# Detecting User Intention at Public Displays from Foot Positions

## Problem and Motivation

Imagine an interactive display at a bus station that always provides content users want to see without requiring explicit input. For example, it might provide quick facts to hasty passersby or entertainment to a waiting user. Public display research aims to use contextual awareness to effectively communicate content using public displays, to catch users' attention, and to motivate more users to interact with the display, all while dealing with the public nature of interaction. Recent research suggests that personal device design paradigms are ineffective for the distracted users of public displays, which would explain the low usage rate of currently deployed systems [1].

Predicting users' intentions is a powerful tool for satisfying their needs. Body language detection can be used to automatically interpret user behavior and react accordingly; gaze can be used to predict the level of attention, or proxemics can be used to measure user awareness. However, public display users' complex behaviors, for example inobservance or maintaining social roles, pose new challenges to sensing and interpreting human behavior. For example, distractions while in public can make gaze insufficient to indicate the focus of a user's attention.

Feet have not been highlighted in public display interactions, even though they richly embody expressivity of people's interests when interacting with other people [2, 3]. Because foot positions are an unconscious expression of body language and thus unaffected by distractions, as a measure of interest they are more robust in a public environment. This work sheds new light on the potential of sensing public display users' intentions by measuring the positions of their feet. Analyzing the positions of users' feet aims at making public displays more adaptive and encouraging more users to interact by enhancing the context-awareness of public displays.

## Background and Related Work

Two steps are necessary for improving interaction between users and the display. First, the human behavior must be carefully investigated in order to construct the public display user model. Second, appropriate sensing technology must be implemented based on the user model.

**User Model Investigation.** Researchers studying interactive public displays are still searching for a common user model that fits their audience. In public display settings, models based on observations in real-work settings, such as the *CityWall* and the *Opinionizer*, focus on the social behavior of the audience [4, 5]. The design objectives mostly focus on attracting, engaging, and motivating the user.

Edward Hall's proxemics theory of interpersonal spatial relationships is widely used to build user models [6]. The user models are derived by defining the users' behaviors and postponing assumptions about the relationship between behaviors and intentions [7]. One of the first proposed systems in this domain was the *Interactive Public Ambient Display*, which interpreted proximity and body orientation as direct signals for display content adaptation [8]. Both modalities were combined to model the users' interest and, with increasing interest, the displayed content became more detailed and private. Similarly, the *Proximity Toolkit* used proximity combined with head orientation to control a multimedia application [9]. However, most of these interaction models have only been tested in laboratory settings using user-mounted technology, which leaves the sensing problem in public settings unsolved.

**Sensing Technology.** A variety of interaction modalities have been developed that use contact-free sensor technology to track user behavior. For example, proximity-sensor arrays were used in a variety of settings to estimate users' interest. Face detection has also been used in vending machines to recommend drinks based on the user's age and gender [10]. Gaze is also promising, and since it is directly related to visual attention it was studied as a measure of attention. *Gaze Tracker* is one such vision-based gaze tracking system, which tracked users' visual attention using eye movement [11]. However, most of the existing gaze-tracking systems require calibration. Foot position tracking is cost efficient and easy to deploy, but previous research on it has focused on mobile input or floor installations [12, 13].

## Uniqueness of the Approach

This project aims to bridge the gap between sensing technology and real-world public display settings by detecting user intention based on the measuring of foot positions. Foot position sensors provide robust, meaningful input for detecting unconsciously expressed user intentions. To reach the goal of context-aware public displays, I have constructed two hypotheses:

1) Public display users' intentions can be measured based on the position and orientation of their feet.
2) The adaptation of the public display content to user intentions, as measured by foot position, helps the display address the user's needs. Such an interactive public display supports fast information searches and motivates users to interact with it.

In the setting for my experiments, visitors either need specific navigational information from the display or they just casually interact with it (without any need for specific information). I distinguish user intentions and perform my experiments in this context.

***Methods***

**Foot Scanner.** A two-dimensional depth map of the floor in front of the display is required to detect foot positioning. A *Hokuyo URG* laser rangefinder was used, because it could be installed at the bottom of the display where it is invisible to the user, and also because it is robust against sunlight. The resulting sensing system enables robust, real-time detection of foot positions out to a radius of 4 meters around the display (see Figure 1b).
The developed algorithm transforms the raw depth data into foot positions. Using background subtraction, three feature points for each foot are determined: the front peak and the two edges (see Figure 1c). These feature points allow for detection of the current angle and proximity of each foot. A technical evaluation showed an angle accuracy of ($\sigma$=7.4°, $\mu$=6.0°) and a foot proximity accuracy of ($\sigma$=4.3cm, $\mu$=3.5cm).

**Display Setup.** The 40-inch multi-touch display with a *PQ-Labs* infrared touch frame attached was placed next to a magazine holder and a water dispenser to draw additional attention from those passing by. The foot scanner was installed at the bottom of the display (see Figure 1b).

**Public Display Content.** The public display application developed for this setup is an information system. It provides a detailed map of the institute's campus and surrounding neighborhood, an entertainment function with a video collection about recent research findings, and a contact list for staff at the institute. Furthermore, the institute's server provided a daily event schedule. A touch trigger symbol is shown as the default screen, to which the system switches back after 15 seconds of inactivity (see Figure 1a). At the bottom of the display, a bar of menu buttons can be used to retrieve the desired content. The application is implemented in *C#* and with *Unity*.

**User Model.** Public display information systems provide information about the local environment to passersby. Depending on the content selected, either the user's motivation lies in an external goal or the tool itself provides a goal. Therefore, the target design of the public display must address two problems. First, users who need information should be able to find it easily; unused functionality must not get in the way. Second, during casual usage, entertaining content should be

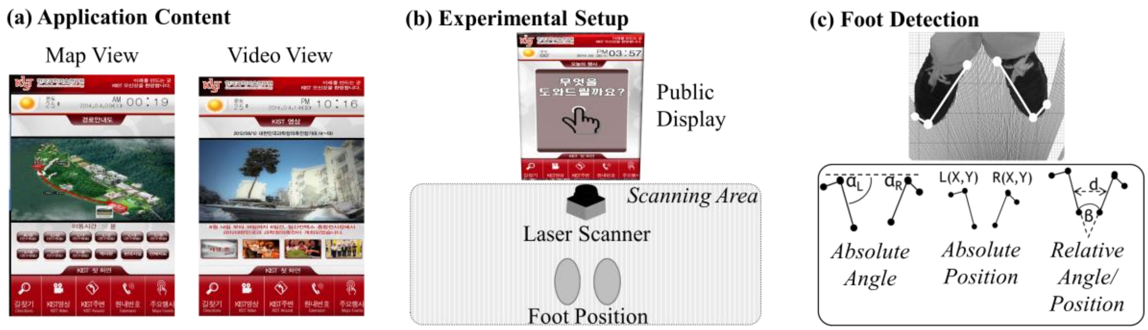


**Figure 1.** (a) Content of public display application. (b) Experiment setup with laser scanner at the bottom of the display. (c) Foot feature points detected by the laser scanner at the front peak and the two edges.

delivered at a lower interaction threshold in order to motivate more users to interact with the system [7]. I pursued an observational study and a case study to test the usefulness of foot position in fulfilling these needs.

## Results and Contributions

### *Observational Study: Foot Position as Indicator of User Intention*

**Experiment Description.** The foot position data, combined with video recordings and the touch log, formed the foundation of the observation results. The resulting dataset consists of two types of user intentions along with the captured foot positions. Foot position features are defined using the angle of each foot and their relative positions. A foot position was logged automatically when it was static and no more than 1.5 meters from the screen. Data from groups of users was sorted out. If a user altered posture during an interaction, the new foot position was logged as new data in the same user category. To simplify the model, the user's understanding of the interface's content and functionality was left out of this study.

User behavior was manually classified into two categories using the video recordings and touch log. If the interaction led to the access of specific information, the interaction was labeled as “information search.” If the user clicked on one of the menu buttons other than navigation, or left without any touch interaction, the interaction was labeled as “casual interaction.”

The labeled data was classified using a support vector machine (SVM) classifier. The SVM configuration was a C-SVC type, using a radial basis function with three degrees. It was trained and tested using a repeated random sub-sampling validation process.

**Results.** I collected data from 115 samples over the course of one calendar week, from Monday to Friday, 9 AM to 7 PM. The average duration of one recorded foot position was 9 seconds, and the average duration of one interaction was 11 seconds. 55 of the 115 samples were classified as “information search” and 50 as “casual interaction.” Figure 2a shows the average of the resulting confusion matrix of the SVM algorithm applied to the 115 samples. The SVM algorithm was trained with feature data (two foot angles and the position) of 30 recorded samples, and tested using the 25 residual samples from information search interactions and 30 samples from casual interactions. The accuracy in detecting user intention using foot position was determined to be 84.4 percent.

**Discussion.** This accuracy is significant and shows that users’ intentions can be detected from the positions of their feet. If a user requests specific information, her feet are orientated towards the screen. If the interaction is casual, her feet are pointing away, indicating an intent to leave. Figure 2b shows frequently captured sample foot positions of interacting users.

**(a) Confusion Matrix**

| | A (Predicted) | B (Predicted) |
|---|---|---|
| **A (Actual)** | 92.0% (23/25) | 8.0% (2/25) |
| **B (Actual)** | 23.3% (7/30) | 76.7% (23/30) |
| | **Accuracy: 84.4%** | |

**(b) Frequently Occurring Foot Positions for each User Intention**


**Figure 2.** (a) Resulting confusion matrix of SVM algorithm applied on labeled foot position data. (b) Frequently occurring foot positions for each user intention. Foot positions during information search tend to be orientated towards the display with a smaller relative angle, whereas during casual interaction users show larger foot orientation angles orientated away from the display.

### *Case Study: Interactive Public Display - Adapting Content to Foot Position*

**Experiment Description.** To inform the applicability of public display content adaptation based on user intentions, I conducted a case study in February 2014, in which the content of a public display was automatically adapted according to the users' foot positions. The system was modified to classify foot positions using the trained SVM algorithm at the beginning of an interaction session. The interaction model was based on adapting content to the foot positioning of users when they walked up. Content was chosen with the dual objectives of optimizing interaction times during information searches and increasing the attention of users during casual usage. Navigation information popped up for users who were classified as needing specific information. An entertainment video collection about recent research findings was shown when casual usage was detected.

In order to interview users, I positioned myself outside of the entrance lobby where the display was located to ensure that users felt unobserved. I approached them only after their interaction with the system was clearly finished. This study resulted in a qualitative analysis of user interviews and captured content- and interaction-related feedback. Interaction time and foot position data were also used as quantitative references.

**Results.** Over the course of ten days, quantitative and qualitative data from 27 users (average age 35, 16 female) was collected. The perceived usefulness of the system was high, and users who needed specific information agreed that they could find the information they needed quickly. During casual usage, the interaction time average significantly increased to 19 seconds, as compared with only 10 seconds in the observational study. 70 percent of users thought the content had been changed based on proximity, indicating the unconsciousness of foot positioning. The accuracy of the user intention classification from foot position was 88.8 percent, similar to the results from the observational study.

**Discussion.** The adapted content was useful for users of both categories. The longer interaction duration and the high content rating during casual usage suggest that users' attention was increased. Their threshold to interact with the system can be reduced by adapting content [4]. The lower interaction time and high perceived usefulness during information searches also suggests that the provided content was appropriate for users in this category.

### *Contribution*

**Summary.** The significant findings of this work show the feasibility of using foot positions to classify the intentions of public display users. The results of the observational study suggest that users' foot positions at public displays are dependent on their intentions. The results of the case study further suggest that the interactive display with content adapted based on user intentions increases engagement and user satisfaction.

**Future Work.** To further enhance this user model, I am planning to integrate other body language indicators. Comparing different indicators of unconscious and conscious interaction modalities will help us further understand user behavior, which is important for further generalization of the user model. The usefulness of the system could also be further enhanced by dynamic usage of the user classification. My goal is to find an application-independent general user model for context-awareness in public displays.

**Conclusion.** These findings make a significant contribution to user intention measurement in public display environments. Because of its unconscious nature, foot position is unaffected by distractions in public spaces. Furthermore, it is simple to sense foot positions using a laser rangefinder tracking system. Finally, user intention classification proved useful for encouraging interaction when showing content to users. Therefore, foot positions are an important modality for detecting the intentions of public display users, and their use could lead to public displays that allow for useful and effective interaction in public.

## References


[1] Huang, Elaine M., Anna Koster, and Jan Borchers. "Overcoming assumptions and uncovering practices: When does the public really look at public displays?." *Pervasive Computing*. Springer Berlin Heidelberg, 2008. 228-243.

[2] Morris, D., Bodywatching: A field guide to the human species. 1985: Cape.

[3] Morris, D., Peoplewatching: The Desmond Morris Guide to Body Language. 2002: Vintage Books.

[4] Brignull, Harry, and Yvonne Rogers. "Enticing people to interact with large public displays in public spaces." *Proceedings of INTERACT*. Vol. 3. 2003.

[5] Peltonen, Peter, et al. "It's Mine, Don't Touch!: interactions at a large multi-touch display in a city centre." *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2008.

[6] Hall, E.T., The hidden dimension. Vol. 6., 1966.

[7] Michelis, Daniel, and Jörg Müller. "The audience funnel: Observations of gesture based interaction with multiple large displays in a city center." *Intl. Journal of Human–Computer Interaction* 27.6 (2011): 562-579.

[8] Vogel, Daniel, and Ravin Balakrishnan. "Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users." *Proceedings of the 17th annual ACM symposium on User interface software and technology*. ACM, 2004.

[9] Nicolai Marquardt,et al. “The proximity toolkit: prototyping proxemic interactions in ubiquitous computing ecologies.” *Proceedings of the 24th annual ACM symposium on User interface software and technology* . ACM, 2011.

[10] Vending Machine Uses Facial Recognition, Retrieved March 26, 2014 from http://geekologie.com/2011/12/vending-machine-uses-facial-recognition.php

[11] Nakanishi, Yasuto, et al. "Vision-based face tracking system for large displays." *UbiComp 2002*: Ubiquitous Computing. Springer Berlin Heidelberg, 2002. 152-159.

[12] Augsten, Thomas, et al. "Multitoe: high-precision interaction with back-projected floors based on high-resolution multi-touch input." *Proceedings of the 23nd annual ACM symposium on User interface software and technology*. ACM, 2010.

[13] Scott, Jeremy, et al. "Sensing foot gestures from the pocket." *Proceedings of the 23nd annual ACM symposium on User interface software and technology*. ACM, 2010.