand gesture recognition module
3. Android client application module

3.1: Image Stabilization
The quality of video frames play an important role in the usability of this system. Image blur due to keeping the camera at a distance less than hyper-focal distance may happen. Jitter in the video stream is also inevitable because of handshake effect. It will be escalated as the zooming level get deeper.

3.2: Hand gesture recognition
In background part, a various methods for detecting hand gestures on mobile devices were discussed. Hand gesture detection process consist of four steps:

1. Hand detection
2. Hand segmentation
3. Extracting hand features
4. Classifying the features using a machine learning algorithm

The work by Song et al (2014) is closely relevant to this study. They introduced in-air hand gestures for mobile devices and described the system usage in different scenario such as gaming, navigating maps, reading digital content or controlling color spectrum in image editing application. However, their system only interacts with the device itself whereas the suggested system in this work tries to build an interface with real objects.

Most of the approaches in implementing in-air communication with mobile devices use skin color detection for detecting hands in a given image or video frame. Skin color detection uses a color range to determine the existence of hand. Due to varied skin color among people or different lightning conditions this approach may not be accurate for in-the-wild scenarios. So, other methods such as optimized Voila-Jones algorithm (Lahiani et al, 2016) are potentially more applicable for this task.

Hand segmentation will be done by drawing contours around the boundaries of detected hand and applying thresholding mechanisms. The feature extraction step is very important because it produces the inputs for machine learning algorithm for classifying operation. Features like fingertips and center of the palm are to be extracted in this step. There are several machine learning algorithm that are used for classification purposes such as support vector machine, decision tree or random forest. I will be using random forest algorithm for feature classification task. Random forest is a collection of decision trees that each one classifies the objects based on some attributes by voting on them. The classification which has the highest vote is chosen.

3.3: Android client application
The image stabilization and gesture recognition modules will be integrated into an Android application. This application serves two purposes:

1. Magnification of printed text
2. Zooming capabilities of objects located in distance

The first functionality provides easy scroll and pan. Usually, handheld magnifiers users need to move the device either horizontally or vertically to continue reading. In our prototype users will be given a magnified window that could be either circular or rectangular in shape. This window will be updated by tracing the fingertip of the user. From observations it has been noticed that when zooming on an object with the phone camera may slip out of the view. Consequently, it is hard to move the mobile device to find the object when high level of magnification is applied. In this case it is better to provide the capability of selective zoom. The second functionality deals with this problem. Users will be to aim at the view that contains the object with any prior consideration such as placing the object at the center. Furthermore, a pinch gesture will be made to enlarge the object.

3.4: Evaluation

The system will be evaluated by conducting a technical evaluation and a user testing process. The method for technical evaluation is to create a confusion matrix to visualize the system performance. Performance indices such as sensitivity, specificity and accuracy will be measured.

To evaluate the usability of our prototype, a user testing process will be conducted, however, It is not determined at this stage that how we are going to recruit participants and how many of them will be recruited. They will be asked to rate the system after doing a specific task using SUS or NASA-TLX evaluation methods.

3.5: Image processing and visually impaired users

Image processing techniques are being broadly used in this study. Promising methods are now available to maximize the residual vision of visually impaired because of advances in image processing. There are considerable number of studies in the literature that aim to explore the value of image processing for low vision subjects. Techniques for enhancing the quality of images or background contrast has improved the reading experience so did the techniques for reducing unwanted jitter and blur. Therefore, a review of high tech visual aids for visually impaired that take advantage of image processing is presented here. This review is based on Moshtael et al (2015) literature review.

Generally vision aids fall into two categories based on how they present information to their users. Sensory substitution aids that provide information in auditory or tactile, actually translate the information to a presentable format. Unlike the first category, other techniques try to render the visual information in way to enhance its visibility, for example through a magnification. The latter category of technologies for vision aids that improves the vision experience are the primary focus of this section.

Optical magnifiers were the dominant method for visual rehabilitation. However, electronic magnifiers have now surpassed them and have become the mostly used type of magnifiers. Image processing provides modern tools to aid visually impaired people. These tools receive an image or a video frame and modify some parameters of that visual input such as illumination or contrast by some mathematical algorithms, then, represent the modified version to the user. This algorithms can be applied to a variety of devices including smartphones, head mounted displays and wearable devices. Sophisticated processing power that is embedded in commercially available devices create a great opportunity for developing new technologies and tools to improve the quality of life for low vision users.

A literature search investigated the studies which tried to explore the benefits of image processing for visually impaired. The areas in which image processing furnishes extensive potentials are listed below:

- Contrast enhancement
- Spatial frequencies
- Edge and contour enhancement
- Background attenuation and scene simplification
- Remapping and retargeting
- Jitter

Image processing application in assistive technologies for visually impaired people is an evolving area of research. In this project we will use image stabilization and hand gesture recognition techniques to implement a magnifier prototype.

4. Plan

Master Thesis Project Plan																	
Tasks	Months																
ID	Description	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	Scheduled Supervision Meetings																
2	Review Related Literature																
3	Formulating Final Project Scope																
4	Formulating Research Questions																
5	Writing the Master Thesis Phase 1 Draft																

6	**Submitting and Presenting Phase 1**													
7	Finalizing Methodology													
8	Image Processing and OpenCV Self-Study													
9	Machine Learning Self-Study													
10	Implementing Image Stabilization Module													
11	Implementing Gesture Recognition Module													
12	Creating Magnifier Application													
13	Creating Selective Magnifier Application													
14	Integrating the Modules with Magnifier Application													
15	Writing the Master Thesis Phase 2 Draft													
16	**Submitting and Presenting Phase 2**													
17	Technical Evaluation													
18	User Testing													

5. Magnifier prototype implementation

The primary focus of the second phase of this master thesis was to create a magnifier prototype. The main module of this prototype is the hand gesture detector which enables the user to interact with the system by performing hand gestures in front of the smartphone’s camera. Hand tracking and segmentation (HTS) are the primary steps for any hand gesture recognition system. HTS algorithms are found to be most efficient to handle the challenges of vision-based system such as skin color detection, complex background removal, and variable lighting condition (Archana G & Gajanan K, 2012). Due to dynamic backgrounds, the segmented image or video stream may contain noise. To remove the unwanted background noise several approaches are suggested in the literature. (Archana G & Gajanan K, 2012) used an edge traversal algorithm for background noise removal. In the following sections the algorithm for providing this functionality is described. In general the algorithm consists of four steps:

- Skin detection
- Skin segmentation
- Detecting the area in the video which contains the hand

- Hand feature point extraction
- Finding fingertips

There are three main steps involved in hand gesture recognition, hand segmentation, hand analysis and gesture detection. Segmentation involves extracting objects of interest from an image or a video frame. First hand candidate regions are searched, then each candidate region is analyzed to find the hand's detailed information such as fingertips and a palm center. Finally, using these information, the gesture is recognized.

Attempts were made to avoid adding invasive constraints such as forcing the user to wear a glove of a special color to provide more flexibility. For the segmentation purpose color information is used. The idea is to segment the video frame based on the skin color. Furthermore, hand features such as the center of the palm and finger tips will be extracted from the segmented area. These features will be used to detect the hand and to be able to define appropriate gestures to interact with the prototype application.

Color is a characteristic of human skin which is easily recognizable and a very powerful descriptor for object detection. Being invariant to rotation and geometric variation of the hand makes skin color a reliable source for real-time hand detection systems. Skin color falls into a specific color spectrum which is more close to red color range. This spectrum can further be studied and analyzed in different color spaces. It is one the skin features that enables us to have robust algorithm and methods for modelling the skin even with the different geometric variations of the hands or face patterns. Choosing the color space, modeling the color distribution and the way of processing the segmented areas in an image or video frame are main problems that a skin color detection algorithm must be able to solve. Each image or video frame is made of pixels. These pixels have different RGB color values. One way to detect skin is to classify the pixels into skin and non-skin individually, independent from the neighboring pixels. This method is called pixel-based skin detection. Region-based methods, in contrast, try to consider the spatial arrangement of each pixel to improve the performance. There are different types of color spaces for processing and storing digital image data. RGB color space is one the most widely used amongst the others which describes the image as a combination of three rays. Hue-saturation based color spaces such as HSV also have several different applications in digital image processing (Vladimir V et al, 2003).

There are various techniques for skin modeling. One way is to explicitly define the skin color boundaries in a color space. As an example (Peer et al, 2003) define the skin in RGB color as following:

(R,G,B) is classified as skin if:

(R,G,B) is classified as skin if:
$R > 95$ and $G > 40$ and $B > 20$ and
$\max\{R, G,B\} - \min\{R,G,B\} > 15$ and
$|R-G| > 15$ and $R > G$ and $R > B$

This method is utilized for the purpose of developing the magnifier prototype. The advantage of this method is its simplicity in defining the skin detection rules which enables the rapid creation of a skin classifier. The OpenCV image processing library provides a function which is able to find a color that is between a given rang. To detect the skin by using this function we need to set up the lower and upper threshold values in the RGB color space. Given the appropriate boundaries it will check for whatever RGB values which belong to that color spectrum. The accuracy of the algorithm is likely to increase with more precise selection of the upper and lower limits.

The hand segmentation is implemented by drawing the contours for the area which contains the detected skin. Precise hand segmentation is crucial for gesture-based Human-Machine Interaction (Ram P & Gyanendra K, 2015). However, the performance will be highly affected in a negative way due to complex background, similar colors of skin and non-uniform illumination. Next step is to apply morphological operations to remove noise and smoothen the detected area which contains the hand. These morphological transformations are simple operations that depend on the image shape. The erosion and dilation operations were used. These two were available only for binary images primarily, but they were further extended for grayscale images too. The basic idea of erosion is to erode away the boundaries of foreground object to have a more smooth image, whilst, dilation operates in the opposite way. To improve the result of applying the morphological operators, a Gaussian blur filter is also applied to ensure the highest amount of possible smoothness. It tries to reduce the noise and decrease the unnecessary details in an image. Gaussian blur filter is a widely used technique in the computer graphic and computer vision domain which is often used for producing photorealistic effects in game engines, photo rendering tools and computer vision software.

The next step is drawing contours around this area of interest to fully segment the hand from the background. Based on the definition from the OpenCV documentation, contours can be explained simply as a curve which joins all continues points that have some intensity or color. In fact contours is a perfect way for shape analysis and performing object detection and recognition. For better performance of the algorithms which are designed for finding contours in a given image it is recommended to use the binary image format (Archana G & Gajanan K, 2012).

Figure 2.1 shows the result of applying the algorithm for skin detection and skin segmentation on video captured by a smartphone camera. The green line around the hand is a visual representation of hand contours.

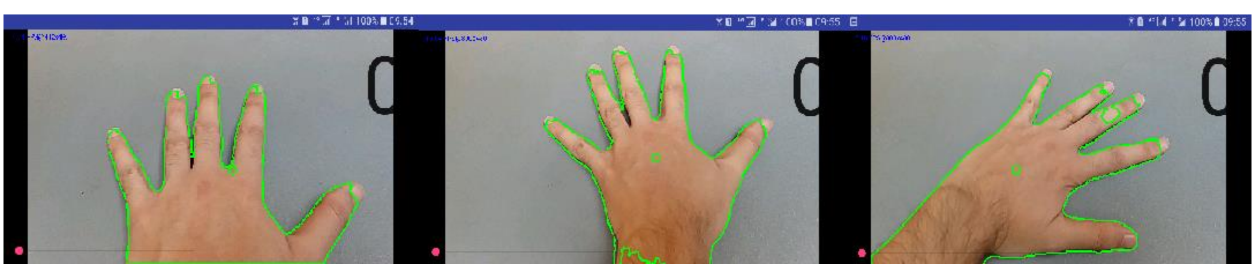

Figure 5.1 The Segmented skin is highlighted.

Based on the threshold values for defining the boundaries of skin color spectrum, the algorithm will produce much more accurate results if the hand is placed against a white or gray background that is similar to those of printed texts. This is because of the contrast between the skin color and these two colors.

Two other concepts in digital image processing will be used to extract the hand features and finally calculate the fingertip coordinates. Furthermore, by using those coordinates, it will be possible to trigger the zooming functionality of the camera and guide the user to read through the text. Convex hull and convexity defects are used for feature extraction purposes. A convex hull is similar to a contour in a sense that it tries to find the curve using the same criteria such as intensity and color, however it only uses the convex (bulged out) point and the final shape around the hand area will be a polygon. The convex hull of a set of points is defined as the smallest convex polygon that encloses all of the points in the set. Convex means that the polygon has no corner that is bent inwards (no concave points). Looking at a set of points, human intuition can help us to quickly figure out, what points are likely to touch the convex hull, and what points will be closer to the center, further away from the border of the convex hull. The Graham scan is one of the mostly used algorithms for finding convex hulls in two dimensions.

Figure 2.2 describes the visual presentation of applying the function which finds the convex hulls.

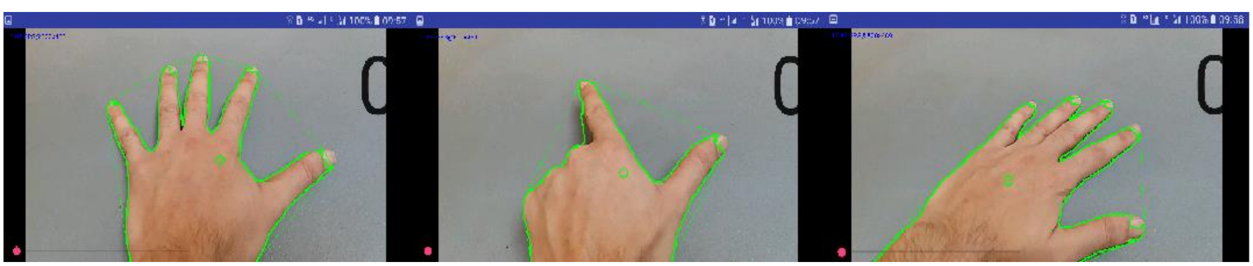

Figure 5.2 Applying convex hull function to detect all bulged out points in the segmented hand area.

In contrast, convexity defects are the points which bulge inside the area between two adjacent convex hull points. By calculating this points we will be able to find the coordinates of the starting and ending points of each finger. The convexity defect algorithm implemented for OpenCV library returns three values to indicate the starting, ending and middle points or coordinates of a defect area. The starting point of the defect area. These starting points are what we are interested in as they signify the finger tips. In figure 3 the convexity defects are shown.

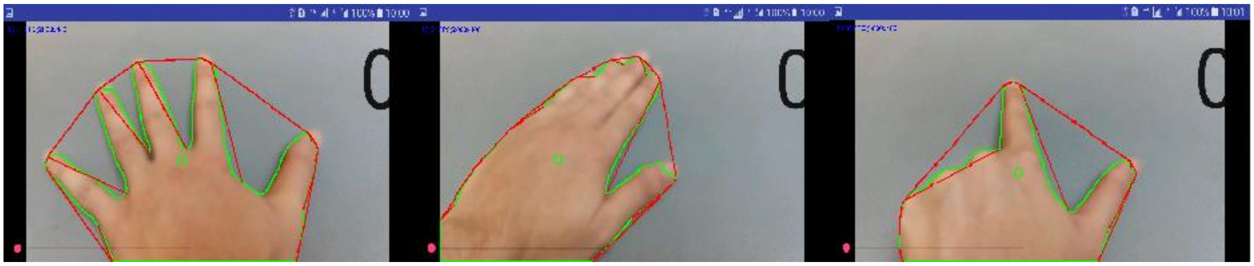

Figure 5.3 Applying convexity defect function to detect all bulged in points in the segmented hand area.

Using the convexity defects function, we get all the defects of the contour around the hand and we save them in another array. These are the lowest points between one finger and the other.

However, to exactly find the finger tips the algorithm needs some adjustments as it returns far too many defect starting point than just the finger tips. We need to find a way to filter these points. Only the ones which are closest to the fingertips should be stored as fingertip coordinates. These points are ready for next step which is the pinch gesture recognition. We filter them based on their distance from the center of the bounding rectangle (which approximately corresponds to the center of the hand), so that only the lowest points between each finger are kept. In order to make this process scale invariant, the height of the bounding rectangle is used as reference.

5.1: Optimizing the algorithm

As mentioned earlier, the algorithm needs some modifications to be able to detect the fingertips with more accuracy. It detects the convexity defects however there are still too many points on the screen which do not necessarily represent the fingertips. Therefore, to enhance the algorithm there is one more step which needs to be taken into account. However, these enhancements are dependent on some assumptions, they will profoundly improve the overall throughput of the algorithm. These assumptions include:

- The inner angle between the two lines of the defect region represent the angle between two fingers and the angle between two fingers belongs to a range within 20 and 120 degrees. It can be calculated by translating the vector to the origin and then calculating the arc cosine of the inner product divided by the norm of the vectors.
- The length of the line from the initial point to the middle point to be above a certain threshold. The lines with this length attribute are more likely to be fingers.

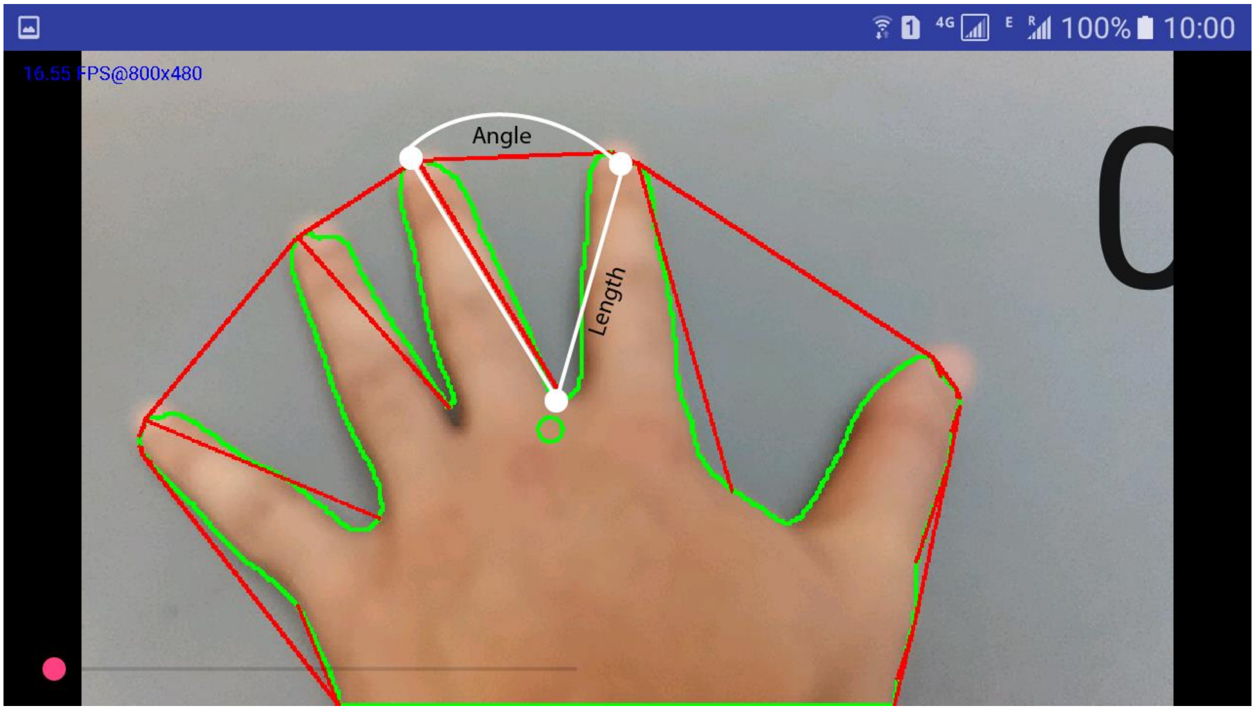

Figure 5.4 An illustration of angle and length

5.2: Technical Evaluation:

A mixed evaluation method has been taken for evaluating the prototype. The focus of the current phase is to perform a quantitative assessment. This will be done by creating confusion matrices to calculate different descriptive measures such as true and negative or accuracy under varying conditions. Confusion matrix enables us to objectively evaluate the prototype. Furthermore, a qualitative approach will be applied for measuring its compliance with universal design principles in the next phase.

A confusion matrix is a table which is often used to describe the performance of a classifier. This is applied on a set of data for which the true values are known. The evaluation of machine learning algorithm is the most common application of confusion matrices. The magnifier prototype basically performs the classification in two different settings. First one is detecting the hand, whether the video frame contains a hand, whereas the second is being able to perform the hand gesture correctly. There are some terms and definitions which worth mentioning before I starting the evaluation with confusion matrices.

- True positives (TP): this tells us how many times there is a hand in the video frame and the algorithm was able to detect it.
- True negative (TN): this indicates how many times we know that there is no hand in the video and the algorithm carries out the task correctly by not detecting any hand.
- False positives (FP): this is opposite to the true positives, meaning the times that a hand was mistakenly detected while there was none present in the video frame.
- False negative (FN): These are the cases in which there is a hand in the video frame but the algorithm is not able to find it.
- Accuracy: this answers the question of how many times the algorithm or the classifier performs correctly. It is calculated by the following formula: (TP+TN)/Total number of tests.
- Misclassification rate: how often the algorithm is producing wrong answers? By the following formula we can calculate Misclassification rate: (FP+FN)/Total number of tests.
- True positive rate: how often the algorithm is able to detect the hand when there is hand in the frame? It can be calculated by this formula: TP/number of tests with a hand present.
- False positive rate: when there is no hand, how often it predicts a hand? The following is the equation to calculate this value: FP/number of tests with no hand present.
- True negative rate also known as specificity: when there is no hand, how often it predicts no hand? It can be formulated as the following; TN/ number of tests with no hand present.
- Precision: when the algorithm finds a hand in the video frame, how often it is correct? TP/total number tests which include a hand
- Prevalence: if we consider the hand presence as our yes condition, how often the yes condition occur in our tests? Number of actual yes/total number of test

5-3: Running tests

This round of tests are conducted to assess the ability of algorithm in finding the fingertips. There are different contributing factors which negatively affect the accuracy though. Apart from the background and lighting condition, both the angle of the wrist and the angle between fingers play a very important role.

The test is carried out against a gray surface around 14:00 under day light situation. In the below picture you see the case where fingertips are not detected despite the hand successfully being recognized:

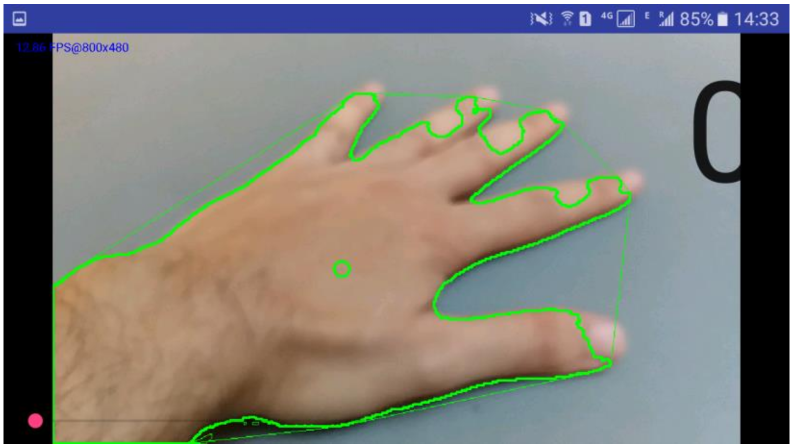

Figure 5.5 Algorithm not being able to detect fingertips.

As you can see fingertips are presented as little green circles so close to the edge of each finger. Apparently, the algorithm is having difficulties with detecting thumb in some situations. It happens in two cases, either the thumb is posed so wide from the center of the palm or it is so close to the center of the palm. It is because of the assumption that fingers length fall into a specific rang. Anything outside of that threshold values for fingers, is not considered a finger.

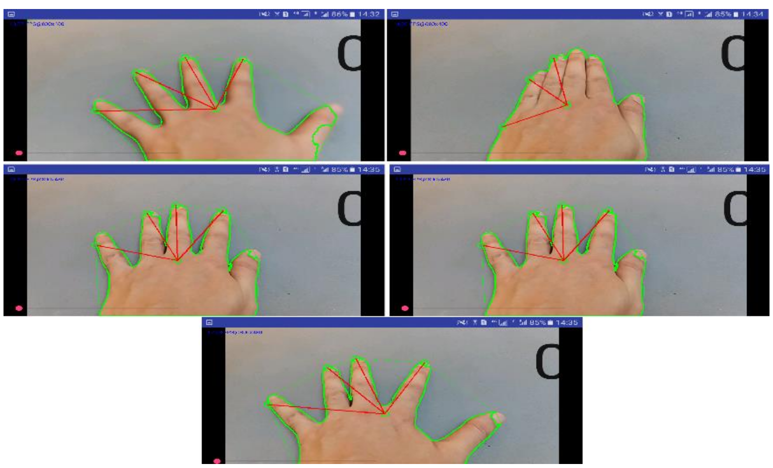

Figure 5.5 Thumb or other fingers which are so close or too far from the center of the palm cannot be detected as fingers.

The detection of fingertips implies that the algorithm was able to segment the skin area, find the convex hull and defects points and calculate the center of the palm. Hence, by evaluating the algorithm’s performance in finding the fingertips, all these aspects are covered. The bellow table demonstrates the confusion matrix for evaluating the fingertip detection. Based on presence or absence of the fingers in a video frame and the ability of the algorithm to correctly distinguish between these two situations, different values such as TP, TN, FP and FN are measured.

Actual Value (as confirmed by experiment)			
Predicted Value (predicted by the test)	NO = 50	Fingertips are detected (YES)	Fingertips are not detected (NO)
	Fingers are present	True Positive (TP): 18	False Negative (FN): 10
	Fingers are not present	False Positive (FP): 13	True Negative (TN): 9

Table 5.1 Confusion matrix for evaluating fingertips detection

Fifty tests were conducted under varying conditions. Various lighting conditions were taken into account as well as different backgrounds. Nevertheless, the algorithm performed better in situations under indirect daylight and gray or white background. The detection rate drops drastically with a noisy background. A noisy background, essentially means a surface or a space with lots of different colors. In the following lines other measures which were mentioned earlier are calculated based on the values from the table.

- **Accuracy: (18+13) / 50 = 0.54**
- **Misclassification Rate: (13+10) / 50 = 0.46**
- **False Positive Rate : 13 / (13 + 9) = 0.59**
- **True Negative Rate:** 9 / (13 + 9) = 0.4
- **Precision: 18 / (18 + 13) = 0.58**
- **Prevalence: 28 / 50 = 0.56**

From the calculated values for accuracy and precision, we can understand that the algorithm is performing on an average level. Therefore, there are possibilities for improvement.

The idea for implementing the gestures was to calculate the distance between two points which are detected as fingertips. However, this idea turned out not to be as efficient as expected due to the constant changes in the video frame ratio after the zooming was triggered. The following table depicts the confusion matrix for executing the pinch gesture:

		Actual Value (as confirmed by experiment)	
Predicted Value (predicted by the test)	NO = 50	Pinch is detected (YES)	Pinch is not detected (NO)
	Pinch is performed	True Positive (TP): 8	False Negative (FN): 20
	Pinch is not performed	False Positive (FP): 15	True Negative (TN): 7

Table 5.2 confusion matrix for evaluating the execution of pinch gesture

- **Accuracy: (8+15) / 50 = 0.46**
- **Misclassification Rate: (15+20) / 50 = 0.7**
- **False Positive Rate : 15 / (15 + 7) = 0.68**
- **True Negative Rate:** 7 / (15 + 7) = 0.31
- **Precision: 8 / (18 + 15) = 0.24**
- **Prevalence: 28 / 50 = 0.56**

The tests for the pinch gesture were conducted using only a gray background under transitive lighting conditions. The decision for an invariant gray background was made since the algorithm had better outcome against such surface. The accuracy and precision show that the algorithm does not have a good performance in detecting the pinch gesture, hence, this aspect needs improvement.

6. Experiment:

In First phase a comprehensive literature study was conducted on the magnifier technologies. The stationary magnifiers known as CCTVs has their own applications and come with scene editing features such as color enhancement, support for color blindness, etc. These magnifiers also offer features like text to speech conversions. However, they impose mobility restrictions on users and they are not affordable by everyone. Portable magnification devices solve the problem of mobility but they are still expensive. Smartphones are a good choice for having a portable magnifier. They have a good processing power and the quality of the smartphone cameras keeps improving.

The problem with portable magnifier whether it is a dedicated portable magnifier or just a smartphone camera is that it could be hard for some users to perform zooming and panning on a text. They have to keep moving the magnifier around the text that they are reading. It can cause fatigue or confusion.

In the second phase a software prototype was developed that can track the user's fingertip and therefore update the zoomed area according to the fingertips position. In the previous phase some technical tests were conducted, using confusion matrices, to evaluate the algorithm ability for detecting fingertips and further executing the pinch gesture. It performed well in case of detecting the fingertips, however the results for pinch gesture was not satisfactory. Hence, I decided to change way that the prototype software was going to be tested.

Instead of pinch by hand gesture, I found a fixed zoom level along with focusing only on implementing the panning by hand gesture is more practical. Therefore, some adjustments were made to the prototype to make it ready for the user testing on panning with fixed zoom level.

6-1: User Testing:

An experiment was conducted to test and compare two panning methods:

- Panning by the software prototype (Hand gesture).
- Using the phone camera and do the panning in a normal way.

At the end of each user testing section, the participants were asked to fill out a questionnaire which was designed based on System usability scale (SUS). In the following the experiment procedure and protocol are described.

The experiment:

It is an experiment designed to test the effects of panning prototype software which operates by hand gestures. By effects, it generally means, whether it enhances the reading experience of the people with visual impairment while reading a text. Ten people from the university participated in this experiment. Unfortunately, it was not easy to find people with visual impairment and I was able to recruit only one person who was using spectacles. To compensate, the texts that users had to read during the experiment as their task were printed in a very small font to make it hard for reading.

Independent variables:

Reading methods are independent variables. There are two panning methods being tested in this experiment. The panning prototype software which is controlled by hand gestures. In addition, using the phone as a magnifier and perform panning by moving the device itself like a normal scenario is the other method.

Dependent variables:

There are two dependent variables. I will measure the time that takes to finish a task with each method. At the end of each run of the test, participants will be asked to answer a questionnaire and state their preferred panning method.

Apparatus and Study Conditions:

The experiment was conducted on a Samsung Galaxy s5 Fashion Edition. For this prototype to work, OpenCV library should be installed on the device. The application will ask for installing OpenCV when it is to be installed for the first time. Participants were asked to hold the device with their dominant hand. This was to prevent any possible unwanted hand shaking. They were also instructed to adjust the distance between the device and the paper, so it capture the entire content. The paper including the task text, was placed against a gray background to create enough contrast between skin color and background color. Under this condition, the algorithm was able to perform better. This was already tested during the technical tests in second phase. All the tests were conducted during the day under the day light condition.

Participants:

Participants were all educated and knew how to use a smartphone and felt comfortable using the app. However, some of the participants needed explanation in order to fully understand the spatial hand gestures and their use in the prototype panning software. Gender did not have any influence on the results of the experiment. So, I tried to recruit participants regardless of their gender and so far eight men and two women took part in the experiment.

The procedure:

Here are the steps of the experiment which will be followed by detailed explanation of each step.

1. Introduction
2. The task, what participants should do.
3. Complete the questionnaire.

Step one (introduction):

During the introduction:

- Participants will get a brief overview of the problem that is being addressed.
- Participants will be thought how to operate the panning prototype.
- Participant will know what happens during the test.
- Participants will know how long the test is going to be and how the duration is divided.

Overview:

This prototype tries to address the issue of panning that users of portable magnifier devices may encounter during reading a text. They want to magnify the text, hence they are required to constantly move the device over the text to continue reading. Instead, holding the phone steady and tracking the finger to update the magnified portion can be a good solution.

How it works?

- Start the application.
- You will see an overlay that has a transparent circle inside it. That is where you can see the magnified text
- Select a predefined zoom level.
- Hold the device in such a distance from the text so it can cover all of it.

- While holding the phone with one hand, try to scan the text with the index finger of your free hand.

What happens during the test?

By asking the participant's consent, only the scene which include the hands, phone and text will be recorded. This is to observe the participant interaction with the panning prototype and see the issues that they may run into. The time that each candidate takes to finish reading a card will be calculated by a stopwatch to compare the time spent on each method.

Step two (The task):

The experiment will be carried out on my phone which is an Android galaxy S5. There are two types of reading methods that are being tested and further compared. The panning prototype which uses hand gesture and the normal phone camera as a magnifier. There are a set of different text cards about general topics like movies, history or cities. The participants will be asked to pick a card for each reading method. To prevent the learning effect, participants will have to choose different cards for each reading method. At the end of reading each card, participants will be asked to answer questions which are directly from the text.

Step three (The questionnaire):

The final step of the task is to complete a questionnaire for system usability testing. This questionnaire is designed based on System Usability Scale standard (SUS).

Additional notes:

- Card formats: cards are printed in an A5 paper in landscape direction.
- Text font size: the font size will be 8.

Duration:

The duration of the test is approximately 20 minutes. The time will be divided as following

- Three minutes for introduction.
- Two minutes for describing the task (refer to step two)
- Five minutes for reading the card by the panning prototype and answering questions.
- Five minutes for reading the card by the normal phone camera and answering questions.
- Five minutes for filling the questionnaire.

In the following you can see two card samples which are used in the experiment:

First card:

Legends of FC Barcelona:

FC Barcelona is a football club in Catalonia, Spain. Some footballer are considered as legends for fans. For example, Carolus Puyol, captain of team that won first treble trophy. Football Players who have played at least 100 games for club are included in the list of legends. This legendry list is for the time between 1899 and May 2015. 958 footballers have played at least one official game with Barcelona. Barcelona has won champions league four times.

Questions:

1. How many players, did play at least one time for FC Barcelona?
2. Who was captain when FC Barcelona won first treble?
3. How many games have FC Barcelona legends played for the club?
4. How many times FC Barcelona has won champions league?
5. What is the topic of the card?

Second card:

Academy awards:

Hollywood celebrates the academy awards every year. People also know academy awards as Oscars. In this ceremony artists are given awards for their artistic activities in the past year. George Stanly was the original designer of the award. The first academy award was held in 1929 at Roosevelt hotel and 270 audiences were present. Jack Nicholson is the three times winner of the academy award best actor and best supporting actor. He is famous for his role in the shinning movie. In the movie Departed he played as an undercover agent for FBI and also was a criminal at the same time.

Questions:

1. When was the first academy award?
2. Where did the first academy award take place?
3. How many people were present in first academy award?
4. Who is the first designer of the academy Award?
5. How many times has Jack Nicholson won academy awards?

7. Results:

In the first two diagrams you see the comparison between the two panning methods in terms of time spent for task completion and participant's preference towards each method:

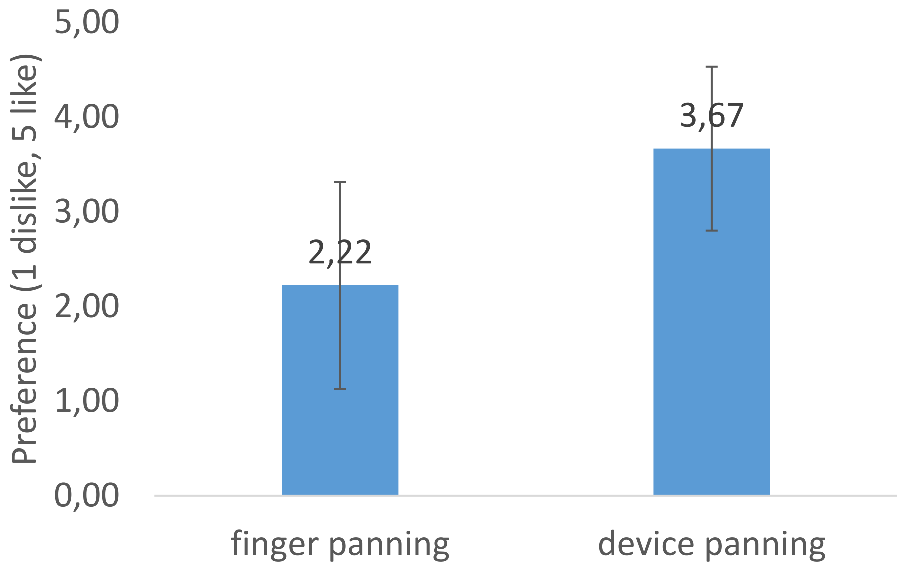

Chart 7.1: Participants preference of each panning method

In this graph the preference of panning methods are compared with each other in scale of one to five.

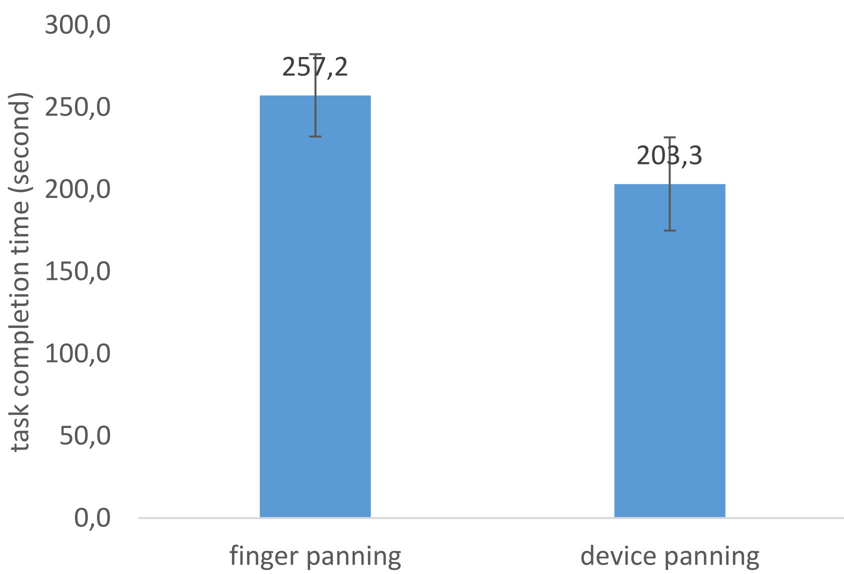

Chart 7.2: Time completion of tasks done by each panning method

In this graph the time of task completion for each method is calculated in seconds and further compared with each other.

From the above graphs we can see that the normal panning performed better and participants were more likely to use this method. However they generally liked the idea of finger panning but in terms of completing the task they finally preferred the normal device panning. In addition, they performed better in completing the task with the normal panning which took them less time to complete the given task.

In the Following tables you see the statistical test for both time and preference:

Paired Samples T-Test

			Test	Statistic	df	p	Effect Size
Panning Software Prototype Task-Timing	-	Normal Panning Task-Timing	Student	4.457	9	0.002	1.410
			Wilcoxon	55.000		0.006	1.000

Note. For the Student t-test, effect size is given by Cohen's *d* ; for the Wilcoxon test, effect size is given by the matched rank biserial correlation.

Assumption Checks

Test of Normality (Shapiro-Wilk)

	W	p
Panning Software Prototype Task-Timing - Normal Panning Task-Timing	0.894	0.187

Note. Significant results suggest a deviation from normality.

Descriptive

Descriptive

	N	Mean	SD	SE
Panning Software Prototype Task-Timing	10	257.2	25.07	7.927
Normal Panning Task-Timing	10	203.3	28.37	8.972

Descriptive Plot

Panning Software Prototype Task-Timing - Normal Panning Task-Timing

Paired Samples T-Test

Paired Samples T-Test

			W	p	Rank-Biserial Correlation
Pannnig software Preference	-	Normal Panning Preferece	2.000	0.029	-0.927

Note. Wilcoxon signed-rank test.

8. Conclusion

Smartphones have changed our lives during past years. They are compact and literally provide enough processing power for daily activities and Ubiquitous computing. The embedded camera of these smartphone offers a great potential for assisting the visually impaired people to accomplish a more independent life by improving their visual experience. People with low vision have the tendency to use their residual vision instead of relying on other sense. Magnifiers are among the solutions that can benefit the visually impaired people to a great extent.

A smartphone based magnifier seems a promising solution. This projects aims to design and evaluate a magnifier prototype to make the reading process easy for low visions. This prototype takes the video captured by smartphone's front camera, stabilize it and enables users to navigate the real world by their fingers. Different studies have explored optimized image stabilization and hand gesture recognition for mobile devices. However, none of those studies offers an integrated solution which specifically targets the needs and requirements of visually impaired.

9. References

Blindness and visual impairment. (2017, October 11). Retrieved from:

http://www.who.int/en/news-room/fact-sheets/detail/blindness-and-visual-impairment

Blindness Prevention. (2015, September). Retrieved from:

http://www.wpro.who.int/mediacentre/factsheets/blindness/en/

Bryan Gerritsen. Electronic Magnifiers and Magnifying Systems. (2018, May 21). Retrieved from: http://www.visionaware.org/info/everyday-living/helpful-products/overview-of-low-vision-devices/electronic-magnifiers/1245

IMU Inertial Measurement Unit. (2018, May 21). Retrieved from:

https://www.xsens.com/tags/imu/

WACAG 2.0. Visual Presentation. (2018, May 21). Retrieved from:

https://www.w3.org/TR/UNDERSTANDING-WCAG20/visual-audio-contrast-visual-presentation.html#fullscreenwindowdef

Tavish Srivastava. Introduction to Random Forest - Simplified. (2014, June 10). Retrieved from: https://www.analyticsvidhya.com/blog/2014/06/introduction-random-forest-simplified/

Magic-Lens. (2018, May 21). Retrieved from: https://www.infoviswiki.net/index.php?title=Magic_Lens

Carlos, A. Eric, C. (2017). Evaluation of a gaze-controlled vision enhancement system for reading in visually impaired people. PLoS ONE 12 (4): e0174910. https://doi.org/10.1371/journal.pone.0174910

Larry Hardesty. Finger-mounted reading device for the blind. (2015, March 10). Retrieved from:
http://news.mit.edu/2015/finger-mounted-reading-device-blind-0310

Crossland, MD. Segurado, Silva R. Macedo, AF. (2014). Smartphone, tablet computer and e-reader use by people with vision impairment. Ophthalmic Physiol Opt 2014; 34: 552–557. doi: 10.1111/opo.12136

Harrison, C.M. (2004). Low-vision reading aids: reading as a pleasurable experience. Published in Personal and Ubiquitous Computing Journal.

Gowases, T. Bednarik, R. Tukiainen, M. (2011). Text Highlighting Improves User Experience for Reading with Magnified Displays. CHI EA '11: CHI '11 Extended Abstracts on Human Factors in Computing Systems

HARPER R, CULHAM L, DICKINSON C. (1999). Head mounted video magnification devices for low vision rehabilitation: a comparison with existing technology
British Journal of Ophthalmology 1999;**83:**495-500.

Stearns, L. DeSouza, V. Yin, Jessica. Findlater, L. E.Forehich, J. (2017). Augmented Reality Magnification for Low Vision Users with the Microsoft Hololens and a Finger-Worn Camera. ASSETS '17: Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility.

Pucihar, K. Coulton, P. (2013). Enhanced Virtual Transparency in Handheld AR: Digital Magnifying Glass. MobileHCI '13: Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services.

Aguilar, C., Castet, E. (2017). Evaluation of a gaze-controlled vision enhancement system for reading in visually impaired people. PLoS ONE 12 (4): e0174910. https://doi.org/10.1371/journal.pone.0174910

Shabab, S., Ashmafee, Md.H. (2016). BLIND READER: An Intelligent Assistant for Blind. 19th International Conference on Computer and Information Technology.

Lamza, A. Wrobel, Z. (2012). New Efficient Method of Digital Video Stabilization for In-Car Camera. International Conference on Multimedia Communications, Services and Security.

Xie, Q.W. Chen, X. Zhang, L.M. Jiang, A. Cui, F. (2016).A Novel Rapid and Efficient Video Stabilization Algorithm for Mobile Platforms. Research Center for Eco-Environment Sciences ,Chinese Academy of Sciences Beijing, 100085, China& Beijing Smarter Eye Technology Co, Ltd.

Hanning, G. Forslow, N. Forssen, P. Rinaby, E. Tornqvist, D. Camller, J. (2011). Stabilizing Cell Phone Video using Inertial Measurement Sensors. 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)

Hsiao, H. Wei, i. (2017) A Real-Time Visual Tracking Technique for Mobile Display Stabilization. Proceedings of the 2017 IEEE International Conference on Applied System Innovation
IEEE-ICASI 2017

Li, Z. Pundik, S., Luo, G. (2013). Stabilization of Magnified Videos on a Mobile Device for Visually Impaired. IEEE Conference on Computer Vision and Pattern Recognition Workshops.

Lee, S. Sanford, J. (2012). Gesture Interface Magnifiers for Low-Vision Users. ASSETS '12: Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility.

Yamada, W. Manabe, H. Hirouki, H. Ochiai, K. Ikeda, D. (2017). CamTrackPoint: Pointing Stick for Mobile Device Using Rear Camera. UbiComp '17: Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers

Xiao, X. Han, Tang. Wang, J. LensGesture. (2013). Augmenting Mobile Interactions with Back-of-Device Finger Gestures. ICMI '13: Proceedings of the 15th ACM on International conference on multimodal interaction.

Butler, A., Izadi, S., and Hodges, S. (2008). SideSight: multi-touch interaction around small devices. UIST '08: Proceedings of the 21st annual ACM symposium on User interface software and technology

Kratz, S. Rohs, M. (2009). HoverFlow: Expanding the Design Space
of Around-Device Interaction. MobileHCI '09: Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services

Ketabdar, H. Yuksel, K. Roshandel, M. (2010). MagiTact: Interaction with Mobile Devices Based on Compass (Magnetic) Sensor. IUI '10 Proceedings of the 15th international conference on intelligent user interfaces.

10. Zhao, C. Chen, K. Auim, T. Patel, S. Reynolds, M. (2014). SideSwipe: detecting in-air gestures around mobile devices using actual GSM signal. UIST '14 Proceedings of the 27th annual ACM symposium on User interface software and technology.
11.
12. Yeo, h. Lee, J. Bianchi, A. Quigley, A. (2016). WatchMI: pressure touch, twist and pan gesture input on unmodified smartwatches. MobileHCI '16 Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services

Kubo, Y. Hizuki, B. Takahshi, S. (2016). Watch Commander: A Gesture-based Invocation System for
Rectangular Smartwatches using B2B-Swipe. UIST '16 Adjunct Proceedings of the 29th Annual Symposium on User Interface Software and Technology

Kratz, S. Rohs, M. Guse, Dennis. Muller, J. Bailly, G. Nischt, M. (2012). PalmSpace: Continuous Around-Device Gestures vs Multitouch for 3D Rotation Tasks on Mobile Devices. AVI '12 Proceedings of the International Working Conference on Advanced Visual Interfaces

Grubert, J. Ofek, E. Pahud, M. Kranz, M. Schmalstieg, D. (2016). GlassHands: Interaction Around Unmodified Mobile Devices Using Sunglasses. ISS '16 Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces.

Song, J. Soros, G. Pece, F. Fanello, S. Izadi, S. Keshin, C. Hilliges, O. (2014) .In-air Gestures Around Unmodified Mobile Devices. UIST '14: Proceedings of the 27th annual ACM symposium on User interface software and technology.

Caballero, M. Chang, T. Menendez, M. Occhialini, V. (2010). Behand: Augmented Virtuality Gestural Interaction for Mobile Phones. MobileHCI '10 Proceedings of the 12th international conference on Human computer interaction with mobile devices and services.

Haratano, R. Kartikasari, A. (2016). Based Real-Time Static Indonesian Sign Language Recognition System Prototype. 2016 8th International Conference on Information Technology and Electrical Engineering (ICITEE)

Lahiani, H. Elleuch, M. (2015). Real Time Hand Gesture Recognition System for Android Devices. 2015 15th International Conference on Intelligent Systems Design and Applications (ISDA)

Houssem, L. Kherallah, M. Neji, M (2016). Hand pose estimation system based on Viola-Jones algorithm for Android devices. 2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA)

Moscovich, M. F. Hughes, J. (2004). Navigating Documents with the Virtual Scroll Ring. UIST '04: Proceedings of the 17th annual ACM symposium on User interface software and technology

Alli, M. Feng, F. (2014). Automatic Page Scrolling for Mobile Web Search. 2014 International Conference on Smart Computing.

Rohs, M. Essi, G. (2006). Which One is Better? – Information Navigation Techniques for Spatially Aware Handheld Displays. ICMI '06 Proceedings of the 8th international conference on Multimodal

Mooser, J. You, Suya. Neumann, N. (2008). Large Document, Small Screen: A Camera Driven Scroll and Zoom Control for Mobile Devices. I3D '08 Proceedings of the 2008 symposium on Interactive 3D graphics and games

Rohs, M. Schoning, J. Raubal, M. Essi, G. Kruger, A. (2007). Map Navigation with Mobile Devices: Virtual versus
Physical Movement with and without Visual Context. ICMI '07: Proceedings of the 9th international conference on Multimodal interfaces

13. Moshtael H, Aslam T, Underwood I, Dhillon B. (2015) .High Tech Aids Low Vision: A Review of Image Processing for the Visually Impaired. Tran Vis Sci Tech. 2015; 4(4):6, doi:10.1167/tvst.4.4.6

Li, L. Ma, M. Zhao, Z. (2014). Real-Time Video Stabilization Based On Fast Block Matching and Improved Kalman Filter. Fifth International Conference on Intelligent Control and Information Processing

Archana G, Gajanan K. (2012). Hand Segmentation Techniques to Hand Gesture Recognition for Natural Human Computer Interaction. International Journal of Human Computer Interaction (IJHCI), Volume (3)

Lilit H. Jo L. Dympna O. Hannah B. (2013). Mobile assistive technologies for the visually impaired. Survey of Ophthalmology 58 (2013).

Frode E, Evelyn E. (2017). Head-Mounted Augmented Reality Displays on the Cheap: A DIY Approach to Sketching and Prototyping Low-Vision Assistive Technologies. International Conference on Universal Access in Human-Computer Interaction .UAHCI 2017: Universal Access in Human–Computer Interaction. Designing Novel Interactions pp 167-186

https://docs.opencv.org/3.3.1/d4/d73/tutorial_py_contours_begin.html

https://docs.opencv.org/3.4/dd/d49/tutorial_py_contour_features.html

https://picoledelimao.github.io/blog/2015/11/15/fingertip-detection-on-opencv/

https://medium.com/@soffritti.pierfrancesco/handy-hands-detection-with-opencv-ac6e9fb3cec1