Title
LookTel --- Computer Vision Applications for the Visually Impaired

Permalink
https://escholarship.org/uc/item/57q1b167

Author
Sudol, Jeremi

Publication Date
2013

Peer reviewed|Thesis/dissertation
LookTel — Computer Vision Applications for the Visually Impaired

A dissertation submitted in partial satisfaction of the requirements for the degree
Doctor of Philosophy in Computer Science

by

Jeremi Sudol

2013
ABSTRACT OF THE DISSERTATION

LookTel — Computer Vision Applications for the Visually Impaired

by

Jeremi Sudol

Doctor of Philosophy in Computer Science

University of California, Los Angeles, 2013

Professor Stefano Soatto, Chair

The modern world is predominantly visual. While most people live and function with vision, many people live their lives with limited or no vision at all. The system presented in this dissertation aims to use computer vision techniques to improve the lives of individuals having severe visual impairments or blindness. By applying real-time image recognition in an accessible manner the system facilitates increased personal independence. The system presented here substantially advances the state of the art both in terms of the speed, accuracy, and robustness to the real-world variability. The real-time on-device recognition speed enables more than just object identification; it permits detection – finding the desired item among others. The high accuracy and robustness expands the ability for people to participate in commerce and expand roles for economic participation in multiple cultures. We present an extensive evaluation demonstration and a real-world public release of the system.
The dissertation of Jeremi Sudol is approved.

Mario Gerla

Zhuowen Tu

Erkki Huhtamo

Stefano Soatto, Committee Chair

University of California, Los Angeles

2013
# TABLE OF CONTENTS

1. PROBLEM STATEMENT AND GOALS .................................................................1

1.1 Scope of project.........................................................................................1

1.2 Visual Impairment and Blindness ..............................................................4
   1.2.1 Medical Model and the Social Model ................................................4
   1.2.2 Definitions of Blindness and Visual Impairment ................................5
   1.2.3 Global Population Statistics...............................................................7
   1.2.4 United States Population Statistics...................................................8
   1.2.5 Causes of Blindness in the United States .........................................10
   1.2.6 The Blindness Experience – The Mars Colonists ............................12

1.3 History and Related Work ........................................................................14
   1.3.1 A Deeper Historical Perspective ......................................................15
   1.3.2 A reality today ...............................................................................19
   1.3.3 A Contemporary Historical Perspective ........................................21

2. LOOKTEL, VERSION 1.0 ...........................................................................26

2.1 Scope of the project ...............................................................................27

2.2 System Architecture ..............................................................................29

2.3 iVisit .......................................................................................................30

2.4 A Real-Time Object Recognition System ..............................................32
   2.4.1 Real-Time Recognition and Micro-Cloud Computing ....................33
2.4.2 Training with the LookTel 1.0 Platform ........................................... 34
2.4.3 Location and landmark recognition .................................................. 36
2.4.4 Tags and Tagging ............................................................................ 37
2.4.5 Object Recognition database management ...................................... 39

2.5 OCR ........................................................................................................ 40

2.6 Remote Sighted Assistance ................................................................. 41
  2.6.1 Mobility Assistance ......................................................................... 42
  2.6.2 Remote Training .............................................................................. 44
  2.6.3 Remote System Assistance ............................................................. 44

2.7 System Accessibility ............................................................................. 45
  2.7.1 Custom Touch Screen Human Interface .......................................... 45
  2.7.2 Mobile Phone Functionality ............................................................ 47
  2.7.3 Screen Reader Compatibility ........................................................... 48

2.8 Hardware Considerations .................................................................... 48
  2.9.1 Details of the implementation ........................................................ 49

2.9 Performance .......................................................................................... 51

2.10 Real-world beta trial of LookTel 1.0 and results .................................. 53
  2.10.1 Scope of the trial ........................................................................... 53
  2.10.2 Beta trial participants .................................................................... 55
  2.10.3 Results and feedback .................................................................... 56
  2.10.4 Conclusions .................................................................................. 61
3. LOOKTEL, VERSION 2.0 .........................................................................................63

3.1 An Accessibility Revolution .................................................................................63
  3.1.1 VoiceOver – a brief introduction ........................................................................64

3.2 LookTel Money Reader ..........................................................................................67
  3.2.1 Public release and results ...................................................................................69

3.3 LookTel Recognizer ..............................................................................................72
  3.3.1 Features and Capabilities ...................................................................................72
  3.3.2 Ongoing updates ...............................................................................................73
  3.3.3 Public release and results ...................................................................................74

4. CONCLUSIONS AND FUTURE WORK ..................................................................75

4.1 Money Reader .......................................................................................................75

4.2 Recognizer ............................................................................................................77

4.3 Remote Eyes ..........................................................................................................78

5. REFERENCES ..........................................................................................................80
INDEX OF FIGURES

FIGURE 1: SAMPLE OF RAISED AND INKED TEXT FROM VALENTIN HAUY. ........................................ 17

FIGURE 2: LOOKTel IN VIDEO SEE-THROUGH MODE. ........................................................................ 23

FIGURE 3: SCHEMATIC OF THE SYSTEM ARCHITECTURE. ............................................................ 30

FIGURE 4: A SCREENSHOT OF BASESTATION IN TRAINING MODE.............................................. 34

FIGURE 5: SAMPLE TAGS................................................................................................................ 37

FIGURE 6: HTC FUZE WITH THE TOUCH SCREEN HUMAN INTERFACE. ....................................... 46
INDEX OF TABLES

TABLE 1: LOOKTEL 1.0 BETA TRIAL PARTICIPANTS.................................................................................. 56
ACKNOWLEDGMENTS

This was a multi-faceted undertaking. Many individuals were and continue to be involved. The key contributors and their roles are listed below:

Orang Dialameh: iVisit CEO and primary inventor of the SeeScan/SeeStar concept.
Chuck Blanchard: Senior Technical Staff, Human Interface and Application Developer.
Eugene Jurinsky: iVisit VP of Software Engineering.
Tim Dorcey: iVisit and networking and communication application development.
Bing Song: Computer Vision Expert
Gary W. Kelly: Human factors engineer, designing and conducting human participant testing and customer evaluations, while assisting with interface design.
David Ross: Veteran’s Administration AREF, Human Subject testing and human participant evaluation lead.
Eyal Binshtock: iVisit, Customer Support.

The following people provided generous advice and guidance: David Lowe (Professor UBC), Stefano Soatto (Professor UCLA), Deborah Estrin (Professor UCLA), Erkki Huhtamo (Professor UCLA), Mario Gerla (Professor UCLA), Zhuowen Tu (Professor UCLA), Mario Munich (CTO Evolution Robotics), and Paolo Pirjanian (CEO Evolution Robotics).

The project was partially funded by the National Institute of Health (NIH) and National Eye Institute (NEI) with the following Grants: 1R43EY016953-01 & 2R44EY016953-0.
VITA

1997 2000  B.S. Computer Science and Psychology
Rensselaer Polytechnic Institute, Troy NY

2000  Visiting Scholar, Computer Science
Swiss Federal Institute of Technology (ETH), Zürich

2001 2004  M.S. Computer Science
New York University, New York, NY

2007 2012  Computer Vision Scientist and Co-Founder
LookTel / IPPLEX, Santa Monica CA

2013  Computer Vision Scientist
LookTel / NantVision / NantWorks, Culver City CA

2013  Adjunct Professor of Inclusive Design
Ontario College of Art and Design, Toronto, Canada

PUBLICATIONS and PRESENTATIONS

J. Sudol, G. Kelly; LookTel Recognizer. International Technology and Persons with Disabilities Conference, CSUN 2013. San Diego, CA

J. Sudol; LookTel Apps and Solutions for Visually Impaired Persons. G3ICT, m-Enabling Health Conference, 2011 Washington DC

J. Sudol; LookTel Money Reader. International Technology and Persons with Disabilities Conference, CSUN 2011. San Diego, CA


1. Problem statement and goals

The modern world is predominantly visual. While most people live and function with vision, many people live their lives with limited or no vision at all. Most tasks of daily living can, with practice and training, be carried out without the use of vision. People around the globe adapt their lives to varied environmental conditions and thrive in diverse environments.

When considering this spectrum of circumstances, we sought to identify several small but concrete examples where recent advances in computer vision and related techniques, combined with current hardware, might improve the lives of individuals having severe visual impairments or blindness. What began as an exercise in the application of technology to improve the lives of these persons, progressed through several iterations and redesigns, before reaching validation in the daily life use of the developed solutions.

1.1 Scope of project

In an effort to increase functional independence for visually impaired people we identified and addressed the following problems: Identification of packaged goods,
identification of currency, way-finding assistance through landmark recognition, ability to read printed text, and accessible interfaces for touch screen-based smartphones. Extended support by allowing real-time human assistance in the form of remote eyesight and remote device control. The system was designed for ease of use and customizable through personalized information. These solutions are facilitated by access to state of the art implementations of object recognition, optical character recognition, VOIP streaming, high resolution image capture, GPS, 3G/4G mobile phones.

The project spans a period of seven years and is closely related to the prior projects of SeeScan and SeeStar developed under sponsorship from the National Institute of Health (NIH) and National Eye Institute (NEI) with the following Grants: 1R43EY016953-01 & 2R44EY016953-0. SeeStar was the working name for remote localization and orientation assistance, as inspired by NavStar solutions of the automobile industry. The SeeScan referred to applications of image recognition technology to assist visually impaired persons.

LookTel 1.0 platform was under development from 2007 until 2010 (SeeStar phase 2 and SeeScan phase 1). LookTel 2.0 applications have been under development from 2010 and are continuing.

While the author has been and remains the system architect, lead developer, and systems integrator for the LookTel 1.0 and 2.0 platforms, many individuals were and
continue to be involved throughout the project. The key contributors and their roles are listed below:

Orang Dialameh: iVisit CEO and primary inventor of the SeeScan/SeeStar concept.
Chuck Blanchard: Senior Member of Technical Staff Human Interface and Application Developer.
Eugene Jurinsky: iVisit VP of Software Engineering.
Tim Dorcey: iVisit and networking and communication application development.
Bing Song: Computer Vision Expert
Gary W. Kelly: Human factors engineer, designing and conducting human participant testing and customer evaluations, while assisting with interface design.
David Ross: Veteran’s Administration AREF, Human Subject testing and human participant evaluation lead.
Eyal Binshtock: iVisit, Customer Support.

The following people provided generous advice and guidance: David Lowe (Professor UBC), Stefano Soatto (Professor UCLA), Deborah Estrin (Professor UCLA), Erkki Huhtamo (Professor UCLA), Mario Gerla (Professor UCLA), Zhuowen Tu (Professor UCLA), Mario Munich (CTO Evolution Robotics), and Paolo Pirjanian (CEO Evolution Robotics).
1.2 Visual Impairment and Blindness

The next sections introduce fundamental frameworks for understanding the disability experience, present a brief overview of blindness and visual impairment in the world and the United States, and introduce characteristics of the blindness experience.

1.2.1 Medical Model and the Social Model

The two major models defining disabilities are the medical and the social model. The medical model begins by defining a “normal” human in terms of biology, and seeks to treat any individual’s deviations from that model. The treatments and solutions focus on minimizing the discrepancies from the generally accepted normal human biology. The medical model is expressed by the international classification of “impairment”, “disability” and “handicap” developed by the World Health Organization in 1980. The WHO defined “impairment” as “any loss or abnormality of psychological, physical or anatomical structure or function”. A “disability” arises when the impairment prevents a person from being able to “perform an activity in the manner or within the range considered normal for a human being.”

The social model takes into account the environmental context of the disability. It presumes a wide range of abilities in the population rather than a deviation from the norm. In this philosophy, the solution is not bringing an individual up to a certain norm,
but providing options to access for everyone. “Impairment” is defined as the functional limitation caused by physical, sensory or mental impairments. “Disability” is then defined as the loss or reduction of opportunities to take part in the normal life of the community on an equal level with others due to physical, environmental or social barriers. The public infrastructure (physical and informational) drives the environmental context that needs consideration here. The core issues very quickly shift from the ability to walk or the ability to see to discussions about access to information and participation. As an analogy, the problem shifts from the wheelchair to the stairs that become barriers to entry; “disability is a cultural phenomenon then, not an individual characteristic” [Kel05]. It should be recognized that “there is not a particular norm or way of being human that is better than others, but that the full range of diversity is equally valid and valuable.” [HMK07]

While we recognize the entrenched dominance of the medical model today, we seek to emphasize the importance of the entire ecosystem of functioning and the acknowledgment of a diverse range of human capabilities, and the need for equal respect and access to participation in the daily life of society.

1.2.2 Definitions of Blindness and Visual Impairment

To discuss issues surrounding visual impairment and blindness, we must first define what is meant by these terms. Unfortunately there is no single standard definition that
qualifies either visual impairment or blindness, though there are several perspectives on these experiences.

Blindness is the lack of visual perception as caused by physiological or neurological factors. For administrative purposes several complex definitions have been introduced to quantify the extent of visual impairment and describe two categories: legal blindness (properly termed statutory blindness) and low vision.

Legal blindness is defined as visual acuity of 20/200 (6/60 in meters) or less in the better eye with the best corrections possible. Practically, it compares to standing at 20 feet (6m) or less from an object, while using best corrective lenses to see with clarity of a normally sighted individual at 200 feet (61m). Another metric addresses the confined field of view, often described as tunnel vision. The statutory classification for blindness is a narrow field of view of 20 degrees or less, where the normal is 180 degrees. In the United States, about 10% of those deemed statutorily blind have lost all vision. They have no vision at all. This is considered total blindness or as having "no light perception" (NLP).

Since visual acuity is a continuous spectrum of perceptive capability low vision covers the range between legal blindness and "normal" vision. Low vision is used to describe those with visual acuities ranging from 20/70 to 20/200 (again, with best corrections possible). The best corrections mean corrective mechanisms available to the individual, such as glasses or contact lenses.
This definition fails to consider visual field restrictions. Visual disorders such as macular degeneration cause a central field visual loss, causing both definitions to break down in some cases. The recent increase in traumatic brain injuries, TBI, causes a further ambiguity when attempting to qualify these persons as blind, or as having low vision.

1.2.3 Global Population Statistics

There are an estimated 285 million visually impaired people worldwide [WHO12]. Of those, 39 million are legally blind and 246 million have low vision. 90% of the world’s visually impaired live in developing countries. 80% of the total global visual impairment is due to preventable causes. Globally the major causes of visual impairment are: uncorrected refractive errors (myopia, hyperopia, and astigmatism), 43%; cataracts, 33%; and glaucoma, 2%. The leading causes of avoidable blindness and visual impairment are: cataracts, trachoma, river blindness, vitamin A deficiency and retinopathy of prematurity.

Up to 80% of visual impairments worldwide are preventable, curable, or could be addressed with glasses or other low vision aids. Projects like Vision2020 focus on reducing the visual impairments from infectious diseases and report steady rates of success.
About 65% of all people who are visually impaired are aged 55 and older. This age group comprises about 20% of the world’s population. With an increasing elderly population in many countries, more people will be at risk of what is termed, “age-related” visual impairments.

An estimated 19 million children are visually impaired. Of these, 12 million children are visually impaired due to refractive errors, a condition that could be easily diagnosed and corrected. 1.4 million are irreversibly blind for the rest of their lives [WHO12].

Visual impairment worldwide has decreased since the early 1990s. This is despite an aging global elderly population. This decrease is principally the result of a reduction in visual impairment from infectious diseases through concerted public health action.

1.2.4 United States Population Statistics

According to the 2011 National Health Institute Survey, 21.2 million American Adults age 18 and older reported experiencing significant vision loss – about 7% of the entire population of the country [NHIS11, SP12]. This survey does not include any institutionalized persons nor homeless persons. The lack of institutional data limits the validity of any estimates for persons 65 and older.
It should be noted that the self-declared visual loss meets no existing definition for low vision or blindness. The relevance of the data is in question, as there is no verification or validity check for this data. It is cited here because it is used prevalently throughout government programs.

Of those reporting visual problems, 12.5 million (59%) are women and 8.7 million (41%) are men. 15.9 million (75%) of those affected are between 18 and 65 years old and 5.4 million (25%) are 65 years and older.

Approximately 10.0 million of Americans who have vision loss are married, 2.4 million are widowed, 3.6 million are divorced or separated, 3.4 million have never married, and 1.9 million live with a partner. Of Americans who have vision loss and are 25 years of age and over, 4.3 million have less than a high school diploma, 5.8 million have a high school diploma or a GED, 5.6 million have some college education, and 3.8 million have a bachelor’s degree or higher.

Approximately 9.6 million people with vision loss in the U.S. have a family income of less than $35,000, 3.1 million have a family income between $35,000 and $49,999, 3.0 million have a family income between $50,000 and $74,999, 1.5 million have a family income between $75,000 and $99,999 and 2.9 million have a family income of $100,000 or more. Approximately 8.3 million Americans with vision loss are poor or near poor.
Few visually impaired persons travel independently. While comprehensive and up-to-date data is lacking, reports from 1992 and 1994 indicate the dramatically low numbers. In 1990, approximately 109,000 people with vision loss in the U.S. used long canes to get around [NHIS92]. Just over 7,000 Americans use dog guides. Annually, approximately 1,500 individuals graduate from a dog-guide program [EE94].

Data on computer use among the visually impaired is also outdated. As of 1999, at least 1.5 million Americans with vision loss use computers. The total number of people ages 15 and older with any "limitation in seeing", who report they have access to the Internet is just over 1.5 million (1,549,000). The number of people ages 15 and older, with a limitation in seeing, who use a computer regularly is a bit under 1 million (979,000). Of those, about 196,000 people with a "severe" limitation in seeing have access to the Internet, and about 102,000 persons with a severe limitation in seeing use a computer on a regular basis. Data on mobile phone usage among the visually impaired population is unavailable at the time of this research.

1.2.5 Causes of Blindness in the United States

The following five causes of blindness account for 90% of all blindness: age-related macular degeneration, diabetic retinopathy, cataracts, glaucoma, and retinitis pigmentosa. All other sources, such as accidents, toxicity, infection, etc. add up to 10% of blindness in America. The following section explains the major agents related
to visual impairment in the United States and the differences from the rest of the world. Most of the infectious causes of blindness have been eradicated in the U.S.

Age-related macular degeneration is the leading cause of vision loss in Americans aged 60 and older, affecting an estimated 10 million people. AMD blurs the sharp, central vision needed for activities such as reading, sewing and driving. It is a painless disease that destroys the macula, the part of the eye that helps one see fine detail [NEI10].

Diabetic retinopathy is the leading cause of blindness in American adults and the most common diabetic eye disease, affecting an estimated 4.1 million adults over the age of 40 in the United States. It is caused by changes in the blood vessels of the light-sensitive retina tissue in the back of the eye.

Cataracts commonly decrease vision in older adults and can be treated with corrective surgery. By age 80, more than half of all Americans have had a cataract or cataract surgery, amounting to 1.35 million cataract operations annually in the United States at a cost of $3.5 billion. Cataracts are a clouding of the lens that is typically related to aging, though not necessarily. [NEI10].

Glaucoma is a group of eye diseases that affect an estimated 2.2 million Americans. Glaucoma occurs when the normal fluid pressure inside the eye slowly rises, which can damage the optic nerve and decrease vision. African Americans over age 40,
everyone over age 60 but especially Mexican Americans, and people who have a
family history of glaucoma are at a higher risk for this eye disease [NEI10].

We would like to point out that this listing of eye disorders covering 90% of all
blindness in America impacts primarily a population of older Americans. It is at odds
with the NHIS data, which suggests that this population is less impacted. The
differences can be partially determined to be due to the ambiguity of definitions, and
protocols used to gather the data for the NHIS.

1.2.6 The Blindness Experience – The Mars Colonists

It is difficult to convey the experience of living one’s life blind. To complicate matters,
the life experiences of blind people are as disparate and fragmented as the different
people with different backgrounds, opinions, educations and philosophies, that
integrate to form life experience. The diversity preempts a common minority-
experience and makes it even harder to communicate to others [Zol81].

Gary Kelly has written an essay that helps bring closer the concepts of living blind to
those unfamiliar with this experience [Kel05]. Using a cultural sciences perspective of
participant and observer, he suggests a thought experiment, where at some point in
the future the United States has colonized Mars and is debating support to the
colonists on Mars. Mars Colonists—the participants, have decidedly differing views
from those on Earth—the observers. The debates touch a number of relevant issues briefly summarized in the following paragraphs. The point to the essay is that Mr. Kelly recommends a community model that focuses on interactions of people with environment, as opposed to having the person be the unit of analysis, which is more traditional to psychology. We highly recommend reading the full essay [Kell05].

1. The experience on Mars may be alien to those on Earth but it is no less human.
2. Without living on Mars, others have no method of direct understanding of the experience.
3. There is no unified profile of a Martian – blindness touches everyone, and exaggerates their individual biases, lacks, and viewpoints.
4. Prevention is good, but does not equate with support of those already affected and living on Mars.
5. Projects often initiated by engineers and scientists generally benefit mostly the authors. Solutions most often collect dust, and the cycle continues.
6. "Mars Colony" is here to stay. It is an artifact of human society
7. Citizenship is earned not conferred. One does not simply become a Mars colonist; one must work hard after arrival to Mars to learn to thrive in the new and harsh environment. It requires adaptation and acceptance to challenges of Mars—bringing about value awareness, value reformation, and a personal adjustment to a new ecology.
8. Majority rule, when a segment of the population is 1%, majority rule amounts to tyranny.
9. Consider building better tools based on requirements and specifications of the Mars colonists.

10. Science needs adequate descriptions and data to solve problems. Most projects fail due to error when gathering the initial descriptions, or the description was inadequate to support the explanations claimed. Acknowledge past failures in terms of their adoption.

The full essay articulates a framework for a programmatic approach design and implementation of solutions based on a descriptive model of observer and participant communication.

1.3 History and Related Work

Throughout history and across different cultures, people who were blind were marginalized and largely excluded from participation in the mainstream society. Centuries of exclusion left a mark on how visually impaired people are perceived today. Social education, tools and access to information, all contribute to slow equalization of these relations. This section introduces a historical perspective and a contemporary review of related projects.
1.3.1 A Deeper Historical Perspective

Ancient cultures have largely excluded persons who are blind from participation. Children who experienced blindness were often killed or abandoned shortly after birth. Disregarded by the community, they were left to fend for themselves mostly living as beggars and outcasts. Some fortunate individuals managed to secure roles as singers, musicians and bards. Lack of access to educational materials significantly limited opportunities to grow and contribute to the fabric of society – and history telling stories and music offered possibilities for such expressions. In China and Japan entire guilds run by persons who were blind ensured that high levels of skill and the necessary education passed on to new members.

In Europe, since the 10th century, private and religious almshouses have been the primary institutions generally associated with assistance to persons regarded as poor or blind. As blindness was often seen as caused by moral shortcomings the extent of education offered was limited in scope if at all considered. In the 13th century, Louis IX of France initiated the first state-supported almshouse, the Almshouse of the Three Hundred, to care for the veterans returning from the Seventh Crusade [Koe04]. The fundamental significance of this institution was secular and unique in that it took a social approach to disabilities and a drive towards social participation and education rather than seclusion and separation [Whe02].
The education remained a problem for a long while, at least partially due to a lack of written materials accessible to those without sight. During the 16th century, Girolamo Cardano, an Italian mathematician developed a system for reading and writing by tracing embossed metal and wooden plates. Subsequently Padre Lana-Terzi evolved it into a regularized system with grids and punctuations, facilitating coded communication [Sta13]. The printing of the first embossed books is generally regarded as the beginning of the modern era of education for persons who are blind. By the end of the 18th century, Valentin Haüy established the National Institute of Blind Youth in Paris and started producing first versions of large-scale embossed print books that would allow people to physically read the written words. Reading these books required of practice and patience since each letter had to be analyzed haptically. Each book was also heavy as the pages were large, thick and each one with two sides glued to it. The font used for the typesetting exhibits unusual ornamentation, possibly to enhance discernment between the letters.
Louis Braille, who was an ambitious student at Hauy’s institution, was an avid reader who quickly mastered all fourteen books available at the school. While at the academy he learned of the Night Writing system of communication used by Napoleon’s soldiers to help disguise their positions at night. In 1839, he quickly identified optimizations that made the system more appealing to people who are blind; 1) using just six pins so that each grouping could be read with one finger touch and 2) using the groupings to represent orthographic letters rather than symbols. While adopted immediately by the colleagues at the institute, the school officials banned it, seeking to replace it with another embossing system. Due to a near rebellion, the Braille system was finally recognized and in 1854 became the official communication system of the blind in France. Despite its many advantages, it didn’t become popularized internationally for many years and even then it remained in competition with several other systems. One
of these was a hybrid between embossed letters and patterns of pins – large but simple shapes – this was the Moon alphabet, developed by Dr. Moon of Brighton.

The light sensitive properties of the chemical element Selenium (Se) – referred to as the Moon element was discovered in the 19th century. While the material itself was first discovered in 1817, Willoughby Smith explored its sensitivity to light and related change in electromagnetic resistance in 1873. Graham Bell used it to develop the first wireless audio communication device in 1880 – the photophone [Bel80]. At the same time Perry and Ayrton among others started publishing proposals for electric vision at a distance – the first attempts at capturing still and moving imagery with electric signals. By 1912, Fournier d’Albe constructed the first optophone – a device for sensing light at a distance and translating it to sound [Alb14]. Subsequently he developed a version that could facilitate reading print illuminated from below [Alb20].

Different sound frequencies would be produced by different letter patterns, essentially becoming a sonification of the letters as an analogue to Hauy’s embossed letters. The concept promised to remove the need for special printing, but had the significant limitation of very slow reading rate. Generally people could achieve a speed of approximately 20 words a minute and after extensive practice a handful managed to read at 60 words a minute – making the system a frustration to use, even for those students. Since then, and to this day different versions of image to sound or image to vibration translation systems continue to be developed – sometimes for reading, but also for sensing of obstacles at a distance and general orientation.
World War I and World War II brought renewed focus to the large numbers of veterans who lost their sight. The Veteran’s Administration was established in 1930 to help administer services to the returning veterans and to advance research. During the 1940’s the Veteran’s Administration pursued active research and development of optophone-like systems – optiphone [Cli31], autophone [Pan57], vidicon reader [Doa79], and many others. Analogous image-to-vibration systems were known as the optacon. These were then actively manufactured from 1970s and distributed until 1996 [Mor75, HZT83].

1.3.2. A reality today

Technology aside, the daily lives of persons with visual impairments and blindness vary depending on age, physical abilities, and the extent of experience living with vision loss. All elements of daily living need to be accounted for – dressing, eating, functioning independently and generally keeping safe. Below are some elements that may convey some characteristics of daily functioning:

- Keeping things organized at home: Items should be kept in predetermined and regular places so they are consistently easy to find and identify.
- Most things will need to be identified by their shape, weight, the sound they make and possibly taste.
- Braille labels, small raised dots, rubber bands, Velcro, or colored tape may be used to mark items for easier identification.
• Many dials for cooking, temperature, microwave can be marked with additional raised dots to indicate critical positions

• To keep money organized, bills are folded differently to indicate denominations – for example fold the $5 lengthwise, the $10 widthwise, the $20 length and widthwise, and leave the $1 unfolded. To identify bills either a custom identifier is necessary or the assistance of another person.

• U.S. fails to provide accessible currency despite the Americans with Disabilities Act mandate.

• Larger transactions are generally performed with credit cards to ensure a paper trail in case of problems.

• For mobility, the white cane is the primary device for sensing the space while walking.

• To prevent falls, remove all clutter, throw rugs, lose wires or cords from the floor at home. Place a slip-proof mat in the bathtub or shower.

• Most common low vision accessibility aids include
  
  o Magnifiers

  o Talking watches or ones that can be opened to touch quietly

  o Large button telephones

  o Perkins Brailler machine or smaller Braille pin

Blindness is an “information” disability – where one channel of signal is missing limiting the capability for access.
1.3.3 A Contemporary Historical Perspective

Recently the applications of computer vision to extend visual accessibility have been blossoming. Nonetheless, we find that most approaches tend to focus on very narrow sub-problems and often within a limited set of constraints that are not representative of the real world.

Bigham et al [BJL10] developed an app that facilitates answering of visually-based questions through mechanical turk solutions [Amz13]. The iOS app called VizWiz allows a person to take an image and record a question about it to be answered by the social network. The preferences allow control over the choice of social networks to deploy the question. They include email, web workers, image recognition engine IQEngines [IQ13], Twitter, or Facebook. The app is simple to use and delivers satisfactory results. Since the web workers are humans participating in an online marketplace for their services, VizWiz has created a complex structure of incentives to motivate them to answer questions in a timely fashion. Without it answers may arrive with delays ranging from seconds to minutes (or even hours) depending on the price of the query. With the managed incentives, the answer arrives in under a minute. If the item in question wasn’t properly framed in the photo, another query must be sent. Individuals concerned about their privacy do not use this application, as there are
limited safeguards around their image data once it’s in the cloud. The solution was introduced in May 2011.

TapTapSee [Tap13] is another object identification app released in January 2013. It features a minimal human interface with just three buttons and a preview of the video. A double tap initiates an image recognition query to an automated image recognition system and should that fail to the mechanical turk network [Tag13]. The simpler interface makes the application easier to use and accessible to a broader population.

Silapachote et al. [SWH05] and Mattar et al. [MHL05] propose a sign recognition system for the visually impaired. It is a preliminary study demonstrating that sign recognition is possible, at least within a moderate subset of images. Their solution relies on a high-resolution still camera, which they suggest is mounted on the head of the participant. Unfortunately that is a very obtrusive solution for people. Additionally, the head orientation of many visually impaired people is a mediocre guide for orientation due to tendencies to sway. The lack of real-time recognition also makes this system cumbersome to use.

For more than 40 years, persons who are blind have protested the lack of esthetics, and have stated consistently that any device that makes them appear strange, unusual, or otherwise attracts undue attention, is a non-starter [Pul09]. This was an issue with the Sonic Guide [Bra82] of the mid 1970s. It has been a consistent advantage for the iPhone, as an iPhone is popular with the general public.
Coughlan and Manduchi [CM09] present an automated mobile way-finding system based on strategically placed colored markers. While the system works within a controlled environment, it is difficult to generalize and extend on a larger scale. Attempting to intervene with an existing environment, especially the private property of others is a tricky and typically very expensive endeavor. We find the infrastructural cost of this approach to be problematic. With a different approach, Kosecka et al [KLY05] successfully demonstrated the feasibility of using scale-invariant (SIFT) features [Low04] for global localization and relative pose estimation. While their work is developed within the context of robotics and a comparably dense dataset, their
results have direct impact in our application. In a related extension, Hile et al. [HGL09] combined GPS information and image-based reconstruction to create a model of the desired environment, to then compute a navigational path for a pedestrian.

Belongie et al. [Bel07, MGB07, WCB10] have been working on a shopping assistant for the visually impaired. Their multi-faceted approach reveals some of the complexities and depth of the problem. In designing the hardware from scratch, they gain the ability for more refined haptic feedback to aid with localization of the items.

Chen and Yuille [CY04] have proposed a framework for real-time text detection and reading text in city scenes. Their work aims to serve text-based information to a visually impaired person while navigating an urban environment. Their approach assumes that the person can anticipate a text region of interest in the environment and is capable of aiming a still camera at it to send to a server for processing. Means of data transmission are not discussed. The results are very impressive; however, text-only processing and reliance solely on still images limit the system. In July 2012, Blindsight released the TextDetective App [Tex13] featuring the text detection and recognition algorithms based on the work of Chen and Yuille.

Google Goggles is an image-based search engine from Google [Goo13]. The engine performs text detection and recognition combined with image-based recognition to determine the content of the image. Google provides the Goggles functionality in their Google app on iOS. Unfortunately, at the time of writing, this capability is not
accessible to the visually impaired people. The implementation does not support the VoiceOver interface and all results are announced as “button.”

Focusing more on the computational limits and tradeoffs in using mobile handsets for computer vision applications, Ta et al. [TCG09] make use of more computationally-friendly feature descriptors (SURFs) and track them between frames. They are able to conserve space and computational power to demonstrate real-time object recognition and label augmentation on a mobile device. While this system works very well for small or medium size databases, the mobile phones runtime memory capacity determines the upper bound on the size of the databases. In response, Chandrasekhar et al. [CCL09] suggest yet a different feature computation, even more efficient and compressible, and then send that information off to a remote server for lookup.

As an alternative to simply sending features over the network, Chen et al. [CTV09] try to capitalize on motion-based compression in video to only send clear and stable key frames to the server for recognition, while performing local real-time tracking on the mobile device. Over Wi-Fi, their results demonstrate successful recognitions in 1 second per stable frame of 320x240 pixels. By comparison, the LookTel system performs the same recognitions at the same resolution at an average of 5 frames per second.

The reality of computer vision applications for the visually impaired has not yet seen significant adoption, with most people relying on low-tech but very reliable
“technologies” such as the cane, guide dogs, folding bills, and using rubber bands and safety pins to identify objects. While there are mobile OCR readers on the market [KNFB13], their cost and limitations make them inaccessible to most. Dedicated laser barcode readers can identify various items, but are bulky, expensive and address only the items with barcodes.

To address these limitations, we combined some of the best implementations of computer vision available today, and present an affordable, accessible, and comprehensive platform for computer-aided visual assistance that runs on a wide range of off-the-shelf smartphones offered by almost all wireless operators.

2. LookTel, version 1.0

We sought to address the problem of computer-aided visual assistance in a real world context, taking into account as many variables as possible—from practical and technical challenges to social implications on design. The context included potential cost, the current state of commodity hardware, connectivity considerations, current low-tech approaches, and the habits and preferences of potential customers who are blind. The system prototype delivered real-time object recognition information, performed OCR for text-based information, and enabled possibilities for sighted
assistance. One purpose of LookTel is to provide an extra "pair of eyes" at critical times to persons who are blind.

2.1 Scope of the project

Based on the results of SeeStar phase 1 surveys and interviews the following five points describe the goals of the LookTel 1.0 platform:

1. Develop the LookTel 1.0 platform, and deploy it with visually impaired subjects and their sighted assistants in order to evaluate the following:
   a. Identification of money denominations
   b. Assistance of a sighted friend or family member to remotely train the system for shopping list items that may not be in the training database.
   c. Finding a credit card that may have been misplaced and mixed with other items on the desk by mistake.
   d. A simulation of walking across a super market aisle and identifying a specific list of 20 items on the shelves such as cans, bottles, jars, boxes, and CDs; getting orientation queues such as to the right, to the left, higher, lower in handheld, chest and head mounted scenarios.
   e. Simulate using the system at home to organize the items in mockup kitchen/bath room scenarios.

2. Develop unique images that can be printed on mailing labels to act as tags
when attached to any object of interest that is otherwise unrecognizable such as frozen food, glass containers, medication bottles, etc.

3. Complete the mobile PDA phone/PC based beta system, and an accessible human interfaces for both the PDA and PC to demonstrate the performance over broadband wireless networks to command and get feedback from the system in performing the assistance tasks identified and validated in items A through E of goal 1.

4. Develop an accessible touch screen human interface, which allowed the use of touch screen devices such as the HTC Touch, HTC Touch Pro or, at that time, the upcoming HTC Quartz.

5. To enable a PDA-Netbook based system with no need for a broadband connection. The PDA connected to the Netbook running the object recognition engine and database over a peer-to-peer Wi-Fi connection.

6. To evaluate a beta system on national wireless networks and Windows Mobile based PDA phones along with Netbooks or personal computers as the LookTel recognition and training servers in a human performance evaluation with selected participants and their assistants. The participant evaluations along with laboratory tests by developers validated the satisfactory performance of the system in real world conditions for the assistance tasks mentioned in Goal 1). Subjects and trainers were allowed to live with the LookTel systems to evaluate them with a wide variety of objects and tags. We evaluated usage patterns and determined effectiveness versus alternative methods used via automatic system usage logging, along with regular, semi-structured interviews
during the evaluation period, to determine individual perceptions of barriers to use and changing perceptions as each individual participant gained experience, confidence, and started to incorporate the LookTel system into daily activities. We asked participants to write personal reports of their experiences within a template of topics that correspond to the tasks that each performed with LookTel regularly. The template encouraged the participants to be responsive to thinking of the system and its use, while providing a free forum for individual experiences and perceptions. Participants only interacted with the research team, and not one another, in order to isolate specific experiences, incidents, and perceptions from others in the group or dominant members of a group. This is consistent with crowd-source methodologies. This methodology was chosen in order to gain a rich description of factors not anticipated by the research team.

2.2 System Architecture

The LookTel 1.0 Platform consisted of two major elements—a BaseStation and a mobile client. The mobile client was responsible for sending live video and still images to the BaseStation for recognition processing, which in turn provided the computed results. Both clients were fully capable of live two-way audio/video messaging.
2.3 iVisit

The LookTel 1.0 Platform was conceptually built around the use of a cell phone camera and high-speed data link to a remote assistant. To facilitate providing an extra "pair of eyes" we sought a capable voice and video streaming solution. The platform is designed so that when a person who is blind wants an “extra pair of eyes” to obtain a useful piece of visual information, that person can take out the cell phone and call a remote assistant. After looking at the person’s surroundings through the lens of the cell phone camera, the assistant provides the requested information, which may be a description of the setting, the location and direction of a salient landmark or destination, a description of the pathway ahead, etc. In addition, the cell phone provides GPS information to the assistant’s computer, which then automatically
displays a Google Map satellite image of the caller’s setting with the caller’s position. Using this, the assistant can help orient the person to nearby locations. The traveler who is blind is an active link in this process, interacting via voice with the remote assistant, providing both a context for what the remote assistants see through the lens, and at times, providing location information, should the GPS fail to do so. Additional information can be provided to the operator based on experiential information possessed by the traveler that the remote assistant may not have, or may not realize from the limited camera images. That knowledge may give the remote assistant enough information to confirm the information through Google Earth, street views, etc.

iVisit [iVi13] is a pioneer in the field of IP video conferencing and has operated one of the first internet IP video conferencing services with over 2 billion minutes and 4 million downloads to date. iVisit currently hosts over 20 million minutes of multiparty video conferencing per month for free using its unique p2p network architecture. The transport protocols are highly optimized for variable- and low-bandwidth connections making them well suited for variability of mobile 3G connections. The efficient computational load of video encoders permits real-time video transmission from mobile phones. Typically, a 160x120 video streams at 8-15 frames per second, while a 320x240 video streams from a mobile at 3-8 frames per second.” iVisit is the core underlying technology powering the LookTel platforms, and the ingredient making this project possible.
2.4 A Real-Time Object Recognition System

The core of the object recognition system consists of a state of the art recognition engine from Evolution Robotics [ER10] in conjunction with a meta-information database. We created tools to capture and annotate new items easily, and to share those existing libraries or elements. Items can be annotated with written text or recorded speech, whichever is easier or more fitting for a specific item.

We have chosen to use a Scale Invariant Feature Transform (SIFT) -based implementation due to its outstanding performance and versatility in a number of circumstances as outlined in the following paragraphs. SIFT works across several scales, allowing flexibility of distance for recognition. It is robust to rotations and moderate viewpoint and illumination changes. The image recognition engine is based on the design by David Lowe [Low04].

The features are stored in k-d tree structured databases, which allow rapid nearest neighbor search while retaining very high recognition rates with minimal false positives, even at hundreds of thousands of items. SIFT consistently ranks at the top among image recognition algorithms.
Besides recognizing products, the functionality can support limited landmark and location recognition. The engine only supports recognition of patterns on 2D planes so any actual 3-dimensional space is approximated by this technique. To overcome this limitation, multiple images from different viewpoints can be combined to represent a single location. We facilitate 3D object recognition by allowing views of a product from all sides and unify the results as a single item.

Many situations not covered by SIFT can alternatively be addressed by use of simple tags—pattern-printed adhesive stickers. These stickers provide the features necessary to correctly identify an item that otherwise may not have enough unique features on its own, such as a clear glass jar or a Tupperware container.

2.4.1 Real-Time Recognition and Micro-Cloud Computing

The incoming video feed is analyzed frame by frame discretely and independently. The LookTel Platform on the PC then computes the features and then looks them up in the currently loaded databases. The data rates were good, as the system was able to get 4-14 frames per second of recognition on the HTC Tilt mobile devices over a 3G network. The frame rate also depends on image resolution (e.g. 160x120 or 320x240) and network conditions. The typically observed round-trip delay times are around 100-300ms.
2.4.2 Training with the LookTel 1.0 Platform

On the LookTel 1.0 platform, adding new items to be recognized consisted of taking images of that item and associating them with textual or audio annotations. Ideally, the images captured the overall appearance of the object, which could involve taking photos from multiple viewpoints. Typically front and back views sufficed. Circular items such as food cans often required three or four images depending on their size.
Once an item was saved in the database, it could be edited or relabeled within our application.

The training process assumed at least partial involvement of a sighted collaborator. The collaborator, who could be remote, was at some stage necessary to guarantee an acceptable capture of the desired object—the framing, general lighting, and quality of the images.

To satisfy a diverse set of use cases, we provided multiple means of training new items, subject to circumstances and convenience:

- Directly from the mobile—A sighted assistant is handed the participant’s mobile device and can capture photos and annotate them directly from the participant’s device. A menu option allows the sighted person to see a live preview through the grid human interface (HI).

- Person on mobile – A sighted assistant operating the BaseStation – allows a physical separation between the partners. The sighted operator of the BaseStation, informed by the live video stream, guides the mobile partner to adjust for capturing the best image. The BaseStation operator then takes a photo remotely and annotates appropriately. The BaseStation operator may be at the partner’s own BaseStation or remotely logged in to the partner’s home BaseStation.

- Decoupled – items – A library of items is requested from a person who is blind via email. The sighted assistant builds a small database of items, and exports the library as a zip file to the requester.
• Experimental – In-house only – There is no sighted assistant. There is a structured camera—the camera mounted on a specialized stand with a predictable field of view, the participants who are blind, try to capture images of items themselves.

• This experiment with open loop control provides for a degree of independence while increasing the error rate, and possibly the frustration and lowering the satisfaction of the person who is blind. Having a human in the loop has its advantages and disadvantages, but overall provides a level of necessary robustness and self-correction to the system. The lack of supervised, top-down enforced quality control over the training data appears problematic. It is likely that as soon as the operator who is blind notices that there are recognition problems with particular items, she/he can investigate, diagnose the issue and often correct through an adjustment to the system.

2.4.3 Location and landmark recognition

As demonstrated by Kosecka et al, SIFT-based recognition is suitable for landmark and location recognition [KLY05]. While the larger issue of generic location recognition (both indoor and outdoor) is a complex and multi-faceted problem on its own, we addressed the problem from the human participant’s perspective. Several indoor locations were used as intermediate waypoints, and were added to the system as if they were generic objects. Their annotations describe their locations. Key landmarks
are sufficient to provide helpful orientation cues—both indoors and out. To help guide a person who is blind through an office building, one takes several intermediate photos along the way and annotates them while walking the route.

### 2.4.4 Tags and Tagging

![Sample tags](image)

Figure 5: Sample tags.

Some items are not suitable for the SIFT-based recognition engine. In most cases this is due to a lack of unique distinguishing features. Examples are Tupperware containers, clear glass jars, or medicine bottles. To overcome this limitation in the
LookTel 1.0 Platform, we created a series of 1.5” and 3” round, reusable vinyl stickers with printed images. The images that were printed on the stickers are already trained in the initial database with a plain description of “new tag”. To apply a tag to an item, one peeled a new sticker from the backing, and applied it to the object. Once the item was first recognized by the software, the person could record an altered audio description to be associated with that item using the mobile phone. When the item was recognized in the future, the newly recorded audio description was played. For consistency, the audio description was stored on the BaseStation and the mobile. In general, any item could be renamed on the fly directly from the mobile. After recognizing an item, the customer was able to simply record a new description, which is automatically associated with that last-recognized item.

The tags struck a balance between being optimally recognizable and being relatively aesthetically pleasing. While we have experimented with various 2D barcodes, as well as random and structured noise patterns, we settled on a selection of images that are meaningful, and even playful to people. We are aware of the importance of having a socially normative outward appearance. While an item with a black and white QR code prominently indicates another level of meta-information, our labels preserve some subtlety and nuance in the aesthetic domain. As for their functionality, they have been chosen for their high fitness both individually and as a collection (local and global), as well as their robustness.
In addition to extending automatic object recognition to difficult-to-recognize objects, the tagging system allows instant usability of the system, right out of the box, without requiring any visual assistance whatsoever. Additional tags and corresponding databases can be added to the system incrementally with database imports. New sticker sets can be compiled by the people using the system, or from a more organized source.

2.4.5 Object Recognition database management

In developing the LookTel 1.0 Platform, it became evident that many advantages might be garnered from having a centralized database of locations and descriptions. We suggest that it is advantageous to engage the community at large to build a comprehensive database of items. As LookTel was lacking the facilities and infrastructure to create and host a global central repository of “everything,” Our application has been designed to make collaboration and contributing new items or collections of items very easy. The XML-based meta-information structure is generic and extensible, and anyone may use it to share more items. Our application facilitates managing, importing, exporting, and subscribing to database updates. We are continuing to investigate various approaches to best grow these databases, including constructive crowdsourcing.
2.5 OCR

Another major component of visually transmitted information is written text. We have integrated the state of the art optical character recognition (OCR) engine from ABBYY [ABB13]. In addition to providing very accurate OCR, the engine includes two very significant components—specialized processing of photo-based images and document structure reconstruction. Photo-based images require very specific pre-processing for making them suitable for OCR such as detection of page orientation, skew-correction, straightening curved text lines, and a recently implemented 3D perspective distortion correction. Reading text and especially structured text allows for much richer interactions with the world on the fly. On the mobile, we’ve created a specific set of controls to quickly navigate a set of digitized pages, down to the character level.

To use OCR, persons who are blind take a still photo using our touch Human Interface, HI. This presents some problems due to accurate pointing and framing of the target text. While we currently do not provide automated guidance for precise alignment of the printed text and the camera, the back-end processing attempts computation at various orientations and allows for much greater flexibility compared with the mobile KNFB reader. Greater computational power allows for better parsing of structured text such as newspaper articles, utility bills, or restaurant menus.
2.6 Remote Sighted Assistance

The iVisit base-platform’s live peer-to-peer video conferencing capabilities allow for various remote sighted assistance tasks. LookTel offers two-way full-duplex audio/video communication over IP. People using the system can call a friend or assistant to show them something with their smartphone, or have them read a handwritten note, by transmitting live video from the smartphone’s camera. The assistant can then use the video feed to guide the caller and get a sense of the scene or particular details of an object. Additionally, the remote assistant can trigger a high-resolution photo capture from the mobile device once the correct view is established. The high-res photo (up to 5 megapixels on some devices) is then transferred to the assistant’s desktop for a closer inspection. The combination of real-time video with two-way audio is the key aspect enabling this “remote-eyes” assistance. While it’s a natural extension of eyesight through technology, it also depends on another human, making it less appropriate for day-to-day assistance due to the high social and economic costs of interaction with other people. Whether the assistants are family and friends, or strangers in a call center, persistent sighted assistance for everyday tasks is not always feasible. This mode of assistance is crucially important in emergency situations where it is especially difficult to provide, particularly when a person is disorientated in an unknown environment, during night hours, or an environment made unknown by rain, wind, snow, or smoke—or even loud noise, as too often may exist in emergency situations. In brief, it is likely that for more than half
of the time when the system is most needed, it cannot be available because of its very nature.

2.6.1 Mobility Assistance

We have performed a series of studies of remote way finding assistance. In these scenarios, a visually impaired person seeks the assistance of a remote sighted assistant for help with one of the following problems: a) identifying their current location and b) establishing a path and direction of travel to a destination, or an incremental waypoint towards one. Once the caller enables the GPS receiver on the mobile handset and the location coordinates are automatically transmitted to the assistant’s computer and displayed in a Google maps environment with full-featured zoom, satellite view, and map view capabilities. Combined with real-time audio/video and the ability for high-resolution image capture, the remote assistant can relatively quickly convey the surrounding environment to the caller. While the GPS element is very useful, it is subject to signal ghosting and deviation in heavily urban areas and should mostly be used for macro localization. For indoor use and fine-resolution location information, the video feed provides the most significant information.

Given the high level of potential noise outdoors, we find that an open ear, sports-designed AirDrives headset performs best in terms of audio quality and volume, while still providing the best comfort and not falling out of people’s ears. The headsets leave
the ear canal unobstructed, allowing for constant monitoring of the surrounding environment, while the small speaker seated just in front of the ear provides clear and ample audio. The headset also has a microphone and a physical volume controller.

We must stress, however, that this should only be used in a stationary fashion, where the caller is standing still. Visually impaired people rely heavily on a learned set of mobility skills that allow them to navigate the environment based on senses other than vision. As such, they need to constantly be monitoring the space around them to guarantee their safety. This means, among other activities, keeping their ears alert, handling a cane, handling a dog, and/or performing echolocation. Speaking to a remote assistant and handling a camera phone is distracting enough that it cannot be performed simultaneously. Additionally, the real-time connection to a sighted assistant gives an illusory hint of security that cannot possibly be guaranteed—thus to initiate any mobility, the caller must retract the mobile device back in their holster, pocket, or purse, and continue to proceed based on their mobility skills. Otherwise there is potential for inadvertent walking directly into a street or into another dangerous situation. Remote sighted way-finding aims to primarily convey location and surrounding environment information to the traveler and explain the next segment of a journey to a destination. The vision the remote assistant has is less than that of a person with low vision. The camera limitations of even the best smartphones are so severe as to leave the remote assistant feeling anxious about the orientation and accuracy of any assessment made through the camera. Disorientation and confusion are real possibilities, as they are with anyone having partial use of vision.
2.6.2 Remote Training

The sighted assistant can be in a remote location while helping the person who is blind to train LookTel to recognize new objects. With the live video stream preview, the assistant can guide the person to move the camera around until the object is properly in view, and then capture the image and annotate it.

2.6.3 Remote System Assistance

Within a managed privilege system, we created the means to remotely control various applications within the system. Once a person delegates someone as her or his assistant, appropriate privileges are granted, and the assistant can log onto the person’s BaseStation computer and make necessary adjustments or set up any application as necessary. In this mode, the full application and computer screen are shared with the assistant, similar to remote desktop software. We’ve added a similar capability to the mobile phone, which allows the remote assistant to see the entire screen of the phone and control it from her or his computer using a desktop mouse and keyboard. This feature has proved critical for support to address arbitrary smartphone idiosyncrasies found in the LookTel 1.0 Platform.
2.7 System Accessibility

With the LookTel 1.0 Platform, we faced several challenges in making both BaseStation and the mobile client accessible for visually impaired and blind persons. We needed to account for diverse hardware including, mobile handsets with various physical button configurations and touch screens, desktops, laptops, netbooks, and mobile internet devices (MIDs).

2.7.1 Custom Touch Screen Human Interface

Today’s smartphones are fairly powerful computational devices, yet they are not designed with accessibility in mind, making them very difficult or impossible to use for visually impaired and elderly persons. The new designs of 2008 were incrementally removing physical button controls in favor of touch screen interfaces. The new touch screen designs do afford unique opportunities. We have developed a simple touch-based human interface based on a standard 12-button dial pad configuration. The interface is a set of 12-button pages arranged in a circular fashion, allowing the operator to navigate from page to page using the bottom three items (left page, home, right page). The remaining 9 items are assignable to various features of the application.

To make the touch-based human interface discoverable (easy to explore and to figure out how to perform desired tasks), we assigned a single tap on an item to announce
its action, and a double tap to activate it. Participants are now able to drag their finger around the screen until they land on an item of their choice. Then a double tap performs the intended action. Such a system aims to prevent accidental activations while making it intuitive and completely accessible. We used a text to speech (TTS) engine from Cepstral [Cep10] to dynamically synthesize speech. Commonly used items are cached for faster interaction.

![Figure 6: HTC Fuze with the touch screen human interface.](image-url)
The home page features the most commonly used items such as start/stop recognition, read text (OCR), call assistant, open mobile address book, open LookTel address book, and get system status. Additional controls, tools and settings are accessed using the right and left navigation items.

For mobile devices with physical navigation joysticks, the joypad controls work in parallel with the touch interface. The left and right directionals navigate to adjacent pages, while up and down directionals traverse individual items within the current page. The center (OK) button activates the currently selected item.

The touch human interface can be augmented with scotch-taped guides or Braille stickers to indicate key items. The human interface is visually displayed mostly for the benefit of assistants (physical and remote) that might be helping with the device.

2.7.2 Mobile Phone Functionality

To reduce the number of dedicated devices a person might carry, we integrated basic phone operations into the LookTel 1.0 platform. Through our touch-based human interface, one is able to access the phone’s contact list, and make and receive standard phone calls. Given that the rest of the phone is not accessible, we did assign a hardware key binding to launch our application as a part of the installer.
We integrated controls for the phone’s audio input and output management, power management, backlight controls, and basic wireless network controls. For Wi-Fi control, participants were able to specify a preferred wireless network by adding in the network details through their desktop BaseStation.

### 2.7.3 Screen Reader Compatibility

The BaseStation desktop client can also act as a standalone recognition device. A webcam was used for video input and recognitions were announced using the Windows Speech API.

### 2.8 Hardware Considerations

The BaseStation component of the LookTel 1.0 platform was designed to run on a Windows platform (XP, Vista, or 7) with a minimum 1.2Ghz processor and 1GB RAM. Machines with more RAM were capable of handling larger object recognition libraries. We have tested performance on more compact devices such as Atom-based portables—the N270/N280 netbooks and Z515/Z520 MIDs—and found their performance to match the full-scale desktop computers.

The mobile client was running on Windows Mobile 6.5 smartphones. These phones all had a touch screen interface and a rear 3 to 5 megapixel camera. Some of the
smartphones feature an LED light in addition to the camera, which in many cases improved the recognition results, and reduced the need for the participants to worry about turning lights on or off when training or recognizing objects.

### 2.9.1 Details of the implementation

The mobile devices were HTC Tilt II devices running the Windows Mobile 6.5 operating system. They were powered by a single core 538 MHz Qualcomm MSM7201A chip with 288 MB RAM and a 512MB ROM. A microSD expansion slot allowed for up to 32GB storage. The camera was a 3.2 megapixel camera, capable of auto focus, digital zoom, and video capture. Some phone models with which we have worked, included a controllable light to enhance camera operation in dark environments. The phone featured an A-GPS unit. Other models we developed for evaluation included the HTC Fuze, HTC TyTn II, HTC TyTn Pro, Tilt, and the HD2.

These smartphones were fairly powerful computational devices, yet they were not designed with accessibility in mind, making them very difficult to use for anyone blind or with a significant visual impairment. The designs were removing physical button controls incrementally in favor of touch screen interfaces. Given that Windows Mobile 6.5 (or any other version) did not contain any accessibility features that would allow a blind or visually impaired person to operate the mobile phone, we had to design and implement an accessible human interface that gave these persons complete control
over the key functions of the mobile phone. The accessibility layer, had to facilitate the
following generic phone functions:

- dialing a new number
  - direct dialing via a dial-pad
  - dialing from the address book
  - managing contacts in an address book
- receiving a phone call (while retaining accessibility as the phone call is in progress)
- powering off, setting to standby and resetting the phone
  - the hardware switch by itself was not enough as otherwise the OS required GUI dialog confirmations
  - reading of battery level, and charging status
- switching between 3G HSDPA+ connections and WiFi
  - WiFi on/off
  - WiFi password capability for protected connections
- volume control
  - hardware button integration
  - switching between speaker (general accessibility) and phone calls (ear piece low volume)
  - support for external headphones
- integration with various physical keys on different model phones
- light on/off (for supporting devices)
• GPS on/off

The human interface had to support the following functions—all the additional functions related to the recognition and remote support, and finally it had to work via a touchscreen. In consultation with both of our human factors engineers and several participants with visual impairments and blindness from the alpha tests, we designed a simple touch-based interface based on a 12-button grid of a telephone dial pad.

2.9 Performance

The absolute performance of the system was difficult to assess given the dimension of variables involved in non-laboratory, “real world” circumstances. Most importantly, our participants deemed typical recognition speeds of around 1 second, acceptable. The system ran at two resolutions on a remote (3G mobile or Wi-Fi) connection, subject to connection conditions:

• 160x120 pixels—typically ran at 8-15 fps, at the expense of lower recognition performance. This mode is most suitable for remote mobility assistance where responsiveness is more important than image quality.

• 320x240 pixels—typically ran at 3-8 fps, and provided significantly more reliable recognition results. When the system was run locally with a good high webcam we surpass 15fps, allowing the system to perform as a detection tool—beyond just recognition, it could be used to locate the items in space. The high frame
rate allowed the participant to zone in on the element in the right region. The recognition rate varied between 80-100% recognized [ER10, MPD06]. The system was robust to viewpoint changes, lighting changes, and compression noise. Even under significant occlusions (at times up to 90%) the system could identify the item correctly. The recognition database could support thousands of distinct items without noticeable impact on performance in speed and recognition [MPD06]. It is difficult to give an absolute comparison metric for the recognition accuracy, since its highly dependent on the contents and quality of the database. Packaged goods and other items with unique graphical labels worked very well, while items lacking distinctive characteristics were more problematic. The tags we selected were rich in discriminating features, resulting in 100% accurate recognition. Initially, it took a short amount of time for the visually impaired and blind participants to aim the smartphone in a manner that accurately captured the desired field of view. The real-time feedback from correct recognitions quickly helps them to develop accurate proprioceptive skills. Another important aspect to consider was the cooperation (and sometimes the lack thereof) of the human participant using the system. In cases of incorrect recognitions, the mistakes were so nonsensical that people disregard them and adjust the viewpoint a small amount to try from another angle. Hearing two different recognitions for one single item was an indication of a potential recognition conflict. In these cases, the participant was to investigate to try to achieve a stable result.
2.10 Real-world beta trial of LookTel 1.0 and results

During the development several blind and visually impaired individuals participated in rapid iteration development cycles to bring the system to a functional state. The next stage was a real-world trial, where a variety of individuals were introduced to the platform without any prior experience with it and evaluated it in light of their everyday experience.

2.10.1 Scope of the trial

The trial attempted to evaluate the technical performance of the overall system and the practical implications from daily use by a variety of people within the context of their normal lives. The trial gauged the level of hands-on support required for successful implementation and use, thus exposing the primary shortcomings of the platform as a practical solution.

The trial lasted two months. Participants conducted the evaluation from their home and familiar environment. Within the Wisdom of the Crowds methodology, absolute independence and isolation of communication between the participants was paramount to preserving valid personal reactions and preventing a “groupthink” or
herd mentality. Each participant received a package containing a cell phone with a voice and 3G data plan, preloaded LookTel Mobile software, a CD with the Windows installation of the LookTel BaseStation software, extensive documentation (HTML), a folder of pre-printed labels, and Air Drives headphones. The trial began with an introductory phone call during which we explained the key aspects of the software for the mobile and the desktop. Half of the participants were requested to perform the initial installation and configuration by themselves, while the other half were remotely assisted step by step to minimize time spent on the configuration and focus on the use of the platform. In both cases we supported the individuals’ efforts to completion. We spent time to ensure proper configuration and to ensure that everything was working on the individual’s personal computer in conjunction with the phone. This step included software installation, accounts creation, access to local wireless network (if available) and phonebook pre-population. With the base system operational we scheduled a series of training sessions, where we worked with the participants to remotely train items around their household. At an agreed time we called their LookTel Mobile account and guided them to point the smartphone camera at the items to train, and assist in capturing accurate photos for subsequent image recognition. We took some of this time for any support questions and necessary explanations of other details of the system.
2.10.2 Beta trial participants

The participants for the beta trial were selected from a pool of volunteers in accordance to wisdom of the crowds principles [Sur05] and with the goal of including the diversity within the population that would potentially benefit from the LookTel system. Given the multi-stage duration of the project we recruited volunteers for interviews in anticipation of such trials. Based on initial questionnaires and follow up interviews, Gary Kelly determined the final membership of the evaluation team.

The final team was comprised 11 participants across the entire United States with one person participating from Canada. The ages of participants ranged from 28-64, with an average age of 47. There were 6 females and 5 males. Most individuals were totally blind as very few had residual vision or minimal light perception. Each person’s self-described level of technical skill varied from below average to high, with the majority of persons at above average. The table below gives a more granular description of the actual breakdown. Most already have owned a mobile phone, though several individuals used a landline only.
<table>
<thead>
<tr>
<th>Age</th>
<th>Gender</th>
<th>Level of Vision</th>
<th>Technical Skill</th>
</tr>
</thead>
<tbody>
<tr>
<td>46</td>
<td>Male</td>
<td>Totally Blind</td>
<td>High</td>
</tr>
<tr>
<td>64</td>
<td>Female</td>
<td>Very low vision</td>
<td>Limited</td>
</tr>
<tr>
<td>56</td>
<td>Female</td>
<td>20/200 - 20/400</td>
<td>Below average</td>
</tr>
<tr>
<td>58</td>
<td>Male</td>
<td>Totally Blind, since birth</td>
<td>Above average</td>
</tr>
<tr>
<td>41</td>
<td>Male</td>
<td>Totally Blind, since birth</td>
<td>Average</td>
</tr>
<tr>
<td>55</td>
<td>Female</td>
<td>NLP left, LP right</td>
<td>Above average</td>
</tr>
<tr>
<td>52</td>
<td>Female</td>
<td>Totally Blind, since birth</td>
<td>Low</td>
</tr>
<tr>
<td>28</td>
<td>Male</td>
<td></td>
<td>High</td>
</tr>
<tr>
<td>41</td>
<td>Male</td>
<td>Totally Blind</td>
<td>Above average</td>
</tr>
<tr>
<td>31</td>
<td>Female</td>
<td>Totally Blind</td>
<td>Above average</td>
</tr>
<tr>
<td>49</td>
<td>Female</td>
<td>Partial (20/400)</td>
<td>Below average</td>
</tr>
</tbody>
</table>

Table 1: LookTel 1.0 beta trial participants.

### 2.10.3 Results and feedback

Below is a listing of general findings of the trial and recommendations for improvement. The subsequent section contains the high level conclusions.
1. Support time for setting up individuals with the LookTel 1.0 platform and getting them to the point of recognizing items was high. Some people required 10 hours or more. Savvy participants or participants with a family member living in the house required significantly less time. This was a minority of our test group.

2. Participants had no concept of how much light is necessary in order to obtain useful images. Explaining the proper lighting conditions in terms helpful to this population will be a necessary tutorial subject. Individuals often would work entirely in the dark or directly against a window in very bright sunlight, creating images that were not useful.

3. Orientation of the device proved difficult to teach. Participants took a long time to grasp how to point the device even after training. This may take extensive training for some participants, especially those blind since birth. This makes text recognition more difficult for these participants without a sighted assistant. Some of them have suggested camera orientation software analogous to the KNFB Reader as possibly helpful to them.

4. Participants accidently locked themselves out on the device. This happened with a decreasing amount of instances as time progressed, but was often disconcerting, causing the person to put the device down and discontinue practicing with it until they were able to speak to LookTel support.

5. Participants may have strong brand loyalty or purchase many food items from the same brand resulting in situations with some labels confusing participants with other items from the same company that haven’t been trained yet.
Participants experienced less than perfect recognition with similar items that were trained.

6. Participants are having a difficult time coming up with what kinds of items they wanted to recognize.

7. Adding labels to round pill bottles alters the image sufficiently to cause recognition to not work, or become unreliable. Some participants were wary of adding the label to the cap because it can get mixed up with other caps from other medicine bottles.

8. Some participants do not want their medication and other private items in the system. One individual wanted to add credit cards, but refused due to concerns about security of the system. Adding a private database that is password protected and sent over encrypted video might be interesting when offered as a premium pay extra feature.

9. The concept of having a featureless or clear background for training images is not something participants grasped easily. A clear or featureless background was often equated with smooth, so granite countertops or patterned wood tables would be presented as potential plain backgrounds. In one instance several sheets of printer paper were arranged to create a featureless surface with lots of guidance via the live video feed. Participants willing to purchase a poster board fared much better.

10. Security pop-ups in Windows have proven to be time consuming to overcome. Often, the participants had the windows security set to the max security
setting (manufacturer default) preventing installation of new software or access to firewall or other settings.

11. Support time to get participants to the ready point for image capture and recognition required a considerable amount of time.
   a. Sending devices pre-logged in, with contact lists pre-filled, and accounts pre setup saved a lot of time – on average 5-6 hours per person.
   b. Less technologically savvy individuals required at least double or triple the support time as more savvy ones.

12. It became clear that participants forgot how to use the mobile interface quickly after a few days of inactivity. Participants constantly using the software became well adept and savvy with the mobile human interface. People initially tended to get frustrated or dislike the human interface, but once they understood it, the consensus was more positive.

13. The majority of participants were either disappointed at the lack of color identification, or suggested adding it as a feature. There is a clear interest in this capability. One individual captured the sentiment well when she said that with color ID and the other features, she can slowly learn to rely on LookTel and wouldn’t need much other technology. This is exciting to her.

14. Referring to the grid as a phone number pad and saying number 4 and number 6 positions allows the participants to understand instantly where an item is. This saved a large amount of support time.

15. Some participants have reported inconsistent ability to connect using the device. Sample comment: “the device would not connect all day the past two
Saturdays, but worked fine with no changes beyond that”. These connection issues may be a result of issues with the device, HTC Tilt 2.

16. Battery life has been a significant problem. The battery would often die while on calls with participants, even though the devices had often been fully charged that morning.

17. Computer issues blocking support or progress were not uncommon with participants.

18. The majority of individuals who tried OCR mentioned that the feature could use improvement. It is difficult to read the full text of a document without a sighted assistant to help with orientation of the device. KNFB has aiming orientation software that guides the person on how to orient the device properly, although it is cumbersome to use.

19. Average support times (phone time only – testing and research not included)
   a. Full documented telephone support time: 80 hours
   b. Average Time to first recognition (currency from imported database)
      i. Initial participants – self-setup: 8.4 hours
      ii. Second round participants – LookTel setup: 2 hours
      iii. Beta group total average: 5.3 hours
   c. Average time to first trained Item recognition
      i. Initial participants – Self Startup: 9 hours
      ii. Second round participants – LookTel setup: 4 hours

20. Probable conversion rates (Beta participants who would or could adopt LookTel in its current state): 25%
2.10.4 Conclusions

The test revealed mixed results in performance and practicality of the system. Most participants felt that the overall system was complicated and fragile. While it’s very wide encompassing in terms of the feature list, the particular implementation and dependence on a broad array of supporting functionality exposes the tight coupling and sensitivity to any variation outside of a thin scope of requirements. Diversity in operating system configuration, wireless configuration, and personal preferences have a strong impact on the eco system. To facilitate access to the recognition technology, a lot of underlying layers had to be developed that are typically managed by the operating system – such as power management, wireless/3G connection management, and volume controls. These fundamental elements proved to require a level of integration beyond the scope of this project for the level of reliability required. Communicating the state of the device including power levels, volume, and critical alerts must happen consistently and dependably and from our experience cannot be developed as an over-the-top solution around the operating system – it needs to be designed to accommodate access from the beginning.

With everything connected properly, the recognition worked accurately as long as the trained items were correctly framed. Most recognition problems occurred when scanning same items of the same brand that have not yet been trained. This happened when there was enough information in the training database to match with the query
item. For example if the Campbell’s Chicken Noodle Soup is the only item in the database, scanning Campbell’s Split Pea Soup will result in a positive match for the chicken soup. The engine will match the common regions and unless every single Campbell’s soup is trained, new items will try to be classified as those already trained. There isn’t a robust method at the moment of determining that the item in question is similar but sufficiently different from the local database to recognize it as a “new and different” item. This issue requires further study and evaluation at larger scales.

The auditory feedback requirements are very personal and highly varied. For some persons there was too much feedback, while for others not enough. Ultimately there need to be two settings: one for amount of feedback/verbosity and the other for speed of speech. Since auditory is the primary information channel for blind and visually impaired persons, those actively working with technology often push the limits of auditory recognition at rates to 300% normal speech. Since some of the text was prerecorded and some synthesized, it was difficult to implement a consistent experience with variable speed speech. Some of these limitations arose out of computational limitations of the mobile devices at the time.

Overall, for regular use it has to be more accurate and less cumbersome than pre-existing systems people have in place. These systems include: rubber bands over certain products, putting products in regular and organized locations, using braille stickers or other physical augments to distinguish the items.
3. LookTel, version 2.0

3.1 An Accessibility Revolution

In late 2009, with the introduction of the iPhone 3GS, Apple also introduced VoiceOver – a layer of the operating system that makes the entire phone accessible through text-to-speech and a specialized set of gestures [Voi13]. Part of its universal access suite, VoiceOver is integrated throughout the OS, ensuring a consistent experience across all components of the system. It is a revolutionary development as for the first time a device designed for mainstream consumers also includes full support for persons with various impairments such as low vision or blindness. The Mac OS integrated VoiceOver a few years earlier making the Mac OS and many applications accessible. Apple extended and improved VoiceOver when porting it over to iOS. Another aspect of VoiceOver is its deep integration into the OS, which most accessibility engines are not permitted to access. This has proven to be a significant advantage for Apple, in that any update to the OS is an opportunity to update VoiceOver and maintain functionality at the infrastructural level.
3.1.1 VoiceOver – a brief introduction

VoiceOver on iOS works by introducing several gestures to mediate access and the focus on various human interface elements. Many gestures are location-sensitive - for example, sliding one’s finger around the screen will announce the visual contents of the screen as the finger passes over each element. This enables blind persons to explore the on-screen layout of an application. A person who is blind can double-tap – in a similar manner to double-clicking a mouse - to activate a selected element, just as if a sighted person tapped the item. Gestures can also be non-location-sensitive – for example, swiping, called flicking right or left, advances the selected element forward or backward without need for careful spatial coordination. A person can turn off the display, with VoiceOver operating, while leaving the touchscreen sensitive to touch, saving battery power, and providing a measure of privacy when used with a Bluetooth earpiece, or headphones. Apple calls this feature “Screen Curtain”.

VoiceOver for iOS is activated using the system’s Settings application. The iOS can be configured in the Settings menu so that VoiceOver can be toggled on and off by a triple-click of the Home button on the device. No matter what the person may be doing a triple click can add or remove the functionality of VoiceOver, without disturbing an app or phone call. This feature permits the possibility of turning off VoiceOver, and instantly having the device operate in a completely familiar way for a sighted assistant or partner. Conversely, that partner or assistant can be using the device, and with a triple click of the home button, have it resume all VoiceOver functionality. This
removes the stigma of specialty hardware systems, and makes it possible to get the assistance of a sighted person with an app or function on the phone.

Integration of VoiceOver at the operating system level means that apps developed by 3rd party developers are often accessible to persons who are blind using iOS devices – whether or not they have any remaining sight. The implications of this design decision cannot be overstated. It is an example of universal design that anticipates the diverse means that people require access the system. Apple has included infrastructural support for access to everything that’s covered by the operating system and many items that are developed by 3rd party developers for the operating system. As a result there is no longer a need for a developer to create a unique solution to make an app accessible via VoiceOver and the iOS. Many developers do not yet take advantage of this, but the situation has improved markedly since 2009, when VoiceOver was introduced. Even the most fundamental elements of the operating system such as copy and paste features carry the impact of accessibility by extending a common platform for information access for all people. VoiceOver also supports communication over Bluetooth with Braille displays and note takers allowing faster interaction possibilities as well as allowing full participation from the community of deaf-blind persons.

VoiceOver includes a speech-rate control for those wishing to hear the audio descriptions at faster or slower rates than regular human speech. It ranges roughly from 25% to 300%. Another option is the variable verbosity control – whereby hints
and additional information may be omitted for those already familiar with the system and seeking faster interaction possibilities.

While VoiceOver is not perfect, and it does not guarantee that everything is fully accessible, it does go far and with sufficient depth, as to have become the platform of choice within the blind and visually impaired community. The touchscreen, previously considered an unlikely accessible component due to lack of reliable and distinguishing buttons, slowly but surely gained popularity as a device of empowerment and democratizing access. It also contributes to personal independence, since more tasks can now be accomplished without need of support from others. The biggest complaint from the blind community is that Apple has no standard requiring that apps be accessible and many popular ones are not. It also does not offer a mechanism for refunding money for apps that prove inaccessible. This pushed the development of accessibility review projects like AppleVis and the VIA app from Braille Institute.

In many instances, adaptive technologies are stigmatizing, as they are vastly different from the devices used by persons who are not blind. The popularity of the iPhone and other iOS devices is in part, due to the ubiquity of an acceptable technology platform that provides many options to everyone, and the fact that many of the same apps provide shared experiences with sighted peers, making the technology an inclusive one.
For application developers this significantly changed the number of issues that must be accounted for when building applications. This is particularly significant when developing solutions that must comprehensively communicate core system information such as battery level, signal strength, wireless network information, and device lock status.

### 3.2 LookTel Money Reader

At this time, currency in United States is not accessible to blind or visually impaired persons. In many other countries the bill designs include a variety of features such as different sizes for each denomination, very different colors and large numbers to help those with low-vision, and in some cases even braille markings. Absence of these elements in the design of United States currency necessitates the use of assistive technologies such as the iBill ($120USD) or the Franklin Bill Reader Money Identifier ($300USD). These are dedicated hardware devices designed solely to identify U.S. currency. They are powered by AAA batteries and are approximately 3”x2”x1” in size. They combine a small camera, an embedded processor and custom software for currency identification. The Franklin Bill Reader runs an embedded version of the Evolution Robotics ViPR engine.

Given the general-purpose computational capabilities of smartphones, coupled with iOS’s VoiceOver accessibility foundation, we sought to address the money
identification by deploying the image recognition engine directly on the iPhone, and related iOS devices. The iOS platform with an accessible method for distributing the app through the Apple App Store created an unparalleled method to distribute the technology at an affordable cost.

The app is simple. Like its physical analogs, it is designed to only do one thing: recognize bills easily and reliably. Once the app is started, the iOS device itself becomes the human interface. One points the camera at the bills to be identified, and listens to the denominations as they are read by VoiceOver. There is a minimal human interface. When used without VoiceOver the app plays pre-synthesized audio descriptions of all bills. When used with VoiceOver, the app synthesizes each description on the fly. A tap on the screen while nothing is actively being recognized triggers an announcement of Money Reader’s readiness. This provides feedback that the system is functioning correctly. This feedback proves invaluable for persons initially learning how to best aim the camera as it gives guidance that the bill in question is not in view of the camera. The speed of Money Reader, 20 frames per second on the iPhone 4, and 30 frames per second on the iPhone 5, combined with the accuracy of the recognition system is good enough to quickly enhance a person’s ability to coordinate through proprioception the relative bill to camera positions.

Many people start by placing the phone flat on the bill and then slowly pulling the phone away. After several successful recognitions, people learn the optimal range for moving the bill into the field of view of the phone’s camera.
3.2.1 Public release and results

Money Reader, version 1.0 was released in March 2011 in the Apple iTunes App Store at $1.99. Within days several thousand copies were downloaded. It was the first publicly released LookTel product. It solved one small problem, but it solved it well, as indicated by its immediate 5 star rating. The app received positive reviews in addition to the rating in the App Store. To this day Money Reader maintains a five star rating. LookTel was selected as the winner of the Federal Communications Commission’s (FCC) Chairman’s Awards for Advancements in Accessibility. The award, which recognizes outstanding efforts to advance communications accessibility for people with disabilities, was awarded in Washington D.C. on October 28th 2011. To date, there have been over ten thousand copies downloaded and the sales continue to increase steadily.

The New York Times published a small article about the LookTel Money Reader application [Bil11]. What followed was a very interesting discussion in the comments section between the general public and the visually impaired community – regarding the possibilities of using a touch-screen phone without sight. That discussion thread gives a glimpse into the perceptions of ability and the far-reaching impact of universal accessibility of the Apple devices – running operating systems from iOS to OSX.

We often receive direct feedback from people who use the Money Reader app. Most comments are positive and express appreciation for making the app so easy to use.
We often hear from instructors and rehabilitation specialists that it is one of the first apps that they recommend people try. These instructors and rehabilitation specialists find the app reliable and intuitive. Most people immediately report positive and encouraging results - "it just works". The rapid feedback with recognition, affords rapid proprioceptive learning. The extension of the hands now offers a fast-enough feedback for muscle memory to learn relative position of both hands while keeping the bill within the view of the camera. The simplicity of use and rapid feedback are very important for people new to the concept of a camera. We’ve even heard from several individuals who have since started working as cashiers with the Money Reader.

Since the original release, several updates have been published. Most notably new currencies have been added along with localized language support. Currencies added in version 2.0 have been the Australian Dollar, the British Pound, the Canadian Dollar, and the Euro. Each of these has been thoroughly tested to prevent false-positive misrecognition. Each of these currencies has also been evaluated to account for the wide range of bill designs that may be in circulation. LookTel Money Reader provides Voice Over support for several languages including English, Spanish, French, Italian, German, Polish, Russian, Korean, Finnish, Danish, Swedish, Norwegian, Japanese, Greek, Hungarian, and Mandarin.

An imminent release of version 2.3 will include newly released bills from Canada and Britain as well as many new currencies. The new currencies are Bahraini Dinar, Brazilian Real, Belarusian Ruble, Hungarian Forint, Israeli Shekel, Indian Rupee,

We conducted extensive research while collaborating with volunteers in many countries to identify and confirm the various designs of bills in circulation.

Several alternative bill identification applications have been released including a solution from the U.S. Bureau of Engraving and Printing (the EyeNote), however the LookTel Money Reader continues to be preferred state of the art application due to its speed and accuracy. Several of the alternatives are free.

The price of Money Reader increased to $9.99 in late 2011. This price increase is still a fraction of what previous and current standalone devices cost. Money Reader remains a popular app, even with free money reading apps being offered. It remains the only app that can read U.S. currency and multiple other currencies in real-time.

Using the same engine we have also developed and released a version of Money Reader for the desktop OSX operating system. While the application technically performs very well it is seldom downloaded. This appears to occur because people owning a Mac often own other iOS devices, and prefer to have Money Reader on their portable device.
3.3 LookTel Recognizer

After the success of the LookTel Money Reader we redesigned the core recognition element within the context of iOS to deliver the simplest mechanism for real-time object recognition. The LookTel Recognizer App is for blind and visually impaired persons who can benefit from a technology to assist in identifying objects that differ visually, but are too similar in other respects for easy identification. Recognizer permits people to store images of objects in a database, and then have the iOS device quickly recognize these items later, when scanning any of the items with the camera. A barcode scanner is included to provide alternative identification and supplemental information helpful for labeling.

Recognizer allows customers to backup and export databases via e-mail. This enables them to restore a database from lost or replaced devices, or share it with other members of the community.

3.3.1 Features and Capabilities

The Recognizer App utilizes real-time recognition capabilities with customization that permits a flexible range of ways to use it – from identifying everyday objects such as packaged goods, identity cards, canned foods, or CDs in a music collection. Following the approach used in Money Reader we minimized the human interface. The app is
built around capturing items for the database, and recognizing them later. On the main screen, the recognition sequence is always running, and will announce a known item anytime it is in the field of view. A double-tap on the main preview screen captures the image for the database, termed a Library, and initiates the recording of the audio label.

The database of trained items is stored locally on the device allowing the editing, adding, removing and relabeling of individual items. The database can be emailed for sharing or backup purposes, as can individual items. Entirely local processing of the image recognition and training removed the major complications encountered with the LookTel 1.0 system. The customer no longer has to manage a backend engine on another computer, nor do they need to rely on any network connectivity. Decoupling these layers from the task of item recognition.

Barcode scanning permits lookups to several sources of meta-data for Universal Product Codes (UPCs). Recognizer is also localized in Spanish and German, and has complete in-app documentation in English, Spanish, and German.

**3.3.2 Ongoing updates**

We are preparing an update release; version 1.3 that includes enhanced categorization capabilities for easier content management within the app and finer control over
sharing capabilities. We have also added more sources of data for barcode lookups for a more successful experience.

In version 1.4 we will be releasing a cloud-based image lookup capability that will allow people to use the app without the need for as much custom training. While custom training will still be possible, it will not be a prerequisite to use the app, as many common item images will be available through the cloud. We recognize that many people who are blind do not have access to a sighted assistant, and in order to maximize independence we need to eliminate any barriers to image acquisition. This improvement is critical for Recognizer to achieve its potential.

3.3.3 Public release and results

The LookTel Recognizer was released to the iTunes App Store in February 2012, and continues to receive positive reviews and has a 4+ star rating. The app costs $9.99. The download rate is slowly increasing. Thousands of people have already downloaded and use the app.

The feedback we receive depends on the level of initial sighted assistance to which the customer has access while getting started. People who have access to a sighted collaborator (parent, spouse, or caretaker) to help with the initial capture of images for items can benefit from this app immediately. Some customers are successfully
capturing their own images for items without visual assistance. We have been contacted in regard to individuals who have correctly identified the male and female restrooms at a campsite. In another situation, the husband trained all the overhead signs of shops at a local mall for his wife, who to her surprise was able to find several new merchants on her next trip to the mall. Another very popular use has been labeling of the common soft drinks at soda fountains, allowing a person to quickly identify the beverage of choice in a public cafeteria. Orientation and mobility (O&M) instructors are evaluating possibilities of using Recognizer as a location identification assistant. By training lots of locations throughout a school building and hanging an iPod Touch like a pendant around the neck of an individual, the Recognizer can give feedback upon arrival to trained locations, and announce the positional information for that spot.

4. Conclusions and Future Work

The conclusions and future work statements follow for each application individually.

4.1 Money Reader

The 2.3 update of Money Reader was released on May 8th, 2013. It offers recognition of 21 currencies, with additional language support. Money Reader is at a stable state and will continue with gradual improvements. More currencies will be added. As the
international monetary database grows it will require additional mechanisms to maintain the quick start without the delay of loading a large database into memory. This has proven to be an issue with some iOS devices, such as the iPod Touch, and older iPhones. Optimization may solve these issues now. For the future, strategies will be necessary that permit a choice of currencies to be loaded.

Some technologies introduced into a culture are key to activities performed in a culture, and can have a disproportional impact beyond the ubiquity of the device or system. Braille is used by less than 10% of persons who are blind. It obviously has a high impact, and is a key technology to many activities from daily living to participation in education and employment for many persons who are blind.

One of the symbols for independence among persons who are blind is the guide dog. While fewer than 7,000 are in service nationally, the guide dog alone has had an enormous impact on the culture of America. Every state, and the federal government, has specific laws governing the use and access to facilities of public accommodation with guide dogs. The use is limited. The impact has been significant beyond the numbers of dogs in service.

The Money Reader joins the guide dog as a symbol of independence, and like the guide dog, provides meaningful independence in a major life activity—access and confidence in handling money independently. The guide dog may provide a service to persons who need travel independence for employment, while Money Reader
provides the means for independent and confident financial transactions involving currency.

While the use of Money Reader is growing, it may never be used by more than a small percentage of the total population that might benefit from it. Other systems such as debit cards, smartcards, and virtual wallets may surpass Money Reader in the number being used, as nations make this possible. Money Reader, or its descendants will have a place in many cultures while currency remains the primary means for accomplishing many transactions.

4.2 Recognizer

Recognizer requires the connection to a cloud-based resource of images to supply the local, real-time recognition. Without this source of data, building of a local database is prohibitive to persons without a sighted assistant. While we have access to one such database we will need to continue updating the content to reflect the content of interest and accuracy.

Recognizer is maturing. The update now under development changes the paradigm for Recognizer, and recognizes the role that recognition technologies will have in the future of persons who are blind in Western culture. The app now includes an elementary library structure, with recognition being one method for adding to the
library, and retrieving information from the library. This paradigm shift is in concert with the growing need of all people in a connected world, to access and retrieve information from anywhere at anytime.

The new Recognizer will lay the foundation for text recognition within the app, and the storage and retrieval of text information into the common library. The human interface will change and grow as customers make more demands on the new capabilities.

It is likely as the internet of everything evolves, that Recognizer will play an important and foundational role for persons who are blind, as it will integrate the visual world into a common library for all daily activities. As image recognition improves, the role of Recognizer can be expected to grow, increasing the probability that the recognition functions will become more sophisticated. The application to recognizing physical landmarks and locations will benefit, as well as the ability to “read” signage.

4.3 Remote Eyes

We will resume work on the Remote Sighted assistant software based on iVisit protocols. The prototype is currently called Remote Eyes. Remote Eyes has a future within the Recognizer to assist with acquiring quality images, and offering assistance with context dependent issues, such as background lighting, background features, etc. As discussed in this document, evaluations of the early system indicate that many
persons who are blind have only a limited understanding of lighting and visual background features that impact quality image recognition. Remote Eyes can assist with training and aid to resolving customer problems.
5. References


[Gib50] James J. Gibson. The perception of the visual world James J. Gibson. 1950


