2012 International Conference on Information and Computer Applications (ICICA 2012)
IPCSIT vol. 24 (2012) © (2012) IACSIT Press, Singapore

Technique of Tangible User Interfaces for Smartphone

Jinsuk Kang+

Jangwee Research Institute for National Defence

Abstract. In this paper, we describe a system for creating natural interaction patterns using tangible objects as input devices. A user's physical interactions with these tangible objects are monitored by analyzing real-time video from a single inexpensive smartphone using variations on common mobile vision algorithms. Limiting tangible object detection to mobile vision techniques provides a non-invasive and inexpensive means to creating user interfaces that are more compelling and natural than those relying on a keyboard and mouse.

Keywords: Tangible User Interfaces, Smartphone, Human Computer Interaction

1. Introduction

In the context of Tangible User Interfaces (TUIs), computer vision is often used for spatial, interactive surface applications because it is capable of sensing the position of multiple objects on a 2D surface in real time while providing additional information such as orientation, color, size, shape, etc. Thus, tag-based computer vision is often used in the development of TUIs. Computer vision TUI systems typically require at least three components: a high-quality camera; a light-weight LCD projector for providing real-time graphical output; and a computer vision software package [1, 2]. In this paper, We describe tree design experiments, implementing interactive systems that explore the technical context of mobile device usage as a potential design target for tangible user interfaces. Also, we utilize a single off-the shelf smartphone camera to passively view a user's interactions with unmarked tangible objects. These interactions are detected and analyzed in near real-time (≈ 10 Hz), and represent application use cases.

2. Tangible User Interfaces

The most common interaction techniques in Human Computer Interaction (HCI) today are based on the familiar mouse and keyboard. While extremely successful, these input devices don't provide natural interfaces to various categories of software. In fact, one may argue that most software developed today cannot live up to its full potential due to the ubiquity of the mouse and keyboard as expected user input devices. In other words, we as software developers and interaction designers are constrained to design simple, often unnatural HCIs [3, 4]. Fig 1 illustrates the model developed in this paper for mobile human computer interaction. The work on this model to this date has drawn from a number of theories and approaches, the most important of which are Model Human Processor (MHP), Distributed Cognition and Activity Theory. Cognitive psychology and information processing theories introduced to the HCI discipline in the early 80s have been the predominant theories ever since. One of the earliest models of HCI is the Card et al [5] Model Human Processor. As Hollan et al [6] explain HCI began as a field at a time when human information processing was the prevailing theory.

+ Corresponding author.
E-mail address: Jskang01@ajou.ac.kr

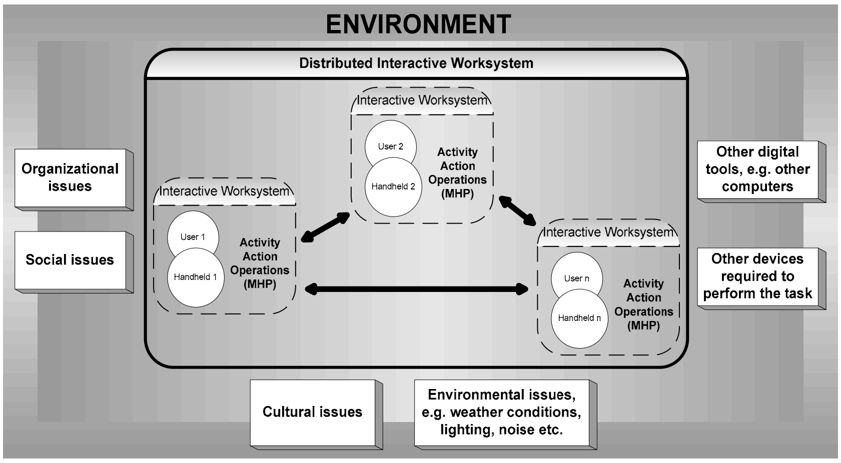
ENVIRONMENT
Distributed Interactive Worksystem
Interactive Worksystem
User 2
Handheld 2
Activity Action Operations (MHP)
Interactive Worksystem
User 1
Handheld 1
Activity Action Operations (MHP)
Interactive Worksystem
User n
Handheld n
Activity Action Operations (MHP)
Organizational issues
Social issues
Other digital tools, e.g. other computers
Other devices required to perform the task
Cultural issues
Environmental issues, e.g. weather conditions, lighting, noise etc.

Fig 1. Mobile interaction model for handheld computer

3. Smartphone Based Object Recognition Model

This approach stores a single object model regardless to the view variations. The invariants depend on the type of camera, such as affine camera and projective camera. The most well-known invariant in projective camera is 5 points invariant based on cross-ratio for planar shapes. This scheme can reduce the search space using efficient indexing, but it is only applicable to well-segmented 2D objects. In contrast, the 3D interface provides visual support for the operator's comprehension of the camera orientation by rendering the image from the smartphone at an angle that corresponds to the camera orientation as shown in Fig 2.

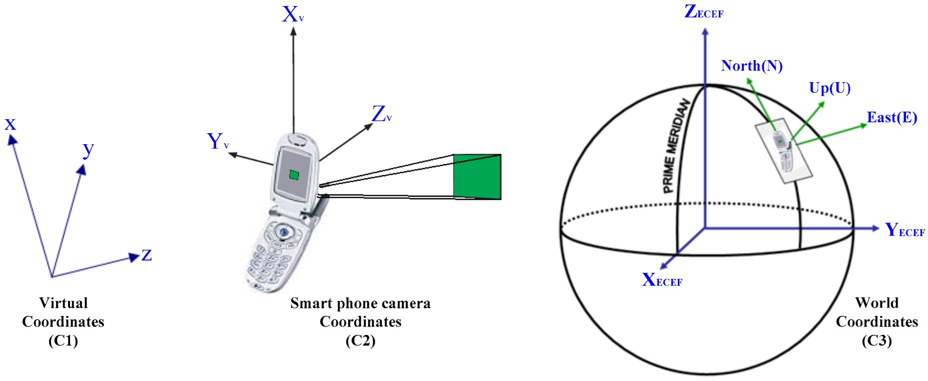

Fig 2. Visualizing the orientation of the smartphone camera using the TUI, (C1) Virtual coordinates, (C2) Smartphone camera coordinates, and (C3) World coordinates.

For this TUIs, each of the captured image sequences has been geo-referenced by using GPS/IMU integrated positioning device. The orientation parameters of each camera coordinate origin are determined with respect to a global coordinate system. By using techniques of photogrammetric intersection, the position of 3D object relative to the camera coordinates is achieved. To eventually calculate the world coordinates of the sign, the following coordinate transformation needs to be implemented, including C1, C2, and C3 coordinate systems.

- Virtual coordinates C1 : This coordinate system is established during the calibration process. The origin of this coordinates is the origin selected in the calibration board. The points used in the calibration process are represented in this coordinate. Consequently, the location of the road sign is first represented in C1 coordinates from the 3D positioning function. The calibration board is purposely put in a plane perpendicular to the smart-phone's y-direction (longitudinal direction). This is to exclude the rotation between the smartphone coordinates and the virtual coordinates. The offsets $(\Delta x, \Delta y, \Delta z)$ between the origin in the virtual coordinates and the origin in the smartphone coordinates are measured during the calibration process.
- Smartphone camera coordinates C2 : While the SmartPhone is moving, its position is determined by the positioning sensors. The origin of the smartPhone coordinates is a fixed point in the smartphone.

For simplification, the location of the GPS receiver is set as the origin. The Y-axis is the forward direction (longitudinal) and the X-axis is point to the passenger's side (transverse). Once the sign location is obtained in the virtual coordinates C1, the task is then to convert it to the smartphone coordinates C2. For example, if point P in the space can be represented as (x_1, y_2, z_3) in the C1 coordinates. It can be converted into C2 coordinates using the following equations:

$$
\begin{aligned}
X_v &= x_1 - \Delta x \\
Y_v &= -(z_1 - \Delta z) \\
Z_v &= (y_1 - \Delta y)
\end{aligned}
\qquad (1)
$$

- World coordinates C3 : With the heading, roll and pitch provided by IMU, the coordinates of point P, (X_v, Y_v, Z_v) in smartphone coordinates can be easily converted to (X'_v, Y'_v, Z'_v) in the ENU (local east, north, up) coordinates. The ENU coordinates can then be converted to ECEF (Earth Centered Earth Fixed) coordinates (X, Y, Z) using Equation (2).

$$
\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} -\sin\lambda & -\sin\phi\cos\lambda & \cos\phi\cos\lambda \\ \cos\lambda & -\sin\phi\sin\lambda & \cos\phi\sin\lambda \\ 0 & \cos\phi & \sin\phi \end{bmatrix} \begin{bmatrix} X'_v \\ Y'_v \\ Z'_v \end{bmatrix} + \begin{bmatrix} X_G \\ Y_G \\ Z_G \end{bmatrix} \qquad (2)
$$

Whereas (X_G, Y_G, Z_G) is the ECEF coordinates of the GPS receiver, i.e. the origin of the smartphone coordinates or the ENU coordinates. λ and ϕ are the geodetic longitude and latitude. ECEF coordinates can be further converted to geodetic coordinates if needed. In the current implementation, the determination of relative distances and sizes of objects in the image pairs is an operation dependent only on the smartphone cameras and its calibration.

4. Snapshot and Semantic Maps Technology

- Smartphone Camera Calibration : Smartphone camera calibration provides us with information about the intrinsic and extrinsic parameters of a camera. The intrinsic parameters are parameters that describe the camera's internals including the focal length, the distortion coefficients of the camera's lens, and the principal point (which is usually the center of the image). The extrinsic parameters define the position and orientation of the camera relative to some world coordinate frame. While the intrinsic parameters of a camera vary from camera to camera, they need only be found once per camera. The extrinsic parameters of a camera are view-dependent. A common technique used to calibrate a camera is to find a number of image-world point correspondences. An object of known geometry such as a book cover pattern is detected in image space. In fact, the book cover pattern as in Fig 3 is the most common calibration object used today.

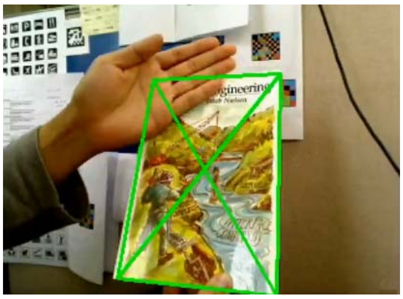

Fig 3. Book cover calibration object detected with point correspondences highlighted

- Snapshot Technology : The idea behind snapshot is that visual images contain a lot of information that is understandable by a human, but not necessarily a computer. In smartphone tasks that involve object recognition and recollection, it is important to aid the user by storing relevant information in the interface, rather than forcing the user to mentally store the information. Consider the case of navigating a smartphone through an environment looking for objects. Suppose that, at the end of the navigation, the operator is required to tell an administrator, for example, where all of the blue boxes in the environment are located. If the environment is sufficiently large, the operator will likely forget where some of the objects were located. To aid the user in search and identification tasks, we have created snapshot technology. Snapshots are pictures that are taken by the smartphone and stored at the corresponding location in a map.

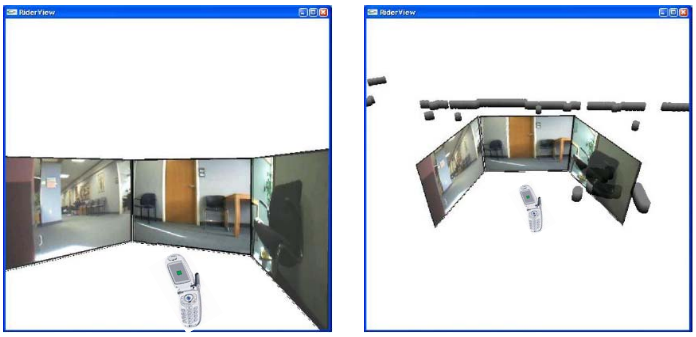

Fig 4. Using snapshots to remember information

5. Performance Evaluations and Discussion

The purpose of this experiment is to compare how well a prototype 2D interface and 3D interface support the operator in a search task where a smartphone zoom camera is used. In particular we are interested in how quickly operators can complete the task and how aware they are of the smartphone and its environment. We hypothesized that the Smartphone Zoon camera is more useful for performing a search task when operators use the 3D interface as compared to a 2D interface. In this experiment, we found that using a movable camera did not improve the time to complete a search task when using the 2D interface. This result is somewhat surprising considering that the environment for the experiment was designed to exploit use of a smartphone camera. The reason for the similar performance with the 2D interface is that operators spend a smaller percentage of their time moving the smartphone (which results in a lower average velocity) when a movable camera is available. With the 3D interface, participants were able to finish the task somewhat faster when using a movable camera because they spent a larger portion of their time navigating the smartphone, even while using the camera. This led to a faster average velocity.

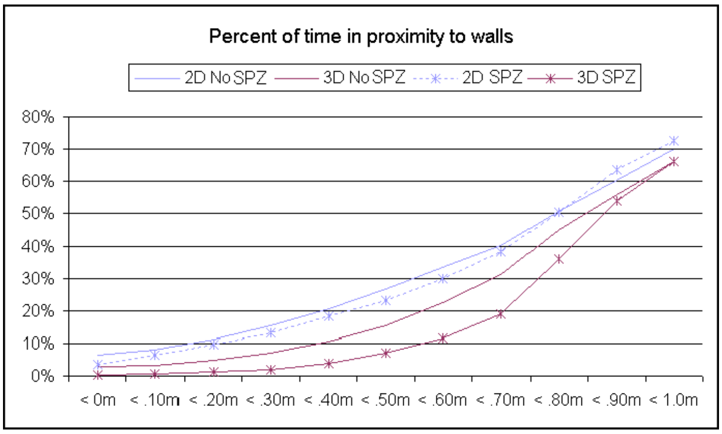

Fig 5. The percentage of time the smartphone is within a given distance of a object

6. Acknowledgments

This research was supported by the Basic Science Research Program through the NRF funded by MEST (No.2011-0026536) and (No. 2011-0015329).

7. References

[1] R. Aish, “3D input for CAAD systems,” Computer Aided Design, Vol.11, No.2, pp.66-70, 1979.

[2] E. Sharlin, B. Watson, Y. Kitamura, F. Kishino, and Y. Itoh., "On tangible user interfaces, humans and spatiality," Personal and Ubiquitous Computing, Vol.8, No.5, pp.338-346, 2004.

[3] Adam Greenfield, "Everyware: The Dawning Age of Ubiquitous Computing," New Riders Publishing, 2006.

[4] Jane Fulton Suri and Ideo, "Thoughtless Acts: Observations on Intuitive Design," Chronicle Books, 2005.

[5] Stuart K. Card, Thomas P. Moran and Allen Newell, "The Psychology of Human-Computer Interaction," Lawrence Erlbaum Associates; New edition, February, 1986.

[6] James Hollan, Edwin Hutchins and David Kirsh, "Distributed Cognition : Toward a New Foundation for Human-Computer Interaction Research," Journal of ACM Transaction on Computer-Human Interaction, Vol.7, pp.174-196, 2000.