A Stereo Vision System for Pedestrian Navigation

Franjo Cecelja, Wamadeva Balachandran, Rommanee Jirawimut
Systems Engineering Department, Brunel University, Uxbridge, Middlesex, U.K., UB8 3PH
Phone: +44 1895 274 000, ext. 2925, Fax: +44 1895 812 556
Email: mailto:empgrrj@brunel.ac.uk *franjo.cecelja@brunel.ac.uk*

Abstract

In this paper we present the application of a stereo vision system for pedestrian navigation. Corner detection, stereo matching, triangulation, tracking, and robust ego-motion estimation are used to estimate incremental ego-motion of the stereo cameras with particular focus on implementation on pedestrians. A novel robust ego-motion estimation algorithm was utilized to eliminate outliers, which are independent moving features, mismatched features in the stereo matching step and incorrect assigned features in the tracking step. We also introduce a new method based on the knowledge of gait analysis to capture images at the same stage of walking cycle. This leads to less winding trajectory, which can be tracked without increasing order and computational cost of the tracker. The whole navigation process has been experimentally verified.

1.1 INTRODUCTION

A navigation system for blind and visually impaired pedestrians is being developed in the Electronic System and Information Technology Research Group, at Brunel University [1] with the main positioning module utilizing a Differential Global Positioning System (DGPS) [2][3]. However, GPS and DGPS share the same problem of low availability in signal-blocked environments. To solve this problem, the solution was found in the application of a dead reckoning (DR) based on stereo vision system as a visual odometer [4]. In this system, features of images obtained from left and right cameras are detected and matched. With the known geometry between the cameras, the displacement (distance and direction) of the cameras can be estimated from the difference in position of the tracked features in successive frames. However, this system is intended to be used in a dynamic scene where the motion of objects seen by the cameras is the result of camera motion as well as object movement. Moreover, the whole system is mounted on a walking user where the shaking and rotations of the cameras can considerably degrade the performance of feature tracking. We have developed a novel algorithm, which, with the help of additional orientation sensor, provides needed accuracy of navigation.

2 PROPOSED APPROACH

The proposed DR system consists of a set of stereo cameras and an orientation sensor in a form of tri-axial accelerometer for pitch and roll, and tri-axial magnetometer for yaw or heading. The stereo cameras are used to capture left and right images. The pitch signal is used as a trigger for capturing image (Fig. 1).

2.1 Feature Detection, Stereo Matching and Triangulation

In our system, we use the *Plessey Corner Detection* [5] to extract features, which are points of interest, from left and right images. A stereo matching algorithm based on *Singular Value Decomposition* (SVD) [6], which does not require a complex searching algorithm, is then used to search for corresponding points detected in left and right images. The algorithm utilizes proximity and feature similarity principles. The possible matches are then chosen by global maximization based on the SVD algorithm, which automatically uses uniqueness constraint, i.e., a one-to-one matching of left and right point features.

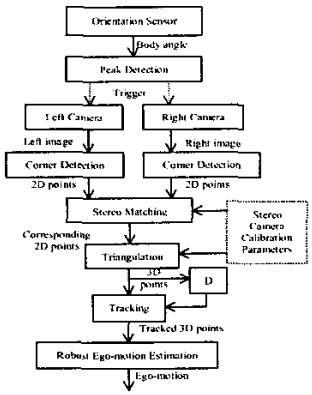


Fig. 1. Overall stereo vision data processing

The stereo cameras are permanently fixed on a rigid base attached to the pedestrian. The left and right images are captured at approximately the same time providing so-called *static stereo analysis* [7]. Hence, both intrinsic and extrinsic parameters can be geometrically calibrated in advance. With these

parameters, the stereo matching algorithms can be improved using equipolar and disparity limit constraint. By doing this, not only the number of correct matched points is increased but the time of computation is also reduced. We use the *optimal triangulation method* [8] to estimate concomitant three-dimensional (3D) position.

2.2 3D Point Tracking

We use 3D tracking based on Kalman-filter multi-target tracking algorithm. The ambiguity of tracking features in clutter (false measurements) gives rise to the *motion correspondence problem*. In the neighborhoods of predicted position, a validation region or gate, based on chi-square distribution of squared *Mahalonobis distance*, is formed to limit searching area [9][10]. From Kalman prediction and update equations, the innovation $\mathbf{v}(k+1)$ is defined as

$$\mathbf{v}(k+1) = \mathbf{z}(k+1) - \hat{\mathbf{z}}(k+1|k) \quad (1)$$

with the innovation covariance $\mathbf{S}(k+1)$ as

$$\mathbf{S}(k+1) = E\{\hat{\mathbf{v}}(k+1)\hat{\mathbf{v}}^T(k+1)\} \quad (2)$$

where $\mathbf{z}(k+1)$ is the true measurement at time $k+1$ and $\hat{\mathbf{z}}(k+1|k)$ is the predicted measurement at time $k+1$ given the past history up to time k. Hence, the searching area can be limited by setting the gate threshold, γ, of the squared Mahalanobis distance, d_m^2. The measurement will be in the gate given by

$$G(k+1,\gamma) = \{\mathbf{z}(k+1) : d_m^2 \le \gamma\} \quad (3)$$

$$d_m^2 = \mathbf{v}^T(k+1)\mathbf{S}^{-1}(k+1)\mathbf{v}(k+1) \quad (4)$$

The problem arises when a detected feature is in more than one gate and when a gate contains more than one detected feature, called the *data association problem* in multi-target tracking. Once the track is initiated, the data association problem has to be solved for track maintenance. The simplest and probably most widely used method is the *Global Nearest Neighbor* (GNN). At each frame, the task is to find the single most likely hypothesis for the assignment of features, according to *Generalized Statistical Distance* (GSD) [10], to existing tracks. The squared GSD, d_g^2, is defined as

$$d_g^2 = \ln\{\mathbf{S}(k+1)\} + d_m^2 \quad (5)$$

The term $\ln\{\mathbf{S}(k+1)\}$ serves as a penalty to a low quality track. The higher GSD implies that it is less likely to be the true feature. In this paper, in addition to the GNN algorithm on 3D-feature, similarity likelihood, which is calculated from the normalized cross-correlation of intensity level of image patches around the 2D-points of the left (and/or right) images in successive frame, is also incorporated. The total posterior is the weighted sum of motion posterior calculated from the squared GSD and similarity posterior calculated from the similarity likelihood. This results in higher detection rate.

2.3 Robust Ego-motion Estimation

In our system the apparent motion of the tracked features in the scene may be the result of (i) the motion from the ego-motion of the cameras on static features in the real scene and (ii) the motion from the ego-motion of the cameras together with the independent motion of the features.

Several algorithms were suggested to detect the independent motion using 2D [11][12][13] or 3D [14][15] features. In addition to the independent moving features, wrong ego-motion estimate may also be the result of mismatched features in the stereo matching step or incorrect assigned features in the tracking step. As a consequence, we can define inliers as features that belong to static objects in the real scene, and outliers as features that belong to moving objects as well as the mismatched and the incorrectly assigned features.

In this paper, we introduce a novel robust ego-motion estimation algorithm, which can virtually segregate all outliers due to mismatched, incorrectly assigned, and independently moving features. This algorithm is robust even though some parameters are not optimized. This is because almost all outliers are eliminated. The principle of the algorithm is simple but effective. Owing to the fact that any sampled set of at least three inliers gives the same 3D-motion estimate (3D rotation and 3D translation), if at least one point in the set is an outlier, the resulting 3D motion will be significantly different. Hence, the RANSAC robust estimator [8] is applied to choose the group with the highest number of 3D features that gives the same 3D-motion estimate. The advantage of this method is that the resulting 3D motion can be directly used in our application as the incremental ego-motion estimate for the DR. Furthermore, the minimum sample size for RANSAC is three instead of five. This means that it is considerably computationally less expensive than the method based on 3D-homography [14].

A widely used method to estimate motion from at least three 3D-points is based on a scalar weighted least square solution [16]. This means that the uncertainty is a constant in every direction at any point in 3D space and this therefore brings about a *spherical* model. In contrast, it is more practical to use an approximate Gaussian distribution *ellipsoidal* model, which results in different uncertainty dependent on the direction and the position of the point in 3D space. As a result of the ellipsoidal covariance, a least square solution is not applicable and a maximum likelihood (ML) solution is therefore used. The following equation shows the cost function to be minimized in ML.

$$c_{ML} = \sum_j \mathbf{e}_j^T \mathbf{C}_j^{-1} \mathbf{e}_j \quad (6)$$

$$\mathbf{e}_j = \mathbf{p}_j^a - (\mathbf{R}\mathbf{p}_j^b + \mathbf{t}) \quad (7)$$

$$\mathbf{C}_j = E\{\mathbf{e}_j \mathbf{e}_j^T\} = \mathbf{C}_j^a + \mathbf{R}\mathbf{C}_j^b\mathbf{R}^T \quad (8)$$

where $\mathbf{R}$ is a 3×3 rotation matrix to be estimated, $\mathbf{t}$ is a 3×1 translation vector to be estimated, and $\mathbf{C}_j^b$ and $\mathbf{C}_j^a$ are the covariance matrices of the j-th before-

movement reconstructed 3D-point, $\mathbf{p}_j^b$, and the j-th after-movement reconstructed 3D-point, $\mathbf{p}_j^a$, respectively. The ellipsoidal Gaussian approximation of $\mathbf{C}_j^b$ and $\mathbf{C}_j^a$ can be found in ref. [17]. Furthermore, instead of using a standard RANSAC distance function, the distance function of each point becomes the squared Mahalanobis distance as shown in equation (9). The RANSAC threshold for discriminating inliers and outliers is determined by the chi-square distribution.

$$d^2_{RANSAC} = \mathbf{e}^T \mathbf{C}^{-1} \mathbf{e} \tag{9}$$

2.4 Image Capturing

Since the stereo cameras are attached to the user, shaking and three-dimensional rotations of the user during walking result in noisy and winding trajectories of the image features [18], which can easily cause a failure of the tracker. The study of gait analysis shows that three rotations of a walking person (roll, pitch and yaw) have repetitive cycles. The rotations on body parts on a transverse plane in one stride cycle are shown in Fig. 2 [19].

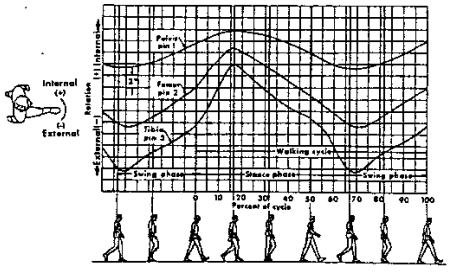


Fig. 2. Rotations of pelvis, femur, and tibia in transverse plane [19]

In this paper, we employ the knowledge of gait analysis to minimize the effect of body rotations. Based on the nature of the repetitive cycles of body angles, the images will be captured only at almost the same stage of the cycle. Hence, the stereo cameras are to capture images at every other trough of the measured pitch signal. It should be noted that we could do this because our application does not need high frame rate. Hence, the captured images will be corresponding to only left (or right) toe-off stages (about 17 or 67 percent of walking cycle), which results in smoother and less winding trajectory of the image features. As a consequence, their features are easier to be tracked by the Kalman filter-based tracker without additional angular velocity states. This again leads to less computational cost.

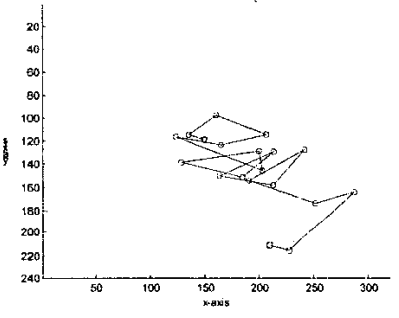
3 RESULTS

In all experiments the stereo baseline, which is the distance between the left and right cameras, was nine centimeters.

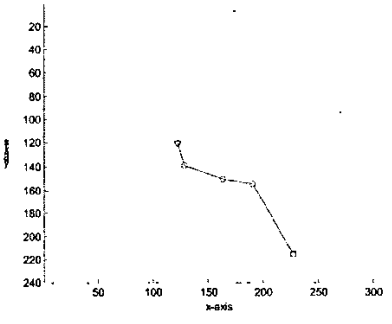
3.1 Effect of winding trajectory

To demonstrate the less winding trajectory, we capture a 320x240 image sequence at 150 ms sampling period. The first image of the sequence with a landmark on the upper left corner of a printer is shown in Fig. 3. The winding trajectory of the landmark in the same sequence re-sampled at 600 ms is shown in Fig. 4(a). In comparison, in Fig. 4(b), the trajectory of the same landmark re-sampled at every other trough is less winding and therefore easier to be tracked.

Fig. 3. The landmark (upper left corner of the printer)



a)



(b)

Fig. 4. (a) The trajectory re-sampled at 600 ms and (b) the trajectory at every even trough

3.2 3D Tracking

This section shows the result of tracking on 640x480 stereo images captured at every other trough of pitch signal. Fig. 5 shows the left images at frame number 1 - 4. In Fig. 6, it can be seen that although the 3D tracking algorithm works well, there are still some incorrect tracks. The problems are the stereo mismatching, which leads to wrong depth estimation, and the incorrectly assigned measurements. After applying the robust ego-motion estimation, the tracks with these outliers are completely removed. Consequently, in the next frame, these incorrect tracks will not increase the chance of incorrect tracking by stealing measurements of other tracks.

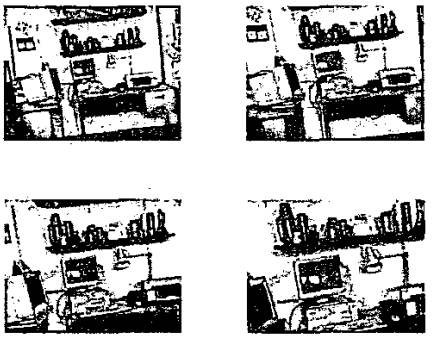


Fig. 5. Left images captured at every other trough: frame number 1, 2, 3 and 4 in order

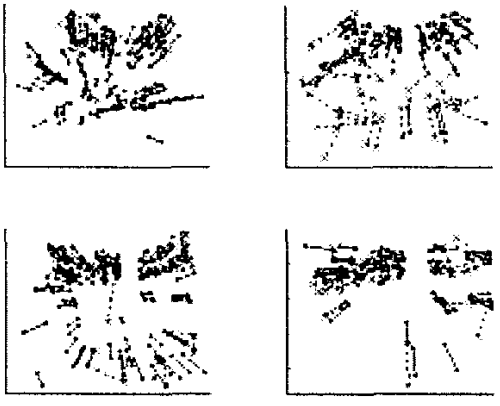


Fig. 6. Result of tracking of image sequence captured at every other trough: dot-solid lines show tracks with inliers, cross-dash lines show tracks with outliers, which are removed. In order, figures show tracks of frame 1&2, frame 2&3, frame 3&4, and frame 4&5

4 SUMMARY

In this paper, we presented the use of a stereo camera system as a visual odometer for pedestrian navigation. The problem of winding trajectory as a consequence of body rotations was reduced using the knowledge of gait analysis. The resulting trajectory was less winding and therefore easier to be tracked. As the system must be used in a dynamic environment containing independent moving object, we introduced our robust ego-motion estimation algorithm to solve this problem. The results showed that not only the tracks with independent moving objects but also those with stereo mismatched and incorrectly assigned measurement are removed. This leads to much more accurate ego-motion estimation.

5 REFERENCES

[1] Garaj, V., F. Cecelja, and W. Balachandran. *The Brunel Navigation System for Blind.* in *ION GPS 2000.* 2000. Salt Lake City, Utah, USA.
[2] Ptasinski, P., et al. An Assessment of DGPS Performance in Personal Navigation and Location Systems. in ION GPS 2000. 2000. Salt Lake City, Utah, USA.
[3] Ptasinski, P., et al., *Brunel Inverse DGPS Positioning System.* NAVIGATION, Journal of the Institute of Navigation, USA (accepted).
[4] Olson, C.F., et al. Stereo Ego-motion Improvements for Robust Rover Navigation. in ICRA. 2001.
[5] Harris, C.G. and M. Stephens. A Combined Corner and Edge Detector. in Proceedings of the 4th Alvey Vision Conference. 1988.
[6] Pilu, M. A direct method for stereo correspondence based on singular value decomposition. in Proceedings. 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Comput. Soc. 1997.
[7] Klette, R., K. Schluns, and A. Koschan, *Computer vision: three-dimensional data from images.* 1998: Springer-Verlag.
[8] Hartley, R. and A. Zisserman, *Multiple view geometry in computer vision.* 2000, Cambridge: Cambridge University Press. 607.
[9] Bar-Shalom, Y., Multitarget-Multisensor Tracking: Advanced Applications. 1990, New York: Artech House.
[10] Blackman, S. and P. Robert, *Design and analysis of modern tracking systems.* Artech House radar library. 1999, Boston, Mass., London.: Artech House. xxxi,1230p. ill. 24cm.
[11] Argyros, A., et al. Qualitative detection of 3D motion discontinuities. in Proceedings of the 1996 IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS 96. Robotic Intelligence Interacting with Dynamic Worlds. IEEE. Part vol.3,. 1996.
[12] Shi, J. and J. Malik. Motion segmentation and tracking using normalized cuts. in Sixth International Conference on Computer Vision. Narosa Publishing House. 1998.
[13] Torr, P.H.S. and D.W. Murray, *Statistical detection of independent movement from a moving camera.* Image & Vision Computing, 1993. **11**(4): p. 180-7.
[14] Demirdjian, D. and R. Horaud, *Motion-egomotion discrimination and motion segmentation from image-pair streams.* Computer Vision & Image Understanding, 2000. **78**(1): p. 53-68.
[15] Zhang, Z. and O.D. Faugeras, *Three-dimensional motion computation and object segmentation in a long sequence of stereo frames.* International Journal of Computer Vision, 1992. **7**(3): p. 211-41.
[16] Weng, J., P. Cohen, and N. Rebibo, *Motion and structure estimation from stereo image sequences.* IEEE Transactions on Robotics & Automation, 1992. **8**(3): p. 362-82.
[17] Matthies, L. and S.A. Shafer, *Error modeling in stereo navigation.* IEEE Journal of Robotics & Automation, 1987. **RA-3**(3): p. 239-48.
[18] Molton, N.D., Computer Vision as an Aid for the Visually Impaired, in Robotics Research Group, Department of Engineering Science. 1998, Oxford University. p. 212.
[19] Inman, V.T., H.J. Ralston, and F. Todd, *Human walking.* 2nd ed. 1994, Baltimore ; London: Williams & Wilkins. 263.