# Poster Abstract: Recognition of Personal Interest using Cooperation of Smartphone and Camera Sensors*

Jeremy Zylberberg, Stephan Rein, and Adam Wolisz
Telecommunication Networks Group
Technische Universität Berlin, Germany
{zylberberg, rein, wolisz}@tkn.tu-berlin.de

***Abstract*—We present a system to detect individuals' potential interest in objects. Cooperating visual and smartphone sensors are employed to gain knowledge of both the items people are interested in and the ID of the mobile devices they are carrying. This can be used to trigger the timely delivery of content to a user's device which is relevant to the object they are interested in. In this paper, we evaluate the efficiency of the approach to matching devices to people. Results demonstrate the system performs with similarity to existing examples, while also being able to offer additional functionality.**

## I. Introduction

Being able to provide a user with timely, relevant content about an object he shows interest in greatly enhances their experience and has a number of practical applications. Such a system could be highly valuable in a museum setting. Objects of interest are museum display items and upon observing objects one can receive relevant content on their personal devices. Some further useful applications include shopping or viewing a timetable at a train station.

We use combined data from a fixed camera and smartphones equipped with accelerometers to detect individual movement patterns. We recognize when a person shows interest in an object by detecting when one stops to observe it in a predefined area. With knowledge of who is showing interest and which device they are carrying, content relevant to the object of interest can then be delivered to them.

The primary contributions of this project include the provision of a working prototype demonstrating both algorithmic and communication functionality and an analysis of the system's performance with view for functionality in real-time.

Interest may be detected by sensing proximity to objects. Using mobile devices alone it is possible to estimate one's position [1], but including a camera will yield better accuracy.

An approach for matching moving people and devices such as in [2] uses a camera fixed to the ceiling, thus the field of view is dependent on room height. We do not require a precise localization and use a wall mounted camera to increase the viewing area. [3] has a ceiling-mounted approach and more energy requirements on the mobile device, while we wish to only use mobiles to transmit data upon request. While [4] has a non ceiling-mounted solution, it relies heavily on image processing in order to extract visual movement information. These examples focus primarily on the matching problem, whereas we take this further by incorporating interest detection with the aim to deliver relevant content.

*This work has been partially supported by the FP7 COAST (FP7-ICT-248036) project, funded by the European Community

## II. System Description

The system is a made up of a number of modules performing either a communication, processing or data delivery task. Figure 1 is a high level representation indicating the control and data flow of the system, highlighting the different tasks performed (numbered 1-5).

The motion tracking and sensor device modules deliver position and motion data to the system. The "Blob Tracking" component of the OpenCV library delivers position data relating to moving objects. Sensor devices are mobile units with smartphone sensors. We implemented software for the devices to enable recording of movement data and transmission of it upon request by the server.

We must be able to handle devices entering and leaving the system, communication errors and delay, and focus on achieving results with a minimum amount of data required. The discovery module (1) recognizes both new devices and new objects entering the area. Module (2) handles additional communication with sensor devices by requesting motion data when matching is performed.

In module (3), we need to match tracked objects to sensor nodes in the area by comparing the similarity of the data. Visual data and accelerometer data need to be transformed such that they have a common representation before being comparable. We estimate the velocity of each source based on an approach in [2] by taking the moving average of a distance measure from video data, and using the magnitude of accelerometer data to calculate its standard deviation. In this form, unique movement patterns are evident and the two data comparable. We use a correlation coefficient to obtain the closest matching accelerometer/object pairs. If no acceptable match is found, subsequent steps in the control flow are redundant and control returns to the top.

Interest detection (4) assesses whether a user is interested in an object. We maintain a set of predefined areas in the room that are centered around objects. Thus, an exact localization

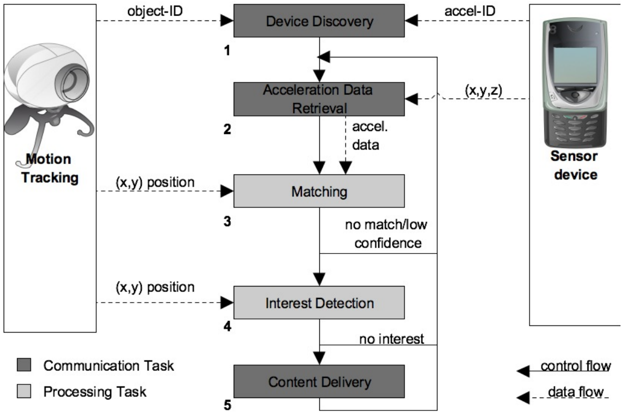


Fig. 1: Representation of system control and data flow

is not required and a single camera can be used to provide an estimation. It has been shown that the lowest point of a person being tracked object correlates to the location of feet on the floor. We use this as a basis to assess whether one is within an area of interest. If the person has a matched device, it follows that content be delivered to it (5).

## III. Evaluation

Initial evaluation involved offline experiments to assess the algorithmic components without real-time constraints. Scenarios of 1-2 minutes each were recorded with up to 5 participants emulating movements typical for a museum visit. At various times, participants showed interest in any of three objects within indoor area of 15×7m, and the camera angled downwards from a height of 4m. The minimum distance between objects was 3m, with the intent to investigate the effects of more closely spaced objects in the future.

Video data was gathered using a digital video camera and recordings transferred to a workstation. Motion tracking was performed to produce sets of positions with timestamps for tracked objects. Accelerometer data was simultaneously recorded using an Irene sensor node transmitting data to a basestation. All devices are discovered by the system at the same time, but people enter the scene separately. For each scenario, object and accelerometer data were used as inputs to evaluate the system. Figure 2 shows an example of four from twelve scenarios with 2 and 3 participants. It demonstrates the calculated time consumed for tasks (1)-(3). In each block, $P_x$ indicates the entrance of the $x_{th}$ person in the area and the respective correlations that were calculated for each device. A successful match – defined here by a cumulative correlation larger than 0.5 – for the first person who enters can be achieved in around 3 seconds. Subsequent matches show longer total times as new data must be retrieved from unmatched devices. Data retrieval time is based on the transmission time of individual packets, which grows as matching time grows. Time required for discovery is smaller than 1 millisecond. Time savings can be achieved by combining multiple measurements in a single packet or for processing time by filtering out erroneous data. Nevertheless, the time span for each entrant is sufficient to proactively trigger content delivery and is similar

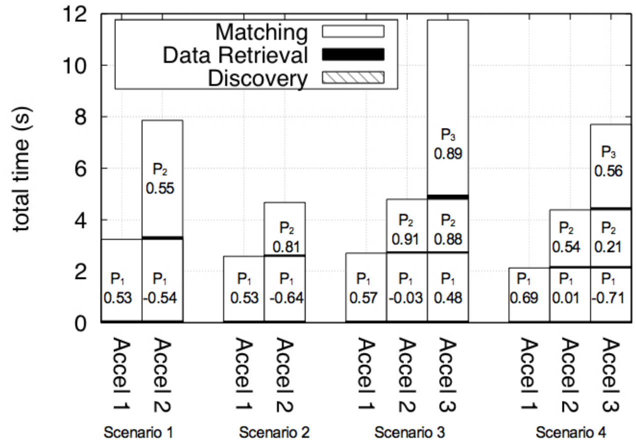


Fig. 2: Time consumption of tasks for scenarios with 2 or 3 persons

to other time and correlation measurements in previous studies [3], [5], while still offering additional functionality.

## IV. Conclusion and Future Work

We present a system which can detect personal interest and deliver content to devices. Results using correlation comparisons alone yield a match within 3 seconds showing potential for a faster response.

We currently optimize the system for real-time constraints. As future extensions, the system can be evaluated with use of smartphones with integrated sensors rather than sensor nodes, or with of a wireless smart camera for motion tracking. Another area of focus may also be an investigation of timely content delivery based on the type of content.

## References


[1] N. Banerjee, S. Agarwal, P. Bahl, R. Chandra, A. Wolman, and M. D. Corner, "Virtual compass: Relative positioning to sense mobile social interactions," in *Pervasive*, 2010, pp. 1–21.

[2] D. Jung, T. Teixeira, and A. Savvides, "Towards cooperative localization of wearable sensors using accelerometers and cameras," in *INFOCOM*, 2010, pp. 2330–2338.

[3] T. Teixeira, D. Jung, and A. Savvides, "Tasking networked cctv cameras and mobile phones to identify and localize multiple people," in *Proc. of 12th Ubicomp*. ACM, 2010, pp. 213–222.

[4] T. Teixeira, D. Jung, G. Dublon, and A. Savvides, "Pem-id: Identifying people by gait-matching using cameras and wearable accelerometers."

[5] ——, "Identifying people in camera networks using wearable accelerometers," in *Proc. of 2nd International Conference on PETRA*. ACM, 2009, pp. 20:1–20:8.