Context Awareness and Discovery for Helping the Blind

Martin Goldberg
CUNY City College

Recommended Citation
Goldberg, Martin, "Context Awareness and Discovery for Helping the Blind" (2012). CUNY Academic Works.
http://academicworks.cuny.edu/cc_etds_theses/90

How does access to this work benefit you? Let us know!
Follow this and additional works at: http://academicworks.cuny.edu/cc_etds_theses
Part of the Computer Sciences Commons

This Thesis is brought to you for free and open access by the City College of New York at CUNY Academic Works. It has been accepted for inclusion in Master's Theses by an authorized administrator of CUNY Academic Works. For more information, please contact AcademicWorks@cuny.edu.
Context Awareness and Discovery for Helping the Blind

Submitted in partial fulfillment of
the requirement for the degree

Master of Science (Computer Science)
at
The City College of New York
of the
City University of New York

by
Martin Goldberg
May 2012

Approved:

__________________________________
Professor Zhigang Zhu, Thesis Advisor

__________________________________
Professor Douglas Troeger, Chairman
Department of Computer Science

- 0 -
# Table of Contents

Abstract ........................................................................................................................................... - 3 -

1. Introduction .................................................................................................................................. - 4 -

2. Background: AT and CV ............................................................................................................... - 5 -
   2.1. Assistive Technology ............................................................................................................... - 5 -
       A. Electronic Travel Aids and Electronic Orientation Aids ..................................................... - 5 -
       B. Assistive Technology Missing Goals: A Lifetime of Mistakes ........................................... - 6 -
   2.2. Computer Vision: Recognizing Rooms and People ............................................................. - 6 -
       A. Sampling Bedrooms .............................................................................................................. - 6 -
       B. Human Detection ................................................................................................................ - 7 -
       C. Face Recognition ................................................................................................................ - 7 -

3. Understanding Context ............................................................................................................... - 8 -
   3.1. Context in Face Detection ..................................................................................................... - 8 -
   3.2. Context-based Object Categorization .................................................................................. - 9 -
   3.3. Augmented Context ............................................................................................................. - 10 -

4. Semantic Context and Ontology ................................................................................................. - 11 -
   4.1. Bridging the Semantic Gap in Image Retrieval ................................................................... - 11 -
   4.2. Ontology Based Visual Information .................................................................................... - 12 -
   4.3. Some Thoughts on Ontology and Context Discovery for the Blind .................................. - 12 -

5. Understanding Prospective Users ............................................................................................. - 14 -
   5.1. The Known vs. the Unknown ............................................................................................... - 15 -
   5.2. First-Hand Experience and Needs ...................................................................................... - 16 -
       A. Background – Physical Facts ............................................................................................. - 16 -
       B. Explicit Difficulties with Seeing ....................................................................................... - 16 -
       C. Examples of Problems Encountered and Solutions Where They Exist .......................... - 16 -

6. User Case Study ........................................................................................................................ - 18 -
   Step 1 – getting out of the train, and orientating towards the path ........................................... - 18 -
   Step 2 – Onto the path and finding the fence ............................................................................. - 20 -
   2a) Finding the path .................................................................................................................... - 20 -

- 1 -
2b) Objects ahead in the subject's path

2c.1) People approaching

2c.2) Multiple people, one coming straight at the subject

2d) Exiting the park

Step 3 – Crossing at the light – on a relatively small crossing

Step 3. Alternative – watching the street for moving vehicles

Step 4 – Starting up 138\textsuperscript{th} street

Step 5 – Continuing up 138\textsuperscript{th} street, approaching the scaffolding

Step 6 – Avoiding the open gate

Step 7 – Getting to the main traffic intersection

Step 8 – Waiting for the light to change, watching the traffic go by

Step 9 – Crossing Amsterdam Avenue

7. Some Practical Problems and Possible Solutions

7.1. Surveillance Cameras and Sensors in Motion: Focus of Attention

7.2. Limits of Processing Power on Transportable Devices: Context in the Cloud

7.3. A Modular Camera System: Context Awareness through Modular Designs

8. Conclusions and Discussions

Acknowledgments

References

Table of Figures
Abstract

The central theme of this thesis is to consider how context can better aid image processing and look at how this can improve computer vision technologies in generating aids for the blind and visually impaired. We look at what has been done in this regard, and then consider, through our use case, how this might be implemented. Through context discovery, we hope to develop robust and effective methods of determining the higher-level features of an image, with focus of attention, context cueing, and data association, so that the vision system can more quickly and efficiently discern and be aware of the context of the targets of interest to help with the detail analysis of the targets. Through a study of existing image processing techniques and approaches, we hope to be able to distill a clear approach to context discovery that can be applied to a number of image processing tasks, such as human detection, landmark recognition and sign reading, particularly for the benefits of visually challenged people and the totally visionless blind.
1. Introduction

We have two converging goals to reach – (1) assisting the blind and visually handicapped through computer vision systems and (2) improving the performance of these systems by utilizing context information. Each of these goals in their own separate fields could be large; it is our postulation that when providing context information to computer vision systems specifically designed to aid the blind, we can more easily define the research problem into a manageable size.

Context, as sociologists such as Brabyn (2011) and Biederman (cited by Galleguillos & Belongie, 2010) have shown that context is critical to human understanding of a given situation and that they support the idea that computer systems would greatly benefit from contextual input. Jain (2010) also shows how context can be utilized to help solve ambiguities.

Through context discovery, we hope to be able to afford a top-down approach to image processing that could help computer vision systems to be more efficient and more effective. Current image processing is mostly a bottom-up process, from low level feature detection, intermediate feature grouping, to higher level object recognition or classification. Through context discovery, we hope to develop robust and effective methods of determining the higher-level features of an image, with focus of attention, context cueing, and data association, so that the vision system can more quickly and efficiently discern and be aware of the context of the targets of interest to help with the detail analysis of the targets. Through a study of existing image processing techniques and approaches, we hope to be able to distill a clear approach to context discovery that can be applied to a number of image processing tasks, such as human detection, landmark recognition and sign reading, particularly for the benefits of visually challenged people and the totally visionless blind.

This work is more about thoughts and ideas on using context information to improving computer vision algorithms for assisting the visually impaired and the blind, rather than the implementations of the ideas. Therefore even in the survey part, we will put down some thoughts on how context could be utilized to make the technologies more useful as assistive technology.

This thesis is organized as follows. Section 2 gives a small literature review of related topics, namely assistive technology for the blind; and computer vision techniques that are important to the blind. Section 3 focuses on the discussion of context awareness and discovery for the blind. Section 4 provides some thoughts on using ontology and context discovery in developing assistive technologies for the blind. Section 5 takes a closer look at the blind subjects themselves, reasoning what we are dealing with so that we can better adapt to their needs, not unnecessarily wasting time on the obvious, while being sure to provide the subjects with an optimal result. In our use case – section 6 – we will look into our image data and extrapolate how context may be discerned and applied. In Section 7 we discuss possible problems and their solutions when dealing with our blind subjects. Finally we will draw some conclusions in section 8.
2. Background: AT and CV

The small background survey will be organized into two groups of topics – assistive technology (AT) and computer vision techniques (CV), each with one or more papers and/or presentations that have been surveyed.

2.1. Assistive Technology

A. Electronic Travel Aids and Electronic Orientation Aids

Farcy et al (2006) make some very interesting observations. In their introduction they show how over the past 40 or so years, while many have claimed great advances in aiding the blind (and visually impaired) through technology, most have failed to have their work accepted into use. They cite the work of Benjamin (1973): the laser cane; Kay (1974): use of sonar in aiding the blind to detect object distances; and Farcy (1996): adaptation of a laser profilometer

Most Electronic Travel Systems rely on ultrasonic echolocation, even the most recent ones (Hoyle 2003, Ultracane) (Miniguide 2005) (Bat ‘K’ sonar cane 2004). The most advanced system from a technological point of view seems to be the Bat ‘K’ sonar cane. They also cite Brabyn (1982), who at the time took a view of a biomedical engineer, in assessing the advances made by that time and realizing their limitations and lack of acceptability.

What the Farcy team has done, is to develop three products, two ETAs (Electronic Travel Aids) and an EOA (Electronic Orientation Aid) which they then use one after the other in a step-by-step process to train users to employ.

The two ETAs (the ‘Tom Pouce’ and the ‘Teletact’) are not of particular interest in context discovery and awareness, save that the authors use the devices in training the user to adopt the EOA and then use them all together in a combination system of some sort. The Tom Pouce uses an infra-red proximeter and the Teletact uses lasers. The EOA applied is an adaptation of a GPS system used for motorcycle navigation. At the time of writing, 2006, the authors claim that they have 10 users of their total system.

In their EOA, orientation (as in understanding the topography of an urban environment) is linked to navigation, since good navigation skills are key to having their system work. As the system is designed for the blind, the experiment uses auditory and haptic output (sense by way of tactile output) as a means of delivering the required information.

What is of specific interest to us then is how the Farcy team has adapted the motorcycle GPS system – in effect adding geographic context to the system – to enhance their system. This is a good, yet limited, example of how context can be employed to better facilitate the image processing process.
B. Assistive Technology Missing Goals: A Lifetime of Mistakes

This mainly slide presentation (Brabyn, 2011) is very enlightening. While this is not strictly an academic paper, this article is of great relevance to our endeavors. In it, Brabyn outlines how assistive technology has missed with the goals it has set itself in the past, elucidating the many factors that go into this failure. The author lists some key “implications and challenges for computer vision”, when he states that these problems might surface:

- Computer vision using remote computing power can also have longer than advertised delays
- Computer hardware gets less expensive all the time, but if custom hardware is needed it can be expensive in small quantities
- What information does he/she really want?
- How should the information best be presented?
- What user controls are appropriate and workable?

These points, among the very many others he makes, all lead us to consider why it has taken this long for anything to be realized and give us reason to believe that our new adoption of old techniques by using context may actually have a chance of success. As a person with the disability myself, and one who has spent some time with other legally blind and totally blind people, I have some first-hand knowledge of what Brabyn is discussing. This presentation therefore is an important cautionary tale of how we need to proceed in order not to repeat the mistakes of the past.

2.2. Computer Vision: Recognizing Rooms and People

Here we just give a few examples that are of importance to the blind. They don’t mean to be complete at all. However we want to use these examples to analyze how vision algorithms perform with and without using context information.

A. Sampling Bedrooms

The paper by Del Pero et al (2011) has applied many theories in generating their method. The Markov Chain Monte Carlo (MCMC) method is used in their statistical evaluation; and more specifically the Metropolis-Hastings algorithm for obtaining a sequence of random samples from a probability distribution for which direct sampling is difficult. In their case they are trying to capture all of the cubic objects in a given room. Their premise is a static camera situated in the room being analyzed.

What is of particular interest to us, though, is the manner in which they derive the room dimensions. From our prospective, in our attempt to discover context and analyze the data in order to assist the visually handicapped and blind users, their way of determining the base cubic structure of the room, regardless of obfuscated corners and other interferences can be of much
use in our efforts.

What is described in this paper (Del Pero, 2011) is a method of finding the main parallel lines through determination of vanishing points. Once the vanishing points are calibrated, the room structure is defined. If some expectation of the known height of the room is available, the dimensions of the room can be calculated easily. This can prove to be very effective in our attempt to derive context.

The secondary aspect from our perspective is the ability to isolate smaller cubic elements within the larger cuboid; this may or may not be of use in that much depends on computing requirements and whether actual use can be made of this information.

B. Human Detection

The paper on Bayesian 3D model-based human detection (Lu Wang et al, 2011) is most interesting and pertinent to what we have in mind. It is mainly for the way that their algorithm is implemented that we find our value in this article. While it is true – like with so many of our reviewed articles – that this paper deals with a surveillance scenario: meaning that the background is generally static and thus the object extraction is made easier because of this fact – we can, nonetheless, learn much from the process.

The object being extracted in this instance is a human being, typically standing. In most cases, there are several human possibilities and the algorithm determines each one. The authors combine a two-step process with a reasoning system which fit very nicely with our endeavors. The two step process is: first, discover possible candidates (the “greedy” method), then, having dismissed the unlikely, extract the humans in the image frame. The human detection is done through decomposing the humanoid object into several base elements – 3D in this case – such as ellipsoids and cylinders. In the background of this process is the reasoning system which applies Bayesian inferences to select what are and are not the objects being sought.

In our scenarios, we are working with cameras in motion, and thus the complexity is increased, but the most basic idea still holds great possibility for us. As we are trying to define context, meaning where the subject is, and who and what they are interacting with, this approach is a good beginning to what we are attempting to do. Human detection is important to our thesis, and the process defined here may go well to aiding us in this endeavor.

C. Face Recognition

The first two articles (Kakadiaris et al, 2007; Brunelli & Poggio, 1993) are useful to us as they both describe using meta-data and templates in facial recognition. The work by Brunelli & Poggio (1993) is nearly 20 years old, but it depicts trends from the 70’s onwards and presents a good agreement for using templates and structural features as an effective way of decoding a facial image. The later article (Kakadiaris et al, 2007) expands this technology into a far more robust usability which we should explore further.
Since it is our goal to create a system to assist the blind and visually impaired, facial recognition is one of our main challenges. Essentially, using meta-data – which is by its nature easier to store and sort through – we hope to be able to build person recognition into our system.

In Chen’s thesis (2006) the author talks about deconstructing the face through a variety of algorithms including the K-means objective and the Snake algorithm. Ostensibly, the idea here is to extract the contours of the eyebrows first, and then of the chin. The rest follows from there.

This may indeed be of value to us. If we are to provide a way for facial recognition to the blind, we need to consider a model which does not keep the image but a non-graphic representation of the face in storage. The structural approach may well be the solution. Unlike a surveillance system which may deal with many unknown faces, the system we are envisioning is to help the non-sighted to know who it is that is in front of them – i.e. well known, documented facial data.

3. Understanding Context

By analyzing two articles, we would like to investigate how and why they use context.

3.1. Context in Face Detection

The first of these – Jain (2010) – has some interesting points: especially about the more effective face detection, although, it deals with classifying photographs, which is different from our needs. We are dealing with a video stream, and do not have captions to rely on. On the other hand, we do have a growing knowledge domain to utilize as we progress, and thus the basis is useful, even though the rest is not.

Jain’s thesis is using contextual information (both from the image and/or captions attached to the image) to better identify faces. His argument is that most ‘good’ face-recognition software relies on a full frontal image and that he has proposed a system for dealing with faces caught at angles other than straight on. The technique used employs a dual 3D ellipsoid construct which when super-imposed on the image should help calibrate the facial dimensions. In a comparison of two images, he shows the director Quentin Tarantino juxtaposed with the tennis player Roger Federer which he shows to be remarkably similar – but by using the background, a director’s chair in the former case, a tennis court in the latter, he suggests that using this context we are able to distinguish the person from the ambiguity. However, it is his use of captions that he accentuates as he is dealing with newspaper clippings.

It is not fully understood if the researcher is able to get the ellipses he uses from the image without human input (the suggestion is no, that he uses annotators, in which case this is even less useful to us).
3.2. **Context-based Object Categorization**

The second article – Galleguillos & Belongie (2010) – looks at how psychology has defined context into five specific groups:

- Interposition
- Support
- Probability
- Position
- Familiar Size

Considering the first two as relating to human perception, not usable by computer systems, they derive the following classes of context within our area:

**Semantic** (Probability): we also call this Domain Knowledge; they give a basic example – given ‘grass’ and ‘sky’, they suggest – based on prior knowledge – the ‘sky’ should appear above the ‘grass’. They then suggest that in the past, ‘professional’ knowledge systems (large, cumbersome and often incomplete) were the prescribed method, noting the trend towards ‘learned’ rules, generated through machine learning and boosting techniques, are more prevalent. They also suggest something we consider particularly likely in our efforts: a hybrid of ‘professional’ and ‘learned’ knowledge. As we now live in an era where through Wi-Fi technology we can upload sample data to servers and receive processed data relatively quickly in return, thus allowing for, over time, large data sets of shared knowledge to be utilized, and this should let us explore both aspects and work with it.

**Spatial** (Position): how objects relate to one another spatially. It has been noted that spatial and semantic often go together.

**Scale** (Familiar size). Galleguillos and Belongie have stated that this is at once both the most helpful, and the most difficult to obtain contextual data. Utilizing techniques developed in the City College Visual Computing Lab (CcvCL, 2012), we may well be able to mitigate these difficulties and netter apply scale context to our endeavors.

**Object relation**: Defined by the article, this requires a previous definition of each object then uses a (Naive) Bayesian approach to suggest what to expect, example of a scene that is recognized as a kitchen, and inside - a kitchen counter top. The expectation of an object being a loaf of bread is much higher than of it being a musical instrument.
Semantic and spatial context are very important. The authors discuss relativity of the global context to the local. They claim that we take the psychological understandings and thus go from the global to the local (top-down) and then process from the image to the expected object (bottom-up).

As a survey this gives us a solid backing to our concepts - that being that others have done, or are doing research into how to manage these methodologies. In our own research we can take these ideas and follow through with them, seeing how we are able to adapt this to our specific target of aiding the blind and visually impaired.

### 3.3. Augmented Context

We can consider a fifth type of context – which we consider separate from those listed above, that is, augmented context. Although it may be thought of as semantic context, it differs in a major aspect: augmented context is provided to the system either through some digital signals other than computer vision - consider a traffic light that emits a signal, letting the system know what color the light is, and perhaps, how long it will take for this to change. Another example of this is the visual noun concept, a situation where symbols which are well-known are added to the environment and the image parsing can first check for them in order to gather a better contextual understanding of the given situation – for this we may consider the signs on restroom doors and the various well known labels that already exist (Molina, et al, 2012).

Augmentation is not a new idea. The novelty is how it is applied. Physical augmentation has been around for over a century: the use of familiar pictograms has been applied to buildings and highways; and, where blind users are concerned, adding Braille to signs at reachable levels has be employed in many cases – public buildings in New York City have be required to Braille up restrooms, stairwells and elevators for decades.

As for electronic augmentation, some has been put into use, such as RFID (Ganz, 2011), QR code (Chang, Tsai & Wang, 2008), etc. other ideas are under research, and it likely that a better, well augmented environment can be put in place. Nevertheless, augmentation has its problems too. To augment a substantial area requires the buy-in of decision makers who must see the benefit of the expense. Then the users must buy-in and adopt the systems, or they become useless. Moreover, it is in maintaining and updating augmentation where the real issues exist. Once the base system is put into use, the project must be expanded, if sensors are to be built to expect the augmentations, the augmentations must be added to any new area that the sensor is to be applied. So, while there are many benefits to augmentation, we cannot rely on this augmentation and thus must look at ways to utilize context in order to facilitate our users – the blind.
4. Semantic Context and Ontology

This small literature survey is intended to review pertinent documentation as it relates to the subject of research of this MS thesis, on context awareness and discovery. There has been much work on Content-Based Image Retrieval (CBIR) (Datta et al, 2008), and image understanding, and image categorization, but generally the limits are bottom-up approaches are usually used. In recent years, Context-Aware Image Management (CAIM) (Karlsen and Nordbotten (2008), Elahi, Karlsen & Akselsen (2009), Nordbotten (2010)) has attracted attention in the field.

4.1. Bridging the Semantic Gap in Image Retrieval

In their article, Zhao and Grosky (2001) are primarily concerned with the top-level aspects of an image: i.e. “this is a picture of a horse”, or “this is a picture of Bill Clinton”. What they try to do is “bridge the semantic gap’ between this top-level understanding and the low-level understanding: i.e. that information derived from analysis of the pixels. This article is relevant to us as the “top-level” aspects of a picture are directly connected to context awareness; however, the authors’ purpose here is to retrieve images from a database based on a query – typical of image retrieval (IR) goals. What I have in mind is to adopt several of these ideas and create an inverse query: that is, use these techniques to deconstruct a single picture based on basic images stored in a dataset - thus inverting the IR process.

This paper connects to us through several other papers. Starting with Datta et al (2008) who led us to Smeulders et al (2000) defined the semantic gap as follows:

“The semantic gap is the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation.” (Smeulders, p. 1353, section 2.4)

Smeulders (2000) has been much cited and his terminology is central to our efforts to achieve context discovery.

Beyond the background support that this article gives our ideas, it becomes tangential to the thesis. In discussions with my fellow researchers, I have suggested that this may be an excellent subject for exploring vision algorithms from this perspective.

The key take-away from this article is the idea that commonality in texture, structure and other similar aspects are important in deriving the top-level understanding and this is the starting point for context discovery – whether it is in comparison of two frames in a video stream, or use of predefined data structures in anticipating the image under analysis.
4.2. Ontology Based Visual Information

What is most interesting about these two articles (Town, 2004; Akdemir et al, 2008) is how they use surveillance to define scenes. The ontology aspect of these papers refers to classifying the image sequences into a formal language that can later be utilized in querying. At this stage, this aspect of the papers is the least useful application to which we might put this research to use.

A second stark difference with what we are trying to do is that – as the papers deal with surveillance – the authors are dealing with cameras in static position (at most, being able to rotate, but no more). We are trying to establish understanding in a series of images generated by a subject in motion.

What we can take away from these papers is the idea of isolating pieces of the images that change from frame to frame and thus disregarding the rest of the image. This can be useful in our endeavors. We will need to consider optical flow as an important measure in what we do, but once we can establish continuity in the visual stream, we should be able to extract that which is noticeable in its discontinuity. This may be very helpful in judging objects in motion, people, vehicles, etc. for which we must be aware of in assisting the visually challenged.

4.3. Some Thoughts on Ontology and Context Discovery for the Blind

It seems, based on the literature that the term ‘ontology’ as applied in computer science is ‘an abstract definition of a concept in such a way as to be used over by various systems’. Thus, ontology, as such, is relevant only in that when describing aids for the blind, we have a cohesive set of terms with which to work. In the past, ontologies have been suggested for surveillance use; in this case, one is typically looking for subtle changes in an otherwise relatively stable environment. In this environment, most of the image is known and easily referred to; thus, when other aspects appear in the image – especially none static changes – one is able – theoretically – to capture the anomalies in the image and focus on the selected images for identification, whether that be of the image or the action – this is thus defined in ontological form. Consider: “Man enters room”, “Man moves inside room”, “Man exits room”.

So, how might this apply to aiding the blind? We might use this as a means of communication particularly in the human-computer interaction: by having a well-defined set of terms and phrases, we believe that computers can be trained to recognize these terms and thus use them in outputting information to the user and also in receiving instructions from the user.

Wayfinding, also known as navigation is common both to low vision and robotics and is the simpler of the two realms; simpler in that less complex systems are able to produce relatively better results. The key to navigation is object detection and avoidance. In most cases, the nature of the object is irrelevant other than perhaps size and possible motion so as to better calculate avoidance. Techniques such as range detection help greatly in this and, in practice; simple navigation can be achieved without the need for computerized analysis. In a non-sensor aided environment the main tool employed is the cane.
Orientation is the more complex area as it requires analysis. In the world of low-vision, the human brain is the primary analytical tool using the cane for minimal environment understanding or a trained guide dog for better input. In the world of robotics, orientation requires both a world model and sensory input and analysis. In this realm, ontological definitions could prove extremely beneficial, especially if small portable systems are used to link up to more extensive servers.

The concepts associated with orientation are context awareness and an understanding of objects in better detail so as to better generate associations between one and another. It is best to simplify the composite environment by developing a set of primitives which can then be used to best describe the complexities.

Let us take for example the computer vision lab; what primitives might we consider? Initial orientation, like navigation, needs to figure out the most basic elements of the environment. Initially we want to know where is up (the ceiling) and down (the floor), once this is defined defining the confines (the walls) and the apertures (the doors) help define the context of the room. The first primitives to look for then would be the points where the ceiling met the walls, particularly where two walls met (i.e. the corner) at which point – utilizing a small knowledge-base) we could calculate whether we were seeing the ceiling or the floor.

Context – after all – is all about references. Where does what lie in respect to something else? Once we have reference, Bayesian inference and similar techniques can be used, coupled with a priori knowledge to ascertain where we might find a table ledge, a monitor and so on.

In my opinion, we could apply ontology of primitives to this scenario. Breaking a primitive down to geometric and other such measurements can help us more easily define the things we are looking for. Ultimately, we need our ontologies to meet with the capabilities of image processing techniques.
5. Understanding Prospective Users

To best understand the people we consider our subjects, i.e. the blind and the visually handicapped, Figure 1 accentuates three types of subjects we are servicing: legally blind, totally blind and robots. The totally blind subject has – naturally – no vision, but is able to understand complex everyday inputs – such as ‘turn left’, or ‘the cup is to the right of your plate’: the concepts ‘left’, ‘right’, ‘plate’, ‘cup’ and ‘turn’ being fully understood by the subject, as opposed to the robot, which needs constant meaningful definition for any such instructions to be used. The low-vision subject has limited vision so can be given less output, mostly in magnified, and contrast enhanced form.

![Blind v. Robots Diagram](image)

Figure 1: Comparing subjects - Low vision, Blind and Robots

Figure 2 looks at navigation needs, showing that wayfinding can be aided without the need of context, but orientation can greatly benefit from context. In the user case study discussed below, this will become clearer.
Another point to be made about dealing with these human subjects is that camera shots (particularly that of the glasses supported camera) can typically be reckoned to be of a certain height above ground (corrected in advance for each individual user), to be used in typical situations (i.e. walking along a street, or down a passageway are considered typical; looking down from an airplane or up from the floor are not); in addition, the frame sequence in a video stream is fairly predictable, and thus calibration is mitigated by these factors.

5.1. The Known vs. the Unknown

When dealing with blind subjects we need to consider how well a situation is known to him. Most humans work better in familiar surroundings, this is even truer for the blind. The well-known requires far less interference from an input device that an unknown environment which requires much input.

Also in terms of semantic context (knowledge domain), when saving ready to utilize data in a resource constrained setting, the familiar needs to be stored nearby: faces of familiar people need to be accessed regularly and thus caching them makes sense. The unknown is always the larger subset, and oftentimes full of ambiguities. To deal with the unknown, more computing power is necessary, typically utilizing huge data sources. To facilitate this, networking with large systems may be necessary. In later work, we might want to discuss how these large data sources should
be built and utilized.

A third subset – that of the ‘generally well known’ – must also be considered. The faces of the President and the local Mayor, for example, fall into this category. In a school setting, the faces of the teaching and administrative staff are candidates for such consideration.

5.2. First-Hand Experience and Needs

I would like to put down my personal experience and needs in a third person description under the title of Subject Zero, before I move to a user case study.

A. Background – Physical Facts

Subject Zero suffers from a condition known as Flecked Retina Syndrome. This takes the form of a collection of dead or dysfunctional cells at the center of the inner retina in each eye. The result is called central scotoma, a situation where the center of the field of vision is lost or severely limited as it becomes very blurred. A genetic syndrome, this often starts in middle age becoming problematic in later years (as is the case with the subject’s mother, aged 70+). The subject has been dealing with this issue since the age of 6, and it reached noticeable proportions during childhood. In the last few years, the disease seems to have progressed. Currently the subject’s vision is measured at 20/400 for the right eye and 20/200 for the left.

B. Explicit Difficulties with Seeing

Subject Zero describes looking at the world without glasses as constantly looking at an oil painting: the forms of the objects in front of him are clear enough for definition, the colors most often discernible, but as items are taken further away, they become more blurry. This happens fairly quickly, unlike with normally sighted people, where this effect does exist, but after great distances; thus the notion of “short-sightedness”.

Glasses serve to clarify the world a lot, bringing the world into better focus, yet there are still limitations to what the subject is able to see.

In addition, it has been shown that Subject Zero has a blind spot in the center of his field of vision, so when objects are moving directly towards him, they may not be visible until the last moment.

C. Examples of Problems Encountered and Solutions Where They Exist

Reading – Subject Zero needs to see things in large print, up close and preferably with good contrast. The ideal keyboard has bold black letting in a heavy and large font on a yellow background.

Where a few years ago, he was able to read paperbacks (with the book held within two inches of his face), he now has trouble with such small print, and feels comfortable reading bold, 24 point fonts.
Most of the subject’s problems are solved by magnification, but the resulting problem is that magnification is not always available as shown in the following scenarios:

- **Paring toe nails**, the act of using both hands on appendages that are too far away for clear visual focus results in the need for guesswork based on touch. A better solution may be found.

- **Finding a book, CD, etc.** from among a stack. Typically, when the subject needs to find a book on the shelf, he needs to grab a number of books in hand, hold them up one by one for inspection, and then replace them on the shelf and loop through the process until the item is found. Having domain knowledge always helps, so knowing the color, expecting a certain graphic etc. can prove to be helpful and is particularly so when searching out items in a supermarket.

- **Reading street signs, shop fronts etc.** Street signs are often too far away to be read until Subject Zero is virtually underneath them, the same goes for shop fronts. Sometimes, being right beneath them is insufficient. This problem is one of distance and mitigated focal length.

- **Seeing traffic light situations across a large road.** The subject can typically see a street light across a two lane road, but across greater distances, finding the light can be difficult and sometimes the light is misread – seeing green instead of red and vice-versa.

- **People recognition.** Subject Zero often mistakes people for others, and sometimes, especially outside, does not recognize familiar faces. People recognition is one of the most difficult issues. Subject Zero has had to deal with it for most of his life.
6. User Case Study

To best understand the concepts being discussed above, a use case of a blind student needing to get from his subway stop to his classroom is to be considered. We use the familiar example, as this is how a typical blind person would. So for this specific case we will have our student get off the No 1 Train in Manhattan, New York, at the 137th Street Station, which will bring him to the City College of New York. In this case he will come up the 138th Street, over Amsterdam Avenue at the cross light and getting to the NAC Building.

**Step 1 – getting out of the train, and orientating towards the path.**

Use of the cane will help the student find his way out of the subway station and above ground. At this point the only orientation required will be anything unexpected – a person in his path, maybe someone he should know, these issues could be solved with the aid of context:

![Figure 3: Looking at Broadway (The opposite direction)](image)
![Figure 4: Looking straight at the subway entrance](image)
![Figure 5: The park comes into view](image)

In the three pictures shown above we have scenes viewed from the starting point in the wrong direction. Figure 3 depicts Broadway which is opposite to where the subject needs to go. How do we know it is Broadway? Initially the system could geo-locate the subject (thus geo-referenced context), using GPS and some recorded domain knowledge of the map of this specific area (the exit from the 137th Street subway station, the corner of 137th and Broadway and the park). The subject would be looking for the park. This image can be said to define a multi-lane road with a traffic island in the center. The store (Duane Reade Pharmacy) can be recognized from its logo and verified through the reading of its name (partially obscured) we also have an MTA symbol visible which can give us more exact orientation (i.e. tell the subject to turn around clockwise). These provide landmark context.

The second image would come into view. People can be found to right of the image and the subway entrance is directly in the foreground of the image. The subway entrance should be easy to recognize as we have descending stairs, and recognizable text and symbols to aid us. Again, we have the temporal context that the user has just come out of the train station, providing a clue here. The park on the very right of the picture comes into view.

The subject is instructed to keep on turning clockwise.
In the third image the park becomes more evident. The grill fence which will define the pathways for us will become evident. So instead of detect the path within the park, the fence might be easier to detect and could be used as a background context to indicating the pathways.

The two images, Figures 6 and 7, both show the park and the path required for our subject to take. (We can see the street – Hamilton Place – in the background and this gives us additional confidence that we are going in the right direction).

It is the appearance of a person in the second image – something that should be easily defined from a comparison of the two images – we could say we use motion as a clue, that is, motion context to detect a moving target, which is of special importance to our subject as the person is not looking at the subject and a collision needs to be avoided. This is typical in an area of much pedestrian traffic. We can define this ontologically as ‘Person crosses in front of subject’.

Knowing that it is a person, not some other object is important:

- They will probably move out of the way quickly,
- They can be asked for help if such is required
- They might be someone the subject should know – especially if the person calls out a greeting of some sort.

The detection of a moving person could also be used as human context information to detect other human images that might be harder to detect. Since the similar illumination and viewing levels of the subject, the detection of other human subjects could be easier (Jain, 2010).
**Step 2 – Onto the path and finding the fence.**

2a) **Finding the path**

![Figure 8: Set 01: Frame 1 - Initial direction](image)

![Figure 9: Set 01: Frame 76 (2.5 seconds later)](image)

![Figure 10: Set 01: Frame 91 - The left side fence disappears from the image](image)

![Figure 11: Set 01: Frame 166 (2.5 seconds later)](image)

In the image selection above we can see how the context does not change much for several frames and then a major change in the images gives us a clue to what has happened.

Between figures 8 and 9, we can see how the grill fence comes closer as the fence on the left slowly retreats revealing the path which needs to be traversed. Here the object-to-object context marries the semantic context (i.e. that we expect a path) and as we get to the frame in figure 10, the context is changed by the disappearance of the fence on the left of the image. As we come to the frame in figure 11, the appearance of the fence crossing the lower edge of the image clues us into the fact that we should turn towards the path’s direction, the blind subject can certainly reach out his/her cane and feel the fence at his/her right.
2b) **Objects ahead in the subject’s path**

- Figure 12: Set 01: Frame 361, objects in the path ahead
- Figure 13: Set 01: Frame 391 - nearing the static object
- Figure 14: Set 01: Frame 436 - the side path becomes visible, the person is closer in relation to the bin

In figures 12 – 14 above, we can see how the two objects (the static blue bin and the walking person) give us some context. The bin (given prior knowledge) gives us the identity of another path, which as approached in figure 16 becomes evident, the break in apparent contiguous fence on the left as it curves away from the expected path. As well we have the motion context of the walking person as (using scale context) we can imply his approach. (He is well to the left of the path while our subject should be on the right of the path.

2c.1) **People approaching**

- Figure 15: Set 01: Frame 766 - Walker approaching
- Figure 16: Set 01: Frame 781
- Figure 17: Set 01: Frame 796 - Walker nears, also noticeable, more behind

In the above set of frames (figures 15 – 17), using object to object relationships, motion context is clearly shown and we can track the progress of a man walking towards the subject. Ontology might declare this as ‘person coming towards subject’. Behind the walker, we can discern more motion; although the vision system may not be able to deduce what it means at this time, a flag, signaling something approaching could be raised here.
2c.2) Multiple people, one coming straight at the subject

In the series of images above (figures 18 – 23), we get a situation of many people coming towards the subject (‘person walking towards the camera’) as we have noticed before; only here we have three individuals passing by, one seemingly coming directly to the subject, as deduced from the first three images – object in motion close to the fence which our subject is using as a guide with the cane. The last three images tell us that the person is crossing over the subject’s path, this uses the object-to-object spatial context as we can measure the space between the moving object (the person) and the static object (the fence) widening. This may also require some scale context as defined by the calibrations of the camera in motion.
2d) Exiting the park

Figure 24: Set 01: Frame 1531 - the end of the path becomes clear as the fences on each side no longer extend beyond the upper bounds of the image.

Figure 25: Set 01: Frame 1711 (3 seconds later) - the street crossing is now very clear to the observer.

In the above images (figures 24 and 25), we derive a clear understanding that we have reached the end of the path. For one thing, the fences at left and right no longer reach the upper bound of the images and with some semantic context, we can ascertain the end of the path. As we know we are at the end of the path - once again, using the context of ‘knowing where we are in general’ – location context - the image parser can presume streets, street corners, etc. Most noticeable are the distinct yellowish areas of the images. Semantic context expects to find pedestrian traffic lights and in the second of the two images, we get a pair, signifying where we should be crossing and guiding the subject towards the one first seen in the first image, which can be traced to the leftmost of the two in the second image.

Step 3 – Crossing at the light – on a relatively small crossing

Figure 26: Set 02: Frame 276 - the light shows a red hand

Figure 27: Set 02: Frame 331 - The light has changed showing a green (or is it white?) walking man

As shown in figures 28a/b, using the distinct colors associated with the pedestrian light – the distinct shade mustard-yellow of the signal box, the luminescent red and white against the near black of the actual traffic light, we can ascertain when the light changes (color specific context). Here the walking people do not help with knowing when to cross (we note that they are crossing on red – strictly unadvisable for blind people as they cannot react in time to vehicles that suddenly appear), however, by tracking the crossing people we can determine the correct
direction in which to go when crossing.

**Step 3. Alternative – watching the street for moving vehicles**

This video stream was taken a little out of context as typically our subject would not turn around to view passing traffic, but we consider these frames to be of value to the discussion.

![Figure 29: Set 03: Frame 421 - Vehicle crossing the street corner](image)

![Figure 30: Set 03: Frame 436 - As the vehicle moves on, we see only a stationary black vehicle in the background](image)

![Figure 31: Set 03: Frame 451 - the vehicle passes leaving open space behind it](image)

![Figure 32: Set 03: Frame 466 - the road seems clear, but the light is red! And the people are not moving!](image)

![Figure 33: Set 03: Frame 481 - The reason becomes clear half a second later - another vehicle comes into view](image)

![Figure 34: Set 03: Frame 492 - the vehicle is crossing, and the light is still red for pedestrians](image)

In the scene described above from figures 29 – 34, we see how motion context can be used to see vehicles crossing, how the light (color context) is utilized to avoid misunderstanding (that the road is clear and can be crossed). Also, the people remain stationary giving added impetus to the ‘do not cross now’ ontology.

**Step 4 – Starting up 138th street**

![Figure 35: Set 04: Frame 1 - orienting ourselves](image)

![Figure 36: Set 04: Frame 76 – some people coming in our direction](image)

![Figure 37: Set 04: Frame 406 – some 12 seconds later, still more people coming our way](image)

In this particular scene we revise some of our previous scenes, (figures 35 – 35), people are...
coming towards us. The wall with graffiti on our right is where we need our subject to get in line with so that the cane can be used to guide him/her forward. The cars are all moving away from the subject, and signify the edge of the road and the sidewalk (object-to-object related context). At this stage, the scaffolding in front of the subject is unidentifiable, perhaps with some geometric context connected to the fact that the subject is outside, the anomaly of the block of dark color of seemingly man-made origin (evidenced by its geometry) needs to be stored so that as it becomes more detailed as the subject nears it, we have the expectation of something we might need to consider at that time.

**Step 5 – Continuing up 138th street, approaching the scaffolding**

As the scaffolding is approached (as shown by the approximately 10 second stream shown by figures 38 – 40), based on semantic context – that of the general construction of scaffolding) we can ascertain the dimensions (spatial context) of the enclosure. When the subject’s cane comes up against the poles holding up the scaffolding, we can assure him/her that s/he can continue forward. As before, we have a crowd in the background and can expect people walking towards the subject. It may be difficult to deduce the stroller in figure 42, but in this case, avoidance is all that is necessary.

Figure 41, above emphasizes a problematic situation, confined by the scaffolding (geometrically calculated before) and having people milling about (using a humanoid template may be required here), a blind subject may need to speak up to request a clear path (doable because these are humanoid – human context). As the subject moves forward, the confines of the scaffolding fall
Step 6 – Avoiding the open gate

Continuing, the open gate is hard to see, notice the pattern of the shadows. The gate is still difficult to discern, but the lack of shadow directly in front can clue us in to its being there. That some obstacle is in front of the subject is becoming clearer, again, the shadow patterns are key.

In this particular scene, the open gate presents a problem common to blind subjects, it is difficult to detect and can quite easily be walked straight into. Using the shadows on the ground, discerning the patterns thrown by the fence, we have a sense of regularity which is then broken by a gap in the shadows, thus using the context of the shadows pattern, we are able to determine the anomaly and hopefully warn the subject of an unknown in the way that when touched by the cane will now make some sense to the subject, hopefully averting mishap.

Step 7 – Getting to the main traffic intersection

Coming up to the curb, the zebra crossing in front should help locate the light across the street. At the curb, locating the light, it is red.

The five second stream shown in figures 47 - 49 reveal how the main intersection is approached. The two sets of zebra crossings in the initial frame reveal the geometry of the intersection. Some structural context of the NAC building – across the street and stretching out to the right of the image can help us co-locate ourselves. Following the white lines that traverse the street (almost vertical in the image) we can trace the point at which the pole supporting the light should be and thus locate the pole.
Step 8 – Waiting for the light to change, watching the traffic go by

Figure 50: Set 08: Frame 106 – The vehicle passes, obscuring the light
Figure 51: Set 08: Frame 121 – as it passes, the whole wall is obscured including the school insignia of City College
Figure 52: Set 08: Frame 136 – the vehicle clears and our contextual points can be rediscovered

This scene, described in figures 50 – 52 above, is not atypical. As a vehicle passes (especially a large one) key markers are obfuscated for a time. However, knowing how we got them in the first place and recalling their relation to one another (spatial context) we should be able to quickly regain the focus, on the light and the zebra crossing. The CCNY insignia on the wall can be utilized as an additional reference in guiding the subject across (branding context).

Step 9 – Crossing Amsterdam Avenue

Figure 53: Set 09: Frame 1576 – as we cross at the green light, we keep the school insignia in sight as a guide.
Figure 54: Set 09: Frame 1591 – As we reach the other side, we notice the person in front is walking at the same rate as us.
Figure 55: Set 09: Frame 1606 – As we near the completion of the crossing the number of white lines in the zebra crossing lessens cuing us into our progress.

As our subject crosses at the light, it is necessary to calculate when s/he has completed the crossing, and be aware of people in front of them, either coming towards them or walking in their direction but slightly in front – this is to avoid collision if the person in front is walking more slowly or halts unexpectedly. Scale context is very relevant in this case as can be seen in the figures 53 – 55 above, where the person ahead remains constant in size, suggesting that s/he is moving at a similar pace to our subject. The lines of the zebra crossing give us an additional clue as to our progress as they lessen in number as one progresses.
7. Some Practical Problems and Possible Solutions

7.1. Surveillance Cameras and Sensors in Motion: Focus of Attention

The key difference in comparing a camera for surveillance to a camera system in motion is that we are unable to rely on the background. With any surveillance system, the main feature of the image captured is the constant background; in this case, the image can also be compared to an image of the pure background, then through a subtraction technique, where one can compare two images and isolate the pixels that are different in the two images, one is able to isolate objects of interest. Even in the case of non-static cameras, this is still possible; yes, a calibration of image stream needs to be calculated, but once done, one can presume sameness over time. At most a surveillance system will need to take into account light changes. When we consider a person (or a robot) in motion, the image becomes more complex immediately. First of all, we must consider the image flow and be able to calibrate the changes as the subject (the person or robot utilizing the system) moves. On top of this, the scene changes quickly and often, such that one cannot rely on pre-processed images. Context, then, replaces the sameness in giving us an ability to isolate objects of interest. It is true that with some cases, when we can recognize a setting, and when we have ready saved images of the setting, we may be able to use the sameness to isolate our objects, but we cannot rely on either being given us at this stage. We might seek to revisit this notion at some later stage.

7.2. Limits of Processing Power on Transportable Devices: Context in the Cloud

Although computing power improves every year, we must still take into account the limitations of transportable devices. While a desktop can run several CPUs each with several cores, smart phones and tablets (our most common choice of transportable devices) are much more limited. It is for this reason that we must consider how to best minimalize the amount of processing required by the computing portion of the transportable system. On the other hand, as we now live in an era where through Wi-Fi technology and cloud computing, we can upload sample data to servers and receive processed data relatively quickly in return, thus allowing for, over time, large data sets of shared knowledge to be utilized, and this should let us explore both aspects and work with it.

7.3. A Modular Camera System: Context Awareness through Modular Designs

It can be inferred, both from the literature and the statements made by users of accessibility aids, that one of the most problematic aspects of current computerized solutions is that the systems are typically made up of parts that are welded together such that there is little room for updating, either by adding or removing physical aspects of the system, or by updating the software within, particularly when we want to incorporate context information into the system. We have taken this to task in considering a more practical solution. At this early stage, we will limit ourselves to
cameras as input devices, but fully expect that we may add other sensors to the system. We have also considered that since the set of users is heterogeneous, the output of the system must also be different as per the users’ needs and requirements.

So, a three-module system is envisioned: a central module which is the central processor unit, the input devices, the second module; and the output devices, the third module. By separating out the three modules, a more robust framework can be created, made easier to upgrade and extend as the system evolves, and to integrate context information into the system.

For now, we suggest a relatively simple system: for input we would have types of interfaces: sensing interface and context interface. The sensing interface could include two cameras, one placed close to the eyes, a pair of spectacles with a camera embedded into one of the arms of the glasses; and a mini-camera worn on a finger, most likely the index finger of the subjects most used hand. The head based camera would service best to “see” the scene, the finger camera being used to aid with discerning detail such as buttons in an elevator, the text on spines of books and/or CD and DVD boxes, etc. The context interface will allow the user to incorporate context information derived from other sources, such as GPS, Internet, human interaction, and other sensing modalities.

The output can be one of two sorts: a) a visual output – typically good for visually handicapped, but still vision dependent subjects, in this case magnification and other image augmentation (such as enhanced contrast) can be applied; and b) non-visual, typically audio and/or haptic output which would be worn by the subject in strategic positions on the body.

The computing unit is the most versatile in terms of what can be applied. For a visually challenged but vision-reliant subject, a tablet PC is probably the best solution, combining the computing aspects of the system with portability and the ability to generate enlarged and augmented images. For the totally blind, a less cumbersome device, probably wearable, is desired, as the vision-less require as hands-free a system as possible. The key to any system, however, is the ability to uplink and download from the internet, preferably in a wireless mode, so that context understanding using cloud computing is possible. Indeed, we would try and develop a system that uses as little wiring to connect the different tiers as possible.
8. Conclusions and Discussions

Annotation and tagging are utilized by many to aid in understanding images. It is our opinion, therefore, that utilizing context, achieved by marrying existing image processing techniques with a reasoning system may achieve even better results. In our scenarios, we are working with a blind subject in a relatively well-known environment. This means that we have much context information we are able to utilize in our reasoning system. As both Farcy (2006) and Jain (2010) have shown, context in image processing has the potential of narrowing the semantic gap. When we combine these two facts, the use of context and the pre-defined environment, by which we mean, we are able – with considerable confidence – to ascertain what is going on around the subject in sufficient detail so as to guide the subject as s/he navigates his/her path and orientates him/herself so as not to collide with either moving or static objects, to be able to predict when extra awareness is required, and in situations where the subject might become confused through lack of visual confirmation, to somehow supplement this loss by giving the subject enough data with which to handle to problem. With any computerized processing of input, references are necessary for automated processing to continue in any meaningful direction. With image processing, this is even more important as the nearly unformatted data that is fed into a computer vision system has proven quite difficult to make sense of at any relevant level.

Acknowledgments

This work is supported by US National Science Foundation Emerging Frontiers in Research and Innovation Program under Award No. EFRI-1137172, and CitySeeds: City College 2011 President Grant for Interdisciplinary Scientific Research Collaborations.

I would like to thank my mentor Professor Zhigang Zhu and his PhD student Edgardo Molina for their help in getting a firmer understanding of the underlying technical issues as they have allowed me to concentrate on this top-down approach.
References


[16]. V. Jain. 2010. “Using Context to Enhance the Understanding of Face Images” Submitted to the Graduate School of the University of Massachusetts Amherst in partial fulfillment of the requirements for the degree of Doctor of Philosophy, September 2010


May 2002.


Table of Figures

Figure 1: Comparing subjects - Low vision, Blind and Robots ........................................ - 14 -
Figure 2: Facets of subjects' needs ................................................................................... - 15 -
Figure 3: Looking at Broadway (The opposite direction) ............................................... - 18 -
Figure 4: Looking straight at the subway entrance .......................................................... - 18 -
Figure 5: The park comes into view .................................................................................. - 18 -
Figure 6: Looking at the path to take (the right direction) ............................................. - 19 -
Figure 7: Almost the same image - but with a person in the way ...................................... - 19 -
Figure 8: Set 01: Frame 1 - Initial direction ...................................................................... - 20 -
Figure 9: Set 01: Frame 76 (2.5 seconds later) ................................................................. - 20 -
Figure 10: Set 01: Frame 91 - The left side fence disappears from the image .............. - 20 -
Figure 11: Set 01: Frame 166 (2.5 seconds later) ............................................................. - 20 -
Figure 12: Set 01: Frame 361, objects in the path ahead ................................................ - 21 -
Figure 13: Set 01: Frame 391 - nearing the static object ................................................ - 21 -
Figure 14: Set 01: Frame 436 - the side path becomes visible, the person is closer in relation to the bin........................................................................................................ - 21 -
Figure 15: Set 01: Frame 766 - Walker approaching ........................................................ - 21 -
Figure 16: Set 01: Frame 781 ......................................................................................... - 21 -
Figure 17: Set 01: Frame 796 - Walker nears, also noticeable, more behind ............... - 21 -
Figure 18: Set 01: Frame 901 (some 7 seconds after the image in figure 17), the individuals approaching become distinct .............................................................................. - 22 -
Figure 19: Set 01: Frame 916: Three individuals walking towards the camera become apparent - 22 -
Figure 20: Set 01: Frame 931: The individual on the right of the image is coming directly at the subject .................................................................................................................. - 22 -
Figure 21: Set 01: Frame 946: The person coming towards the subject is up close - can we ascertain that she has turned to avoid the subject? .................................................. - 22 -
Figure 22: Set 01: Frame 961 - The person in front is very close now, but moving away from the fence ........................................................................................................................................ - 22 -
Figure 23: Set 01: Frame 976 - The person in front of the subject has crossed over leaving the way clear ................................................................................................................................... - 22 -
Figure 24: Set 01: Frame 1531 - the end of the path becomes clear as the fences on each side no longer extend beyond the upper bounds of the image ........................................... - 22 -
Figure 25: Set 01: Frame 1711 (3 seconds later) - the street crossing is now very clear to the observer ................................................................................................................................ - 23 -
Figure 26: Set 02: Frame 276 - the light shows a red hand .............................................. - 23 -
Figure 27: Set 02: Frame 331 - The light has changed showing a green (or is it white?) walking man .................................................................................................................................................. - 23 -
Figure 28a, b: Actual high resolution representation of the red and green lights as they appear in the images .................................................................................................................. - 23 -
Figure 29: Set 03: Frame 421 - Vehicle crossing the street corner .................................................. 24 

Figure 30: Set 03: Frame 436 - As the vehicle moves on, we see only a stationary black vehicle in the background .......................................................................................................................... 24 

Figure 31: Set 03: Frame 451 - the vehicle passes leaving open space behind it .................. 24 

Figure 32: Set 03: Frame 466 - the road seems clear, but the light is red! And the people are not moving! .................................................................................................................................................. 24 

Figure 33: Set 03: Frame 481 - The reason becomes clear half a second later - another vehicle comes into view ........................................................................................................................................... 24 

Figure 34: Set 03: Frame 492 - the vehicle is crossing, and the light is still red for pedestrians 24 

Figure 35: Set 04: Frame 1 - orienting ourselves ................................................................................................................... 24 

Figure 36: Set 04: Frame 76 - some people coming in our direction ..................................... 24 

Figure 37: Set 04: Frame 406 - some 12 seconds later, still more people coming our way ... 24 

Figure 38: Set 05: Frame 181 - approaching the scaffolding .................................................. 25 

Figure 39: Set 05: Frame 271 - coming beneath the scaffolding, people approaching ...... 25 

Figure 40: Set 05: Frame 301 - the pole on the right needs noticing ...................................... 25 

Figure 41: Set 05: Frame 436 - coming up to the milling crowd of people ......................... 25 

Figure 42: Set 05: Frame 571 - Navigating the crowd, the exit from the scaffolding in view. - 25 

Figure 43: Set 05: Frame 576 - Still people milling about, but now clear of the scaffolding... 25 

Figure 44: Set 06: Frame 751 - Continuing, the open gate is hard to see, notice the pattern of the shadows .................................................................................................................................................. 26 

Figure 45: Set 06: Frame 796 - the gate is still difficult to discern, but the lack of shadow directly in front can clue us in to its being there .................................................................................................................. 26 

Figure 46: Set 06: Frame 841 - that some obstacle is in front of the subject is becoming clearer, again, the shadow patterns are key ................................................................. 26 

Figure 47: Set 07: Frame 856 - coming up to the big traffic intersection and some more traffic lights .................................................................................................................................................. 26 

Figure 48: Set 07: Frame 901 - coming up to the curb, the zebra crossing in front should help locate the light across the street .................................................................................................................. 26 

Figure 49: at the curb, locating the light, it is red ...................................................................... 26 

Figure 50: Set 08: Frame 106 - The vehicle passes, obscuring the light .............................. 27 

Figure 51: Set 08: Frame 121 - as it passes, the whole wall is obscured including the school insignia of City College .......................................................................................................................... 27 

Figure 52: Set 08: Frame 136 - the vehicle clears and our contextual points can be rediscovered 27 

Figure 53: Set 09: Frame 1576 - as we cross at the green light, we keep the school insignia in sight as a guide .................................................................................................................................................. 27 

Figure 54: Set 09: Frame 1591 - As we reach the other side, we notice the person in front is walking at the same rate as us .............................................................................................................. 27 

Figure 55: Set 09: Frame 1606 - As we near the completion of the crossing the number of white
lines in the zebra crossing lessens cuing us into our progress.

Figure 56: Camera mounted on glasses

Figure 57: A small camera that can be mounted on the finger