Applications of Vision-Based Attention-Guided Perceptive Devices to Aware Environments

Bogdan Raducanu and Panos Markopoulos

Faculty of Industrial Design, Technical University of Eindhoven
Den Dolech 2, PO Box 513, 5600MB Eindhoven, The Netherlands
{b.m.raducanu, p.markopoulos}@tue.nl

Abstract. This paper discusses a computer vision based approach for enhancing a physical environment with machine perception. Using techniques for assessing the distance and orientation of a target from the camera, we detect user presence and estimate whether an object of interest is the focus of attention of the user. Our solution uses low-cost cameras and is designed to be robust to lighting variations typical of home and work environments. We argue why this approach is a useful component for the incremental construction of aware environments and discuss some practical applications of using such a system.

1 Introduction

1.1 Overview

A key component of the Ambient Intelligence vision [1] is the capability to interact with a computational environment in natural and personalised ways. One way to achieve naturalness is by means of *implicit input* [2], Implicit input entails that our natural interactions with the physical environment provide sufficient input to a variety of non-standard devices without any further user intervention. Such automatically captured input contrasts the current model of interaction where the user has to perform several secondary tasks relating to the operation of the computing device in order to achieve their primary task. Attaining this capability involves in endowing everyday objects with machine perception capabilities. By machine perception we refer to the whole class of sensing and pattern recognition techniques that can be deployed to sense and interpret aspects of activities taking place within the physical environment of interest.

Potentially, machine perception offers significant advantages to users. It can help disambiguate input, e.g., knowing who's talking to whom, which display is attended to so as to route system output, etc. Machine perception can be achieved with a variety of technologies (pressure sensors, video cameras, radio-frequency tags, fingerprint readers), which are integrated in objects commonly found in our environment (chairs, tables, displays). Each of these technologies has their respective advantages and

disadvantages concerning robustness, complexity, costs and the type of inferences that can be made about user activity. This paper discusses some applications of computer vision techniques to support implicit input in ubiquitous computing environments. In particular, we discuss the detection of person proximity to an object of interest and whether this person faces towards the object. This helps to estimate whether that objects becomes user's focus of attention. The intuitive idea that people will face objects they are interested in is supported by some empirical research at Microsoft [4]. Computer vision offers several advantages over competing technologies (e.g., movement sensors, ultrasound) to answer this question.

1.2 Indoor Person Presence Detection

The problem of indoor person presence detection can be addressed at different scales and varying resolutions [3]. At a building level, one might be interested to know which room people are in, e.g. for supporting an Intercom type of application [12]. At a room level, we might wish to broadly detect which part of the room a person is at (near the window, door, table). At sub-room level we might want to know more precisely the coordinates of a person, or whether this person is directing his attention to a particular object of interest. Depending on the type of detection we address, several ranges of error are tolerable. For example, an error of a few meters is considered a good approximation for presence detection at building level, while for other tasks, e.g., whether a user is attending to a small display, the error cannot be greater than a few centimetres.

Current solutions for building level detection are based on wireless communication devices. A classical demonstration of location awareness is the Active Badge system [15] developed at the Olivetti research labs, based on infrared emitting badges. In [3], they present a system that relies upon RF identifiers placed in the shoes of users and floor-mat antennas. Based on a history of information recorded by the receiver the system can estimate whether the person is in a room or not. A more recent method based on ultrasound active badges is presented in [10].

For room-level presence detection, in [13] they created a "smart floor", whose purpose is to identify and track the people stepping on it. The floor contains force measuring load cells. In [3], [7] and [9] they use computer vision in order to track several persons in real-time. The persons' location in the room is established based on a combination of knowledge of the cameras' relative location, fields of view of the cameras and heuristics on the movements of people. Compared to computer vision based solutions, such technologies are robust to varying lighting conditions but do not help with verifying focus of attention (where a person is facing to).

Several computer vision based techniques have been proposed for sub-room level detection. These rely either on localizing a badge carried by the user [8] or on face localization [5], [11] and [14].

This paper discusses a method for estimating the attention of a person towards an object, by detecting whether a special badge carried by the person is facing towards the camera and whether the distance between the badge and the camera is below a certain threshold. It is assumed that the camera is placed in such a position in order to

optimise the interaction between the user and the device that is attached to. We want to avoid situations in which, for objects of very large dimensions (like wall-mounted displays, for instance), when the user is facing towards the object the badge cannot be visible by the camera.

The paper is structured as follows. Section 2 describes the system developed. In section 3 we discuss some potential applications that are currently under way. Finally, in section 4, we will present our conclusions and draw some guidelines for future work.

2 System Descriptiom

2.1 Overview

Our system consists of two components (see figure 1):

- a perceiving component, represented by a Logitech™ webcam with an infra-red filter and an array of infra-red LEDs placed around the camera in form of a ring;
- a passive component, represented by a badge, which has reflector tape patches attached on it.

The role of the filter is to let pass only those frequencies of the light that are close to the infrared rays spectrum. By using this kind of filter, the camera will perceive mostly the light that is reflected from these reflector patches. In consequence, background information is discarded from the beginning, making the image analysis process much simpler.

(a) (b)

Fig. 1. Experimental hardware setup: (a) a Logitech webcam provided with an IR-light source and IR filter and (b) the first author wearing the target with IR reflector material and facing the camera

The badge is a rectangle made of rigid paper and has attached patches of reflector tape (in shape of discs), on each of its four corners. A very important property of this material is that it reflects the infrared light back, on the same direction it came from the source (infrared LEDs). The size of the badge is 10x15 centimetres and the discs have a diameter of 2 centimetres.

2.2 Description of the Method

The standard approach for estimating the 3D coordinates of a point in space, a stereo vision system is needed. Thus, the 3D coordinates are estimated based on the pixel disparity, i.e. the difference in object pixels' location in the two images. One of the disadvantages of using a stereovision system (besides its higher cost and necessity for specific hardware) is that accidental changes in the orientation of one of the cameras require a recalibration of the whole system. This makes stereo-vision based solution insufficiently robust for dynamic environments like the home or the office.

Here we explore an alternative, i.e. to estimate the 3D coordinates based on the information about the points and lines whose perspective projection we observe. Such relations with the perspective geometry constraints can often provide enough information to uniquely determine the 3D coordinates of the object. Such knowledge can come about when we have a model of the object being viewed in the perspective projection. The technique to inference the 3D point coordinates when knowing its 2D coordinates on the image plane of the camera is called "inverse perspective projection". In our case, we use a technique that is described in [6], which allows the 3D reconstruction based on the observed perspective projection of two parallel line segments (the lines connecting the centres of the two patches situated along the vertical edge of the badge, for instance). This method does not use any information regarding camera's absolute orientation in the scene. We always express the relative position of the person with respect to the camera. In consequence, small modifications in the camera position will not affect system's performance.

2.3 System's Performance

Since we looked for a low-cost solution to our problem, we tested the algorithm on a PC of modest performances with 128 MB of RAM and a processor of 730 KHz. We set the frame rate of the webcam at 10 frames/sec, which is acceptable for real-time tracking.

The experiments performed were intended to assess the accuracy of the distance measured by the proposed algorithm. In the case of infrared technology, the only factor that can affect the performances of the algorithm described above is the amount of infrared radiation presented in the environment. Empirical experiments demonstrated the robustness of our system in case of diffused natural light, considered during different moments of the day, and also artificial light. In our view, these are the most likely scenarios in an office or home environment. We found that only direct sunlight, presented in the scene covered by the camera, can affect system's performance.

We estimated the accuracy of the distance measured when the target was positioned at orientations of 0, 30 and 45 degrees with respect to the camera. The distance range was set between 40 centimetres and 2 meters. The error in distance estimation, at 2 meters (which is the maximum distance perceived by the system in its current configuration), was about 4%.

For most applications inside a home or a small office space, it can reasonably be expected that the threshold distance relating to when a person is paying attention to a specific device, should fall within the mentioned range. The low-end of the distance range would corresponds when the user is in front of a PC, while the high-end would correspond for the case of "wall mounted display".

3 Applications

In what follows, we sketch out some applications of this system that we are currently developing to demonstrate and to test the concept of attention detection.

3.1 Magnifying Map

The first demo of our system is envisioned as an application for info-kiosks, which may be found at tourist offices, for instance, and offer practical information for the newcomers. When there is no person in front of the camera, the system will display a neutral screen. Once the system notices the presence of a person (for a couple of seconds), it assumes that as an "attention-getting" event and displays a map on the screen. The level of details shown on the map is depending on the distance of the person with respect to the camera. If the person is situated afar (about 2 meters), then the system displays a general map of the city. As the person approaches, the level of details changes accordingly, so that at a closer distance (half meter), the current neighborhood is highlighted on the map. When the person moves away, then the system returns to its stand-by mode, waiting for the next visitor. This behavior can help overcome the limited resolution of large-scale electronic displays when compared to printed paper (for which more detail can be attained simply by walking up to the map).

3.2 "Follow-Me" Displaying Message System

This application is intended to illustrate how the contextual information can be used in order to have the incoming messages displayed at the most convenient location. We install our system on several workstations (WS). These WSs, on their turn, are connected to a message delivery server. In figure 2 we depicted a very simplified sketch of the configuration described.

Each time a WS notices the presence of a person nearby, it sends an event to the server (the WS can send any ID, maybe the most easy way is to send its own network

address). The server keeps track of the last WS where the person "has been seen", so that when a message destined to that person arrives to the server, it knows which is the terminal closest to him/her. This way, the person doesn't have to go to a specific location in the building in order to check for new messages. This is a realistic scenario, taking into account that is very common that a person can be present, throughout the day in several locations, not only in his/her office. As an example, we can refer to the university environment, where the researchers, besides their office, often has to go to a lab to do some experiments or have to attend a discussion session in a meeting room.

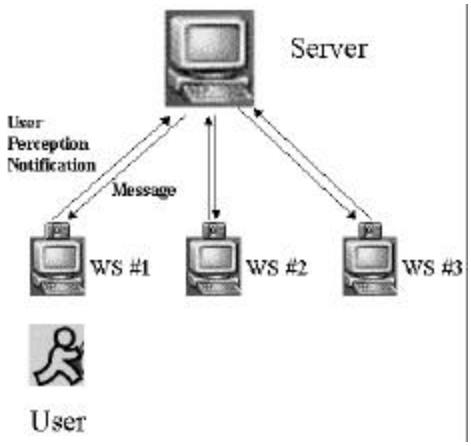

Fig.2. A sketch of the system architecture used for the distributed messaging application

On the other hand, this application shows that the creation of an aware environment can be addressed incrementally, starting with one perceiving device and dynamically add others, as they become available.

3.3 Discussion

The presented applications are limited to a single (anonymous) user. Other applications that can be envisioned (using this limitation) are related with a museum environment. Detecting a gallery visitor, who wears a badge, in front of a painting, may then trigger the playback of recorded audio information related with it.

The fact that all users of the system wear a unique badge and are indistinguishable offers advantages and disadvantages. In some cases, anonymity may be preferred by users, for instance in the museum example. In other cases, correct allocation of a display would benefit from user identification, to show a personalized message on the

monitor he is facing to. The current limitation can be overcome, by extending the current badge with a new one, having encoded (through a number of dots) the identity of the person who carries it. This way, we could enhance the contextual information, by adding person's identity. This thing would be very useful for the second application mentioned in this section ("Follow-Me"), because the system could know which person actually stays in front of it. By delivering only personalized messages, user's privacy can also be protected.

In order to experiment with an 'aware display' that will be used in the context of messaging applications, we have constructed a prototype screen enhanced with the camera as shown in figure 3. This device, which is under construction, will enclose a single-board PC [16] and will offer PC functionality from its touch-screen. This packaging of our system is crucial to enable field testing of awareness applications, so that they will be acceptable to install temporarily in people's home. It also serves as a crude prototype of the devices we expect to furnish an aware home.

Fig. 3. A "transparent" representation for an ubiquitous computing component: a touchscreen enhanced with perceptive capability due to the embedded webcam

4 Conclusions and Future Work

In this paper we proposed a new, low-cost solution to detect user's presence at sub-room level resolutions. This approach presents a high robustness against varying lightning conditions. The detection range is between 40cm and 2m, which makes it suitable for a large variety of applications. In consequence, this will allow a redefinition of the term "near", depending on the context the application will be developed for. Applications that are currently under development have been discussed and also other potential ones have been proposed.

While the presented method gave some very encouraging results, it obligates the person to wear this target attached to his close. In everyday context this can be an onerous obligation for the user. On the other hand, it provides a direct mechanism to the user to control when his activities are monitored and responded to: the user can simply remove the badge.

There are several alternative mechanisms to detect user proximity, e.g., using RFID tags, or using ultrasound signals. These approaches can work very accurately in domestic environments and can be very robust when there is no interference with other electronic devices. However, computer vision is better suited for the specific problem of detecting the direction the user is facing in. In our next step we shall investigate the feasibility of detecting user's attention without the need for reflecting badges, by directly detecting the human face and head pose in the scene.

Acknowledgments

The authors want to thank to Martin Boschman for the design of the infrared circuitry and to Charles Mignot for making the "aware display".

References

1. Aarts, E., Harwig, R., Schuurmans, M. Ambient Intelligence. In: Denning P.J. (ed.) The Invisible Future. McGraw Hill New York (2001) 235-250
2. Abowd, G.D., Mynatt, E.D.: Charting Past, Present and Future Research in Ubiquitous Computing. ACM Transactions on Computer-Human Interaction (2000), 7(1):29-58
3. Aware Home Research Initiative. Georgia Institute of Technology, http://www.cc.gatech.edu/fce/ahri/
4. Brumitt, B., Cadiz, J.J.: Let There Be Light: Comparing Interfaces for Homes of the Future. Proceedings of Interact'01, Japan (2001) 375-382
5. Darell, T., Tollmar, K., Bentley, F., Checka, N., Morency, L.-P., Rahimi, A., Oh, A.: Face-Responsive Interfaces: From Direct Manipulation to Perceptive Presence. Proceedings of Ubicomp, Sweden (2002) 135-151
6. Haralick, R.M. and Shapiro, L.G.: *Computer and Robot Vision*. Addison-Wesley, New York (1993)
7. Intille, S.S., Davis, J.W., Bobick, A.F.: Real-Time Closed-World Tracking. MIT Media Lab Tech Report TR-403 (1996)
8. de Ipina, D.L., Mendonca, P.R.S., Hopper, A.: TRIP: A Low-Cost Vision-Based Location System for Ubiquitous Computing. Personal and Ubiquitous Computing Journal, Springer (2002) 6(3):206-219
9. Krumm, J., Harris S., Meyers B., Brumitt, B., Hale, M., Shafer, S.: Multi-Camera Multi-Person Tracking for EasyLiving. Proceedings of 3rd IEEE International Workshop on Visual Surveillance, Dublin, Ireland (2000) 3-10

10. Krumm, J., Wiliams L., Smith, G.: SmartMoveX on a Graph – An Inexpensive Active Badge Tracker. Microsoft Research, Technical Report MSR-TR-2002-70 (2002)
11. Nakanishi, Y., Fujii, T., Kiatjima, K., Sato, Y., Koike, H.: Vision-Based Face Tracking System for Large Displays. Proceedings of Ubicomp, Sweden (2002) 152-159
12. Nagel, K., Kidd, C.D., O'Connell, T., Dey, A.K., Abowd, G.D.: The Family Intercom: Developing a Context-Aware Audio Communication System. Proceedings of Ubicomp, Atlanta, USA (2001) 176-183
13. Orr, R.J., Abowd, G.D.: The Smart Floor: A Mechanism for Natural User Identification and Tracking. GVU Technical Report GIT-GVU-00-02, Georgia Institute of Technology (2000)
14. PosterCam Project: Compaq Research: http://crl.research.compaq.com/vision/interfaces/ppostercam/default
15. Want, R., Hopper, A., Falcao, A., Gibbons, J.: The Active Badge Location System. ACM Transactions on Information Systems (1992) 10(1):91-102
16. Workbox Computer: http://www.zerez.com/producten/workboxp3/techspecs.htm