2011 International Conference on Indoor Positioning and Indoor Navigation (IPIN), 21-23 September 2011, Guimarães, Portugal

Improvement of Inertial Sensor Based Indoor Navigation by Video Content Analysis

T. Bernoulli*, U. Dersch**, M. Krammer*, U. Walder*, and K. Zahn**

* Institute for Building Informatics, Graz University of Technology (TU Graz), Graz, Austria. bauinformatik@tugraz.at

** CC Innovation in Intelligent Multimedia Sensor Networks, Lucerne University of Applied Sciences and Arts (HSLU), Lucerne, Switzerland. ulrich.dersch@hslu.ch

***Abstract*—Foot-mounted inertial systems for indoor positioning and pedestrian guidance are an elegant and cheap solution to track first responders within buildings and underground structures. They can run completely autonomous, i.e. they do not require preinstalled infrastructure installations. But there are some difficult problems to solve: Starting from a known position the positioning error increases with the travelled distance, if only a double integration of the accelerations is performed. Therefore it is necessary to cut down the integration intervals and to reposition the system from time to time vis-à-vis a known landmark. Several algorithms have been developed at TU Graz to reduce these errors. The precise recognition of the motion patterns allows for performing zero velocity updates (ZUPT) to achieve a high accuracy in distance. Further improvements are gained by the fusion of additional sensors like barometer or GPS. Different map matching algorithms are used to perform periodic repositioning. Since ground floors often are not available, HSLU has developed a video content analysis based repositioning method that increases the accuracy of the heading of the IMU system considerably.**

***Keywords*—foot mounted IMU, sensor fusion, video content analysis**

I. Introduction

Foot-mounted sensors are getting more and more popular in indoor positioning. The reasons are the increased comfort and precision due to better, small and wireless hardware, the fusion with additional sensors, as well as new algorithms to recognize zero velocity updates [1]. Different map matching solutions allow for a further improvement of the positioning accuracy [2]. The falling costs, resulting from the fact that the basic hardware components are also part of modern smartphones and the high autonomy of inertial based systems, are other major advantages of this technology.

Though the precision in distance is meanwhile in the range of centimeters even for longer distances (>100m), the stabilization of the heading causes major problems in environments where the earth's magnetic field is heavily disturbed. Automatic gyro-stabilization as used in high-end sensors reduces this problem significantly, but is not stable enough to hold an initial heading for a long time, especially in the case where ground floors are not available for a periodic adjustment of the heading by map matching techniques (Fig. 1). Normally GPS data fusion is used to enhance the inertial data outdoors. But in urban areas with high buildings the resulting precision is often too poor to guide pedestrians precisely through streets and public areas.

Figure 1. Map matching in a environment with disturbed earth's magnetic field caused by a railway track alongside the building

In case that a facility can be documented in advance by digital images, the technology of video content analysis is a promising solution to adjust the heading in real-time. From only a few image contents staying constant, the algorithms developed at HSLU can calculate the position of the camera with high accuracy (<1m). The fusion of camera and inertial data allows for stabilizing the heading and to hold it for the time of a typical deployment e.g. of fire fighters in mixed indoor-outdoor environments.

In the following the used technologies implemented in the research and application framework AIONAV (Autonomous Indoor and Outdoor NAVigation) and achieved results of sensor fusion and video content analysis are described and discussed in more detail.

II. The AIONAV System (Autonomous Indoor and Outdoor NAVigation)

The AIONAV system developed at TU Graz mainly for research purposes, but also for practical applications like first responders support, consists of four main parts: The sensor based positioning system, a building information model (BIM) based position verification and improvement system, a user interface, and a subsystem to keep multiple AIONAV system instances synchronized over networks.

The system is designed to provide position information not only to the first responder wearing the AIONAV device but to synchronize the available information between multiple system instances. This allows first responders to coordinate more efficiently. The user interface is twofold: One for the person on the field having only a small touchscreen and not needing a lot of options and one for the command center where a common computer is used. The touchscreen based interface allows quick access to important information as well as setting the current position if the position determined by the system does not fulfill the precision requirements.

The AIONAV sensor sub-system is a set of algorithms and I/O operations processing data from sensors. Inertial measurement units (IMUs) providing linear and rotational accelerations, the magnetic field, as well as temperature and pressure data, are the basis of the positioning system. To handle the described long term accuracy problem additional sensors can be used. Some of those, e.g. GPS receivers, already provide positions in the global coordinate system.

As soon as different sensors act simultaneously and provide complementary data, sensor fusion becomes important. If each sensor is handled individually the algorithms can be tailored to reflect the characteristics of a sensor. For example different IMU based algorithms such as a motion pattern dependent zero velocity updates (ZUPT) have been developed [1] to compute positions from the sensor data. The data of sensors already providing position information (e.g. GPS) may be used unprocessed or processed dependent on their accuracy and reliability. Every sensor type has its advantages and disadvantages regarding error-stability, reliability in specific situations etc. The aim of sensor fusion is to use characteristics of one sensor to counteract problems of another one.

There exist multiple approaches to fuse sensor data. Raw data of different sensors can be combined using specialized algorithms to get positions. AIONAV has the requirement to be completely flexible. A user should be able to connect sensors supported by the system and AIONAV computes one ideal fused position out of these multiple, inhomogeneous data-sources. To provide this flexibility the position data is used to perform position sensor fusion. The approach provides high flexibility of combining fusion algorithms. AIONAV supports all kinds of fusion techniques - from general Kalman filtering to specialized hand tailored fusion algorithms in case tweaking is necessary (Fig 2).

Besides sensor fusion the system implements BIM based position verification and improvement algorithms. If floor plans of a building are available they can be easily enhanced to a simple BIM. For indoor navigation this provides a big improvement as long term heading problems can be corrected using map matching.

This paper focuses on a third way of improving indoor navigation: The use of video content analysis which has the advantage not needing even a simple BIM.

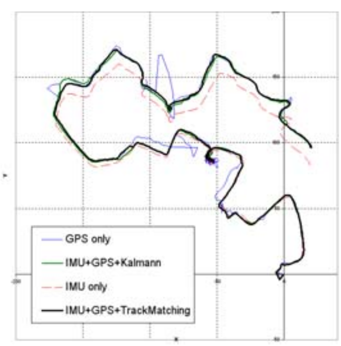

Figure 2. Different sensor fusion methods with best result for track matching algorithm (<2m indoor and outdoor for 900m)

The AIONAV system has a great potential to make the deployment of first responders more efficient. But this is not the only group of people to profit from this system. With the video content analysis it has the potential to improve e.g. the way how blind persons move in public spaces helping them to find the tactile guiding systems or build a bridge if these are not laid everywhere. Another field of application is for military positioning and navigation systems used in urban warfare training and deployment. Tests with the Swiss Army in this field showed very promising results and will be discussed in the full paper in more detail.

III. Video Content analysis

The idea of fusing video and inertial data is motivated by the human perception: a human can “navigate” with closed eyes rather precisely over some ten meters using the vestibular system, the human inertial sensor, situated in the middle ear. However, at longer distances the heading will diverge and has to be recalibrated using the human vision sensor. Here a similar approach is used: MEMS based inertial sensors of the AIONAV system will provide information sufficiently precise to allow for dead reckoning at shorter times scales up to minutes. At longer times, when inertial based navigation tends to accumulate integration errors video image correlation techniques are used to recalibrate the system. Therefore the AIONAV system is extended by a small portable camera taking live pictures, which are processed in real time.

As a first step images of the facility are taken in advance either along certain paths or at “strategic” points and are stored, together with the positions where these images were taken, in a database on the AIONAV system. During the navigation the live images taken with the video camera are correlated in real time to the images in the database. In case a match between the live and a preregistered image is found the current position estimation from the inertial navigation is checked for consistency with the position of the image in the database and the former is corrected based on the visual information if necessary.

Obviously the goal should be to store as few images as possible in the database. Therefore the image correlation algorithm has to be flexible enough to allow for the recognition of images not only close to but also in a certain vicinity of those positions, where preregistered images were taken. On the other hand the correlation must be reliable enough such that “false positives” i.e. false correlations are highly unlikely. Therefore a two-step correlation algorithm has been developed. In a first step so called feature points are determined in both images, which are known to be insensitive to lighting and contrast variation and invariant under scaling, rotations and translations [3]. The identification of corresponding feature points between the live and a database image provides a first estimate of a correlation coefficient. In Fig. 3 one example is shown, where a live image (top left) could be matched to a database image (top center). Both images show the same corridor, with a camera displacement of approximately 2.5 m mainly along the optical axis. Despite this displacement corresponding features in both images could be identified and are shown on the live image (red rectangles) together with the difference vector (blue lines) indicating the feature’s shift onto the corresponding position on the database image.

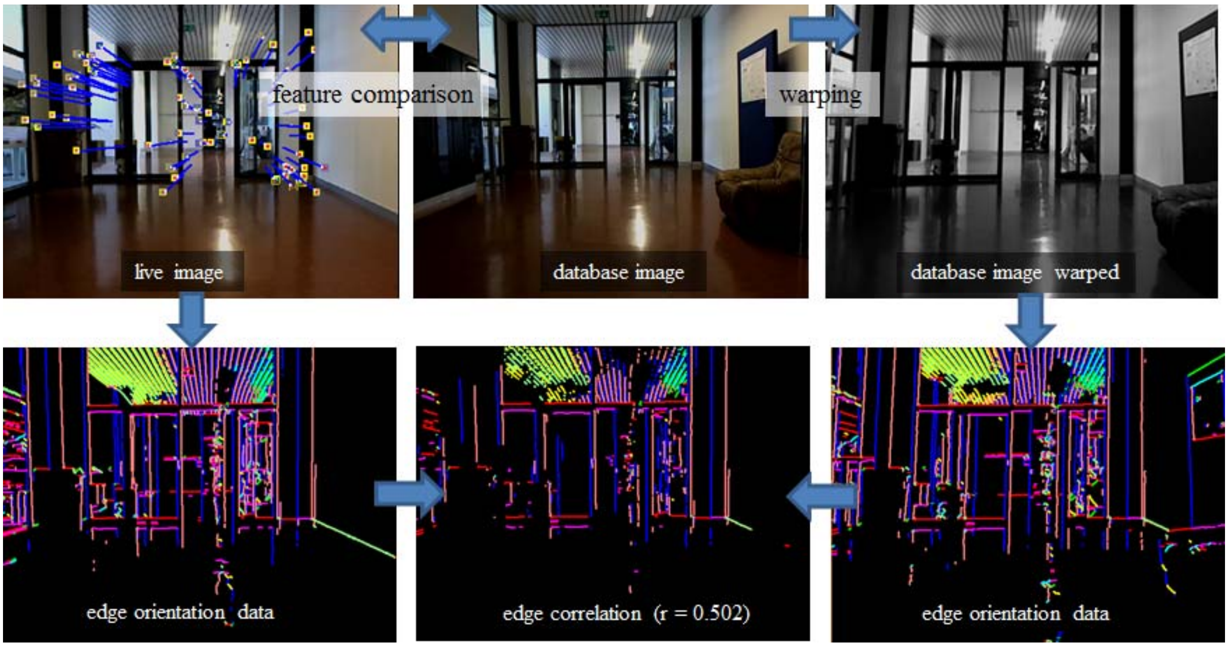

Figure 3. Common features in the live image (top left) and the database image (top center) are used to warp the database image on the live image frame (top right). Then the respective edge orientations (represented by the pseudo colors) are extracted (bottom left and right) and their correlation is determined (bottom center).

The feature matching is based on heuristic limits for the maximum differences between feature descriptors (in order to be considered as equal) as well as a minimum required number of feature matches between the live image and the database image. The feature point correspondence gives a first hint on a possible match of the live image with a preregistered database image. However this criterion is not reliable enough especially in larger facilities where structures tend to be similar and feature matches even between non-corresponding images are likely. This problem is enhanced by the fact that the feature matching is tuned to be fairly tolerant in order to cope with varying illumination conditions.

Therefore, in a second step, a matching algorithm based on global geometric image properties was implemented. Therefore we apply the so called weak perspective conditions, i.e. we assume that all feature points lie on a single plane. Then by standard techniques [4] the homography matrix, which relates the coordinate system between the two planes (respective images) can be extracted. In fact, as the extraction of the homography matrix is based on a random sampling (RANSAC) approach for the feature points, only the subset of points fulfilling the condition to be in (or close) to a single plane will be selected. Using the homography matrix the database image is transformed ("warped") onto the coordinate system of the live image (Fig. 3, top right). Now both for the warped database image as well as for the live image the edges are extracted through a Hough like transform and classified according to their directions (Fig. 3, bottom). For the edge direction binning methods similar to the HOG classification scheme [5] are used, represented by the different pseudo colors for the different edge orientations. Then edges of the same directions are correlated according to their spatial proximity and their orientation, and the matching edge segments between the two images are determined. Fig. 3, bottom center, represents those edges from the live image, which could also be identified in the warped preregistered image. This gives a heuristic overall correlation coefficient and allows for a robust discrimination of possible matches between the current image and images in the database.

Once a live image is identified to correspond to the same scene as a picture in the database, the position of the preregistered image is compared to the current position determined from the inertial sensors of the AIONAV system. If both values are not consistent the current position will be corrected accordingly. Here the complementarity of the combined system becomes obvious. As we do not have precise scale information from the camera's perspective transform, we cannot estimate the position along the optical axis from the video image with high precision. However, as the motion is always in direction of the optical axis of the camera (the camera is fixed to the navigator's body) and the error in the IMU data is mainly governed by the heading drift, i.e. the error perpendicular to the optical axis, we can rely on the distance information from the IMU data and only correct the heading drift using the position obtained from the database image. This is illustrated in Fig. 4. There a person wearing the sensor fusion system (AIONAV plus camera) was moving several times back and forth the aisle of a larger office building (the blue arrows (top) correspond to the actual path). The system was initialized with preregistered images of distance granularity of about 2 m. Therefore the same sensor fusion system could be used where the images, together with the manually corrected IMU positions are stored in real time. For the navigation test a large heading drift of the IMU was caused intentionally by decalibrating the magnetic field sensors with a parasitic field created by a small permanent magnet. In real applications this kind of error is frequently

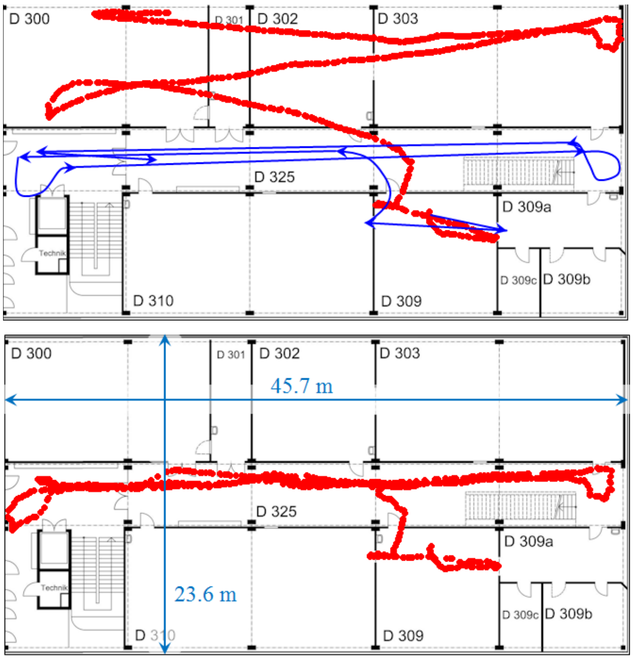

Figure 4. (top) IMU based navigation with heading deviation induced by a parasitic magnetic field (top). The same data corrected with the heading information obtained from the image correlation technique (bottom). The real path (for both cases) is shown as blue arrows (top).

caused by the magnetic fields lines of electrical installations inside the buildings. The induced heading drift leads to strongly corrupted position estimations (red dots, top). At the bottom the same dataset was evaluated using the image correlation technique in conjunction with the discussed heading correction. There the position estimation is always close to the real path with a maximum error of the order of 1 m. The important point to notice is that the position correction based on the image correlation techniques is not deteriorated at long time scales and would work equally well “forever”.

Conclusions

A sensor fusion system was presented combining inertial and video sensors for a semi-autonomous navigation system. It relies at short time on the position data obtained from the AIONAV while at longer times image correlation techniques are used to compensate the heading drift of the inertial sensors. Therefore live video images are identified with preregistered scenes from a database. A two-step correlation and identification process is performed based on feature correspondence identification and subsequent geometric matching using HOG-like edge direction classification. The fusion system was shown to resolve the inherent long time instability problem of inertial sensors and the position estimates remain correct at basically any time scale. In addition the fusion system is auto configurable because the first recognition step of the live camera image with a preregistered database image can be used as position setup value. Potential applications are for navigation or guidance situations where the path or facility can be documented by preregistered images.

Acknowledgment

Indoor positioning research at TU Graz is kindly supported by a group of industry partners (RUAG Electronics AG, Securitron AG, MIGROS, Speedikon Facility Management AG and Building Insurance of Berne), the Swiss Army (armasuisse), the Austrian Army and the Ministry for Traffic, Innovation and Technology, Austria.

References

[1] U. Walder, T. Bernoulli, “Context-Adaptive Algorithms to Improve Indoor Positioning with Inertial Sensors“, Proceedings of the 2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Mautz, R., Kunz, M. and Ingensand, H. (eds.), IEEE Xplore, page 968-973

[2] G. Glanzer, T. Bernoulli, T. Wießflecker and U. Walder, “Semi-autonomous indoor positioning using MEMS-based inertial measurement units and building information,” 6th Workshop on Positioning, Navigation and Communication (WPNC'09), Hannover, 2009.

[3] H. Bay, T. Tuytelaars, L.V. Gool, "SURF: Speeded Up Robust Features", Proceedings of the ninth European Conference on Computer Vision, May 2006.

[4] Z. Zhang, “A flexible new technique for camera calibration,” IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (2000): 1330–1334.

[5] N. Dalal, B. Triggs, “Histograms of oriented gradients for human detection”, in: Computer Vision and Pattern Recognition, San Diego, CA, June 20–25, 2005.