Computer Vision Based People Tracking for Motivating Behavior in Public Spaces

by

Jacob A. Hyman

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

August 2003 [September 2003]

Author . . . .
Department of Electrical Engineering and Computer Science
August 22, 2003

Certified by . . . .
Stephen S. Intille
Changing Places / House_n Technology Director
Thesis Supervisor

Accepted by . .
Arthur C. Smith
Chairman, Department Committee on Graduate Students

Computer Vision Based People Tracking for Motivating Behavior in Public Spaces

by

Jacob A. Hyman

Submitted to the Department of Electrical Engineering and Computer Science
on August 22, 2003, in partial fulfillment of the
requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

In this work a system that automates the process of people counting to determine what effects "just-in-time" messages have on motivating behavior is described. The system is designed to permit automatic study of the impact of motivational messages on people's stair use. A projector presents a point-of-decision message to passers-by choosing between a set of stairs and an escalator while a computer vision algorithm counts each type of traffic. Preliminary results of the effects of messages displayed in a Boston area subway station are discussed. The system is designed to be easily moved to different locations with minimal change to the setup and algorithm. Results from an initial trail showed a 4.3% increase in stair usage ($p < .001$), demonstrating both the viability of the measurement technology and the potential of point-of-decision messaging to change behavior.

Thesis Supervisor: Stephen S. Intille
Title: Changing Places / House_*n* Technology Director

Acknowledgments

I would like to thank my advisor, Stephen Intille, for his support and guidance over the last year and a half. Without his vision, this project would never have happened. Thank you for donating so much of your time and effort in helping me complete this thesis.

I would also like to thank my parents, Sandra and Derry, and my brother Aaron for all that you have done for me over the last 23 years. You have always encouraged me in all of my pursuits, and it is because of you that I was able to make it through MIT. Any words I could think of would not reflect how much you have meant to me. Thanks for your continued love and support for the last 5 years I have been away from home.

House_n has provided a wonderful work environment for me over the last year and a half. Thank you to all the graduate students, undergrads, research staff, and support staff for providing such a great place to do research. Special thanks to TJ and Tyson for all your help with "The Box" and to John for annotating all those hours of video.

My time at MIT was defined by the friends I made along the way. Thank you for making my stay here so much better.

I would also like to thank the Metro Boston Transit Authority for allowing us to install counting systems in 3 of their stations.

This work was funded in part by the Robert Wood Johnson Foundation (Improving methods for measuring stair use in public spaces, #044256) and by the National Science Foundation (ITR/PE: Using context recognition for preventive medicine in the home, #0112900).

Contents

1 Introduction **17**

1.1 The Need for Preventative Health Technologies 17

1.2 Measuring Behavior in Public Spaces 18

1.3 Contributions of This Work 19

2 The Problem **21**

2.1 Motivating Behavior Change in Public Spaces 21

2.2 Scenarios for Proactive Messaging 23

2.3 A New Opportunity 24

2.4 Measuring the Impact of Technology on Behavior 26

3 Theory and Rationale **27**

3.1 Point of Decision Messaging 27

3.2 Previous Stair Studies 28

3.2.1 Observations on Previous Stair Studies 33

3.3 People Counting Algorithms 33

3.3.1 Goals of Stairs/Escalator Counter 38

4 Design and Implementation **39**

4.1 System Overview 39

4.2 System Components 40

4.3 The People Counting Algorithm 43

4.3.1 Image Acquisition and Pre-Processing 43

4.3.2 Frame Differencing 43
4.3.3 Background Subtraction 44
4.3.4 Image Stacks 45
4.3.5 Analysis of Image Stacks 47
4.3.6 Tracking 47
4.3.7 Data Storage 49
4.3.8 General Performance Characteristics 49
4.4 Presentation of Motivational Messages 51

5 Evaluation **53**
5.1 Installation Sites 53
5.1.1 Kendall Inbound 53
5.1.2 Kendall Outbound 54
5.1.3 Davis 56
5.2 Obtaining Test Data 57
5.3 People Counting Algorithm Performance 57
5.4 Impact of Motivational Messages on Stair Usage 63
5.5 Computing Significance of Results 69
5.6 Evaluation as a Cost-effective Measurement Tool 71

6 Conclusion **73**

A Enclosure Box Diagrams **75**

B Camera Turret Diagram **83**

C Projector Control **85**

D Parts List **89**

E Calibration **91**

F Interesting Daily Graphs **97**

G People Count Data **103**

H Annotation Software **107**

I Data Collection Procedures **109**

I.1 Communicating with the Laptop 109

I.2 Test Video 109

I.3 Daily Counts 110

J Letter from the MBTA **111**

K Data Set Formats **113**

K.1 Annotation Data Format 113

K.2 Escalator File Format - AVI 113

K.3 Escalator File Format - Live Video 114

K.4 Stair File Format - AVI 114

K.5 Stair File Format - Live Video 115

L Matlab Simulation Code **117**

List of Figures

2-1 The motivational sign used by Brownell et al. in their original stair study [8]. 22

2-2 Results scanned from the Brownell et al. follow up study [8] showing an increase in stair usage during intervention that returned to slightly above baseline in 3 months. 22

2-3 Everywhere Display providing information about energy use. 24

4-1 The system in Kendall outbound station. The box has a projector to display the sign and a camera directed at the stairs/escalator and connected to a laptop to count passers-by. 40

4-2 The view of the stairs and escalator from the Kendall inbound camera (320 x 240 resolution). 41

4-3 Flowchart describing people counting algorithm. 42

4-4 Images from Kendall inbound station with a crowd of people on the stairs and escalator. 42

4-5 Real image and corresponding difference image. 44

4-6 Real image and corresponding background subtraction image. Note that background subtraction is only performed in areas of interest that are highlighted. 44

4-7 Two image stacks. The left is created from the background subtraction images, and the right is created from the difference images. 46

4-8 Scan lines places along the stairs and escalator during calibration of an installation site. 46

4-9 Trackers tracking a person's movement down the stairs. Because of arm movement, some of the trackers move slightly up. A threshold is set on how far a tracker needs to move in either direction to avoid having these trackers count pixels moving in the wrong direction. . . 48

4-10 The algorithm must perform in conditions that can vary as much as these images. The image on the left is from when the sun is shining directly into the station. The image on the right is from a dark evening. 50

5-1 View of Kendall inbound system and projection. The second image shows the distance between the turnstiles and the stairs/escalator. . . 54

5-2 View of Kendall outbound system and projection. 55

5-3 Side view of the Kendall outbound station. 55

5-4 View of Davis system. The sign will be projected against the brick wall to the right of the escalators. 56

5-5 Graphs showing the typical traffic flow through Kendall inbound for both the escalator and the stairs. The graphs also show how the algorithm estimated the number of people throughout the day. The graph on the left shows a weekday, and the graph on the right shows a weekend. 60

5-6 Graphs showing the typical traffic flow through Kendall outbound for both the escalator and the stairs. The graphs also show how the algorithm estimated the number of people throughout the day. The graph on the left shows a weekday, and the graph on the right shows a weekend. 62

5-7 Graphs showing the typical traffic flow through Davis for both the escalator and the stairs. The graphs also show how the algorithm estimated the number of people throughout the day. The graph on the left shows a weekday, and the graph on the right shows a weekend. . . 64

5-8 Intervention sign (original Brownell study [8] with escalator heart modified). 65

5-9 Stair usage percentage for weekdays before and after the intervention in Kendall inbound. Stair usage increased from 39.3% to 43.7% ($p < .001$). 66

5-10 Stair usage percentage for weekend days before and after the intervention in Kendall inbound. Stair usage increased from 23.0% to 26.0% ($p < .001$). 67

5-11 Stair usage percentage for weekdays before and after the intervention in Kendall outbound. Stair usage increased from 38.8% to 39.6% ($p < .0155$). 67

5-12 A comparison of the view of the signs from the turnstiles in both Kendall inbound (left) and Kendall outbound (right). The top of the sign in Kendall outbound is obscured. 68

5-13 Stair usage percentage for weekend days before and after the intervention in Kendall outbound. Stair usage increased from 21.8% to 24.6% ($p < .001$). 69

F-1 The cumulative traffic through Kendall inbound on 7/27/03. The escalator has a barrier in front of it until 11:25 am. 98

F-2 A cone is placed in front of the escalator on 7/27/03 which is later replaced by a barrier. 98

F-3 The cone which replaced the barrier is eventually removed and people resume using the escalator. 99

F-4 The cumulative traffic through Kendall inbound on 7/28/03. The escalator was broken until 7 pm. 99

F-5 Workers arrive to fix a broken escalator and put a barrier in front of the entrance on 7/28/03. 100

F-6 The barrier is removed, but the workers continue fixing the panel in front of the escalator on 7/28/03. 100

F-7 The escalator is eventually fixed on 7/28/03. 100

H-1 Video annotating software. The control panel on the left has speed and position sliders, buttons to annotate people going up the escalator, down the stairs, and up the stairs, and a pause button. The panel on the right shows the video. 108

J-1 Letter from the MBTA giving permission to perform experiment in Kendall inbound, Kendall outbound, and Davis. 112

List of Tables

5.1 Table showing the manual vs. automatic counts for 4 days at Kendall inbound. For each day, the stair use percentage for manual counts and automatic counts is also shown. 58
5.2 Table showing the manual vs. automatic counts for 4 days at Kendall outbound. For each day, the stair use percentage for manual counts and automatic counts is also shown. 59
5.3 Table showing the manual vs. automatic counts for 4 days at Davis. For each day, the stair use percentage for manual counts and automatic counts is also shown. 61

E.1 Table showing the threshold values used in both Kendall stations and Davis. 95

G.1 Table showing the counts collected in Kendall inbound and outbound during the intervention period. 104
G.2 Table showing the counts collected in Kendall inbound and outbound during the baseline period. 105

Chapter 1

Introduction

1.1 The Need for Preventative Health Technologies

As a result of the baby boom in the 1940's and 50's, the percentage of the world's population over 60 is projected to more than double in the next 50 years [23]. An already overwhelmed health care system needs to prepare for the burden of an expanding elderly population. Encouraging healthy habits and investing in preventative health care could be a major factor in combating this problem.

One way to contribute to the goal of creating healthier lifestyles is to encourage exercise in public spaces. The Surgeon General has endorsed a Centers for Disease Control recommendation of 30 minutes or more of moderate intensity exercise every day [32]. Incorporating exercise into daily routine makes achieving that level of daily exercise easier, because people do not have to build time specifically for exercise into their schedules. Point-of-decision messages can inform people of choices that lead to easy exercise and ecourage healthier behavior. Stair usage is a good example of this type of activity [8, 25, 20, 7, 6, 3]. Additionally, stairs are present in many public spaces, and climbing a set of stairs produces a similar expenditure of energy to aerobic dancing, bicycling, or in-line skating per unit time [1].

Most prior work in the area of using point-of-decision interventions has focused on simple, static presentations of information. However, new technologies show promise to increase the relevance and effectiveness of "just-in-time" messaging by monitoring and responding to events as they occur [18]. The looming health care crisis is creating a need to study how to encourage healthier lifestyles and keep the aging population out of doctor's offices. Incorporating sensing and display technology into environments potentially provides an inexpensive way to positively affect the behavior of large numbers of people.

1.2 Measuring Behavior in Public Spaces

By creating artificial environments to perform behavioral experiments, an element of bias is introduced to any experiment. Brownell et al. ran one of the first studies to examine the impact of "just-in-time" messages on physical activity in real environments [8]. In many situations, it is possible to discreetly observe people's decisions in a natural environment. For example, when an escalator and staircase are placed side by side, people have the option of two methods of ascent. By observing this choice, a researcher can determine patterns of physical activity in a non-laboratory setting. This type of experimental design allows a researcher to observe thousands of subjects at intervention points to collect statistically significant data on point of behavior messaging and motivation of physical activities.

Brownell et al. [8] determined that an intervention in the form of a small sign encouraging stair use placed at the base of stair/escalator pairs in a commuter station, bus-terminal, and shopping mall led to an increase of nearly 7% in the number of people taking the stairs during the intervention period. Additionally, this increase in stair usage lasted up to three months after the intervention had been removed. These results inspired many similar experiments that studied the effects of interventions on stair usage because of the potential for a cost-effective way to make a large impact in people's behavior. Studies using signs [25, 26, 6, 3, 9], stair-riser banners [20], and artwork and music [7] as interventions all found increases in stair usage from 2% to 9% overall during intervention periods.

While these studies have shown new ways of generating behavior change in public spaces, none of them have demonstrated new ways to measure this change. Even though the interventions in these experiments are cost effective, the actual experiments are expensive to perform. In nearly all experiments human observers have manually counted each person's decision on choosing the stair or escalator. Consequently, research that attempts to study behavior changes over long time periods or in multiple locations becomes difficult. When measurement requires a human observer, examining how the impact of an intervention wears off and the long-term habituation of healthier behavior becomes costly. However, in this work, technology is developed that can replace direct observation with automated monitoring. The flat cost of an automated system for observation means longer studies are only marginally more expensive in both time and money than short studies. Additionally, a well designed system could easily be moved to a different location to make future experimentation easy. Most importantly, automated monitoring creates the opportunity to study context-sensitive and dynamic interventions.

1.3 Contributions of This Work

This work uses computer-vision based people counting and video projection to automate the process of behavior change experimentation in public spaces. With this system, the impact of multiple motivational messages on influencing active behavior in public spaces can be compared. Further, studies can run for any length of time with little incremental cost in dollars or labor. A computer vision algorithm for counting people crossing a line is presented that has been tested in 3 public transit locations. The algorithm does not rely on a specific camera angle and has shown consistent performance in the presence of natural lighting and multiple occlusions. This system can be easily installed in different public spaces with little to no modification required in the hardware or software. This technology will enable new public health studies on the impact of "just-in-time" messaging on behavior in public spaces.

Chapter 2

The Problem

2.1 Motivating Behavior Change in Public Spaces

The study that inspired this project and all preceding stair studies was performed over 20 years ago by Brownell et al [8]. The original experiment was designed to monitor physical activity in natural settings and see what effect a small sign might have on this activity. Subjects were observed making a choice between adjacent stairwells and escalators in a Philadelphia area shopping mall, train station, and bus station. This first study found that a .9 by 1.1 meter poster encouraging stair use placed on an easel at the base of the stairs/escalator (see Fig. 2-1) increased stair usage from 5.3% to 13.7% in the first trial and from 7.1% to 15% during the second trial. A location specific effect was also seen in the trials. A larger number of people used the stairs at a shopping mall location than at two commuter stations (train and bus) during the baseline phase. Additionally, the increase in stair usage at the shopping mall was higher (10.2%) than at the train station (7.1%) and the bus station (4.7%).

These results led to a follow-up study examining the effects of repeated exposure and long term effects of an intervention. 24,603 total observations were made at a commuter station in downtown Philadelphia on weekdays from 7:30am to 9:00am. First, a baseline observation was made for 5 days, followed by 15 days of intervention. Afterwards the intervention was removed and observation remained for a 10-day period. There was then a 5 day follow up with no intervention one month and three

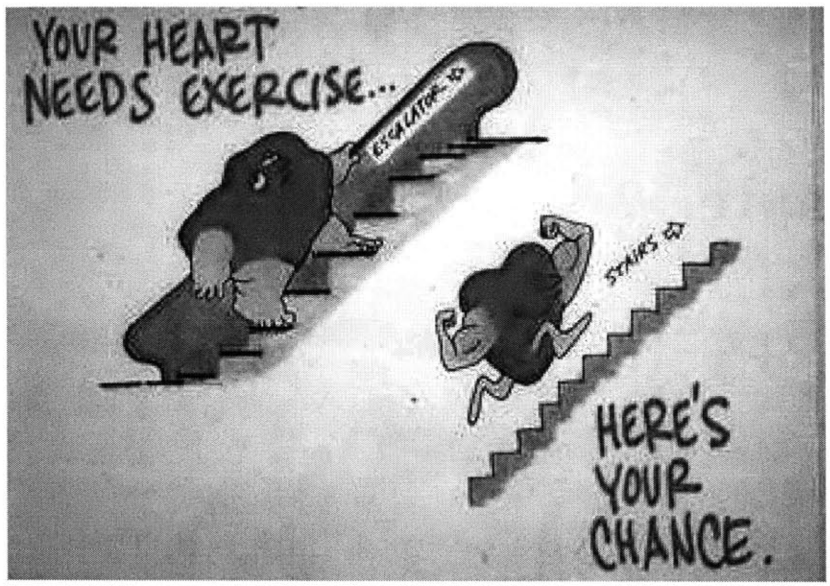


Figure 2-1: The motivational sign used by Brownell et al. in their original stair study [8].

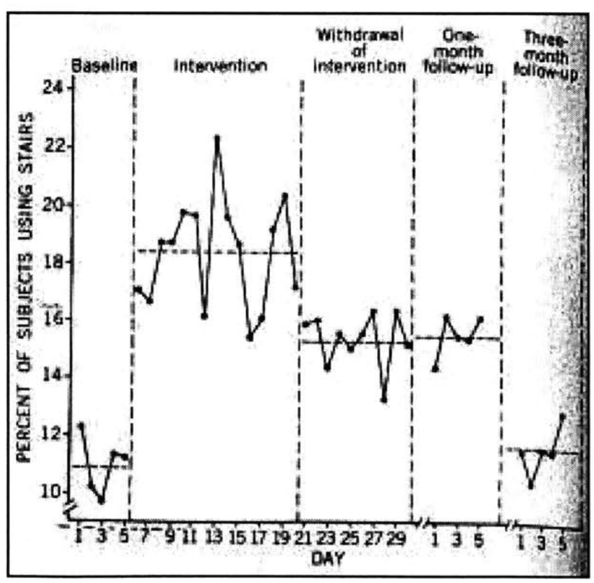


Figure 2-2: Results scanned from the Brownell et al. follow up study [8] showing an increase in stair usage during intervention that returned to slightly above baseline in 3 months.

months later. Stair use was observed to have increased 11.6% to 18.3% from the baseline to intervention period. In the week after intervention, the usage level remained elevated at 15.5% and one month later was still at 15.6%. After 3 months, stair usage had returned to near baseline at 11.9% (see Fig. 2-2).

This experiment showed that an inexpensive sign can significantly increase the amount of physical activity performed in public spaces. Brownell et al. observed that past research has shown 30-50% of persons starting a rigorous exercise program stop before finishing. An even larger number do not completely adhere to the exercise program ([4, 33] as reported in [8]). Large scale changes in behavior are hard to institute, but Brownell's study has shown that small changes are easy to initiate and maintain by minor modifications to public spaces. Simple, clearly-stated messages can be added to environments without creating disruption in traffic flow. When properly placed, the messages present information about an upcoming decision at the moment of choice. People moving through the station are alerted to the decision they can make and given an incentive to take the stairs (it is healthier), and a small behavior change is created. Over time an attitude change can occur leading people to identify other small changes they can make towards a healthier lifestyle. More research on point-of-decision messaging in public spaces may lead to environments that encourage healthy behavior.

2.2 Scenarios for Proactive Messaging

In the near future, distributed computing techniques combined with smaller and more prevalent sensing technology will enable both public and private spaces to be outfitted with proactive messaging technologies.

For example, suppose an office worker is on a bus commuting to work. He pulls out his PDA/cellualar phone device and makes a call. After he finishes his call, the device immediately vibrates, and he glances at the screen. "Take the stairs when you get to the office" is displayed on the screen. Because the worker has just finished a phone call, he is not interrupted from another task. The PDA determined that based on

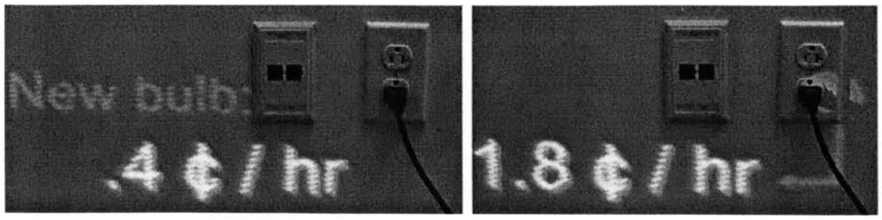


Figure 2-3: Everywhere Display providing information about energy use.

the time, GPS location information, and the worker's typical schedule that he would likely arrive at work in the next few minutes. The ability to present information without creating an interruption is exploited to present a "just-in-time" reminder to motivate physical activity.

In addition to personal devices being able to recognize behavior and provide feedback, display technologies embedded in environments will be able to better convey these messages. For example, the Everywhere Display [22] is a device that can project information onto any flat surface in a small room without perspective distortion. This technology can be used to provide messages in the home environment intended to modify behavior. For example, many people do not know the amount of energy used by common electrical devices such as lamps or how much money they could save by switching to energy efficient bulbs. With the Everywhere Display, when a homeowner turns on a lamp, he can be informed of this information in an unobtrusive way (see Fig. 2-3) that may gradually help the homeowner learn the consequences of small decisions. By presenting messages with useful information at appropriate times, the computer system may encourage the occupant to adjust behaviors, such as switching to energy efficient bulbs.

2.3 A New Opportunity

Using computers and technology presents a new opportunity to motivate behavior change in an effective and inexpensive manner. Previous studies have already shown that simple messages displayed at the right time and place can influence behavior in public spaces [25, 26, 6, 3, 9]. However, this research has focused on static messages that do not change based on the surroundings. The scenarios previously described suggest that using current (or soon to be available) technology, computer systems will have the capability to present "just-in-time" context-sensitive feedback to users with the goal of motivating behavior change. The ability for computers to persuade is a new research area. BJ Fogg has identified seven topics that further research on captology [1] needs to explore, five of which are directly examined by the work in this thesis [11]:

- Interactive technologies that change behavior
- Adapting theories and frameworks from other fields
- Using interactive technologies that are specialized, distributed, or embedded
- Using technologies where the designer's goal is to persuade
- Focusing on "what is" and "what could be." This means not only focusing on potential persuasive methods, but also exploring technology with respect to current persuasive techniques.

Using sensors to detect context and new display technologies (whether built into PDA/phone-like devices or embedded into the environment), researchers can study "point of decision" messaging where interventions are personalized to individuals and their activities. For example, if a system notices that the escalator has become very crowded but the stairs remain empty, it can change the message that is displayed to passers-by. Instead of a 'Take the Stairs" message, a new message stating "Save time, don't wait for the escalator, take the stairs" can be displayed. This ability to tailor messages to decisions as they occur is what makes "just-in-time" messaging so potentially powerful for affecting behavior.

[1] Use of computers as persuasive technologies.

2.4 Measuring the Impact of Technology on Behavior

Most previous studies on "just-in-time" messaging to motivate physical activity have used large amounts of man power to observe behavioral choices in real environments. Sensing technologies can greatly reduce the man power required and enable context-sensitive interventions. The ability to tailor messages to specific situations as they occur has the potential to significantly increase the impact of a behavior change message presented in a public space.

In past stair studies, a person had to be physically present to count each usage of the stairs or escalator. For a transit station that is open 20 hours a day, that would require 140 hours of labor to collect data for just one week. This type of comprehensive measurement is prohibitively expensive. Many previous stair studies only observed stair usage for 1-2 hours a day [6, 8, 9, 19, 20, 25], while a few made observations up to 5 hours a day [7, 26].

The system described in this thesis allows for studies that explore some of the currently unknown attributes of point of decision interventions. Because the system contains a computer that, in conjunction with sensors, can monitor the environment, the effect of altering messages based on an environmental situation can be observed. Studies can cost-effectively examine the effects of messaging over periods of time longer than a year. The effects of different types of messages can be examined without manual intervention in the study.

Chapter 3

Theory and Rationale

3.1 Point of Decision Messaging

Studies involving point of decision messaging have primarily focused on workplace production or preventative health. Regardless of the target population, the research shares a common goal of providing information to people at the proper time to influence their behavior. In both areas of application, point of decision information has been shown to be a useful tool.

For a system to be effective in motivating behavior change, it needs to present messages that are easy to understand at the right time and place in a non-annoying way. Display technologies such as the Everywhere Display [22], portable displays on cell phones, and embedded LCDs can be used to display the messages. Conveying appropriate messages at the right time requires sensing technology that can determine the context of how the user's actions or activity in the environment may effect behavior. By combining sensing and display technology, the full potential of "just-in-time" messaging can be achieved. Advanced technology is not a necessity for point of decision messaging. Studies using simple technology have created an improvement in safety in the workplace [21, 14], encouraged seat belt use [12], increased public recycling [12, 13], and reduced electricity consumption [38, 29].

For example, in an attempt to improve healthy decisions at grocery stores, the Pawtucket Heart Health Program attempted to use point-of-purchase labels as part of a larger community interaction designed to prevent cardiovascular disease [17]. Over four years from 1984 to 1988, they placed labels in four different supermarkets identifying foods as "low-fat," "low-sodium," "low-fat, low-sodium," and "fat ratio OK." Once a year, exiting shoppers were randomly selected and those that had made a purchase were asked if they had seen one of four sets of labels in the store. Those that correctly identified the labels were asked if they were encouraged to purchase the labelled foods. Awareness of the labels increased from 11% to 24% over the course of the study and the percentage of people who said they were motivated to purchase labelled foods increased from 36% to 54%. From these results, the authors of the study estimated that 636 people per week were encouraged to purchase the labelled products.

3.2 Previous Stair Studies

As a follow up to Brownell, Stunkard, and Albaum's [8] original study involving point-of-decision intervention for stair usage, researchers have studied the impact of other types of motivational messages. Blamey, Mutrie, and Aitchison performed a similar study that produced comparable results overall [6]. In their study, conducted in a Scottish underground station, an increase from 8% baseline stair use to 15-17% stair use during intervention was seen. Multiple signs saying "Stay Healthy, Save Time, Use the Stairs" were posted for 3 weeks. Observations were made for an additional 2 weeks and then again after 4 weeks and 12 weeks. The researchers used stepwise logistic regression with stair/escalator usage as the response variable and subject sex, week of study, and their interactions as explanatory factors. The results showed that men were more likely to take the stairs at all times (12% baseline and 21% intervention vs. 5% and 12% respectively for women) and stair usage still remained elevated 12 weeks after the intervention. A downward trend over the weeks after the intervention suggest that stair usage my eventually reach baseline levels again. The overall trends seen in this study match the results of Brownell et al. [8]

Coleman and Gonzalez examined the effects of a culturally targeted intervention in the predominantly Hispanic community of El Paso, Texas [9]. Their results did not show a significant improvement in stair usage when using an image promoting health "for the life of your family" compared to a promotional image targeted at individuals. They did find a wide variety of variance across their intervention sites. While stair usage increased among women at all sites, stair usage by men actually decreased at a public library and office building. At both these sites, subjects were choosing between an elevator and stairs (leaving more time for those planning on taking the elevator to make a decision). Additionally, a larger percentage of people were already taking the stairs during baseline. This indicates future exploration might be needed when determining how baseline stair usage and the time available to make a decision affect the results of point of decision motivation. The study noted similar increases in stair usage to other prior work [3, 6, 8, 7].

In a study conducted in a Baltimore/Washington D.C. area shopping mall, Andersen et al. examined if messages targeted at health-promotion or weight-control had a differing impact on the stair/escalator choice [3]. Overall, they did not determine a significant difference between the two signs because stair usage increased from 4.8% at baseline to 6.9% and 7.2% respectively for the health and weight-control signs. However, they did notice a difference in stair usage among ethnic groups. The signs did not cause any significant increase in stair usage for black shoppers.[1] Additionally, in contrast to Brownell et al. [8], they found baseline stair usage for those under and over 40 was equivalent. They did find that stair usage increased more for the older group than the younger group. For those over 40, stair usage increased from 5.1% to 8.1% with the health-benefits sign and 8.7% with the weight-control sign. For those under 40, usage increased from 4.6% to 6.0% with the health-benefits sign and 6.1% with the weight control sign. While this study confirmed the observation that obese persons take the stairs less during the baseline phase noted in [8], it showed no difference between the two groups during intervention phase. The study also found that obese subjects increased their usage of the stairs slightly more with the weight control sign (7.7%) as compared to the health-benefits sign (6.3%) from a baseline of

[1] Only 2146 observations were made.

3.8%. These results indicate that for certain groups, targeted messages may improve stair usage compared to more general messages.

Russell, Dzewaltowski, and Ryan decided to use messages that discourage the use of an elevator instead of encouraging the use of the stairs in their research [25]. During their intervention, they placed a 20cm x 20cm sign reading "Elevator for physically challenged and staff use only, others use stairs please" near the elevator. Over the 5-week intervention phase, stair usage increased from 39.7% at baseline to 41.9% during the intervention. When the results were examined by sex, men had a significant increase in stair usage (42.4% to 45.9%) while women did not show a significant increase (35.1% to 36.7%). Additionally, stair usage actually decreased on Fridays (36.1% to 28.7%) during the intervention phase. This anomaly is potentially explained by a larger number of staff in the library on Fridays relative to students. Unlike the prior study of Anderson et al. [3], subjects over 30 years of age did not show significant increase in stair usage. However, more staff members fit into the above 30 category, and the intervention was not directed at staff.

Russell and Hutchinson continued the previous study by comparing the impact of deterrent to health-promotion prompts in a regional airport in the mid-western United States [26]. The health promotion sign had an image of a fit looking heart and the text "Save time, keep your heath healthy, use the stairs," while the the deterrent sign featured the text "Please limit escalator use to staff and those individuals unable to use the stairs" and a picture of a businessman using the stairs. In this study there was no significant difference between the types of interventions with an increase from 8.2% at baseline to 14.9% and 14.4% with an encouraging sign and deterrent sign respectively. When broken down by sex, they found that both younger and older women exhibited higher stair usage than their male counterparts (although young males used the stairs more than older females) contrary to the results of their prior study.

Kerr et al. examined the effects of stair-riser banners over a six-month period in a United Kingdom shopping mall [20]. The banners contained the phrases "Keep fit," "Daily exercise," "Work your legs," "Free exercise," "Stay Healthy," "Easy Exercise," "Be active," and "Exercise your heart." The message "Take the stairs" appeared on 3 banners. A previous study by the same authors had shown banners placed all the way up the stairs to be be more effective than a simple sign placed at the base of the stairs [19]. In the six-month study, results similar to Brownell et al. [8] were noted. Stair usage increased during the intervention phase from 5-12% (depending on the group) and then gradually declined to a level that was 2-5% above baseline after the intervention was removed. The effects of the intervention were not the same across different age and sex groups. In the initial six weeks of intervention, women's stair usage increased more than men's usage. In the second six weeks of intervention, men under 60 years of age maintained their increased stair usage, while women under 60 years of age showed a decrease in stair usage (although still elevated compared to baseline). For both men and women over 60, stair usage continued to increase over the second six week intervention period. Across all groups, stair usage remained elevated compared to baseline over the entire ten week follow-up period. These results suggest the importance of long-term studies examining how people change their responses to interventions. The researchers also determined that stair usage increased as the overall traffic over the stairs/escalators increased. This effect was taken into account in the results of their study.

Titze et al. utilized both manual counts and automated counts over six office buildings in Switzerland [36]. The manual counts were done by an observer known to subjects, and the automatic measurements were recorded by a break-beam sensor on the stairs and an open/close counter on the elevator doors. Two of the buildings had to be removed from the study: one because of automated counting malfunction, and one because of abnormally high percentage of people taking the stairs during baseline (95%). The study utilized a wide variety of interventions including offering fruit and games of chance to those who took the stairs. The intervention period lasted for four months. Overall, the study found an increase from 61.8% baseline stair usage to 67.1% stair usage when using manual counting. The automatic counting saw a small increase from 68.8% to 71.4%, but this was not statistically significant. Additionally, the pattern of stair usage recorded by automatic sensors was not consistent with manual counts. Possible reasons for the poor results of the automatic counting include observational bias (the subjects were aware of the manual counts), people testing the stair sensor during baseline, and the inaccuracy of both the stair and escalator counting systems. For instance, the escalator sensor could not count how many people were in the elevator, but simply door openings and closings. The break beam sensor had an occlusion problem where multiple people could walk through only breaking the beam once, or one person could walk through breaking the beam several times. Two important issues were raised by this study. First, inexpensive interventions were shown to have a significant impact on stair usage. Additionally, simple methods of automatic counting using mechanical means were shown to be ineffective in practice. A well-designed computer vision algorithm could perform much better than a break beam sensor or open/close door counter. Finally, a large difference was noted in the baseline stair usage among the four buildings. Baseline stair usage was highest in the newest buildings which contained well lit, attractive, and short sets of stairs. The lowest baseline usage occurred in an older building with poorly lit staircases and a large number of stairs between landings. These results indicate that the visual appeal of stairs can impact their usage.

Another study by Boutelle et al. examined the effects of making the stairs more aesthetically attractive in an 8-story building on the University of Minnesota campus [7]. There were two intervention phases in this study. The first consisted of signs encouraging people to "take the stairs for your health," while the second consisted of artwork placed in the stairwells and music that was audible throughout the entire stairwell. Signs alone encouraged an increase in stair usage to 12.7% from a baseline level of 11.1%. The music and artwork in addition to the signs led to even a higher percentage of stair usage at 15.5%. Because the music and art in the stairwell were regularly changed, the increase may have partially been from an interest in the change, not the actual aesthetic appeal of the modified stairs.

3.2.1 Observations on Previous Stair Studies

Many of the studies discussed previously saw the effects of signs vary over different age groups, different sexes, and different ethnic groups. For the purpose of this thesis, only the overall effects of the sign are being examined. In all of these studies an overall increase in stair usage ranging from 2.1% to 9% was seen. An overall increase in stair usage within that range was expected in the Boston area subway stations.

3.3 People Counting Algorithms

Work on the problem of counting people using computer technology ranges from prototype systems tested only in labs to systems that are used in real-world environments. The Spanish Railway Company did a market survey of several methods of people counting, as reported by Albiol, Mora, and Naranja [2]. Mechanical counters such as turnstiles are very accurate. However they create a barrier to traffic flow. In most places, such barriers are impractical. For instance, the Metro Boston Transit Authority (MBTA) did not want the counting method to obstruct people as they exited the station. Break-beam sensors do not impede exiting passengers, but they have an occlusion problem. If the break-beam is broken while multiple people walk through, there is no way to determine how many people passed during that time. Additionally without a complicated multiple beam setup, there is no way to determine the direction of passing. Differential weight systems and sensitive carpets can be accurate, but they require heavy modification of the environment and significant amounts of maintenance. They also do not provide an easy way of determining directionality of passers-by. None of these systems provide the accuracy necessary for experimentation without requiring heavy modification to the environment or impeding traffic.

Computer-vision based people counting offers an alternative to these other methods. One of the common problems all computer vision systems face is having to separate people from a background scene. Several proposed counting systems use multiple cameras to help with this process. Using stereo differencing and an overhead camera view, Terada et al. [34] created a system that can count people and determine direction of movement as they cross a measurement line. The top-down view avoids the problem of occlusion as groups of people pass through the camera's field of view. Their system also uses a space-time image to help determine directionality. However, the system was only tested with 43 people in a controlled environment (lobby of an office building). No error rates were given. Additionally, occlusion is dealt with by requiring a specific camera angle.

Beymer and Konolige also use stereo vision in people tracking, but they relax the camera position restrictions of Terada et al [5]. Their system uses continuous tracking and detection to handle occlusion. Template based tracking is able to drop detection of people as they become occluded, eliminating false positives in tracking. However, when a person is no longer occluded, they are detected as a new instance. This method of handling occlusions would lead to double counting in the stair/escalator application. Additionally, the performance level quickly drops as the number of people and occlusions in a scene increase. With a small test set of 5 people and 28 occlusions a tracking rate of 70% was achieved.

Hashimoto et al. tackle the problem of people counting using a specialized imaging system of their own design [16]. Using IR sensitive ceramics, mechanical chopping parts, and IR-transparent lenses, they developed an array based system that could accurately count passers-by at a rate of 95%. The system uses background subtraction to create thermal images that are then analyzed. This system was not considered appropriate for the stair/escalator counting because it requires a direct overhead view of the point of passing that cannot be obtained in locations with typical ceiling heights. Additionally, the system requires a distance of at least 10cm between passing individuals to properly count them as two separate people, and large movements from arms and legs were seen to create problems in counting. The high density traffic that occurs when a mass of people exit a rush hour train would likely lead to significant counting error in this setup.

Tesei et al. use image segmentation and "long-memory" to track people and handle occlusions [35]. Background subtraction is used to highlight areas of interest (blobs) by subtracting a reference frame from the current frame and thresholding the result.

Using features such as blob area, perimeter, bounding box area, height and width, mean gray level, and center position, the blobs are tracked from frame to frame. By keeping track of this information over time, the algorithm handles the merging and separation of blobs that occurs from occlusion. By storing information about what blobs combined to form a new blob during occlusion, the blobs can be assigned to their original labels when they separate. While some of the ideas from Tesei et al. are used in the people counting system described in this thesis, the algorithm would not be able to handle the occlusion that occurs in the transit station. Many times an people enter the field of view occluded and stay partially obscured until they leave the field of view. The system would have no way of accurately counting when a large number of people exit the station at once.

To improve on background segmentation algorithms, Shio and Sklanksy use extra cues to simulate the perceptual grouping that occurs in the human vision system [31]. They first calculate motion estimations from consecutive frames and use that information to help segment people from the background and determine the boundary between people in cases where there is occlusion. They observe that while parts of people move in different directions, over a few seconds time, all the parts of a person move as a group. The actual segmentation uses a probabilistic object model that incorporates width, height, and direction of difference motion and a merging/splitting step to segment individual people in a moving picture sequence. The paper demonstrates that using extra information such as an object model can improve segmentation and provide a possible way to deal with occlusion. However, this system would likely not perform well for stairs/escalator counting because using motion as a "perceptual-grouping" cue would not help with occlusion when a large group of people are all moving in the same direction at the same speed.

Schofield et al. utilize yet another method for separating people from a background image to determine a count [27]. They train RAM-based neural networks to perform background segmentation and then pass the processed images off to be analyzed. To deal with occlusions and overlapping blobs, the system uses a dynamically adjusted spacing threshold. While the neural network based background segmentation enabled the algorithm to deal with varying lighting conditions, the algorithm only dealt with counting people within a specific image. Tracking or counting people over time was not considered.

As an alternative approach, Sexton et al. use a simplified segmentation algorithm and manage to get error rates ranging from 1% to 20% in their trials in a Parisian railway station [30]. They use background subtraction with a constantly updated reference frame to isolate people from the background. The resulting blobs are then tracked frame to frame by simply matching blobs with the closest centroids. Much like the previous systems, they use an overhead camera angle to reduce occlusions. Counting performance ranged from 1.5% error to 21% error. Also, larger crowds caused a frame-rate drop while processing leading to a higher error. With much faster processors in use today than those available at the time of this study, frame-rate problems could likely be minimized.

Like Sexton et al., Segen and Pingali concentrate on image processing after segmentation [28]. They use standard background segmentation techniques to isolate areas of interest, and then identify and track features in those areas between frames. The paths of these features are then merged into clusters that represent the motion of a person over time. The paths could be used to determine how many people crossed a particular line in the scene, and in what direction the people were passing. However, their implementation is only tested to run in real-time with up to 8 people in the scene and makes no attempt to deal with occlusions.

Haritaoglu and Flickner approach the problem of real time tracking of people to determine shopping groups in stores by using temporal information to improve segmentation and tracking [15]. To segment silhouettes from the background, they adopted a background subtraction model that utilizes color and intensity of pixel values over time to classify pixels as foreground, background, or shadow. The pixels are also filtered over time to remove moving pixels from the background model. Foreground groups are then segmented into individual people using temporal and global motion constraints, and individuals are tracked using an appearance model based on color and edge densities. Experimentation using this system was focused on determining shopping groups and how many individuals make up the shopping groups. Much like other systems that use motion to help with identifying people, a situation in a transit station where a large number of people are moving in the same direction at the same speed will likely give the algorithm trouble.

Conrad and Johnsonbaugh attempt to simplify segmentation and the entire people counting process using an overhead camera. Instead of using background subtraction, they use consecutive frame differencing to avoid the problems of changing illumination. Their algorithm also only examines a small window of the full scene that is perpendicular to the flow of traffic. The window is broken into gates. Using assumptions about the minimum/maximum width of people and the amount of noise in their images, they are able to determine the number of people in the window at a given time. Using the center of mass of the images in the window over time, they are able to determine direction of travel. They achieved a 95.6% accuracy rate over 7491 people with a quick and simple algorithm. However, they rely on an overhead camera view to reduce occlusion problems, and the accuracy of the algorithm as described would decrease with constant streams of people moving through the window.

Finally, Albiol et al. developed a system to count people in high density traffic exiting and entering a Spanish public transportation train [2]. Their system uses a camera in a fixed position above the door mechanism on the train itself. While people enter and exit the train, the images that the camera captures are reduced to image stacks (also known as space-time images). Using scan lines, three image stacks are created. The image stacks are images themselves, where each row corresponds to the image data from one of the scan lines on the original image. The horizontal axis of the stack image is the same as the horizontal dimension in the original images and the vertical axis corresponds to increasing time. Instead of background subtraction, a gradient function is used to perform segmentation from the background. Once the train doors close, the image stacks are then processed by a computer as the train moves towards another station.

As people move into and out of the train, they leave distinctive "prints" behind in the image stacks. The researchers were able to train a computer algorithm to analyze these stacks and determine how many people crossed the threshold of the door. To complete the counting, an optical flow algorithm was used to determine direction of passing. Over 149 test stops this system counted 318 incoming passengers and 379 outgoing passengers when the real numbers were 321 and 385 respectively for an overall accuracy of 98.7%.

3.3.1 Goals of Stairs/Escalator Counter

In the context of creating an automatic people counting system to use in stair/escalator studies, several restrictions were placed on the experimental setup. An algorithm that could handle the large crowds moving through a subway station was the most important goal. Since the algorithm was designed to count people exiting a station, being able to work properly with natural lighting changes was also very important. Additionally, because the system was intended to be placed in several locations it was not possible to require a specific top-down view for the camera to minimize occlusion (which requires unusually high ceilings). Finally, while accuracy in counting raw numbers of people on the stairs or escalator was desirable, the ratio of people taking the stairs compared to people taking the escalator was the number the system was most concerned with determining. Stereo camera heads were avoided because the additional depth information they might provide would be unlikely to significantly improve segmentation of large clusters of people.

The idea of training an algorithm to identify prints in a image stack used by Albiol et al. [2] was initially tried in the Boston area subway station. However, the lack of an overhead camera view and occlusion problems created difficulties with this technique. Still, the image stacks were maintained in the final algorithm. The algorithm described in Albiol et al. processes the image stacks as the train travels between stations. The algorithm described in this thesis has to process image stacks while still continually capturing video data. The area of interest also ranges over the entire width of a stair/escalator combination. Additionally, for this work, it was not possible to install equipment on all trains. A device at a fixed location was necessary.

Chapter 4

Design and Implementation

With the preceding goals in mind, a self-contained people counting/message projecting system was created.

4.1 System Overview

The system is designed to be low profile and easily adaptable to many situations. A housing box was built to protect the equipment inside from the various hazards in a transit station (see Appendix A). The box containing all the system parts is mounted above the base of the stairs and escalator. Passengers leaving a train and walking towards the main exit of the station see a projection of the intervention message on a 1m x .66m piece of foam core directly above the stairs/escalator (see Fig 4-1). The enclosure was designed to be as discrete as possible and not obstruct the view of the projected sign to passengers exiting a train. As the persons exiting the station choose which method of exit to use, their decision is recorded by a computer vision algorithm running on a laptop connected to a webcam. The counts are aggregated over 10 minute time periods and written to a file on the laptop's hard drive that can later be downloaded via a wireless connection and analyzed off-site. A more complete description of each component follows.

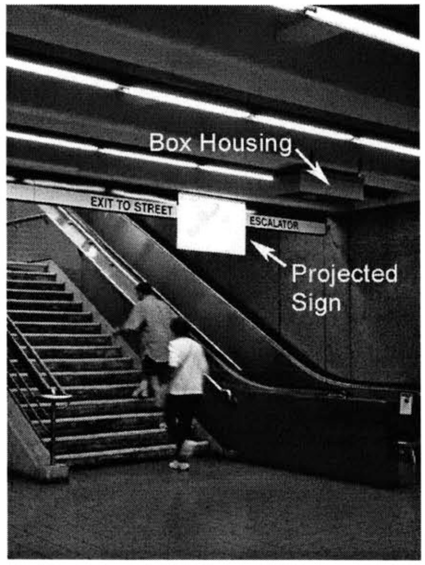


Figure 4-1: The system in Kendall outbound station. The box has a projector to display the sign and a camera directed at the stairs/escalator and connected to a laptop to count passers-by.

4.2 System Components

The first component of the system is a modified Orange Micro Ibot2 USB2.0 webcam. The camera has been removed from its plastic housing and mounted in a turret that allows it to be rotated and moved up or down in viewing angle to accommodate different installation sites (see Appendix B). Additionally, the camera has been fitted with a wide angle lens to allow a complete view of the stairs and escalator (see Fig. 4-2). The camera captures images in YUV format with a resolution of 160x120 at 15fps for processing by the laptop. All autogain and autoexposure features are turned off.

The next component is a ViewSonic PJ500 projector that is used to display motivational messages on a foam board sign or wall. The foam board is suspended from the ceiling just above the center of the bottom of the stairs and escalator. The projector displays images at 800x600 resolution with an intensity of 1200 lumens. This is bright enough for good visibility even in a well-lit transit station. The projector also has a serial input that allows it to be controlled by a computer.

The final part of the system is a laptop computer that controls both the camera

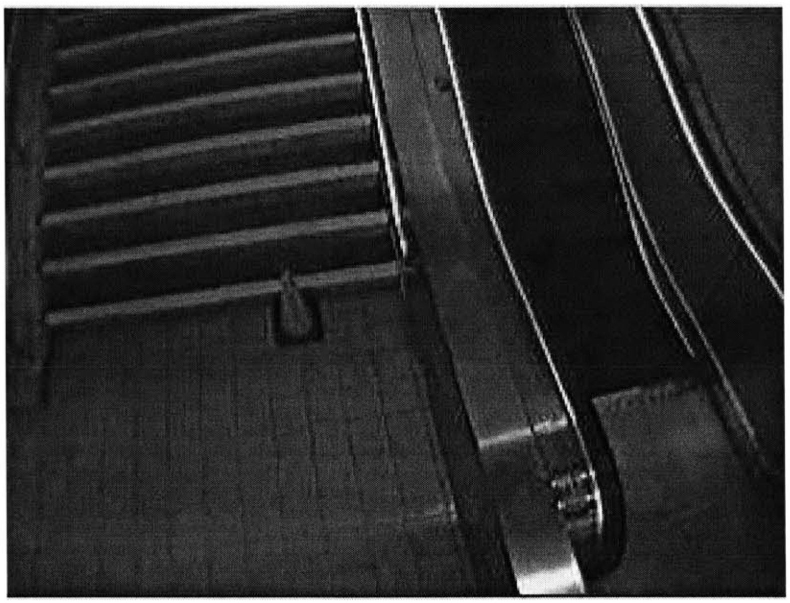


Figure 4-2: The view of the stairs and escalator from the Kendall inbound camera (320 x 240 resolution).

and projector. The computer receives input from the webcam and determines how many people are using the stairs and escalator. The computer stores each day's counts in text files that record how many people pass in each 10 minute period. For each period, the computer also records an image as seen from the camera in jpeg format for future analysis if any inconsistencies are noticed in the count data (see Appendix F for examples). The computer also provides a video source for the motivational messages that are displayed. By connecting the computer to the projector's control port, the projector can be periodically monitored and restarted if it turns off for any reason. The computer is equipped with an 802.11 wireless card which is set up to use an ad-hoc network. This allows a researcher to use another laptop with a wireless card to download data, upload new messages, and debug problems. WinVNC [24] was used to provide remote control of the laptop in the transit station.

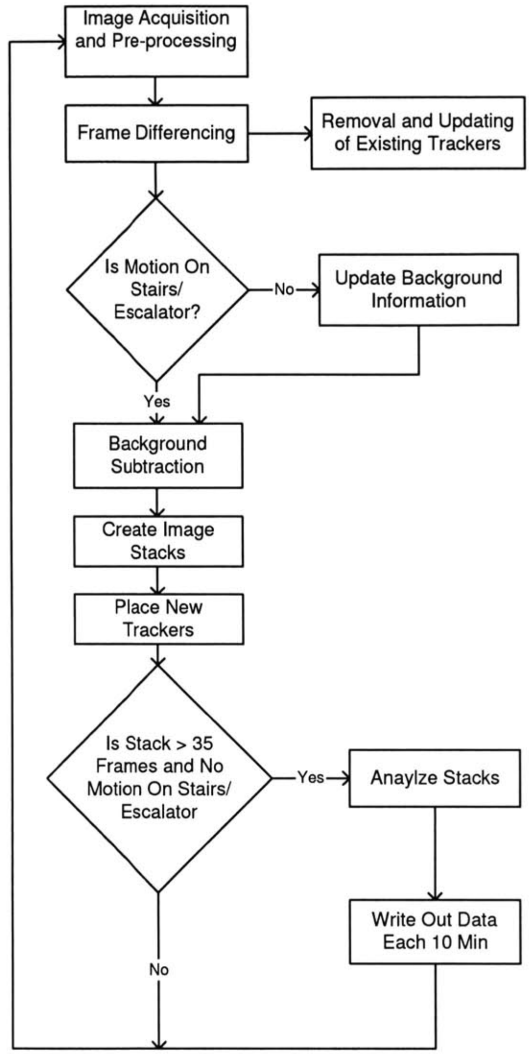


Figure 4-3: Flowchart describing people counting algorithm.

Figure 4-4: Images from Kendall inbound station with a crowd of people on the stairs and escalator.

4.3 The People Counting Algorithm

Prior work illustrates the difficulty of tracking individual people in the transit station due to the large number of occlusions that occur as high-density groups of people exit the station (see Fig. 4-4). The occlusion problems become even worse when a directly overhead camera view is not available. Therefore, in this work an algorithm is developed that operates on the assumption that each person moving through the frame generates a measurable amount of activity in a difference image. By dynamically determining the average activity per person, the total amount of activity can be converted to an estimate of the number of people that have passed through the image. Over the course of a full day with thousands of observations, people who generate more activity and people who generate less activity average out to a number that approaches the actual count.

Fig. 4-3 shows a flowchart view of the algorithm. The individual steps are described in the following sections.

4.3.1 Image Acquisition and Pre-Processing

Images are acquired in RGB24 format at 160x120 pixels. The camera grabs a new image buffer 15 times per second, and the algorithm requests a new frame whenever it can start processing another frame. The pre-processing consists of convolving a 2x2 averaging matrix with the image to blur the image slightly and help eliminate noise from the background segmentation process.

4.3.2 Frame Differencing

The first step in the people counting algorithm is frame by frame differencing. Consecutive frames are compared pixel by pixel in the Y, U, and V channels. Differences greater than a threshold (determined at each location by trial and error) are highlighted as areas of interest in the frame (see Appendix E for details on how to calibrate the system). Frame differencing is useful because areas that have motion from frame to frame are identified, but global lighting changes only show up very briefly. People

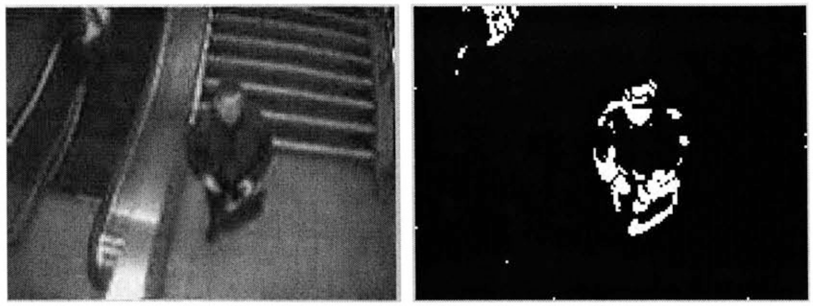


Figure 4-5: Real image and corresponding difference image.

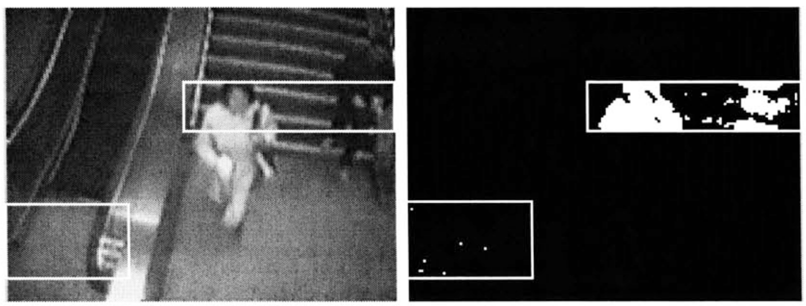


Figure 4-6: Real image and corresponding background subtraction image. Note that background subtraction is only performed in areas of interest that are highlighted.

moving through the frame typically show up in the difference image as outlines (see Fig. 4-5).

4.3.3 Background Subtraction

While frame differencing is effective in determining motion, it is not effective in highlighting foreground objects that are stationary from frame to frame. Additionally, frame differencing has a tendency to only highlight the edges of objects in the foreground, which can make image analysis more difficult. Consequently, the algorithm uses a background subtraction [37] step as a supplement to frame differencing.

By examining the stair and escalator area of the difference images over time, the algorithm can determine when there is no activity in one or both of these regions. Allowance is made for the periodic motion of the escalator by ignoring small amounts of pixels that are highlighted by the differencing. When there has been no activity detected from frame differencing, the current captured image is likely to contain only the empty transit station, and it is added to a statistical model of the background. The model keeps track of 75 image frames. When a new frame is added to the model, the oldest is removed. A mean and variance value are calculated for each channel (Y,U, and V) of each pixel in the image.

For the subtraction step, pixels in the current frame are compared against the background model. For each of the three channels in a pixel, the mean value for the same channel in the corresponding pixel in the background model is subtracted from the current value. The result of this subtraction is then divided by the variance for the same channel in the corresponding pixel in the model. This process is equivalent to calculating a 1-dimensional mahalanobis distance. For the U or V channels, if this distance is greater than a certain threshold, the pixel is highlighted as foreground. The threshold is kept the same for both the U and V values, and it is set during a calibration step at each site (see Appendix E). The threshold value for the Y channel can be different from the U and V threshold. If the current Y channel value is greater than its threshold, it is treated slightly differently to avoid selecting shadows as foreground. If only the Y channel value would cause the pixel to be selected as foreground (meaning the mahalanobis distance of the current U and V values are within their threshold), the Y intensity is compared to a threshold that describes the minimum intensity needed for the pixel to be considered shadow. If the value is below that threshold, the pixel is highlighted as foreground. If the pixel intensity is greater than that threshold, the intensity difference is contributed to a shadow, and the pixel is classified as background. Background subtraction creates images with blobs that correspond to foreground objects (see Fig. 4-6).

4.3.4 Image Stacks

To actually calculate the number of people using the stairs or escalator, the algorithm uses a strategy that assumes each person passing through the frame creates a measurable amount of highlighted pixels in the differencing image. By determining the overall number of difference pixels seen and dividing by the average number of pixels attributed to a single person, the algorithm can estimate how many people

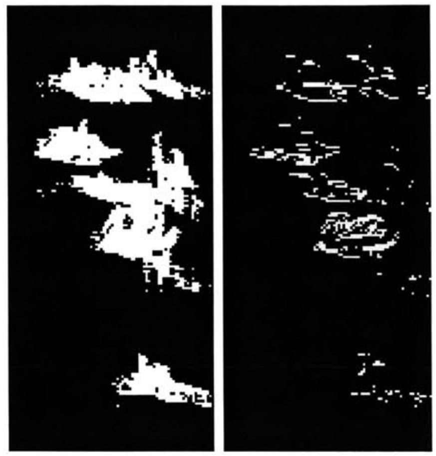


Figure 4-7: Two image stacks. The left is created from the background subtraction images, and the right is created from the difference images.

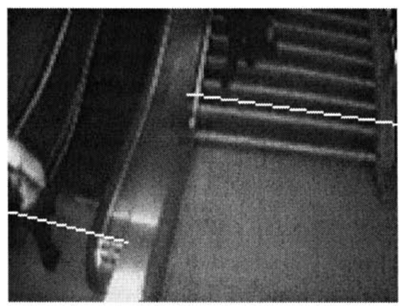


Figure 4-8: Scan lines places along the stairs and escalator during calibration of an installation site.

have passed through its field of view. The difference images work well for this strategy because they only highlight changes from the last image (primarily edges). This outlining leads to a much smaller variance in the number of pixels a large person and a small person highlight as compared to the variance of pixels created in the background subtraction images. Additionally, the difference image is not as susceptible to sudden changes in lighting (eg. due to a cloud suddenly obscuring the sun) because only one or two frames are affected.

Monitoring activity in specific areas of the frame requires a temporal history of the scene. To facilitate in this process, a single scan line is placed across both the stairs and escalator during calibration at each site (see Fig. 4-8). For each frame, the pixels on the scan lines are copied onto two stack images: one for the background subtraction, and one for the difference image (see Fig. 4-7). In the stack images, the horizontal dimension corresponds to the horizontal axis in the original frame, while the vertical axis corresponds to time. This is the same type of image stack (or space-time image) used in Albiol et al. [2]. Examining the stack image constructed from the background subtraction data, it is easy to see that people leave distinctive blobs as they pass a scan line.

4.3.5 Analysis of Image Stacks

Since the algorithm is running in real time, there is a determination made on when to stop capturing images so the current stacks can be analyzed. When the frame differencing indicates that there is no activity on the stairs or escalator and the stack images have reached a minimum size (35 pixels in height), analysis is performed on that set of stacks. The analysis usually takes less than the amount of time required to capture another image, leading to little or no loss of information.

The first step in analyzing the stacks is to determine when a single person has crossed the scan line. Using minimum and maximum dimensions for blob width and height, single person blobs are identified. A bounding box is created around the blob. The corresponding area in the difference image stack is then examined. The number of pixels in the difference stack in the area defined by the bounding box is then counted and added to a running average. The running average keeps track of the number of pixels in last 25 blobs selected. For each blob selected, the people counter is incremented by 1 and the pixels in that blob are subtracted from the difference stack. The remaining number of pixels in the difference stack is divided by the current running average and that number is added to the people count.

4.3.6 Tracking

The counts from the image stacks do not have any directionality associated with them. This lack of information is fine for the escalator because over the course of several days of manual observation only a handful of people used the escalator in the

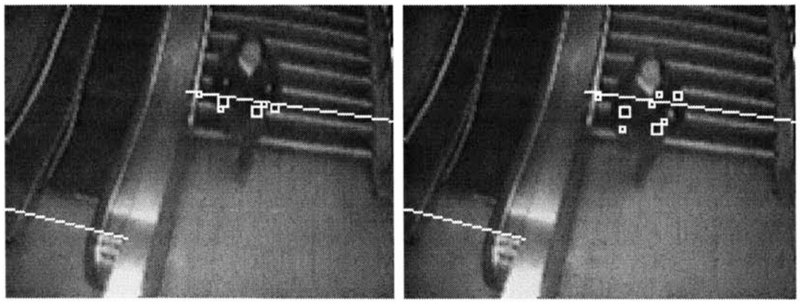


Figure 4-9: Trackers tracking a person's movement down the stairs. Because of arm movement, some of the trackers move slightly up. A threshold is set on how far a tracker needs to move in either direction to avoid having these trackers count pixels moving in the wrong direction.

wrong direction. However, people travel the stairs in both directions. To determine the direction of travel of people on the stairs, trackers are used. The trackers use pixel correlation in the tracking process. During each frame, consecutive pixels in the difference image that are on the stair scan-line are grouped together. Clusters that are larger than two pixels have a tracker assigned to them. The trackers are initially placed at the center of each cluster with a width equal to that of the cluster. Using the pre-processed image, the trackers examine the intensity level of each of their pixels in the Y, U, and V color-space channels. In the next frame, the trackers examine the pixels in an area that is 2 pixels wider than the cluster on each side, and 4 pixels up or down. For the every examined location, the difference in values for corresponding color channels in each of the pixels is found, and the total difference for that location is determined. The differences in the intensity channel are weighted 4x as much as the differences in the U or V channels. The tracker then moves itself to the location that results in the smallest total difference.

Each tracker has a lifespan of 10 frames (.6s). After 10 frames the final position of the tracker is compared to the starting position, and a determination is made if the pixels moved up or down in the frame. If the pixels move more than a threshold distance up or down, they are counted as moving in that direction. The thresholds are set by using the manually annotated video and trial and error to determine what combination of thresholds creates the most accurate estimate of people moving up or down the stairs (see Appendix E). The percentage of pixels moving up the stairs is kept track of for each 10 minute aggregation of data. The total number of people on the stairs in that time is multiplied by the percentage of pixels moving up to determine what number of people went up the stairs in that time period.

Correlation tracking is used because it only requires subtracting 3 values per pixel. This operation can be performed very quickly and does not significantly affect the frame-rate of the algorithm. Complex methods such as optical flow are much more computationally intensive, and they did not lead to significantly better estimations of the percentage of people walking up the stairs. Histogram backprojection [10] was also more computationally intensive, and it did not produce better results during testing.

4.3.7 Data Storage

The results from the people counting algorithm are written to text files for each day of counting. The files contain a time stamp for when each period of analysis ended and the number of people counted in that time period (see Appendix K). When data is written to the file, a snapshot of the current view from the camera is also recorded to disk. This helps to determine if anything unusual happened during the day. For example, in four of the days of counting in one of the transit stations, the escalator was either broken or closed for maintenance.

4.3.8 General Performance Characteristics

The algorithm was designed to handle many of the situations that occur in the transit station. Because frame differencing is used for the actual count estimation, sudden changes in lighting should not have a large effect on the system. Additionally, the background model is constantly being updated, and a running average is kept for the number of pixels a single person leaves on the difference stack. These features reduce the effect of slower changes in lighting such as those caused by slow moving clouds and the sun's position over time. The system was tested in conditions that varied

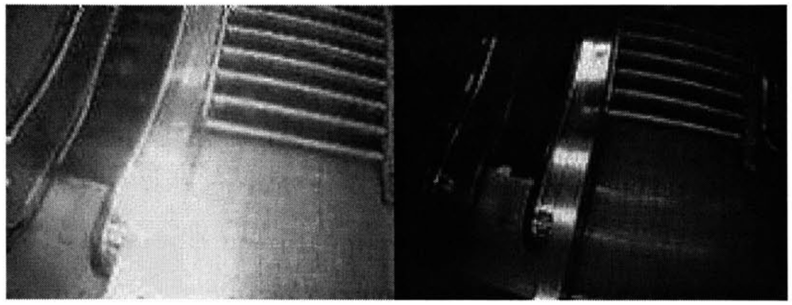


Figure 4-10: The algorithm must perform in conditions that can vary as much as these images. The image on the left is from when the sun is shining directly into the station. The image on the right is from a dark evening.

greatly from dark to extremely sunny (see Fig. 4-10).

While changes from lighting and weather can be compensated for, certain characteristics of people passing through the station can confuse the algorithm. People carrying large bags or other objects up the stair will likely be counted as more than one person. However, heavy winter clothing should not affect the counts very much. The dynamically calculated pixels per person value should increase to reflect any extra pixels in the difference image caused by thicker clothing. In addition to what people are carrying or wearing, their behavior can affect the algorithm. If a person sits or stands on the stairs without much movement, they will not be picked up in the difference image and not be counted until they continue moving up the stairs. However, if a person paces back and forth along a step, or moves up and down the steps, they will show up in the difference image and cause over-counting for the stairs. Mechanical problems with the escalator also present a challenge for the algorithm. When counting people, the algorithm has no way to distinguish between people working on a broken escalator or people actually using the escalator. Additionally, a broken escalator will result in an abnormally high stair usage percentage which would effect the results of the study. When mechanical problems with the escalator do occur, the effect is usually clearly visible in the daily graphs of stair and escalator usage. Images captured from throughout that day can then be manually inspected to confirm if the escalator was broken. In the future, a simple feature could be added to the algorithm to detect the periodic motion of the escalator. If the detector senses a lack of motion on the escalator, it can add a note into the data file indicating that the escalator may have been stopped. Appendix F presents an example of two days during the experiment in which the escalator was closed for maintenance and the effect it had on traffic flow through the station.

4.4 Presentation of Motivational Messages

After testing the counting technology and collecting baseline data on stair and escalator use at each site, a motivational message is displayed by the projector. The projector is connected to the video output of the laptop, so that any image that is displayed full screen on the laptop will be projected to the sign. The program that displays the images could receive input from the counting algorithm to produce images that are animated based on people's stair usage. The display program also periodically checks the projectors status using a serial link between the projector and the laptop. The projector is automatically turned off between 1am and 5am to extend the lifetime of the bulb. With the projector on 20 hours a day, each bulb should last for 200 days.

Chapter 5

Evaluation

5.1 Installation Sites

5.1.1 Kendall Inbound

Kendall inbound was selected as the first installation site. The stairs at this site consist of a set of 16 steps, a landing, and another set of 14 steps. The station has many advantages for this project. A large percentage of the people going through the station are either professionals who work in the area or students attending MIT. Since the population moving through the station consists of mostly the same people every day, the effect of motivational messages over time can be measured. While there are two exits from the station, there is one exit that the majority of the traffic in the station flows through. Additionally, the geometry of the station created a good location to install the box containing the system. There are several beams running perpendicular to the stair/escalator entrance that provide an ideal mounting spot for the box. The camera looks downward from the bottom of the box and the projector projects out the front of the box onto a foam board suspended from an existing sign in the station. Finally, the Kendall station is located within one block of the House_n office. This location was convenient for maintenance and dealing with problems that appeared during development and testing.

Figure 5-1: View of Kendall inbound system and projection. The second image shows the distance between the turnstiles and the stairs/escalator.

5.1.2 Kendall Outbound

The Kendall outbound station was chosen as the location for the second system. The physical layout of the station is similar to the inbound side of Kendall. The stairs also consist of a set of 16 steps followed by a landing and another set of 14 stairs. No modification was necessary for the system parts or the housing design. All of the reasons that apply to the inbound station selection apply to the outbound side. Because the beam location, the camera view for this system is closer to a completely overhead view than the inbound camera.

Placing another system on the outbound side of the Kendall station did raise some concerns. There is a possibility of interaction between motivational messages on both sides of Kendall. Any changes in physical activity patterns may not as easily be attributed to one motivational message. However, the messages are only addressed at people leaving a station. Regular commuters only exit one station and enter the other. There are no escalators going down into either station, so all traffic entering must use the stairs or a separately located elevator. These considerations reduce the the possibility of an interaction between signs when examining the effect on people's

Figure 5-2: View of Kendall outbound system and projection.

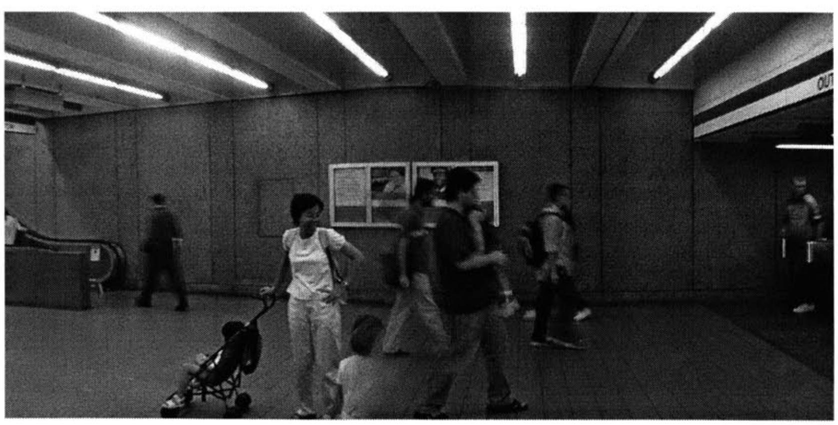


Figure 5-3: Side view of the Kendall outbound station.

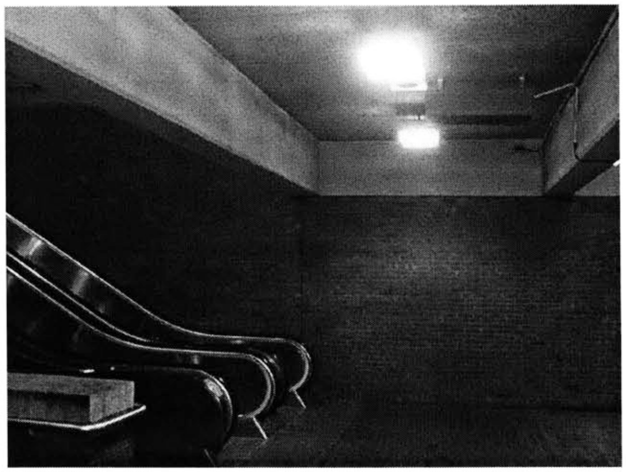


Figure 5-4: View of Davis system. The sign will be projected against the brick wall to the right of the escalators.

behavior. The positive aspects of the Kendall station outweighed concerns about sign interaction.

5.1.3 Davis

The College St. exit of the Davis station was chosen as the final location for a system. This location provided a stair and escalator right next to each other and a large wall to project a motivational message on (see Fig. 5-4). The stairs consist of two sets of eleven steps separated by a landing. As people turn a corner to approach the stairs, the motivational message is clear to see before a choice on whether to take the stairs or escalator is made. This site also has a lower level of traffic than either of the Kendall stations. The algorithm performance has been slightly better in less crowded situations, so this was a benefit of choosing Davis. Additionally, Davis is located 3 stops away from Kendall allowing for easy access.

To mount the system in the Davis station required minor modifications to the box housing the equipment. Because the projection is not directed above the stairs/escalator, but rather to a side wall, the camera needed to be facing a different direction than the projector. Because the camera is mounted on a turret, it was easy to rotate the camera and leave a slightly larger window in the bottom of the box for a complete view of the stairs/escalator.

5.2 Obtaining Test Data

In evaluating the effect of motivational messages, the accuracy of the people counting is very important. As the accuracy of the people counting algorithm increases, smaller differences in stair use will be able to be seen.

To test the algorithm, video was captured in uncompressed RGB format from each of the locations. At each location, four days of video was collected. The days consisted of 17 hours of video from 6:30 am to 11:30pm. A program was then created that allowed a researcher to playback the video at different rates and manually count the people on the stairs and escalator (see Appendix H). This data set of video and the corresponding annotations was used as a standard against which algorithm performance was measured.

To account for annotation error, 1.1 hours of random sections of video were annotated by another observer and their counts were compared. The inter-observer annotation agreement rate for people taking the stairs versus the escalator was 98% and 99% respectfully.

5.3 People Counting Algorithm Performance

For the Kendall inbound station, results of the vision algorithm testing are summarized in Table 5.1. The escalator counting error for the computer-vision algorithm ranged from -14.4% to -1.2% compared to the number of manually counted people. For the two weekdays, the range was -1.8% to -1.2% and for the weekends the range was -14.4% to -12.0%. The error in algorithm's stair counts ranged from -16.5% to +5.0% over the four days. For the weekdays, the accuracy ranged from -16.5% to 5.1%, and for the weekend, the rate ranged from -9.9% to -3.0%. While there was an overall downward bias in the computer counts, it appeared to affect the escalator and stairs about equally for a given day. Lighting conditions and other factors that

Kendall Inbound Algorithm Performance
Manual vs. Algorithm Counts

Method of Counting	Escalator	Wednesday Stairs Total	Stairs Up	Escalator	Thursday Stairs Total	Stairs Up
Manual	1789	6538	1228	1744	6555	1069
Computer Algorithm	1756	5740	1026	1724	6739	1123
Error	-1.84%	-12.21%	-16.45%	-1.15%	2.81%	5.05%
Method of Counting		Stair Usage			Stair Usage	
Manual		40.70%			38.00%	
Computer Algorithm		36.88%			39.45%	
Difference		-3.82%			1.44%	

Method of Counting	Escalator	Saturday Stairs Total	Stairs Up	Escalator	Sunday Stairs Total	Stairs Up
Manual	1135	3474	322	743	2304	201
Computer Algorithm	999	2987	290	636	2055	195
Error	-11.98%	-14.02%	-9.94%	-14.40%	-10.81%	-2.99%
Method of Counting		Stair Usage			Stair Usage	
Manual		22.10%			21.29%	
Computer Algorithm		22.50%			23.47%	
Difference		0.40%			2.17%	

Table 5.1: Table showing the manual vs. automatic counts for 4 days at Kendall inbound. For each day, the stair use percentage for manual counts and automatic counts is also shown.

contribute to the error in counts create similar conditions for both the stairs and the escalator. The goal of this work was to estimate the correct percentage of people using the stairs, not create the most accurate estimate of the specific number of people on either. In evaluating the algorithm's performance on calculating stair usage, the difference between calculated stair use percentage and actual stair use percentage ranged from -3.8% to +2.2%. The mean error for the difference in percentage of people using the stairs was +.05% with a standard error of 1.3%.

Traffic patterns vary through the inbound station on weekdays and weekends. During weekdays, there is a large flow of people from just after 6:00 am until approximately 10:30 am. After the morning rush hour, the amount of traffic slowly tapers off. This pattern is likely do to the nature of the Kendall Square area. The area consists of businesses and the MIT campus. The increase in traffic in the morning is from people arriving for work. There is not a corresponding increase in the evening, because the residential population of the area is primarily students who stay on campus all day. For weekends, there is no increased traffic during the morning hours. The

Kendall Outbound Algorithm Performance
Manual vs. Algorithm Counts

Method of Counting	Escalator	Wednesday Stairs Total	Stairs Up	Escalator	Friday Stairs Total	Stairs Up
Manual	3099	5187	1949	2813	4657	1730
Computer Algorithm	2706	5124	1825	2395	4504	1597
Error	-12.68%	-1.21%	-6.36%	-14.86%	-3.29%	-7.69%
Method of Counting		Stair Usage			Stair Usage	
Manual		38.61%			38.08%	
Computer Algorithm		40.28%			40.01%	
Difference		1.67%			1.92%	

Method of Counting	Escalator	Saturday Stairs Total	Stairs Up	Escalator	Sunday Stairs Total	Stairs Up
Manual	1285	2104	614	1332	1522	335
Computer Algorithm	1108	2037	539	1089	1373	284
Error	-13.77%	-3.18%	-12.21%	-18.24%	-9.79%	-15.22%
Method of Counting		Stair Usage			Stair Usage	
Manual		32.33%			20.10%	
Computer Algorithm		32.73%			20.68%	
Difference		0.39%			0.59%	

Table 5.2: Table showing the manual vs. automatic counts for 4 days at Kendall outbound. For each day, the stair use percentage for manual counts and automatic counts is also shown.

traffic flow stays relatively constant until it starts tapering off in the late afternoon. Fig. 5-5 shows the cumulative traffic through Kendall inbound for both a weekday and weekend. The graphs also demonstrate how well the automatic counting algorithm estimates the cumulative number of people exiting the station over the course of the day.

For the Kendall outbound station, results of the computer vision algorithm are summarized in table 5.2. The automatic escalator counting error ranged from -18.2% to -12.7% compared to the manual counts. For the two weekdays, the range was -12.7% to -14.9%, and for the weekends the range was -18.2% to -13.8%. The stair counting error of the algorithm ranged from -15.2% to -6.4% over the four days. For the weekdays, the error was ranged from -7.7% to -6.4%, and on the weekends the range was from -15.2% to -12.2%. Again, there was an overall bias downwards in the count that appeared to effect the escalator and stairs nearly equally. As a result, the difference between calculated stair usage percentage and actual stair usage percentage ranged from +.4% to +1.9%. The mean error for percentage of people using the stairs

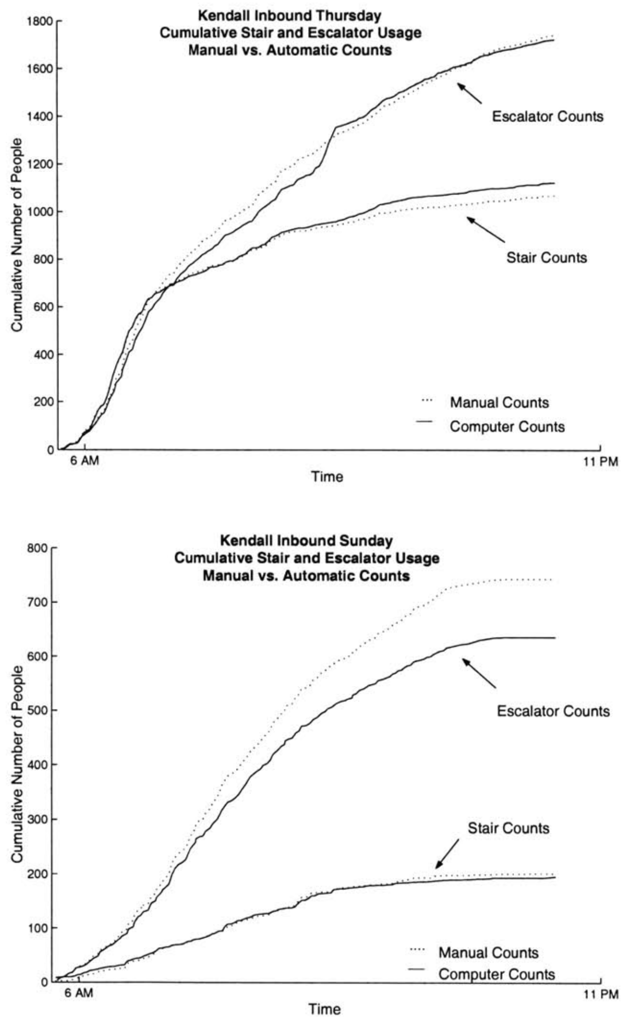


Figure 5-5: Graphs showing the typical traffic flow through Kendall inbound for both the escalator and the stairs. The graphs also show how the algorithm estimated the number of people throughout the day. The graph on the left shows a weekday, and the graph on the right shows a weekend.

Davis Algorithm Performance
Manual vs. Algorithm Counts

Method of Counting	Tuesday Escalator	Tuesday Stairs Total	Tuesday Stairs Up	Friday Escalator	Friday Stairs Total	Friday Stairs Up
Manual	3217	2295	1170	3261	2190	948
Computer Algorithm	3167	2455	1293	3384	2391	1082
Error	-1.55%	6.97%	10.51%	3.77%	9.18%	14.14%
Method of Counting		Stair Usage			Stair Usage	
Manual		26.67%			22.52%	
Computer Algorithm		28.99%			24.23%	
Difference		2.32%			1.70%	

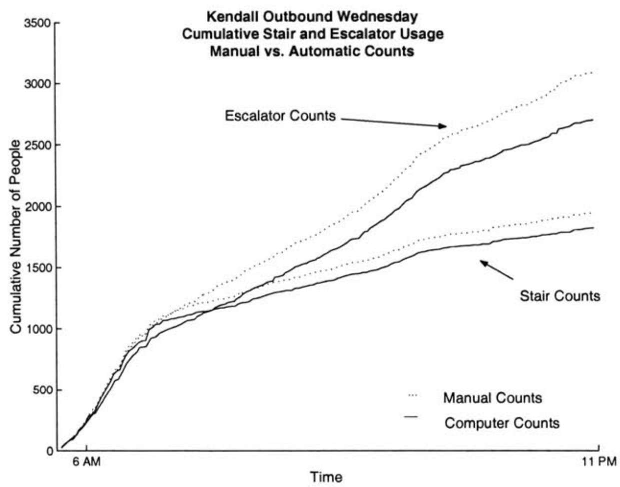
Method of Counting	Saturday Escalator	Saturday Stairs Total	Saturday Stairs Up	Sunday Escalator	Sunday Stairs Total	Sunday Stairs Up
Manual	2098	914	370	1670	626	317
Computer Algorithm	2060	1042	446	1745	739	381
Error	-1.81%	14.00%	20.54%	4.49%	18.05%	20.19%
Method of Counting		Stair Usage			Stair Usage	
Manual		14.99%			15.95%	
Computer Algorithm		17.80%			17.92%	
Difference		2.81%			1.97%	

Table 5.3: Table showing the manual vs. automatic counts for 4 days at Davis. For each day, the stair use percentage for manual counts and automatic counts is also shown.

was +1.1% with a standard error of .4%.

Traffic patterns also vary through the outbound station on weekdays and weekends. During weekdays, there is a large flow of people leaving the station from just after 6:00 am until 10:30 am. After the morning rush hour, the amount of traffic slows quickly and remains fairly constant for the rest of the day similar to the inbound station. For weekends, traffic starts out leaving the station at a fairly low rate. Around 11am, the rate of people exiting the station increases and stays approximately constant for the rest of the day. Fig. 5-6 shows the cumulative traffic through Kendall outbound for both a weekday and weekend. The graphs also demonstrate how the automatic counting algorithm estimates the cumulative number of people exiting the station over the course of the day.

For the Davis station, results of the computer vision algorithm are summarized in table 5.3. The automated escalator counting error ranged from -1.8% to +4.5% compared to the manual counts. For the two weekdays, the range was -1.6% to +3.8%, and for the weekends the range was -1.8% to +4.5%. The stair counting error of the



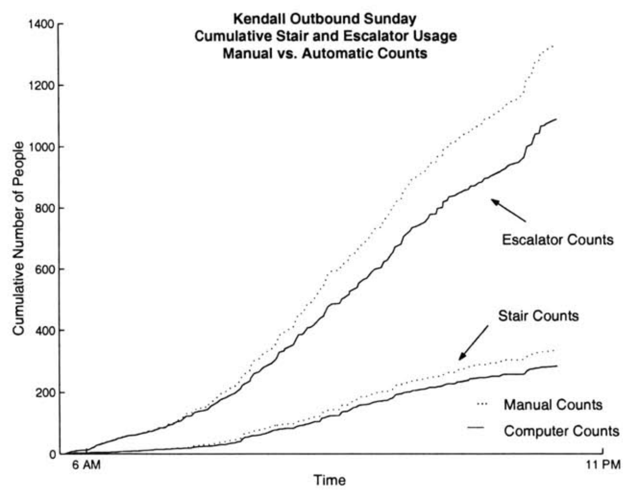


Figure 5-6: Graphs showing the typical traffic flow through Kendall outbound for both the escalator and the stairs. The graphs also show how the algorithm estimated the number of people throughout the day. The graph on the left shows a weekday, and the graph on the right shows a weekend.

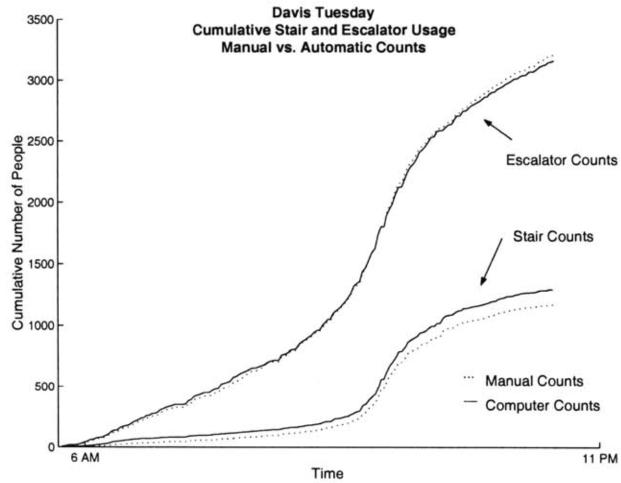
algorithm ranged from +10.5% to +20.5% over the four days. For the weekdays, the error range was +10.5% to 14.1%, and on the weekends it ranged from +20.2% to +20.5%. The difference between the algorithm's calculated stair usage percentage and actual stair usage percentage ranged from +1.7% to +2.8%. The mean error for percentage of people using the stairs was +2.2% with a standard error of .24%.

The traffic patterns in Davis are different than those in either of the Kendall inbound or outbound stations. During weekdays, there is a fairly constant flow of traffic leaving the station at a slow rate. During the evening rush hours, the amount of people leaving the station increases quickly. After the evening rush hour, the traffic flow slows down again to a similar rate as during the morning and early afternoon. This traffic pattern creates a S-shape (see Fig 5-7). This shape is likely due to the residential nature of the area. For weekends, the pattern is similar to the other stations. There is a constant slow amount of traffic until around 10 am when the rate increases. For the rest of the day, the traffic flow remains fairly constant out of the station. Fig. 5-6 also demonstrates how the automatic counting algorithm estimates the cumulative number of people exiting the station over the course of the day.

Over all three sites, the algorithm was tested on 204 hours of video with 24,186 people exiting using the escalator and 10,253 people exiting up the stairs. The difference between calculated stair usage percentage and actual stair usage percentage ranged from -3.8% to 2.8%. The mean error for percentage of people using the stairs was 1.1% with a standard error of .65%.

5.4 Impact of Motivational Messages on Stair Usage

After baseline data was collected for Kendall inbound and outbound, a motivational message was displayed at both locations. This message is the same message that was used in the study by Brownell et al. [8] except that the heart on the escalator was modified to look less obese due to concerns that it might be mildly offensive to some



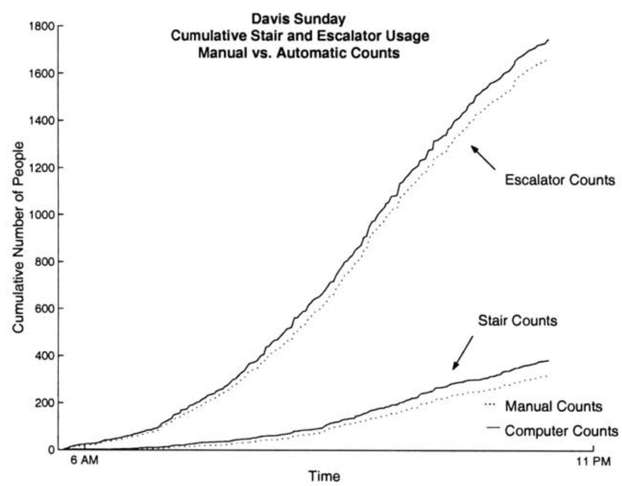


Figure 5-7: Graphs showing the typical traffic flow through Davis for both the escalator and the stairs. The graphs also show how the algorithm estimated the number of people throughout the day. The graph on the left shows a weekday, and the graph on the right shows a weekend.

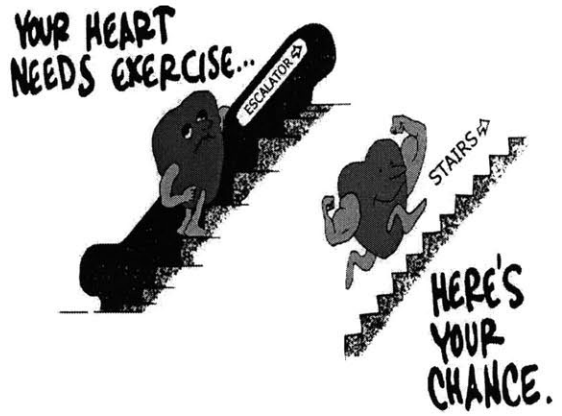


Figure 5-8: Intervention sign (original Brownell study [8] with escalator heart modified).

individuals (see Fig. 5-8). Many of the other studies discussed in this work used signs based on this poster. Additionally, when this sign was used in the Brownell et al. study, it produced one of the larger increases in stair usage. This sign was chosen because it was likely have a large impact on stair usage, and it allows for a good comparison between the results of this study and the previous research.

Weekday stair usage in Kendall inbound showed a statistically significant increase of 4.3% ($p < .001$) in response to the intervention. The average stair use increased from 39.3% during the seven baseline days to 43.7% during the nine days that the sign was projected (see Fig 5-9). Over the baseline period 20,284 observations were made, and during the intervention period 26,158 were made. The last two days of the intervention period produced a drop in the stair usage. There is not enough information at this time to determine if these points represent the start of a downward trend or if the following day's stair usage would have been closer to the intervention average. Kerr et al. previously found that in males under 60 years of age, stair usage begins to decrease as the intervention period continues [20]. It is possible that a similar effect is occurring, and people are beginning to ignore the intervention. Further research needs to examine the possible loss of effectiveness of an intervention over time, and determine if animation or other forms of feedback can maintain a

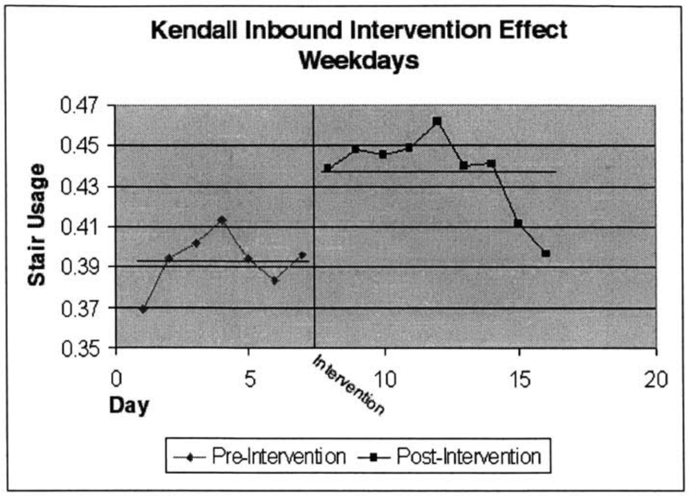


Figure 5-9: Stair usage percentage for weekdays before and after the intervention in Kendall inbound. Stair usage increased from 39.3% to 43.7% ($p < .001$).

higher stair usage level.[1]

Weekend stair usage in Kendall inbound showed a statistically significant increase of 3.0% ($p < .001$) in response to the intervention. The average stair use increased from 23.0% during the four baseline days to 26.0% during the 4 days that the sign was projected (see Fig 5-10). In the four baseline days 4685 people were counted leaving the station, and during the four intervention days 4824 people were counted exiting the station. There appears to be a downward trend in the stair usage over the intervention period, which more data may support. Both the weekend and weekday graphs showed possible downward trends of stair usage during the intervention period, so it is possible that the effectiveness of the sign decreased over time for this location.

Weekday stair usage in Kendall outbound showed a statistically significant increase of .7% ($p < .0155$) in response to the intervention. The average stair use increased from 38.8% during the ten baseline days to 39.6% during the ten days that the sign was projected (see Fig 5-11). 45,203 observations were made during the baseline phase, and 45,582 were made during the intervention phase. Unlike the inbound station, the stair usage over the intervention period had an upward trend. One

[1]The system will continue to run for at least 6 months and likely as long as 1 year. Many of the questions about the current results will be resolved over this time.

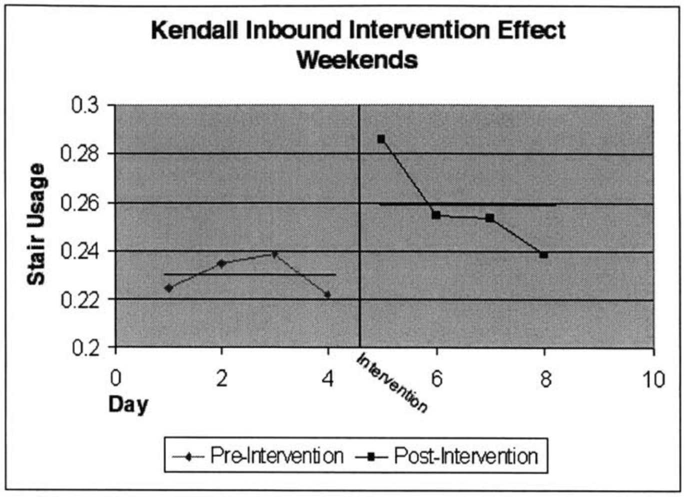


Figure 5-10: Stair usage percentage for weekend days before and after the intervention in Kendall inbound. Stair usage increased from 23.0% to 26.0% (p < .001).

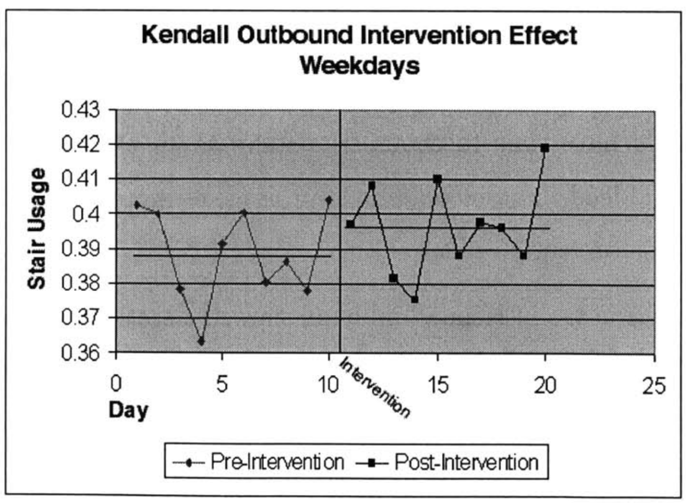


Figure 5-11: Stair usage percentage for weekdays before and after the intervention in Kendall outbound. Stair usage increased from 38.8% to 39.6% (p < .0155).

Figure 5-12: A comparison of the view of the signs from the turnstiles in both Kendall inbound (left) and Kendall outbound (right). The top of the sign in Kendall outbound is obscured.

hypothesis for this trend is that because of the geometry of the Kendall outbound station, the top the projected sign is somewhat obscured from the view of a person at the turnstiles (see Fig. 5-12). This positioning makes the "Your Heart Needs Exercise" text unreadable. As people move quickly through the station, by the time they can read the entire message they may have already committed to using the stairs or escalator. On their next trip through the station, if they had previously noticed the sign, they would have time to choose the stairs as an alternative to the escalator. This effect would lead to an increase in stair usage as more people begin noticing the sign during the intervention phase.

Weekend stair usage in Kendall inbound showed a statistically significant increase of 2.8% ($p < .001$) in response to the intervention. The average stair use increased from 21.8% during the four baseline days to 24.6% during the four days that the sign was projected (see Fig 5-13). 9,209 people were counted leaving the station during the baseline phase, and 7,530 were counted during the intervention. Unlike the weekday data for the outbound station, a trend of increased stair usage over time is not seen in in the weekend data. This lack of a trend may involve the different traffic patterns seen on weekdays and weekends. On weekends when there is no rush

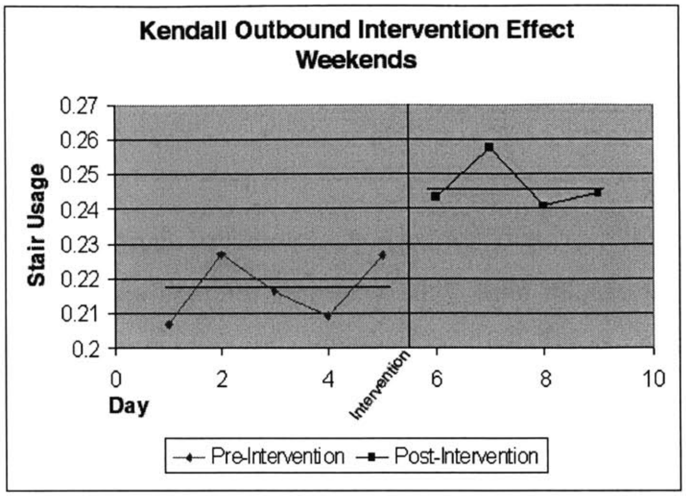


Figure 5-13: Stair usage percentage for weekend days before and after the intervention in Kendall outbound. Stair usage increased from 21.8% to 24.6% ($p < .001$).

to leave the station, people have more time to read the sign and alter their path to take the stairs.

Stair usage was significantly higher during the week in both stations than during the weekend. During the week, the majority of traffic passes through the station during rush hours. When a large number of people exit a train at once, the escalator quickly fills up with people. Consequently, more people take the stairs to avoid waiting. On the weekends, traffic flow is more consistent throughout the day and the escalator rarely becomes backed-up. There is no time saved by using the stairs, so the overall stair usage is lower.

5.5 Computing Significance of Results

There are several sources of potential error when examining the results of this work. The first source involves the human counts used for testing the algorithm. These numbers were used to measure the counting algorithm performance. To determine an error rate for human counting, all annotation was done by one person. A second person then annotated 1.1 hours of video that had already been annotated (.5% of the total test video). A level of 98% and 99% agreement was seen between annotators for the stairs and escalator respectively. Because the annotation agreement was high and it was done using software that allowed both slowing down and pausing the video, the annotation numbers were considered to be the actual numbers of people passing through the station.

Previous studies primarily used several standard significance tests to judge how meaningful their results were. The two most common approaches were t-tests and chi-squared tests. For the t-tests, the daily stair percentages were grouped into pre-intervention and post-intervention, and a p-value for the difference between the means of each group was calculated. For a chi-squared test, the presence of a motivational message was the experimental condition, and the choice of taking the stair or escalator was the outcome of a trial. A p-value for these numbers can be computed by using a 2x2 contingency table and performing the chi-squared contingency test. For this work, there is a larger observational error than the previous studies because of the use of computer vision for counting. Standard significance tests may underestimate p-values (and overestimate the significance) of the results.

To determine if the results were statistically significant, a simulation was created. The simulation works as follows: For each person that passes through the station, a random number is chosen from a uniform distribution between zero and one. Using the baseline stair use percentage as a cut-off, people with random numbers below the cut-off are considered to have taken the stairs, and people with a random number above the cut-off are considered to have taken the escalator. Using the four days of annotated video for a given station, detection rates of the algorithm for the stairs and escalator are determined. These are determined by dividing the number of people the algorithm counted on the escalator(stairs) over the four days by the number of manually counted people on the escalator(stairs) over the same four days. Another random drawing from the uniform distribution is done for each person in the simulation. If the person was designated as taking the escalator, they are classified as detected by the algorithm if their random number is less than the escalator detection rate. For people designated on the stairs, the stair detection rate is used as a cut-off. To simulate one trial, the above procedure is done once for each person that exited the station during the baseline phase. The process is then performed once for each person exiting the station during the intervention phase. The number of people exiting the station during a phase is calculated by dividing the number of people that the algorithm counted on the escalator by the escalator detection rate and adding that to the number of people detected on the stairs divided by the stair detection rate. A stair usage percentage for the baseline group and the intervention group can now be determined from the simulation data. The difference between the two percentages is calculated and the trial is complete.

The simulation runs 100,000 trials recording the difference between baseline and intervention stair use percentage in each trial. To determine a p-value for the simulation, the number of times this simulated difference in stair use percentage was greater than the difference between the observed stair usage means is calculated. This number is divided by 100,000 to determine the likelihood that the difference seen in real life stair usage was caused by chance. Appendix L has the matlab code used to run this simulation.

This method is computing the same value that a t-test or chi-squared test computes using mathematical formula by determining what percentage of the time the results seen in experimentation would occur by chance. By augmenting the model with information about the algorithm's detection rates for the stairs and escalator, the simulation should produce more accurate p-values. When the simulation results were compared with a chi-squared test over the same data (without accounting for the error of the algorithm), the simulation produced p-values equal to or slightly higher than those of the chi-squared test.

5.6 Evaluation as a Cost-effective Measurement Tool

Appendix D shows the cost breakdown of constructing one of the automatic counting systems described in this work. The cost for parts for one box is $3,820. At the standard MIT undergraduate research assistant wage of $10 per hour, 118 hours of labor for construction and maintenance would bring the total cost to $5,000. If the

$5,000 was strictly budgeted to paying undergraduate research assistants for people counting, it would fund 500 hours of people counting. In the experiments running in the Kendall inbound and outbound stations, the systems count people for 20 hours a day. If an undergraduate research assistant was to count for 20 hours a day, the $5,000 would fund 25 days worth of counting.

In testing the effect of an intervention on stair usage in the Kendall stations, 50 days worth of counting was performed. Since two systems were in use, the total cost was approximately equivalent to paying two undergraduate research assistants to count for the same amount of time. If the study were increased in length by any amount, the automated system would quickly become more economical than employing human counters.

Additionally, stationing a human observer in a transit station for 20 hours a day, especially in harsh weather conditions, is impractical. Consequently, it would take a human observer significantly longer than 50 days to accumulate 1,000 hours of counts. The system is planned to continue running for another 6 to 12 months, providing more observations at a lower cost than a human counter.

Chapter 6

Conclusion

In this work, a system that automates the process of people-counting to determine the effects of "just-in-time" messaging on behavior in public spaces was described. A computer vision algorithm was developed that could determine the percentage of people taking the stairs as they left a transit station with an error range of -3.8% to +2.8% and a standard error of .65%. The algorithm operates in real time and does not require a specific overhead placement to operate properly. The entire self-contained system designed to run the algorithm and project signs in a transit station consists of a camera, laptop, and projector. An enclosure was designed to protect the equipment and allow for easy installation in different transit stations. Three systems were built and deployed in Boston area transit stations. Two of the systems ran for nearly four weeks each. They collected data on stair and escalator usage with and without an intervention encouraging stair use. In both sites, a statistically significant increase in stair usage was seen on weekdays and weekends. The results from these experiments agree with the results of many previous stair/escalator studies.

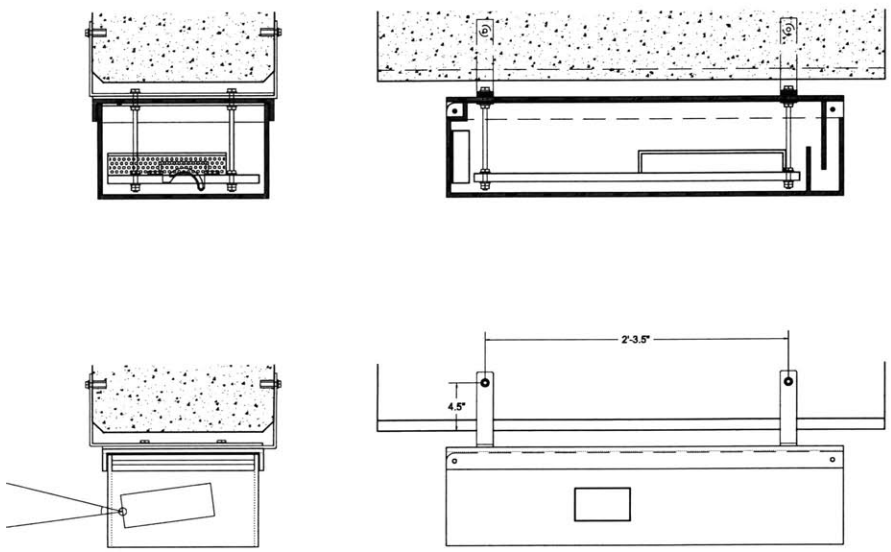
In performing an experiment on the effects of an intervention on stair usage, several avenues for future research were identified. The results from this work indicate that the effectiveness of an intervention can vary over time. The system developed can alter the message it projects based on things happening in the environment, and it is cost-effective to run for a long period of time. Consequently, it presents the perfect opportunity to examine long-term effects of interventions. Studies involving changing messages periodically or using animated messages may lead to information on how to prevent people from becoming desensitized to interventions designed to increase healthy behavior. Messages that provide feedback from the environment, such as a sign that shows the stair usage for a location over the past six hours, may also provide beneficial to increasing healthy behavior in public spaces. This system also provides the ability to display interventions that are based on the state of the environment in which the system is installed. For example, a system that switches from a health-related message to a message that says "Save time, take the stairs," when the escalator becomes crowded may prove more effective because it presents a message that is relative to that location at that specific point in time.

An automated counting system that can measure the effects of "just-in-time" messaging is viable and cost-effective. By building on this work, new research on encouraging healthy behavior in public spaces is possible. This research may one day lead to longer, healthier lives for many people.

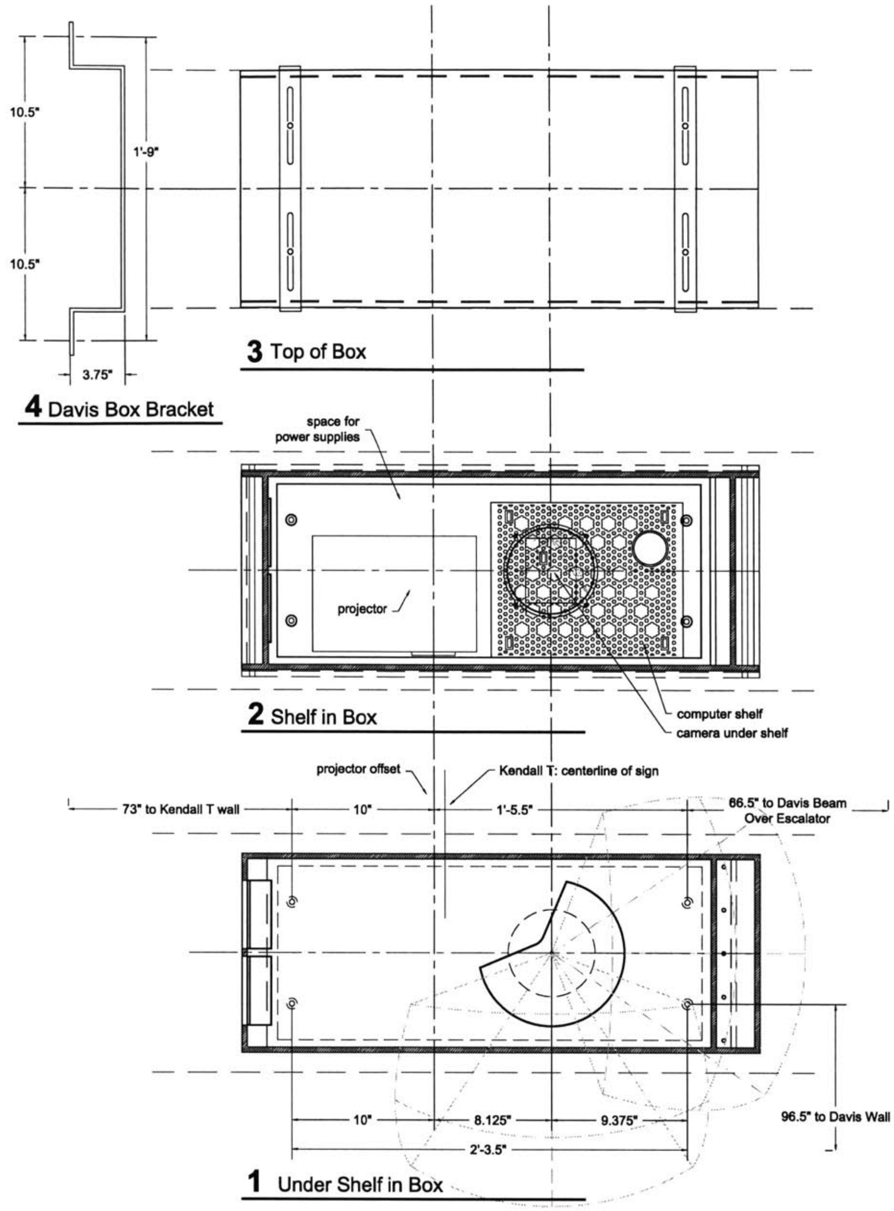
Appendix A

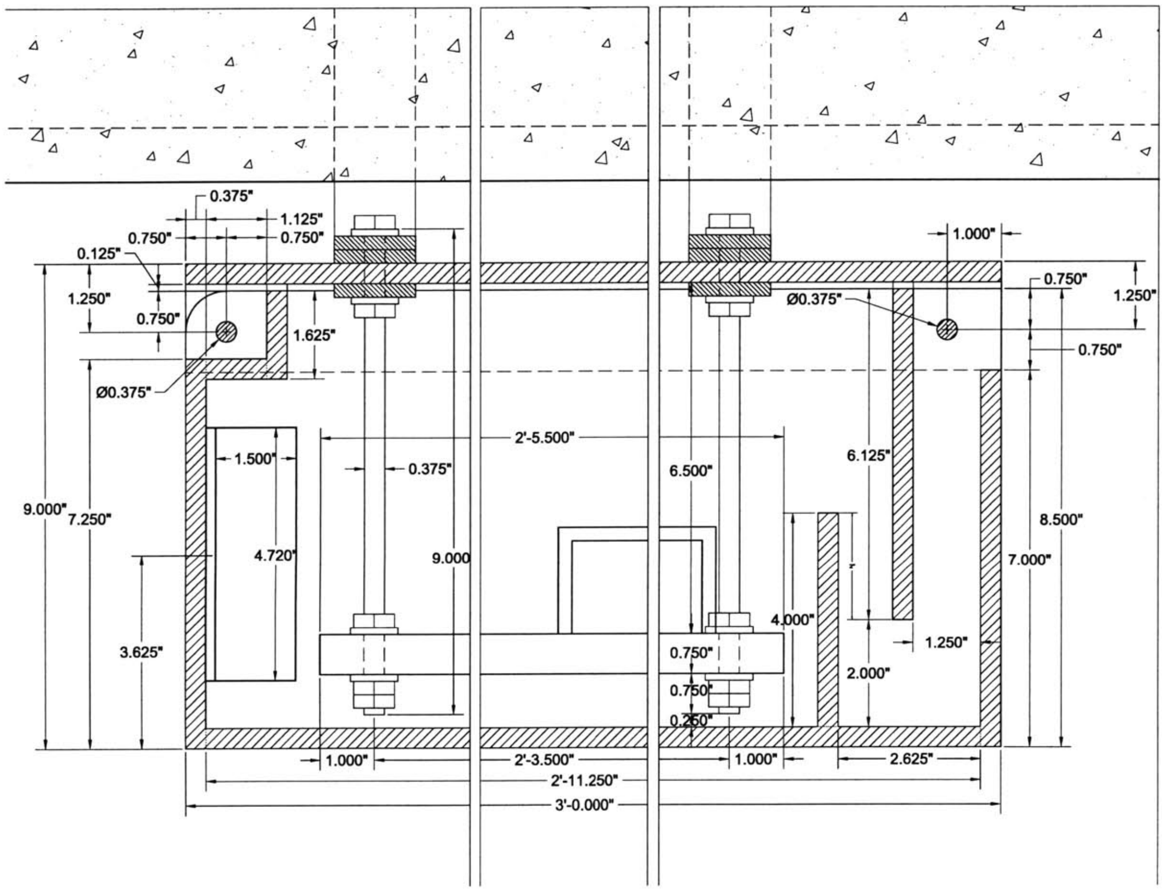
Enclosure Box Diagrams

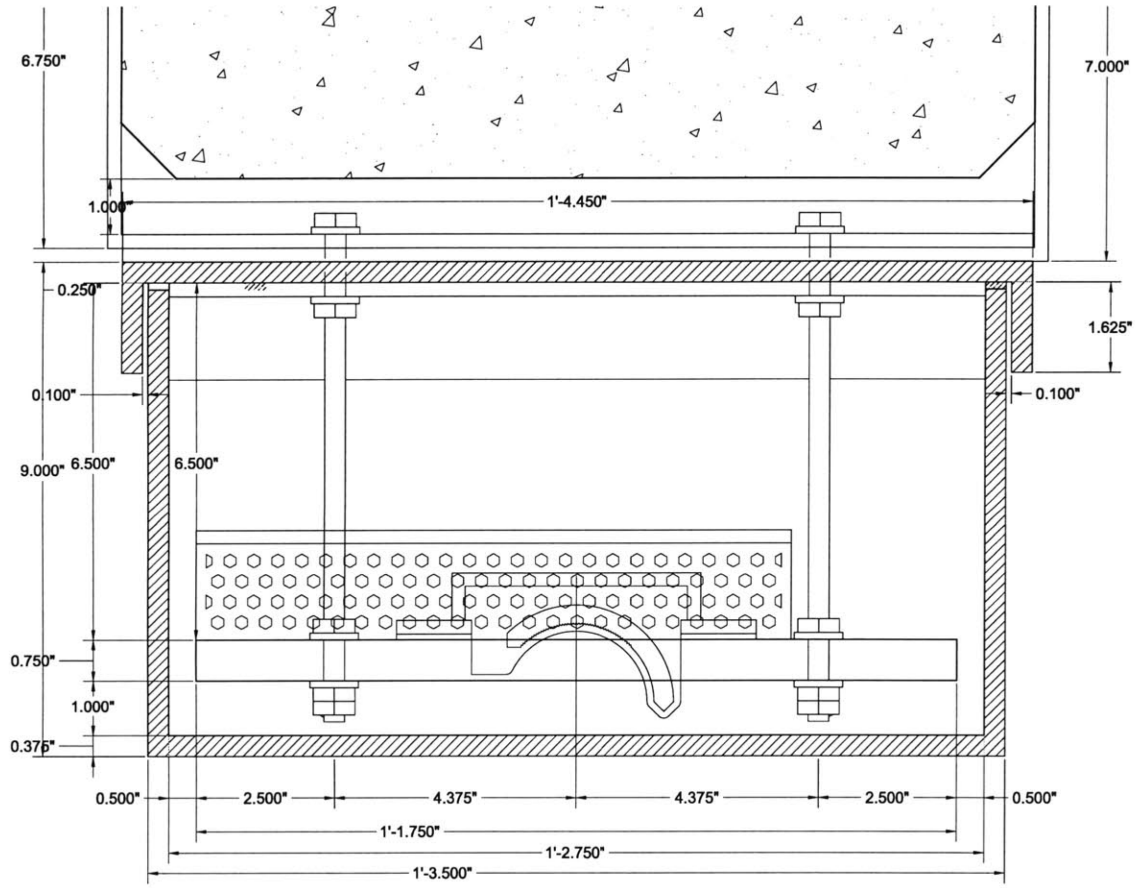
The enclosure box was constructed out of 3/8" Lexan polycarbonate. Individual parts can be found in Appendix D. The box was constructed separately from the mounting brackets so the way in which the box is secured to the ceiling can be easily changed. For example, the two Kendall locations have brackets that attach to the side of a beam, while the Davis location has a bracket that is bolted directly up into the ceiling.

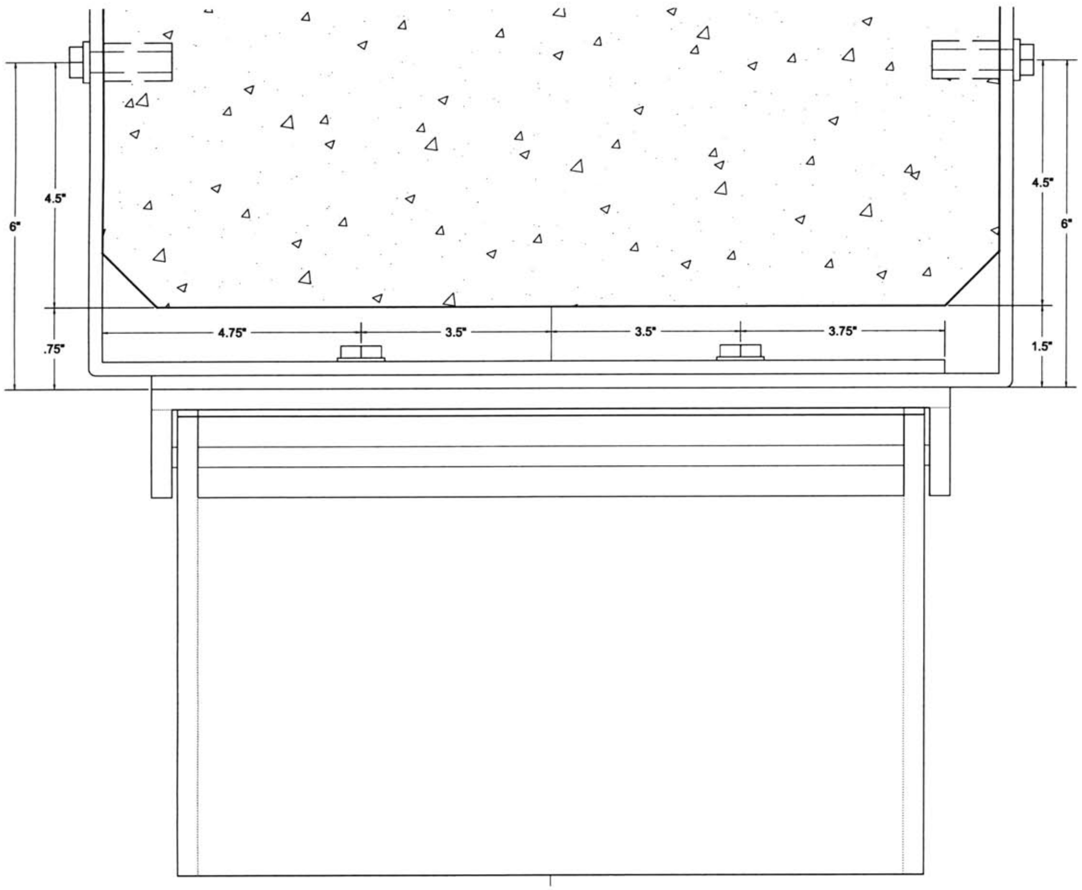


2'-3.5"
4.5"

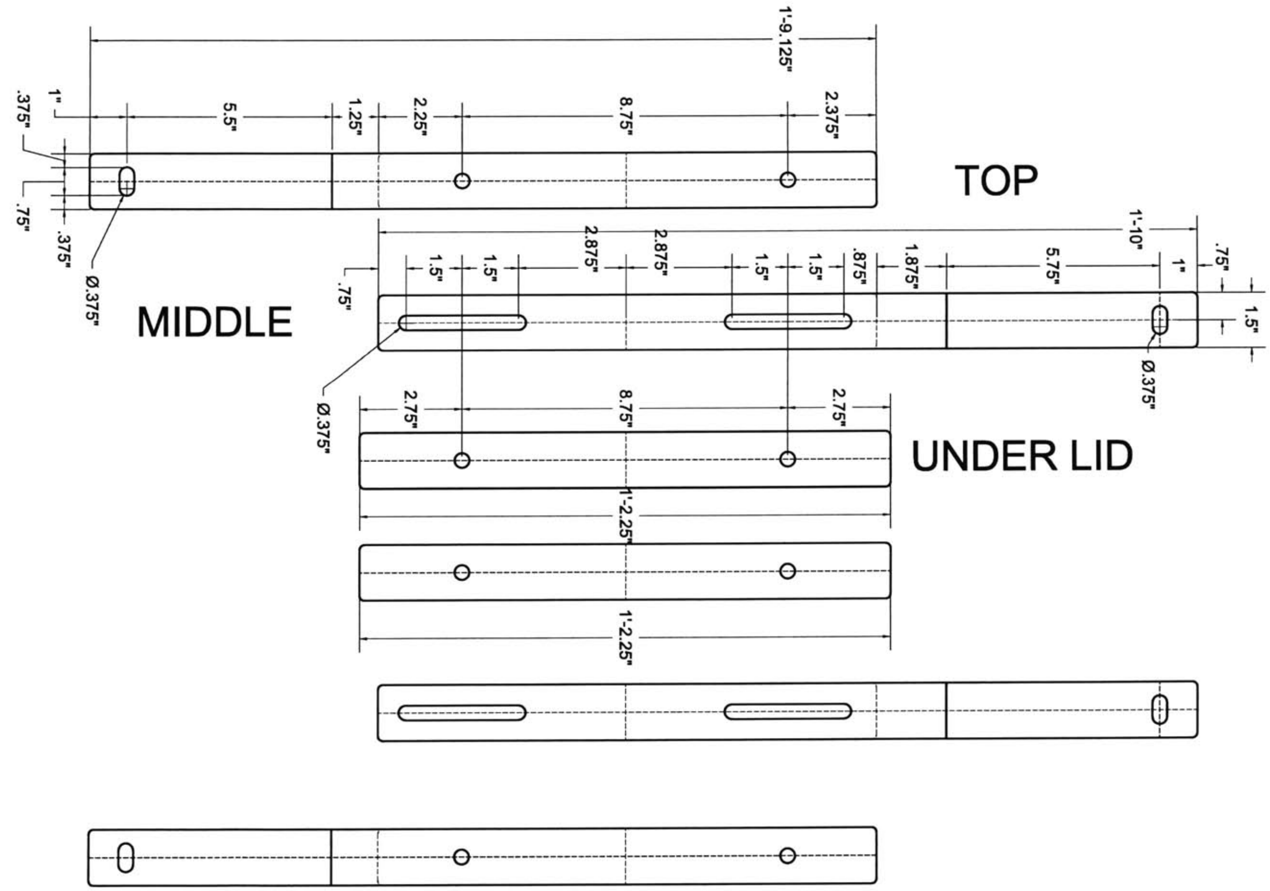

10.5"
1'-9"
10.5"
3.75"
4 Davis Box Bracket
3 Top of Box
space for power supplies
projector
2 Shelf in Box
computer shelf
camera under shelf
projector offset
Kendall T: centerline of sign
73" to Kendall T wall
10"
1'-5.5"
66.5" to Davis Beam Over Escalator
10"
8.125"
9.375"
2'-3.5"
96.5" to Davis Wall
1 Under Shelf in Box


0.375"
1.125"
0.750"
0.750"
0.125"
1.250"
0.750"
1.625"
Ø0.375"
1.500"
4.720"
3.625"
9.000"
7.250"
2'-5.500"
0.375"
9.000
6.500"
0.750"
0.750"
0.250"
4.000"
2.000"
6.125"
1.250"
1.000"
0.750"
Ø0.375"
1.250"
0.750"
8.500"
7.000"
1.000"
2'-3.500"
1.000"
2.625"
2'-11.250"
3'-0.000"


6.750"
7.000"
1'-4.450"
1.000"
0.250"
1.625"
0.100"
0.100"
9.000"
6.500"
6.500"
0.750"
1.000"
0.375"
0.500"
2.500"
4.375"
4.375"
2.500"
0.500"
1'-1.750"
1'-2.750"
1'-3.500"



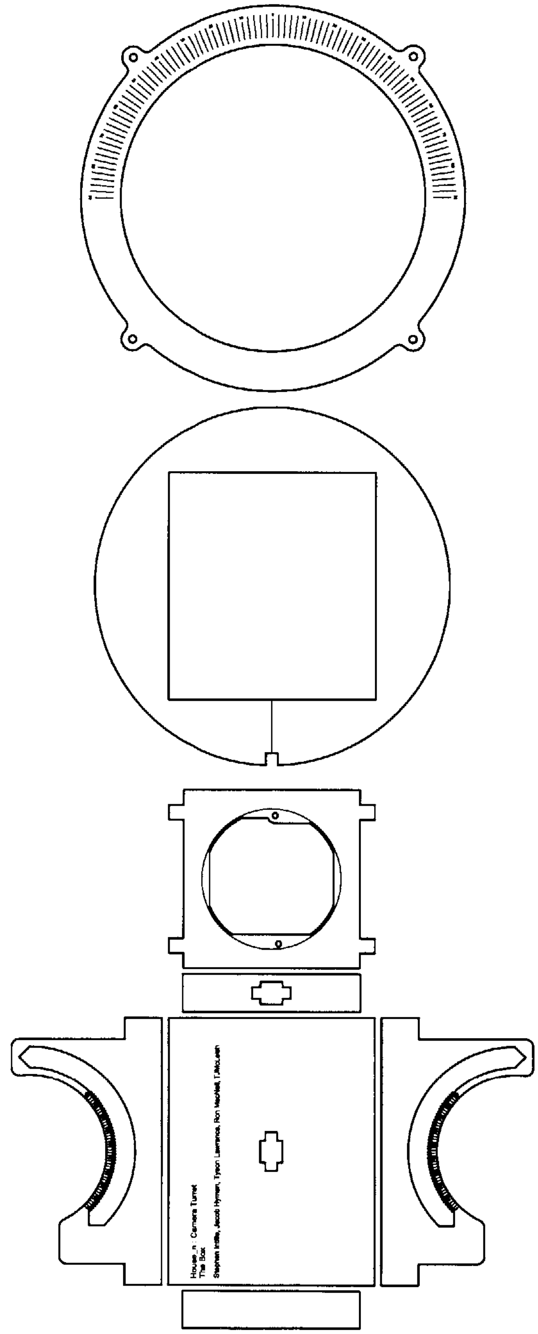
6"
4.5"
.75"
4.75"
3.5"
3.5"
3.75"
4.5"
6"
1.5"


TOP
MIDDLE
UNDER LID
1'-9.125"
1"
.375"
5.5"
1.25"
2.25"
8.75"
2.375"
.75"
.375"
Ø.375"
1'-10"
2.875"
2.875"
1.5"
1.5"
1.5"
1.5"
.875"
1.875"
5.75"
1"
.75"
.75"
1.5"
Ø.375"
Ø.375"
2.75"
8.75"
2.75"
1'-2.25"
1'-2.25"

Appendix B

Camera Turret Diagram

The camera turret was constructed out of 1/4" black acrylic. The round piece with tick marks on it is made out of 1/8" black acrylic. The turret rests on this piece allowing it to spin smoothly.


House_n : Camera Turret
The Box
Stephen Intille, Jacob Hyman, Tyson Lawrence, Ron MacNeil, TJ McLeish

Appendix C

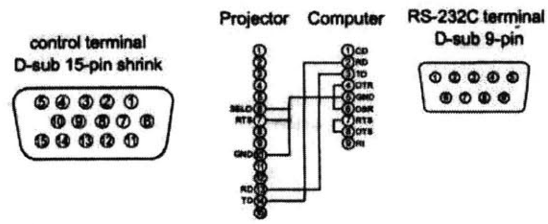
Projector Control

Diagrams from ViewSonic Corporation.

This diagram shows how to construct a serial cable to communicate with the projector. It also describes the command protocol and serial port settings necessary.

13. RS-232C communication

(1) Turn off the projector and computer power supplies and connect with the RS-232C cable.
(2) Turn on the computer power supply and, after the computer has started up, turn on the projector power supply.



Communications setting
19200bps, 8N1

1 Protocol
Consist of header (7 bytes) + command data (6 bytes).

2 Header
BE + EF + 03 + 06 + 00 + CRC_low + CRC_hige
CRC_low : Lower byte of CRC flag for command data.
CRC_high : Upper byte of CRC flag for command data.

3 Command data

Command data chart

byte_0	byte_1	byte_2	byte_3	byte_4	byte_5
Action		Type		Setting code	
low	high	low	high	low	high

Action (byte_0 - 1)

Action	Classification	Content
1	SET	Change setting to desired value.
2	GET	Read projector internal setup value.
4	INCREMENT	Increment setup value by 1.
5	DECREMENT	Decrement setup value by 1.
6	EXECUTE	Run a command.

This page describes what the projector returns when issued various commands.

Requesting projector status (Get command)

(1) Send the request code Header + Command data ('02H'+'00H'+ type (2 bytes) +'00H'+'00H') from the computer to the projector.

(2) The projector returns the response code '1DH'+ data (2 bytes) to the computer.

Changing the projector settings (Set command)

(1) Send the setting code Header + Command data ('01H'+'00H'+ type (2 bytes) + setting code (2 bytes)) from the computer to the projector.

(2) The projector changes the setting based on the above setting code.

(3) The projector returns the response code '06H' to the computer.

Using the projector default settings (Reset Command)

(1) The computer sends the default setting code Header + Command data ('06H'+'00H'+ type (2 bytes) +'00H'+'00H') to the projector.

(2) The projector changes the specified setting to the default value.

(3) The projector returns the response code '06H' to the computer.

Increasing the projector setting value (Increment command)

(1) The computer sends the increment code Header + Command data ('04H'+'00H'+ type (2 bytes) +'00H'+'00H') to the projector.

(2) The projector increases the setting value on the above setting code.

(3) The projector returns the response code '06H' to the computer.

Decreasing the projector setting value (Decrement command)

(1) The computer sends the decrement code Header + Command data ('05H'+'00H'+ type (2 bytes) +'00H'+'00H') to the projector.

(2) The projector decreases the setting value on the above setting code.

(3) The projector returns the response code '06H' to the computer.

When a command sent by the projector cannot be understood by the computer

When the command sent by the projector cannot be understood, the error command '15H' is returned by the computer. Some times, the projector ignores RS-232C commands during other works. If the error command '15H' is returned, please send the same command again.

When data sent by the projector cannot be practice

When the command sent by the projector cannot be practiced, the error code '1cH' +'xxxxH' is returned.
When the data length is greater than indicated by the data length code, the projector will ignore the excess data code.
Conversely, when the data length is shorter than indicated by the data length code, an error code will be returned to the projector.

NOTE:

- Operation cannot be guaranteed when the projector receives an undefined command or data.
- Provide an interval of at least 40ms between the response code and any other code.
- The projector outputs test data when the power supply is switched ON, and when the lamp is lit. Ignore this data.
- Commands are not accepted during warm-up.

This chart shows the command strings to send the projector to get it perform various functions including turning on and off.

Command data chart

Names	Operation type	Header	CRC	Action	Type	Setting code
				Command data		
Magnify	Get	BE EF 03 06 00	7C D2	02 00	07 30	00 00
	Increment	BE EF 03 06 00	1A D2	04 00	07 30	00 00
	Decrement	BE EF 03 06 00	CB D3	05 00	07 30	00 00
Auto off	Get	BE EF 03 06 00	08 86	02 00	10 31	00 00
	Increment	BE EF 03 06 00	6E 86	04 00	10 31	00 00
	Decrement	BE EF 03 06 00	BF 87	05 00	10 31	00 00
Brightness Reset	Execute	BE EF 03 06 00	58 D3	06 00	00 70	00 00
Contrast Reset	Execute	BE EF 03 06 00	A4 D2	06 00	01 70	00 00
V.Position Reset	Execute	BE EF 03 06 00	E0 D2	06 00	02 70	00 00
H.Position Reset	Execute	BE EF 03 06 00	IC D3	06 00	03 70	00 00
H.Size Reset	Execute	BE EF 03 06 00	68 D2	06 00	04 70	00 00
Color Balance R Reset	Execute	BE EF 03 06 00	94 D3	06 00	05 70	00 00
Color Balance B Reset	Execute	BE EF 03 06 00	D0 D3	06 00	06 70	00 00
Sharpness Reset	Execute	BE EF 03 06 00	C4 D0	06 00	09 70	00 00
Color Reset	Execute	BE EF 03 06 00	80 D0	06 00	0A 70	00 00
Tint Reset	Execute	BE EF 03 06 00	7C D1	06 00	0B 70	00 00
Keystone Reset	Execute	BE EF 03 06 00	08 D0	06 00	0C 70	00 00
Auto	Execute	BE EF 03 06 00	91 D0	06 00	0A 20	00 00
Blank on/off	Set off	BE EF 03 06 00	FB D8	01 00	20 30	00 00
	Set on	BE EF 03 06 00	6B D9	01 00	20 30	01 00
	Get	BE EF 03 06 00	C8 D8	02 00	20 30	00 00
Error Status	Get	BE EF 03 06 00	D9 D8	02 00	20 60	00 00
		(Example of Return) 00 00 (Normal) 01 00 (Cover-error) 02 00 (Fan-error) 03 00 (Lamp-error) 04 00 (Temp-error) 05 00 (Air-flow-error) 06 00 (Lamp-Time-over) 07 00 (Cold-error)				
Power	Set OFF	BE EF 03 06 00	2A D3	01 00	00 60	00 00
	Set ON	BE EF 03 06 00	BA D2	01 00	00 60	01 00
	Get	BE EF 03 06 00	19 D3	02 00	00 60	00 00
Input Source	Set RGB1	BE EF 03 06 00	FE D2	01 00	00 20	00 00
	Set Video	BE EF 03 06 00	6E D3	01 00	00 20	01 00
	Set SVideo	BE EF 03 06 00	9E D3	01 00	00 20	02 00
	Set Component	BE EF 03 06 00	AE D1	01 00	00 20	05 00
	Get	BE EF 03 06 00	CD D2	02 00	00 20	00 00
Volume	Get	BE EF 03 06 00	31 D3	02 00	01 20	00 00
	Increment	BE EF 03 06 00	57 D3	04 00	01 20	00 00
	Decrement	BE EF 03 06 00	86 D2	05 00	01 20	00 00

Appendix D

Parts List

Parts listed are for construction/set-up of one system with 2003 estimated prices in parenthesis. Total cost of parts per installation unit: $3,820

Computer Related Parts:

- Dell Inspiron 5100 Laptop Computer ($1,800)
- USB to Serial Adapter ($30)
- Kanguru 120 GB USB2.0 Hard Drive ($200)

Camera Related Parts:

- Orange Micro IBot2 Camera ($80)
- Lens Holder (M12x0.5, 16.2, centered, Part number CMT103 from Sunex Inc.) ($6)
- Reflective IR cut-off filter (Part Number IRC30-10R from Sunex Inc.) ($17.50)
- Wide-Angle Lens (Part Number V-4302-2 from Marshall Electronics) ($29)
- 1/4" Black Acrylic for Camera Turret (see Appendix B). ($40)

Projector Related Equipment:

- ViewSonic PJ500 Projector ($1,000)
- Cable for laptop/projector communication (see Appendix C) ($15)

Enclosure Related Equipment:

- 3/8" Lexan Polycarbonate for box construction ($320)
- 4 18-8 Stainless Steel Hex Head Cap Screw, 3/8" - 16 Thread, 9" Length (fully threaded) ($10)
- 12 18-8 Stainless Steel Hex Heavy Thin (Jam) Nut, 3/8" - 16 Screw Size ($10)
- 8 Spring Lock Washers, 3/8" Screw Size ($6)
- 12 Flat Washers, 3/8" Screw Size ($8)
- Aluminum Rod, 3/8" Diameter ($20)
- 2 Cotter Pins, 3/8" Diameter, 1-1/4" Length ($5)
- 2 Small Locks (for cotter pins) ($5)
- Foam Rubber Strip, 1/8" thick, 3/8" width (Part Number 93725K52 from McMaster-Carr) ($8)
- 3/4" Plywood (for tray in box) ($15)
- 2 115VAC Fan, 120 x 38mm (Part Number 259-1093-ND from Digikey) ($28)
- 2 Power Cords for Fan (Part Number CR365-ND from Digikey) ($3)
- Six Outlet Power Strip ($8)
- 1/4" Aluminum Strips for Mounting Brackets (see Appendix A) ($150 includes welding and cutting)
- 1/4" Velcro Strips (to mount items on tray) ($5)
- 1/4" Acrylic (for Computer Stand)

Appendix E

Calibration

Because of varying lighting conditions, camera angles, and stair/escalator arrangements, the system needs to be calibrated to each location that it is placed in. The first step in calibration is to determine where to place the scan lines. One line is placed along the entrance area to the escalator and the other is placed across the gray area between the edges of steps on the stairs (see Fig 4-8). Placing the scan lines in areas that get the least amount of natural light will result in better algorithm performance. A rectangular area around each of the scan lines needs to be defined for background subtraction to be performed. The area should start at least one pixel the left of the scan line's start and continue one pixel past its end. The area should also start several pixels above the scan line's top point and end several pixels below its bottom point.

To calibrate the threshold for frame differencing, a process of trial and error is used. Starting with a very low value, the threshold is gradually increased. Once the majority of the noise in the difference image is gone, but the edges of people are still clearly visible, the current value for the threshold should be used.

To calibrate for the background subtraction, a similar trial and error method is used. There are three background subtraction thresholds (one for each channel). Experience has shown that keeping the Y channel threshold slightly higher than the U and V channel thresholds produces good results. The U and V channel thresholds should be equal. To determine final thresholds, the values should initially be set low for all the channels. With these low values, there will be a lot of noise in the background subtraction image. Additionally, shadows will be selected as foreground creating blobs that are a lot larger than the actual foreground object. The thresholds should all be set to values that are perfect squares. When incrementing the values, all the thresholds should be increased to the next largest perfect square until values are reached that produce blobs only slightly larger than the actual foreground object. The blobs should be mostly solid, but a few empty patches of pixels is OK. The minimum shadow threshold should be set to the same number as the Y channel threshold. The background subtraction thresholds are set separately for the stair and escalator areas.

To set the thresholds for how far a tracker has to move to be considered as travelled up or down the stairs is a little more difficult. This calibration needs to be done using manually annotated test video. First a value of 5 is selected for both the up and down threshold, and the algorithm is run for twenty to thirty minutes. The percentage of trackers moving up is then compared to the percentage of people going up the stairs during that time. If too high a percentage of trackers are moving up, the up threshold can be increased or the down threshold can be decreased. The process is repeated until the percentage of trackers moving up is approximately equal to the percentage of people moving up the stairs. Using a combination of increasing one threshold and decreasing the other works better than just increasing or decreasing one threshold by a larger value.

The number of pixels to ignore because of escalator motion is easy to set. Simply observe the number of difference pixels that appear per frame in the escalator area of interest (the area for background subtraction) when the escalator is empty. This threshold should be set slightly higher than the typical number observed.

Noise level thresholds need to be set for both the escalator and stairs. These thresholds are the number of pixels that need to appear in a difference image stack for it to be analyzed by the algorithm. These can be set by having the algorithm output how many pixels appear in each image stack that is created. By examining the video as the algorithm runs, the typical level for noise in an empty difference image stack can be determined. The threshold should be set just higher than the typical noise level observed. This needs to be done for both the stair and escalator separately.

The determination of whether or not a print in the background image stack is a single person relies on minimum and maximum sizes for the blob. To determine the numbers to use for the minimum and maximum height and minimum and maximum width requires trial and error. Initially, the minimum width should be set to 10, and the maximum width should be set to 1/2 the width of the stairs (in pixels). The minimum and maximum height should be set to 10 and 40 respectively. The algorithm can be run over segments of video, and it can output the background image stacks indicating which blobs were considered single people. By comparing the blobs selected by the algorithm to the actual set of single person blobs, the numbers can be tweaked until the algorithm selects most of the correct blobs. Additionally, if the algorithm is consistently under-counting throughout the day, the maximum and minimum numbers can be increased. Conversely, if the algorithm is consistently over-counting throughout the day, the numbers can all be decreased. These numbers need to be set for both the escalator and the stairs separately.

The final threshold to set is the minimum number of difference pixels (minclustsize) for a person to be added to the running average of pixels per person. Occasionally, a person will show up in the background subtraction image but not leave many pixels in the difference image. If the number of difference pixels for this person was added to the average of pixels per person, it would artificially lower the running average. To compensate, a minimum number of pixels needs to appear in the difference image stack for a person to be added to the running average of pixels per person. This threshold can be set by examining the background subtraction and difference image stacks produced by the algorithm. The number should be set a little higher than the average number of pixels in the difference stack seen for people that produce a distinct print in the background subtraction image stack, but very few pixels in the difference image stack. As long as this number is within a certain range, it does not have a significant effect on the performance of the algorithm.

The actual thresholds used in both Kendall stations and Davis can be seen in

Table E.1.

Threshold	Kendall (Both)	Davis
Difference Threshold	12	12
Stairs		
Background Subtraction Y	36	49
Background Subtraction U	25	36
Background Subtraction V	25	36
Shadow Min	25	36
Escalator		
Background Subtraction Y	16	36
Background Subtraction U	9	25
Background Subtraction V	9	25
Shadow Min	9	25
Tracker Up	6	3
Tracker Down	1	4
Escalator Motion to Ignore	15	15
Stair Noise Level	15	10
Escalator Noise Level	16	10
Stairs Single Person Print Selection		
Min Width	15	7
Max Width	50	35
Min Height	10	7
Max Height	38	30
Escalator Single Person Print Selection		
Min Width	10	7
Max Width	10	35
Min Height	50	10
Max Height	50	40
Stair minclustsize	30	25
Escalator minclustsize	10	10

Table E.1: Table showing the threshold values used in both Kendall stations and Davis.

Appendix F

Interesting Daily Graphs

In analyzing the data from the Kendall transit stations a few days displayed unusual stair usage percentages. For example, on July 27th, the stair usage percentage for Kendall inbound was over 33%, which is more than 10% higher than normal for a weekend during the baseline phase. Looking at the graph of traffic through inbound that day shows that there were no people using the escalator before 11:30 am (see Fig F-1). Examining the captured images for that day reveals that a cone was placed in front of the escalator early that morning (presumably because it was broken) (see Fig. F-2). The escalator had a barrier in front of it until 11:30 am (see Fig. F-3). At this point, the increase in traffic on the escalator is rather dramatic.

While the escalator was eventually fixed on the 27th, it broke again between midnight and 5am on July 28, 2003 (see Fig. F-4). The graph shows an interesting traffic pattern of people leaving the station. The majority of the people exiting the station are taking the stairs until about 8pm. This disparity is because the elevator was broken. However, unlike the previous day, there is still some traffic on the escalator, especially during morning rush hours. This is because there is no barrier in front of the escalator, so when the station is crowded people will still use the broken escalator as if it was another set of stairs. After rush hour, traffic on the escalator is very low again. Examining the captured images from this day show that the escalator was indeed broken. Near 3pm, workers begin working on the elevator and their presence causes a small increase in the escalator count. By 3:25pm, they

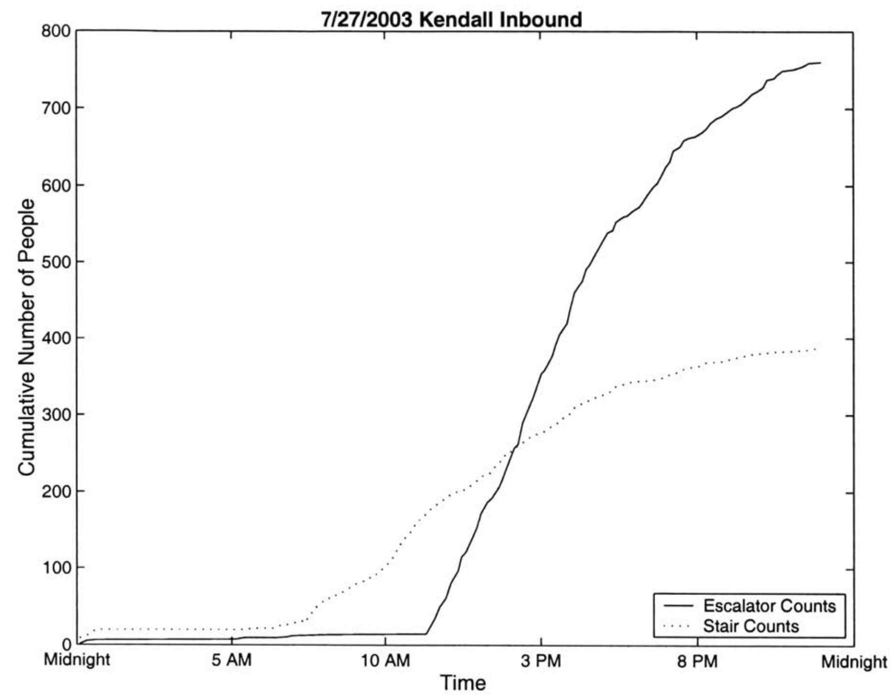


Figure F-1: The cumulative traffic through Kendall inbound on 7/27/03. The escalator has a barrier in front of it until 11:25 am.

Figure F-2: A cone is placed in front of the escalator on 7/27/03 which is later replaced by a barrier.

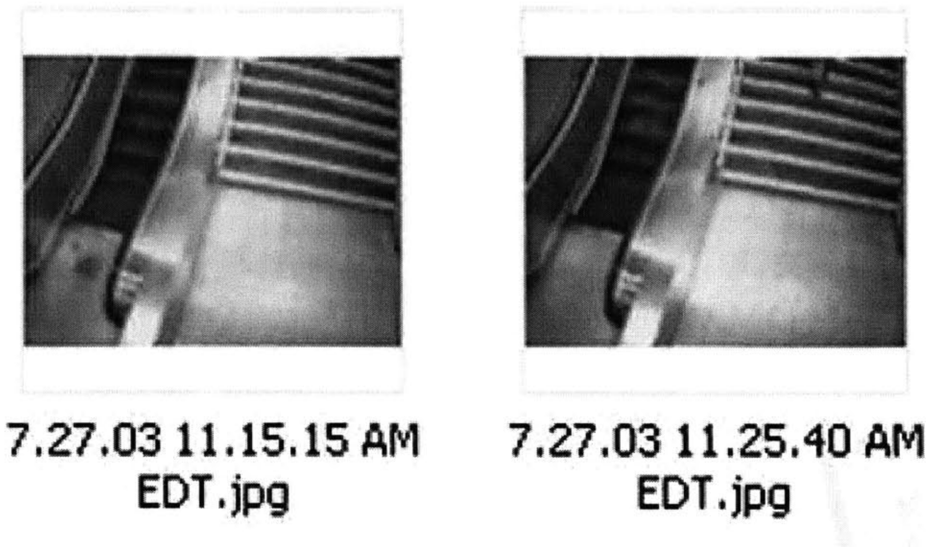


Figure F-3: The cone which replaced the barrier is eventually removed and people resume using the escalator.

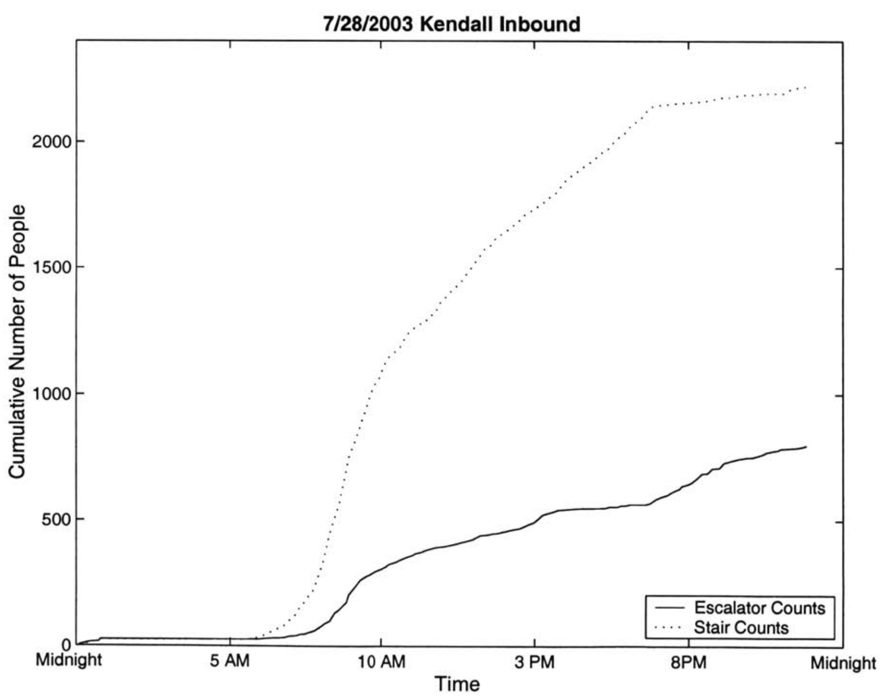


Figure F-4: The cumulative traffic through Kendall inbound on 7/28/03. The escalator was broken until 7 pm.

Figure F-5: Workers arrive to fix a broken escalator and put a barrier in front of the entrance on 7/28/03.

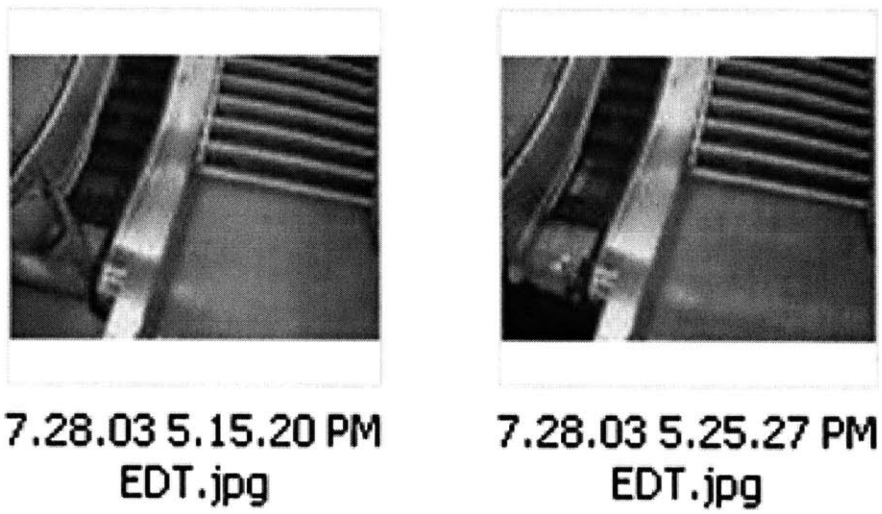


Figure F-6: The barrier is removed, but the workers continue fixing the panel in front of the escalator on 7/28/03.

Figure F-7: The escalator is eventually fixed on 7/28/03.

have placed a barrier in front of the escalator (see Fig. F-5). Between 5:15pm and 5:25pm, the barrier is moved, but they are still working underneath the panel in front of the escalator (see Fig. F-6). The escalator eventually reopens between 6:45pm and 6:55 pm (see Fig. F-7). The graph of traffic through the station shows a return to the normal trend at this time. The rate of people exiting on the escalator increases to levels normal for this time and the number of people leaving on the stairs nearly flattens out again (see Fig F-4).

Appendix G

People Count Data

Counts from the intervention period can be seen in Table G.1. Counts from the baseline period can be seen in Table G.2.

Intervention Counts

Date	Station	Counts			
		Escalator Counts	Stairs Total	Stairs Up	Stair Percentage
8.19.03	inbound	1824	6076	1201	39.70%
	outbound	2594	4840	1872	41.92%
8.18.03	inbound	1692	5727	1183	41.15%
	outbound	2534	4406	1608	38.82%
8.17.03	inbound	776	2070	243	23.85%
	outbound	1258	1419	407	24.44%
8.16.03	inbound	1032	2846	350	25.33%
	outbound	1545	1811	490	24.08%
8.15.03	inbound	1593	6213	1256	44.09%
	outbound	2649	4695	1737	39.60%
8.14.03	inbound	1579	6287	1240	43.99%
	outbound	2757	4846	1820	39.76%
8.13.03	inbound	1287	7227	1853	59.01%
	outbound	2908	4891	1843	38.79%
8.12.03	inbound	1510	6292	1294	46.15%
	outbound	2712	5033	1884	40.99%
8.11.03	inbound	1577	6285	1285	44.90%
	outbound	2852	4743	1713	37.52%
8.10.03	inbound	719	2403	245	25.41%
	outbound	1141	1550	396	25.76%
8.9.03	inbound	1042	3532	417	28.58%
	outbound	1735	2188	558	24.33%
8.8.03	inbound	1592	6835	1280	44.57%
	outbound	2822	4886	1741	38.15%
8.7.03	inbound	1640	6909	1333	44.84%
	outbound	2786	5073	1921	40.81%
8.6.03	inbound	1729	6604	1350	43.85%
	outbound	2913	5046	1916	39.68%

Table G.1: Table showing the counts collected in Kendall inbound and outbound during the intervention period.

Pre-Intervention Counts

Date	Station	Counts			
		Escalator Counts	Stairs Total	Stairs Up	Stair Percentage
8.4.03	Inbound	1801	5926	1180	39.58%
	Outbound	2651	5315	1798	40.41%
8.3.03	inbound	900	2334	257	22.21%
	outbound	1321	1585	387	22.66%
8.2.03	inbound	1072	3070	336	23.86%
	outbound	1731	2179	458	20.92%
8.1.03	inbound	1768	6224	1100	38.35%
	outbound	2787	5358	1693	37.79%
7.31.03	inbound	1790	6520	1165	39.42%
	outbound	2956	5311	1859	38.61%
7.30.03	inbound	1725	6357	1214	41.31%
	outbound	2842	5201	1745	38.04%
7.29.03	inbound	1741	6264	1171	40.21%
	outbound	2748	5295	1834	40.03%
7.28.03	inbound	798	7333	2220	73.56%
	outbound	2678	5070	1722	39.14%
7.27.03	inbound	760	2581	387	33.74%
	outbound	1329	1657	367	21.64%
7.26.03	inbound	940	3795	579	38.12%
	outbound	1734	2175	509	22.69%
7.25.03	outbound	3008	5191	1718	36.35%
7.24.03	outbound	2884	5253	1757	37.86%

Table G.2: Table showing the counts collected in Kendall inbound and outbound during the baseline period.

Appendix H

Annotation Software

To facilitate the manual annotation of the large amounts of test video collected, an annotation application was developed (see Fig. H-1). The application allows the user to count people moving up the stairs, down the stairs, and up the escalator. There are three buttons corresponding to each possibility, and each button updates a queue that records the frame number on which the button is clicked. The buttons are also accessible by keyboard shortcuts: 'a', 's', and 'd'. Additionally the video can be paused and un-paused by using the pause button or the spacebar. A speed control that ranges from 1/4 to 5x real time is controlled by a vertical slider on the control panel. The '[' key and ']' allow the speed to be decreased or increased respectively by using keyboard commands. Finally, the slider along the bottom of control panel allows a user to rewind the video back to a previous point if an error in annotation is made. All counts are reset to to what they were at the point to which the video was rewound. The slider does not move forward in a video file, so care should be taken to not rewind to far back.

Due to limitations in the Java Media Framework, test video must be split into segments of less than 20 minutes each when working with this video format. To compensate for this, a full day of video needs to adhere to the following naming convention: video.avi, video-2.avi, video-3.avi, ... When opening a video file for annotation the only file that should be opened is the first file in the sequence (the one without a part number). Based on the last count recorded, the software will

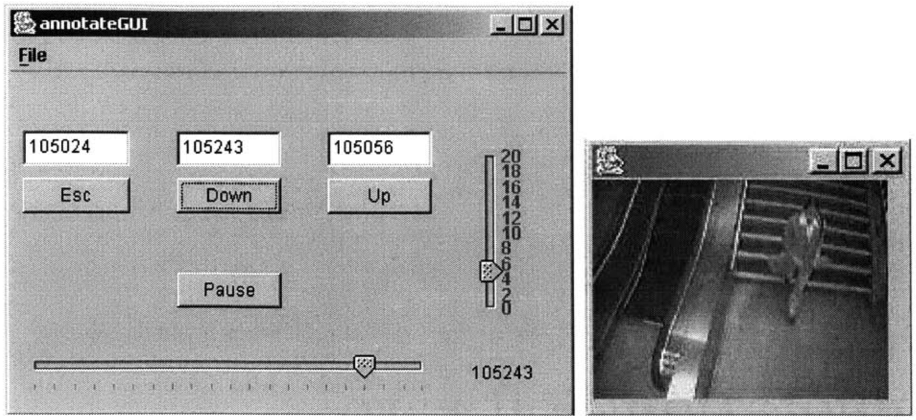


Figure H-1: Video annotating software. The control panel on the left has speed and position sliders, buttons to annotate people going up the escalator, down the stairs, and up the stairs, and a pause button. The panel on the right shows the video.

automatically move to the correct frame in the correct video file.

The actual count queues are stored in the c:\ directory with the names "upcounts.txt," "downcounts.txt," and "esccounts.txt." These files are text files where each line lists a frame number that a person was seen. The software always appends to any file with one of the previous three names. To start a new day of annotation, these files must be moved from the c:\ directory. The next time the program is started, it will create new files to store the annotation data.

Appendix I

Data Collection Procedures

I.1 Communicating with the Laptop

The laptop in the transit station is equipped with a 802.11b wireless card. The card is setup to use the ad-hoc network "KendallT." It is assigned a static ip address (192.168.100.100 for Kendall inbound and Davis, and 192.168.100.102 for Kendall outbound). WinVNC [1] was used for remote access of the laptop. Additionally a folder entitled "Subway" was selected to store data files and images. It is made available to the network as a Windows shared folder.

I.2 Test Video

The test video was collected using the Microsoft Amcap video capture suite. In all cases, the camera was set to 160x120 resolution and captured video at a frame rate of 15 fps. The camera was set to capture in the 24-bit RGB format and store video files in uncompressed RGB24 format. The capture was set to last for 61200 sec or 17 hours. It was started between 6 and 6:30 am for each day of captured video. The video was captured directly to an external USB2.0 drive that could easily be taken in and out of the housing and brought to the lab for manual annotation.

[1] http://www.uk.research.att.com/vnc/

I.3 Daily Counts

All the daily counts and image capture files are saved to the directory "C:\subway". To download the files from the laptop, a researcher simply brings another laptop equipped with 802.11b wireless to the transit station. This laptop can be configured to use the ip address 192.168.100.101 over the wireless ad-hoc network "KendallT". To access the files, the researcher needs to just copy them from the shared folder to their own laptop. The data files can later be deleted from the laptop in the transit station using WinVNC.

Appendix J

Letter from the MBTA

Massachusetts Bay Transportation Authority

Mitt Romney *Governor* · *Kerry Healey* *Lt. Governor* · *Daniel A. Grabauskas* *Secretary and MBTA Chairman* · *Michael H. Mulhern* *General Manager*

Stephen Intille
Massachusetts Institute of Technology
1 Cambridge Center, 4th floor
Cambridge, MA 02142

June 23, 2003

Dear Mr. Intille:

This is to certify that the Massachusetts Bay Transportation Authority authorizes the bearer of this letter to install and maintain equipment in Kendall Station Southbound, Kendall Station Northbound, and Davis Station Outbound for your project.

All work will be done in the unpaid area of the station and maintenance of this equipment should be done as to not interfere with the normal flow of customers both in and out of the station.

If additional information is required, I can be reached at 617-222-4752.

John J. McSweeney, Jr.
Director of Operations Support

cc: C. A. Terzakis
M. Shirkus
S. Wolfson

Massachusetts Bay Transportation Authority, 500 Arborway, Jamaica Plain, MA 02130

Figure J-1: Letter from the MBTA giving permission to perform experiment in Kendall inbound, Kendall outbound, and Davis.

Appendix K

Data Set Formats

K.1 Annotation Data Format

Results from manually annotated video are stored in text files. For each video there are three text files: upCounts.txt, downCounts.txt, and escCounts.txt. The files keep track of people moving up the stairs, down the stairs, and up the escalator respectively. Each file consists of a single number per line. Each number represents a frame number that someone took the stairs or escalator. The number of lines in a file is equal to the number of people counted for that category.

K.2 Escalator File Format - AVI

The output for escalator data on an avi file is a text file of the following format. Each entry in the text file consists of 4 lines of data:

- Line 1: Frame number
- Line 2: Number of difference image pixels highlighted in that time segment
- Line 3: The running average of pixels per person
- Line 4: The number of people on the escalator for the entire video up to this frame number

There is no special spacing between entries, so the first four lines are the first entry, the second four lines are the next entry, etc.

```
18848
898
180.11111
8.0
25499
140
176.1
9.0
```

K.3 Escalator File Format - Live Video

For live video there are only 2 lines per entry:

- Line 1: Time stamp
- Line 2: The number of people on the escalator during that time period

```
7.30.03 7.30.46 AM EDT
15.947148
7.30.03 7.39.02 AM EDT
13.393211
7.30.03 7.47.28 AM EDT
22.664145
```

K.4 Stair File Format - AVI

The output for escalator data on an avi file is a text file of the following format. Each entry in the text file consists of 6 lines of data:

- Line 1: Frame number

- Line 2: Number of difference image pixels highlighted in that time segment
- Line 3: The percentage of difference pixels moving up in the frame
- Line 4: The running average of pixels per person
- Line 5: The number of people in total on the stairs during the entire video up to this frame number
- Line 6: The number of people moving up the stairs during the entire video up to this frame number

There is no special spacing between entries, so the first six lines are the first entry, the second four lines are the next entry, etc.

```
53831
126
0.022727273
120.2
9.008364
2.1439137
77627
324
0.057471264
127.0
11.024112 2.2597613
```

K.5 Stair File Format - Live Video

For live video there are only 3 lines per entry:

- Line 1: Time stamp
- Line 2: The number of people total on the stairs during that time period

- Line 3: The number of people moving up the stairs during that time period

```
7.30.03 7.35.42 AM EDT

42.05064

18.889101

7.30.03 7.45.25 AM EDT

51.9684

16.703573
```

Appendix L

Matlab Simulation Code

This is the code used to perform a simulation of the experiment and determine p-values for a set of data. To run the code, simply provide the number of people who passed through the station during baseline phase, the number of people during the intervention phase, the stair use percentage for baseline, the stair detection rate of the algorithm for that station, the escalator detection rate of the algorithm for that station, the difference in average stair use percentage between baseline and intervention, and the number of trials to perform. A p-value is returned.

```
function p = simulatesignificance(baselineObs, interventionObs,
                                  baselinePercent, stairPercent,
                                  escPercent, percentdiff, numTrials)

%function p = simulatesignificance(baselineobs, interventionobs,
                                   baselinepercent, stairpercent,
                                   escpercent, percentdiff, numtrials)
%performs a simulation and computes a p-value for
%the results of a stair intervention with the camera system

%baselineobs -     number of people total in baseline observations
```

```
%                    (should scale up to reflect real number, not observed number)

%interventionobs - number of people total in intervention observations
%                    (should scale up to reflect real, not observed number)

%baselinepercent - the percentage of people taking the stairs
%                    during baseline (0-1)

%stairpercent -     the percentage of people on the stairs observed
%                    by the algorithm (from video tests)

%escpercent -       the percentage of people on the escalator observed
%                    by the algorithm (from video tests)

%percentdiff -      the difference in stair use percentage from
%                    baseline to intervention

%numtrials -        the number of trials to run (probably around 100,000 is good)

%p -                the p-value for the simulation

%diffs -            the vector containing the stair use percentage difference
%                    for each trial


% reset the state of the random generator
rand('state',sum(100*clock))

baselineUsage = zeros(1); interventUsage = zeros(1);
```

```
diffs = zeros(1);

for x = 1:numTrials

    % randomly select which people take the stairs
    % or escalator using the baseline percent
    basePeople = rand(1,baselineObs);
    interventPeople = rand(1,interventionObs);
    baseStairPeople = find(basePeople <= baselinePercent);
    baseEscPeople = find(basePeople > baselinePercent);
    interventStairPeople = find(interventPeople <= baselinePercent);
    interventEscPeople = find(interventPeople > baselinePercent);

    % now figure out how many of stair people
    % were actually observed
    basePeople(baseStairPeople) = rand(1,length(baseStairPeople));
    interventPeople(interventStairPeople) =
         rand(1, length(interventStairPeople));
    observedbaseLineStairPeople =
         find(basePeople(baseStairPeople) <= stairPercent);
    observedInterventStairPeople =
         find(interventPeople(interventStairPeople) <= stairPercent);

    % now figure out how many of esc people were actually observed
    basePeople(baseEscPeople) = rand(1,length(baseEscPeople));
    interventPeople(interventEscPeople) =
         rand(1, length(interventEscPeople));
    observedbaseLineEscPeople =
         find(basePeople(baseEscPeople) <= escPercent);
    observedInterventEscPeople =
```

```
        find(interventPeople(interventEscPeople) <= escPercent);

    baselineUsage(x) = length(observedbaseLineStairPeople) /
                           (length(observedbaseLineStairPeople) +
                            length(observedbaseLineEscPeople));
    interventUsage(x) = length(observedInterventStairPeople) /
                           (length(observedInterventStairPeople) +
                            length(observedInterventEscPeople));
end

diffs = interventUsage - baselineUsage;

p = length(find(abs(diffs)>= percentdiff)) / numTrials;
```

Bibliography

[1] B.E. Ainsworth, W.L. Haskell, M.C. Whitt, M.L. Irwin, A.M. Swartz, S.J. Strath, W.L. O'Brien, Jr. Bassett, D.R., K.H. Schmitz, P.O. Emplaincourt, Jr. Jacobs, D.R., and A.S. Leon. Compendium of physical activities: An update of activity codes and MET intensities. *Medicine & Science in Sports & Exercise*, 32(9 Suppl):S498–504, 2000.

[2] A. Albiol, I. Mora, and V. Naranjo. Real-time high density people counter using morphological tools. *IEEE Transactions on Intelligent Transportation Systems*, 2(1524-9050):204–218, 2001.

[3] R.E. Andersen, S.C. Franckowiak, J. Snyder, S.J. Bartlett, and K.R. Fontaine. Can inexpensive signs encourage the use of stairs? Results from a community intervention. *Annals of Internal Medicine*, 129:363–9, 1998.

[4] D. Ballantyne, A. Clark, G.S. Dyker, C.R. Gillis, V.M. Hawthorne, D.A. Henry, D.J. Hole, R.M. Murdoch, T. Semple, and G.M. Stewart. Prescribing exercise for the healthy: assessment of compliance and effects on plasma lipids and lipoproteins. *Health Bulletin*, pages 169–176, 1978.

[5] D. Beymer and K. Konolige. Real-time tracking of multiple people using stereo. In *Proceedings of the IEEE Frame Rate Workshop*. Corfu, Greece, 1999.

[6] A. Blamey, N.M. N, and T. Aitchison. Health promotion by encouraged use of stairs. *British Medical Journal*, pages 289–90, 1995.

[7] K.N. Boutelle, R.W. Jeffery, D.M. Murray, and M.K. Schmitz. Using signs, artwork, and music to promote stair use in a public building. *American Journal of Public Health*, 91(12):2004–6, 2001.

[8] K.D. Brownell, A.J. Stunkard, and J.M. Albaum. Evaluation and modification of exercise patterns in the natural environment. *American Journal of Psychiatry*, 137:1540–5, 1980.

[9] K.J. Coleman and E.C. Gonzalez. Promoting stair use in a US-Mexico border community. *American Journal of Public Health*, 91(12):2007–9, 2001.

[10] F. Ennesser and G. Medioni. Finding Waldo, or focus of attention using local color information. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 711–712. New York, NY, 1993.

[11] B.J. Fogg. Persuasive computers: perspectives and research directions. In *Proceedings of CHI 98*, pages 225–232. 1998.

[12] E.S. Geller. A delayed reward strategy for large-scale motivation of safety belt use: a test of long-term impact. *Accid. Anal. and Prev.*, 16:457–463, 1984.

[13] E.S. Geller, R.A. Winett, and P.B. Everett. *Saving the Environment.* New York, 1982.

[14] S.J. Guastello. Do we really know how well our occupational accident prevention programs work? *Safety Science*, 16:445–463, 1993.

[15] I. Haritaoglu and M. Flickner. Detection and tracking of shopping groups in stores. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages I–431–8 vol.1, 2001.

[16] K. Hashimoto, K. Morinaka, N. Yoshiike, C. Kawaguchi, and S. Matsueda. People count system using multi-sensing application. In *International Conference on Solid State Sensors and Actuators*, volume 2, pages 1291–1294, 1997.

[17] M.K. Hunt, R.C. Lefebvre, M.L. Hixson, S.W. Banspach, A.R. Assaf, and R.A. Carleton. Pawtucket Heart Health Program point-of-purchase nutrition education program in supermarkets. *American Journal of Public Health*, 80(6):730–2, 1990.

[18] S.S. Intille. Designing a Home of the Future. *IEEE Pervasive Computing*, April-June:80–86, 2002.

[19] J. Kerr, F. Eves, and D. Carroll. Encouraging stair use: stair-riser banners are better than posters. *American Journal of Public Health*, 91(8):1192–3, 2001.

[20] J. Kerr, F. Eves, and D. Carroll. Six-month observational study of prompted stair climbing. *Preventive Medicine*, 33(5):422–7, 2001.

[21] J. Komaki, K.D. Barwick, and L.R. Scott. A behavioral approach to occupational safety: pinpointing and reinforcing safe performance in a food manufacturing plant. *J. Appl. Psychol.*, 63:434–445, 1978.

[22] C. Pinhanez. The everywhere displays projector: a device to create ubiquitous graphical interfaces. In G.D. Abowd, B. Brumitt, and S.A.N. Shafer, editors, *Proceedings of the Conference on Ubiquitous Computing*, LNCS 2201, pages 315–331. Berlin Heidelberg, 2001.

[23] M.E. Pollack. Planning technology for intelligent cognitive orthotics. In *6th International Conference on AI Planning and Scheduling*, 2002.

[24] T. Richardson, Q. Stafford-Fraser, K. Wood, and A. Hopper. Virtual Network Computing. *IEEE Internet Computing*, 2(1):pp. 33–38, 1998.

[25] W.D. Russell, D.A. Dzewaltowski, and G.J. Ryan. The effectiveness of a point-of-decision prompt in deterring sedentary behavior. *American Journal of Health Promotion : AJHP*, 13(5):257–9, ii, 1999.

[26] W.D. Russell and J. Hutchinson. Comparison of health promotion and deterrent prompts in increasing use of stairs over escalators. *Perceptual and Motor Skills*, 91(1):55–61, 2000.

[27] A.J. Schofield, T.J. Stonham, and P.A. Mehta. A RAM based neural network approach to people counting. In *Fifth International Conference on Image Processing and its Applications*, pages 652–656. IEE, 1995.

[28] J. Segen and S.G. Pingali. A camera-based system for tracking people in real time. In *Proceedings of the 13th International Conference on Pattern Recognition*, volume 3, pages 63–67, 1996.

[29] C. Seligman, L. Becker, and J.M. Darley. Behavioral approaches to residential energy conservation. *Energy and Building*, 1978.

[30] G. Sexton, X. Zhang, D. Redpath, and D. Greaves. Advances in automated pedestrian counting. In *IEE ECOS 95*, volume IEE publication No. 408, pages pp 106–110, 1995.

[31] A. Shio and J. Sklansky. Segmentation of people in motion. In *Proceedings of the Workshop on Visual Motion*, pages 325–332. 1991.

[32] O.o.t. Surgeon General. Physical Activity and Health: A Report of the Surgeon General. Technical report, 1996.

[33] H.L. Taylor, E.R. Buskirk, and R.D. Remington. Exercise in controlled trials of the prevention of coronary heath disease. In *Fed Proc*, volume 32, pages 1623–1627, 1973.

[34] K. Terada, D. Yoshida, S. Oe, and J. Yamaguchi. A method of counting the passing people by using the stereo images. In *International Conference on Image Processing*, volume 2, pages 338–342, 1999.

[35] A. Tesei, A. Teschioni, and C.S. Regazzoni. "Long-Memory" matching of interacting complex objects from real image sequences. In *Time Varying Image Processing and Moving Object Recognition*, Firenze, 1996.

[36] S. Titze, B.W. Martin, R. Seiler, and B. Marti. A worksite intervention module encouraging the use of stairs: results and evaluation issues. *Sozial-und Praventivmedizin*, 46(1):13–9, 2001.

[37] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers. Wallflower: principles and practice of background maintenance. In *Proceeding of the International Conference on Computer Vision*, pages 255–261. Corfu, Greece, 1999.

[38] R.C. Winkler and R.A. Winett. Behavioural interventions in resource management: A systems approach based on behavioral economics. *American Psychologist*, 1982.