Vision based People Tracking for Ubiquitous Augmented Reality Applications

Christian A.L. Waechter* Daniel Pustka† Gudrun J. Klinker‡

Technische Universität München
Fachgebiet Augmented Reality

ABSTRACT

The task of vision based people tracking is a major research problem in the context of surveillance applications or human behavior estimation, but it has had only minimal impact on (Ubiquitous) Augmented Reality applications thus far. Deploying stationary infrastructural hardware within indoor environments for the purpose of Augmented Reality could provide a users' devices with additional functionality that a small device and mobile sensors cannot provide to its user. Therefore people tracking could be expected to become an ubiquitously available infrastructural element in buildings since surveillance cameras are widely used. The use for scenarios indoors or close to buildings is obvious. We present and discuss several different ways where people tracking in real-time could influence the fields of Augmented Reality and further vision based applications.

Keywords: Augmented Reality, People Tracking, Sensor Fusion

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities; I.4.8 [Scene Analysis]: Sensor fusion—Tracking;

1 MOTIVATION

A new generation of small and powerful hardware, like Ultra Mobile PCs (UMPCs) and Netbooks, offer the possibility to transfer AR applications to wearable platforms. Recent advances in the hardware specific optimization of tracking algorithms for mobile phones [8] allow for a completely new quality of Ubiquitous Augmented Reality (UAR) applications on mobile devices. Yet, although these platforms show increasing potential to present virtual information, there is still a need to combine the sensor information from the mobile clients with information that is retrieved from stationary systems.

In robotics and surveillance applications, various stationary tracking environments based on GPS, RFID, infrared or UWB tags have been merged with mobile sensor data. Schulz et al. have presented a fusion system that combines an anonymous laser range-finder providing highly accurate position data with infrared and ultrasound badges, providing more rough position data [6]. The system allocates correct personal ids to the gathered trajectories. The correspondence algorithm uses a Rao-Blackwellized particle filter approach to determine the position as well as the id of an object. This approach is used to localize persons on a map of a building. Graumann et al. have presented a multi-level framework to combine different sensor data and encapsulate certain properties within certain layers [2]. They showed a location based application with switching sensors for outdoor and indoor scenarios. As a fusion system they use a mobile laptop that integrates the different device related estimated data. The mobile device is equipped with GPS, Wifi and UC Berkeley sensor motes that gather user specific position data and fuses them on a local bases. The fused data is used for a navigation application.

Use of these systems requires the mobile units to carry special targets, such as RFID, GPS or Wifi units. Vision based people tracking systems with stationary cameras, e.g. for surveillance applications ([3], [1]), offer the possibility to track persons without requesting such special tags. Using people tracking for AR applications has not yet been researched thus far. In this paper, we present a system that allocates and fuses anonymous position information of a stationary vision-based people tracking system with inertial and sporadic optical tracking data of a hand-held device to augment a real-time video stream from a mobile camera.

*e-mail: christian.waechter@cs.tum.edu
†e-mail:daniel-pustka@cs.tum.edu
‡e-mail:gudrun.klinker@cs.tum.edu

2 VISION BASED PEOPLE TRACKING

Our people tracking system [7] uses a model-based Sequential Monte Carlo (SMC) Filtering approach to track persons in an observed area. We use a single, ceiling mounted camera that observes the scenery from a bird's eye view. Each person entering the area is detected by analyzing the pre-processed, background-subtracted camera images and assuming that BLOBs which cannot be associated with already detected persons indicate a new person. Persons who are already recognized by the system are tracked by individual Particle Filters that use 300 weighted hypotheses of the object's position to represent its uncertainty. All persons are represented in three-dimensional Cartesian space as spheroids with a height of 180 and a width of 40 cm. For each hypothesis we use the spheroids and the knowledge about the image formation process to generate virtual images that show a possible appearance of the person on the image plane. To evaluate the weight for each hypothesis our measurement likelihood function compares the region around the hypothesized ellipsoidal area of that person in the image with the current camera image. The likelihood measurement function takes also dynamic occlusion from the other persons into account and uses the three-dimensional representation to reliably estimate the weight of each hypothesis. The ceiling mounted camera provides a resolution of 320×240 at $15Hz$ and has a 35° FOV.

3 MOBILE HARDWARE

We use a laptop equipped with a VGA-webcam running at $30Hz$ as wearable computing platform. The camera is used for the initial pose estimation from a binary square marker as well as to capture the live video stream for showing the augmentation. An inertial sensor is attached and registered to the mobile device [5]. The device connects to a central server using a WiFi connection to send own position data and request the position information of the people tracking system.

4 REGISTRATION

In order to use the position information of the people tracking the user has to register his device to the environment: By pointing the webcam of the user worn device toward an optical square marker, which was previously registered to the environment, the exact pose information of the user's device within the building can be estimated. The location information of the pose links the user to his individual trajectory observed by the people tracking system at real-time. An identity mapping client associates the identities of the people tracking with the user information by estimation of the minimal distance of the two positions [4].

5 Possible Applications

Beside surveillance and behavior estimation we introduce some topics where people tracking could be of interest and further applications are planned.

Indoor Navigation Multiview approaches to people tracking generally provide two dimensional position information of all observed objects. Since the trajectory of persons can be determined over a longer period of time and overlapping camera views this enables the possibility for indoor navigation. After the registration of the user's device within an environment map of the building it would be possible to lead the user to specific places which might be of interest for him. A possible scenario is an application that guides the user to interesting areas within a museum by using people tracking from installed surveillance cameras. An additional application for navigation would be within an office building to lead the user to the next location the user must visit following certain processes.

Human Computer Interaction Position data from the people tracking system can be used as input data to applications that require human interaction. The advantage of differentiating between several persons can lead to gaming applications that distinguish distinct persons or team members. Intelligent rooms can react to certain human behavior since individuals are continuously tracked.

Ubiquitous Augmented Reality Since Augmented Reality applications use a complete pose (6DoF) to display three dimensional graphics perspectively correct, people tracking, as introduced here, can only be used in combination with other sensors. In our case the people tracking from the bird's eye view associates the two dimensional position information of the person with the device's position and it lacks the possibility to estimate the height of the device. To overcome this shortcoming we take the last known z-position of the device, normally estimated during the registration process using the optical square marker. The fusion of the position information with a gyroscope applied in our scenario enables to use the device to show perspectively correct rendered 3D graphics.

6 Example Demonstration

The people tracking observes the area of the waiting room at our office. The systems starts tracking persons as soon as they enter the observed area and stores their trajectories for the requesting clients.

Two visitors enter the room without any device and are individually tracked by the people tracking system. A third person carrying the previously described mobile hardware enters the room and is also being tracked (Figure 1). As this user points his webcam toward a wall mounted square marker the respective position of the device is used to register it to the nearest position within the people tracking system. An 3D arrow is then augmented near the marker in the camera image. When the user points his webcam away from the marker, the new pose of the device is determined by the complementary fusion of the people tracking system and gyroscope: the arrow can still be displayed. Turning around in the direction of the arrow, the user gets to see a virtual sheep augmented to the middle of the room (Figure 2). As the user moves through the environment the sheep remains roughly at its position and keeps its correct orientation.

7 Future Work & Conclusion

Although there are restrictions, especially the inaccurate z-position, we suggest people tracking to be considered as a possible solution for indoor location estimation, that is independent of the user's hardware. The fusion of anonymous position information of the people tracking and personal orientation information shown here demonstrates a huge potential for UAR applications on small devices with sensors that cannot determine the device's complete pose information by themselves. Intelligent environments with an integrated vision based people tracking could provide the users with additional information on a basis that also respects privacy issues.

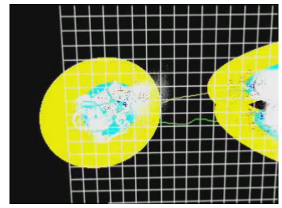

Figure 1: People Tracking View

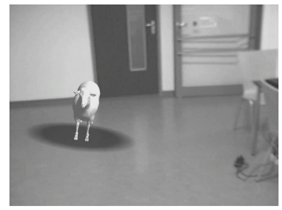

Figure 2: Mobile Client View

Acknowledgements

This work was supported by the PRESENCCIA Integrated Project funded under the European Sixth Framework Program, Future and Emerging Technologies (FET) (contract no. 27731).

References

[1] F. Fleuret, J. Berclaz, R. Lengagne, and P. Fua. Multi-camera people tracking with a probabilistic occupancy map. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):267–282, Feb. 2008.

[2] D. Graumann, W. Lara, J. Hightower, and G. Borriello. Real-world implementation of the location stack: The universal location framework. In *Proc. Fifth IEEE Workshop on Mobile Computing Systems and Applications*, pages 122–128, 2003.

[3] S. Kong, C. Sanderson, B. Lovell, and A. Cavallaro. Classifying and tracking multiple persons for proactive surveillance of mass transport systems. In *IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS'07)*, page 159. IEEE, 2007.

[4] D. Pustka, M. Huber, C. Waechter, F. Echtler, P. Keitler, G. Klinker, J. Newmann, and D. Schmalstieg. Ubitrack: Automatic Configuration of Pervasive Sensor Networks for Augmented Reality. Submitted to Pervasive Computing, 2009.

[5] D. Pustka and G. Klinker. Dynamic Gyroscope Fusion in Ubiquitous Tracking Environments. In *Proc. 7th International Symposium on Mixed and Augmented Reality (ISMAR)*, 2008.

[6] D. Schulz, D. Fox, and J. Hightower. People tracking with anonymous and id-sensors using Rao-Blackwellised particle filters. In *International joint conference on artificial intelligence*, volume 18, pages 921–928. LAWRENCE ERLBAUM ASSOCIATES LTD, 2003.

[7] C. Waechter, D. Pustka, and G. Klinker. A Perspective Approach Toward Monocular Particle-based People Tracking. Submitted to British Machine Vision Conference (BMVC'09), September 2009.

[8] D. Wagner, G. Reitmayr, A. Mulloni, T. Drummond, and D. Schmalstieg. Pose tracking from natural features on mobile phones. In *Proc. 7th International Symposium on Mixed and Augmented Reality (ISMAR)*, 2008.