Dead reckoning supports stereo vision in pedestrians tracking

Tullio Salmon Cinotti, Luigi Di Stefano, Giuseppe Raffa, Luca Roffia, Marina Pettinari, Martino Mola
ARCES - University of Bologna
{tsalmon, ldistefano, graffa, lroffia, mpettinari, mmola} @arces.unibo.it

Abstract

This "Work In Progress" paper reports on a novel approach to pedestrian tracking in confined areas which is based on a combined use of a stereo vision tracking system and a set of sensors available on a mobile platform. This platform is a hands-free, sensory augmented, wearable computer, designed to support visits to museums and archaeological sites. The system is going to be demonstrated within the framework of a Cultural Heritage application at the "Musei Universitari di Palazzo Poggi" in Bologna"[1].

1. Introduction

Localization and tracking are key features in many pervasive computing applications and cultural heritage sites have proved to be interesting and challenging environments for demonstrating innovative ideas in this domain [1]. Building a pervasive environment in historical or preexistent sites is difficult. Adding spread of sensors across the area is normally not allowed by the cultural heritage authorities, unless they are unobtrusively integrated within the hosting architecture. Furthermore constrains and precision levels may change significantly throughout the site. Experiments in museums and archaeological areas show that stand-alone localization systems like GPS or WLAN based indoor localizations techniques may not meet the user needs in case the resolution and the precision required by the localization system be high, due to the high density of artifacts to locate. The same stand-alone solutions fail when an uninterrupted operation is required. In these cases solutions based on a combination of co-operative location and tracking subsystems should be investigated.

The research we are proposing, aims to demonstrate a technique to continuously track the user in "difficult" environments which is based on the co-operation between two complementary systems: the first one is an Inertial Tracking Platform (ITP) embedded in a wearable navigator (par. 2), while the second one is based on small stereo pairs hidden in the environment. In the solution we are proposing, the ITP is used to provide tracking information in the gaps within the areas covered by the high precision Visual Tracking System (VTS) (par. 3). Such an approach is already well known in other contexts where other absolute positioning systems are used [2] instead of the VTS. The VTS is needed where the density of the exhibits is high and the localization requirements are extremely strict. In areas where the requirements are more lease, the ITP may be used. The issues related to the management and integration of the proposed tracking system components need to be solved with a distributed architecture where the intelligence is divided between the mobile devices and a central tracking server (par. 4).

2. Inertial "on-board" Tracking Platform

Inertial tracking is provided by a step detection algorithm running on a multi-sensor board (MSB) embedded in a wearable platform based on a *Mobile Intel® Pentium® III UltraLowVoltage Processor at 500 MHz.* The platform is WHYRE®, the wireless personal museum navigator made by Ducati Sistemi [3]. The wireless connection can be either a WLAN, a GPRS or a UMTS module. Hence, different kinds of sites can be covered. The MSB provides

[1] *This project is funded by the University of Bologna and by EPOCH, a 6thFP European Network of Excellence.*

azimuth, roll, pitch and step information. Its main components are described in [3]: a microcontroller retrieves data by a three bit digital compass, a two axes accelerometer and a gyroscope. In order to allow communication with common devices, Bluetooth and ZigBee interfaces are currently being developed. A GPS receiver with integrated antenna is mounted on a separate small form factor PCB for outdoor applications, while the WLAN card is used for coarse indoor localization.

Literature on human walking analysis [5] shows the way a body moves during a gait cycle: the human body acceleration can be, in first instance, resolved into a vertical and a frontal (horizontal) component. The two-axes accelerometer detects the frontal component only, since in the MSB the accelerometer axis are approximately parallel to the movement plane. The compass and the gyroscope detect the heading. The ITP algorithm is based on the model described in [5] and its main goals are:

- Estimate the number of steps;
- Estimate the length of each step;
- Evaluate the heading of the walk.

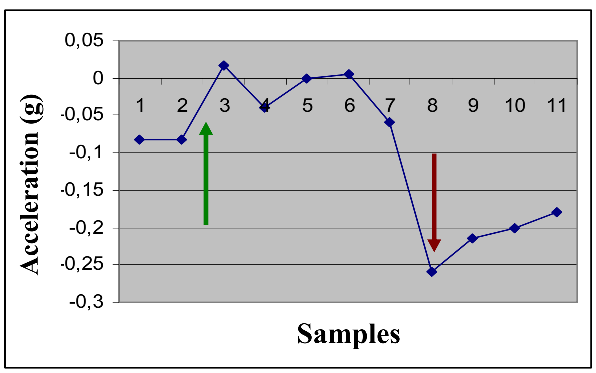

Figure 1. Frontal acceleration during a step

As shown in figure 1, when the step begins the frontal acceleration has a positive derivative (green arrow). The end of the step is represented by a strong negative derivative of the frontal acceleration (red arrow). A finite state machine recognizes the sequence of these two events and detects the steps. The algorithm is able to provide basic but useful pedestrian activity information, i.e. still or motion. Furthermore, it empirically estimates the step length as a function of the maximum and minimum accelerations during each step [6]. This system has already been tested outdoor in an archaeological environment (Pompeii) and has already been compared to the GPS system. Preliminary results (figure 2) show that the inertial tracking (red path) is smoother than the GSP trajectory (green path) but it suffers from the following inaccuracies:

- The number of steps and the step length (figure 2: segment “A”);
- The heading drift after a long walk on a rough and bumpy path (figure 2: segment “B”).

Figure 2 shows an average 7% error in the path length measure and an accumulated heading drift of 12 degrees after a 200 m walk (approximate walk time: 215 sec).

Figure 2. ITP compared with GPS

The ITP is useful if an infrastructure for absolute positioning is not available or not deployable, but it needs to be periodically resynchronized because of its inherent accumulated error going up with time. One of the goals of this research is to resynchronize the ITP with scattered around VTSs.

3. Visual Tracking based on stereo vision

The capability to visually track several moving persons over long time spans through occlusions and across complex backgrounds represents a crucial and still largely unresolved issue. Many methods for real-time multi-person detection and tracking based on the analysis of images supplied by a static camera have been described in the literature. Unfortunately due to the overlapping of people in the 2D images as well as to rapid changes in their appearance it is extremely difficult to robustly maintain tracks integrity over time. On the other hand, 3D scene reconstruction using stereo cameras holds the potential for dealing with these issues. In fact, yielding the 3D coordinates of image points, stereo allows to build an orthographic view of the scene with the projection plane parallel to the ground plane. In such a projection plane people tend not to overlap significantly and their appearance cannot change quickly. An orthographic view that has proven to be very effective for reliable multi-person detection and tracking can be obtained by means of plan-view statistics (i.e. occupancy and height maps [7], [8]). Our Visual Tracking System is based on the concept of plan-view statistics and deploys a down-looking digital stereo camera (i.e. an STH-MD1 by VIDERE DESIGN) mounted overhead at a height of approximately 3 m from ground. Synchronized image pairs coming from the cameras are first processed by means of a real-time correlation-based stereo matching algorithm [9] in order to continuously gather dense disparity maps. Since the stereo camera is fully calibrated, the disparity information associated with each pixel is converted into 3D coordinates with respect to the ground reference frame. Then, a 2D orthographic view of the scene is built by proper projection of the 3D coordinates associated with moving persons into the occupancy and height maps. People tracking takes place into the 2D orthographic view and relies on Kalman filtering to predict the new positions of tracked objects. An effective heuristic handles in a coherent and straightforward way complex events such as dynamic/static partial or total occlusion and the general case of multiple people merging. Inputs to the tracking algorithm are the list of tracked persons until time *t-1* and the current set of objects detected within the 2D view at time *t*. Each entity (tracked person or new object) has a bounding-box located at its centre of mass. Based on the Kalman filter, at time t the algorithm considers the intersections among the predicted bounding-boxes of tracked persons and the bounding-boxes associated with the currently detected objects. The simple case of a one-to-one correspondence between a tracked person and a new object is solved by matching the current object with its associated person (this is called straightforward tracking). When two or more predicted bounding boxes fall onto the same detected object the algorithm assigns the pixels in the 2D view to the objects adopting a distance-based cost function that enforces the following principles:

- People sizes cannot change rapidly;
- People heads cannot overlap in the 2D view;
- The merged blob must be partitioned among objects keeping the resulting individual objects as compact as possible.

A near-optimal solution of the resulting NP-hard problem can be found using an approach based on Lagrangian relaxation. Novel objects appearing in the scene are firstly compared with recently lost objects: if a match is found then the algorithm assigns the identity of the object otherwise a new people entity will be instantiated.

Experimental results have demonstrated the robustness, the accuracy (i.e. objects are located with an uncertainty of *5 cm * 5 cm* at the ground) and the real-time performance of the proposed Visual Tracking System.

4. System integration

Switching between the ITP and the VTS and associating an unambiguous ID to each detected user are the major issues to be solved by the proposed tracking approach. If information from both systems is available, priority is given to the VTS because it provides absolute and high precision positioning information. Once the VTS doesn't detect the device anymore, its position is provided by the ITP, using as starting point the last position detected by the VTS. Since the system has to localize several users, it is necessary to associate an identifier (ID) to each device that enters in an area covered by the cameras. Such an ID must be known by the device, in order to obtain its own position on the client side. The VTS already manages several users and associates to each one an impersonal sequential ID. This ID has to be related to a unique device identifier (e.g. MAC, IP). To this end an approach based on a Central Tracking Server that communicates with all the devices is being considered. Each client knows its own location and sends its ITP position to the server when it expects to be close to the area monitored by the VTS. The server uses this information to perform an algorithm based on the minimum distance between the received position and the detected positions, leading to the association of the impersonal ID with the right device. If such user identification process is inadequate, because, for example, the accumulated ITP position error is found to be too large, then identifiable markers embedded in the wearable platform shall be considered.

The precision of the ITP tracking algorithm significantly impacts the inertial tracking process in VTS-uncovered areas, and it also affects the reliability of the association between the device and the impersonal ID given by the VTS involved in the ITP/VTS switching mechanism. After preliminary laboratory testing, current results (see par.2) show that the ITP precision depends on the ITP wearing mode and on the user gait (walking or just stepping forward, backward or side); it also depends on the user physical structure. We are trying to increase the ITP accuracy level with improved algorithms, with new sensors (for example with a 3 axis accelerometer) and with new wearing modes. To this end the ITP may become a node of a WPAN (Wireless Personal Area Network) driven by the mobile device. Eventually, in order to compensate the types of inaccuracies shown in par.2, investigations shall be carried out to verify if the ITP error model can be evaluated. This could be achieved comparing the precise trajectory detected by the VTS with the one simultaneously observed by the ITP.

5. Conclusions

Museums and archaeological sites are challenging test-beds for always-on high resolution, real time tracking systems. In this paper, a system based on the combined use of a high-precision Visual Tracking System and an inertial tracking platform was presented.

This research is in progress and it is far from being completed; however the main components of this system have already been developed, deployed and tested in-field.

The final integration is currently being carried out and the target system is expected to be tested at the "*Musei Universitari di Palazzo Poggi*" during 2006.

6. References

[1] N. Ryan, T. Salmon Cinotti, G. Raffa "Smart Environments and their Applications to Cultural Heritage", Proceedings, *Ubicomp05 Workshop*, Tokyo, 2005.

[2] V. Gabaglio, Q. Ladetto, B. Merminod, "Kalman filter mechanization for GPS-INS Pedestrian Navigation", *GNSS 2001*, Sevilla (Spain), pp. 8-11, May 2001.

[3] T. Salmon Cinotti, N. Raviprakash, G. Mincolelli, F. Sforza, G. Raffa, L. Roffia, "WHYRE: a Context-aware Wearable Computer for Museums and Archaeological Sites", Proceedings *ISWC'04*, Arlington, 2004

[4] G.A. Cavagna, R. Margaria, "Mechanics of walking", *Journal of Applied Physiology*, vol. 21, pp. 271-278, 1966.

[5] Q. Ladetto, V. Gabaglio, B. Merminod, P. Terrier, Y. Schutz, "Human Walking Analysis Assisted by DGPS", *GNSS 2000*, Edinburgh (sc), 1-4 May 2000.

[6] H. Weinberg, "Using the ADXL202 in Pedometer and Personal Navigation Applications", *APPLICATION NOTE*, Analog Devices, 2002

[7] D. Beymer, "Person counting using stereo", Proceedings of *Workshop on Human Motion 2000*, December 2000, pp. 127–133.

[8] M. Harville, "Stereo person tracking with adaptive plan-view templates of height and occupancy statistics", *Image and Vision Computing*, Vol. 22, pp. 127–142, 2004.

[9] L. Di Stefano, S. Mattoccia, M. Marchionni, "A fast area-based stereo matching algorithm", *Image and Vision Computing*, Vol. 22, pp. 983–1005, 2004.