Visual Odometer for Pedestrian Navigation

Rommanee Jirawimut, Simant Prakoonwit, *Member, IEEE*, Franjo Cecelja, and Wamadeva Balachandran, *Senior Member, IEEE*

***Abstract*—This paper presents a visual odometer system using stereo cameras for pedestrian navigation. A novel method for pedestrian navigation based on the knowledge of gait analysis and robust ego-motion estimation is proposed. Two major problems of implementing the system on a pedestrian are stated. Firstly, the features collected from cameras attached on a walking pedestrian normally have winding trajectory resulting in inaccurate tracking. Secondly, the observed object moving independently leads to incorrect ego-motion estimation. Using gait analysis, capturing images at the same stage of the walking cycle produces a less winding trajectory that allows tracking without stabilizing the images. Robust ego motion is also introduced to eliminate outliers that are independently moving features, mismatched features in the stereo matching step and incorrectly assigned features in the tracking step. Data processing techniques including corner detection, stereo matching, triangulation, tracking, and ego-motion estimation are employed. The outcome is the estimated incremental ego motion of the stereo cameras. The approach not only enables the system to operate on walking users but also improves the accuracy of ego-motion estimation.**

***Index Terms*—Blind and visually impaired, human gait analysis, pedestrian navigation, stereo vision, visual odometer.**

I. Introduction

A navigation system for blind and visually impaired pedestrians, namely Brunel Navigation System for Blind (BNSB), is being developed in the Electronic System and Information Technology Research Group, Department of Systems Engineering, Brunel University [1], [2]. The main positioning module utilizes a differential global positioning system (DGPS), which is the global positioning system (GPS) operated in a differential mode, in order to provide more acceptable positioning accuracy for the blind user [3]–[5]. However, GPS and DGPS share the same problem of low availability in signal blocked environments. To solve this problem, a dead reckoning (DR) system has been integrated with the DGPS for determining positions of the user in such environments [6], [7].

In order to estimate the current position of a pedestrian from the last known GPS or DGPS position using DR calculation, the distance and direction have to be known. In general, the direction is obtained from magnetic compass and/or rate gyroscope. The distance is calculated by multiplying the step size average, which is estimated when the GPS signal is available, with the number of steps, which is obtained from an electronic pedometer [6], [7]. However, the problem is that the user may not be walking at the same velocity as when the GPS signal was available. It also suffers from step size error when the user is walking up/downstairs or along sloped terrain. Moreover, the magnetic compass is not robust against magnetic disturbance.

It is, therefore, necessary to integrate another set of sensors to enhance the accuracy of the direction and distance information. One of the possible choices is to use a stereo vision system as a visual odometer (incremental stereo ego motion [8]). In this system, features obtained from left and right images are detected and matched. With the known geometry between the cameras, the positions of the features relative to the cameras can be calculated. Hence, the ego motion (rotation and translation) of the cameras can be estimated from the difference in position of the tracked features in successive frames. In long distance navigation, the visual odometer was developed for a wheeled robot to navigate itself to a destination, which was not in the scene. Combining this visual odometer with periodic updates from an absolute orientation sensor, the system can provide a robust positioning solution with a very low rate of accumulated error growth. However, this robot is developed for a static scene on Mars where no moving object exists in the scene. On the contrary, our system is designed for a dynamic environment where independent moving objects are typical. Consequently, the motion of objects seen by the cameras is the result of camera motion as well as object movement. The problem is also complicated by the shaking and rotations of the cameras caused by the walking user. Therefore, the performance of unmodified algorithms for wheeled robot may be degraded considerably in such circumstances.

In addition to the visual odometer, information from the stereo cameras is also useful in the navigation system for the blind and partially sighted. For example, the stereo vision system can be used for obstacle detection [9], kerbs, steps and staircase detection [10], and zebra crossing detection [11]. Images can also be transmitted back to a human operator at the navigation service center for both security and navigation purposes [1].

II. Proposed Approach

The proposed visual odometer system consists of a set of stereo cameras, an orientation sensor (tri-axial accelerometer for pitch and roll, and tri-axial magnetometer for yaw or heading), and a small portable computer. The stereo cameras are used to capture left and right images according to the pitch signal (see Section II-D). The overall data processing is shown in Fig. 1.

A. Feature Detection, Stereo Matching, and Triangulation

In our system, we use the *Harris* or *Plessey Corner Detection* [12] to extract features, which are points of interest, from

Manuscript received June 15, 2002; revised February 5, 2003. This work was supported by the Electronic Systems and Information Technology Research Group, Department of Systems Engineering, Brunel University, Uxbridge, Middlesex, U.K.

The authors are with the Systems Engineering Department, Brunel University, Uxbridge, Middlesex, U.K.

Digital Object Identifier 10.1109/TIM.2003.815996

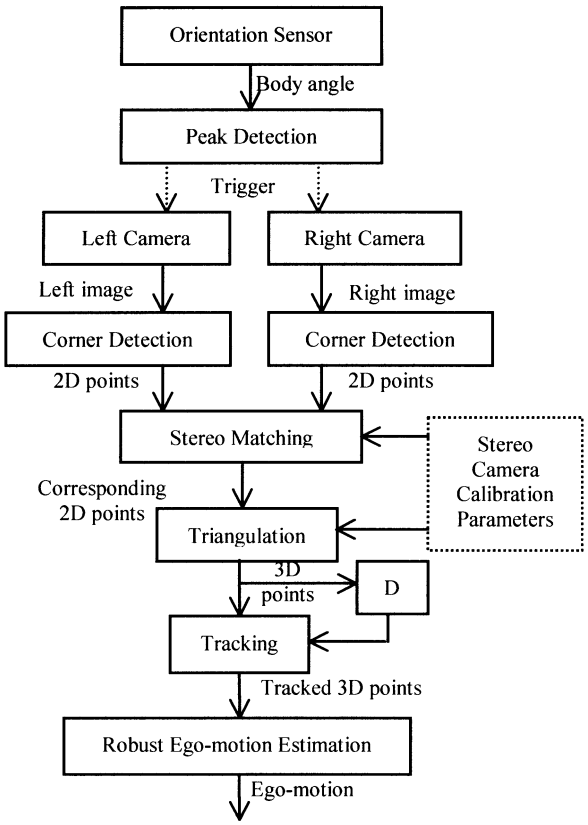


Fig. 1. Overall data processing.

left and right images. A stereo matching algorithm is then used to solve *stereo correspondence problem*, which is the search for the optimal corresponding features detected from left and right images. In this paper, a stereo matching algorithm based on *singular value decomposition* (SVD) [13], which does not require a complex search algorithm, is chosen. The algorithm utilizes proximity and feature similarity principles. The possible matches are then selected by global maximization based on the SVD algorithm, which implicitly incorporates uniqueness constraint, i.e., a one-to-one matching of left and right features.

In our system, the stereo cameras are permanently fixed on a rigid base attached to the pedestrian so that the geometry between the stereo cameras cannot be changed as shown in Fig. 2. In addition, the left and right images are captured at approximately the same time. This is so-called *static stereo analysis* [14]. Hence, both intrinsic and extrinsic parameters [15] can be geometrically calibrated in advance. With these parameters, the stereo matching algorithms can be improved using epipolar and disparity limit constraints [14]. By doing so, not only the number of correct matched points is increased but the time of computation is also reduced.

Given the matched points, the three-dimensional (3-D) position of the matched points with respect to the cameras can be estimated using a triangulation algorithm. In this paper, the *optimal triangulation method* [16] is used.

B. Three–Dimensional Point Tracking

In a stereo vision system, the features can be tracked in either two dimensions (2-D) [17]–[21] or 3-D [22]. Since the underlying motion of features appeared to the cameras are in 3-D, it is more accurate to perform 3-D point tracking. Consequently, the estimated dynamics can be used effectively in Kalman-filter based multi-target tracking algorithms.

The ambiguity of tracking features in clutter (false measurements) gives rise to the *motion correspondence problem*. In the neighborhoods of a predicted position, a validation region or *gate*, based on squared *Mahalonobis distance*, is formed to limit the search area [23], [24]. From Kalman prediction and update equations, the innovation $\boldsymbol{\upsilon}$ is defined as

$$\boldsymbol{\upsilon}_{ij}(k+1) = \mathbf{z}_i(k+1) - \hat{\mathbf{z}}_j(k+1|k) \tag{1}$$

where $\mathbf{z}_i(k+1)$ is the i^{th} true 3-D position measurement at frame $k+1$ and $\hat{\mathbf{z}}_j(k+1|k)$ is the predicted 3-D position measurement of the j^{th} track at frame $k+1$ given the past history up to frame k. Hence, the search area can be limited by setting the *gate threshold* γ, of the squared Mahalanobis distance d^2. The measurement $\mathbf{z}_i(k+1)$ is in the gate G_j of the j^{th} track given by

$$G_j(k+1,\gamma) = \left\{\mathbf{z}_i(k+1) : d_{ij}^2(k+1) \le \gamma\right\} \tag{2}$$

$$d_{ij}^2(k+1) = \boldsymbol{\upsilon}_{ij}^T(k+1)\mathbf{S}_j^{-1}(k+1)\boldsymbol{\upsilon}_{ij}(k+1) \tag{3}$$

where $\mathbf{S}_j(k+1)$ is the innovation covariance of the j^{th} track.

The *gate probability* (P_G), which is the probability that the true measurement will fall in the gate, is obtained from the cumulative probability distribution of the chi-square distribution for the gate threshold γ and with the degree of freedom n_z equal to three (3-D position measurement), as shown in the following equation:

$$P_G = F_{n_z}(\gamma) = \int_0^{\gamma} \chi_{n_z}^2(\xi)d\xi. \tag{4}$$

Defining the gate probability P_G, which is typically chosen as 0.95, the gate threshold γ can then be calculated from the inverse function as $\gamma = F_{n_z}^{-1}(P_G)$. Nevertheless, problems arise when a detected feature is in more than one gate or when a gate contains more than one detected features. This is so-called the *assignment problem*. The simplest and probably most widely used method to this multiple target tracking problem is the *global nearest neighbor* (GNN) [24]. At frame k, a *data association* matrix with rows and columns representing the detected features and the existing tracks is formed by calculating the squared *generalized statistical distance* (GSD) of the detected features to the tracks. The squared GSD for the i^{th} feature to the j^{th} track defined as

$$g_{ij}^2(k+1) = \ln\{|\mathbf{S}_j(k+1)|\} + d_{ij}^2(k+1). \tag{5}$$

The term $\ln\{|\mathbf{S}_j(k+1)|\}$ serves as a penalty to the j^{th} track since it grows with the uncertainty of the track. The higher GSD implies that it is less likely to be the true feature, i.e., less probability to be assigned to the track. In this paper, extension to the GNN algorithm based on texture similarity is introduced. The data association matrix is replaced by a new score matrix. Element of the matrix is based on motion and texture similarity scores. The motion score m_{ij} is defined in (10) and the texture

Fig. 2. Mounting locations of stereo cameras and orientation sensor (3-DM module) on the pedestrian.

score s_{ij} in (8). The texture score is essentially the normalized cross-correlation coefficient of intensity level of image patches around the 2-D points of the left images in frames k and $k+1$. The total score p_{ij} in (6) is the weighted sum of normalized motion score $\overline{m}_{ij}$ and normalized texture score $\overline{s}_{ij}$

$$p_{ij} = w\overline{s}_{ij} + (1-w)\overline{m}_{ij} \tag{6}$$

$$\overline{s}_{ij} = \frac{s_{ij}}{\sum_{i=1}^{I}\sum_{j=1}^{J} s_{ij}}, \quad \overline{m}_{ij} = \frac{m_{ij}}{\sum_{i=1}^{I}\sum_{j=1}^{J} m_{ij}} \tag{7}$$

$$s_{ij} = \frac{\sum_{u=x-\frac{P}{2}}^{x+\frac{P}{2}}\sum_{v=y-\frac{P}{2}}^{y+\frac{P}{2}}\left[\left(I_{uv}(k)-\mu(k)\right)\left(I_{uv}(k)-\mu(k+1)\right)\right]}{P^2\sigma(k)\sigma(k+1)} \tag{8}$$

$$\mu(k) = \frac{1}{P^2}\sum_{u=x-\frac{P}{2}}^{x+\frac{P}{2}}\sum_{v=y-\frac{P}{2}}^{y+\frac{P}{2}} I_{uv}(k),$$

$$\sigma(k) = \sqrt{\sum_{u=x-\frac{P}{2}}^{x+\frac{P}{2}}\sum_{v=y-\frac{P}{2}}^{y+\frac{P}{2}}\left(I_{uv}(k)-\mu(k)\right)} \tag{9}$$

$$m_{ij} = \max\left\{g_{11}^2,\ldots,g_{IJ}^2\right\} - d_{ij}^2 \tag{10}$$

where w is a weight parameter, $P \times P$ is the size of image patched around the detected 2-D point (x, y), $I_{uv}(\mathrm{k})$ is the image intensity at the point (u, v) at time k. The normalization in (7) limits the score to zero to one. Equation (10) shows that the motion score is monotonically decreasing with the squared GSD given in (5). With this score matrix, maximization on the assignment is required. Using SVD based algorithm as described in Section II-A gives the optimal assignments of the detected features to the existing tracks.

C. Robust Ego-Motion Estimation

It should be noted that the main difference of our system to other systems based on stereo cameras is that the apparent motion of the tracked features in the scene may be the result of two cases. The first is the motion from the ego motion of the cameras on static features in the real scene. Second is the motion from the ego motion of the cameras together with the independent motion of the features, which brings about wrong ego-motion estimate.

Several algorithms were suggested to detect the independent motion. Some of these that can be used for either 2-D or 3-D features, utilize 2-D optical flow [25], [26], which is computationally expensive. Some algorithms are based on the assumption that the underlying motion of the detected 2-D feature can be approximated by a linear or affine vector field [27].

On the other hand, in a stereo camera system, once the features are reconstructed, their 3-D properties can be further employed for motion segmentation. An ego-motion and motion segmentation algorithm based on 3-D homography and RANSAC robust estimator [16] was suggested in [28]. The disadvantage of this algorithm is that the sampled set of RANSAC must have at least five tracked 3-D pairs. This leads to a high number of required samples and therefore computational cost. Moreover, the estimated 3-D homography is not meaningful or of any use in our application. Another algorithm that utilizes 3-D properties is reported by Zhang [29]. The principle is based on the fact that 3-D structures having the approximately similar kinematic parameters should belong to the same underlying object and can be grouped together for estimating the ego motion. The kinematic parameters include 3-D translation velocity, 3-D translation acceleration and 3-D angular velocity.

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 incorrectly assigned features in the tracking step. As a consequence, we can define inliers as

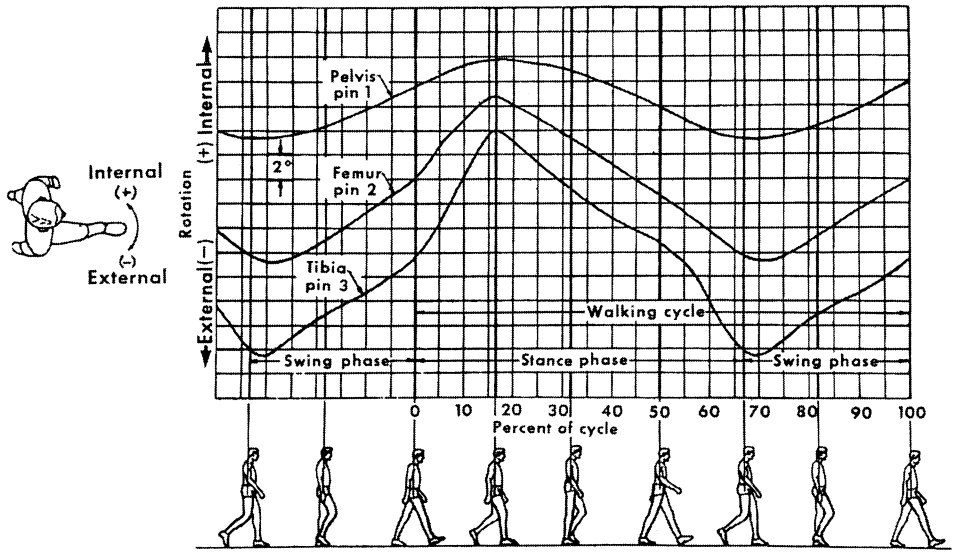


Fig. 3. Rotations of pelvis, femur, and tibia in transverse plane: composite curves of 19 young male adults.

features belonging to static objects, and outliers as features belonging to moving objects as well as the mismatched and the incorrectly assigned features.

In this paper, we introduce a robust ego-motion estimation algorithm, which can virtually segregate all outliers due to mismatched, incorrectly assigned, and independently moving features. This algorithm is robust against these groups of outliers. The idea is simple but effective. As any sampled set of at least three inliers gives the same 3-D motion estimate (3-D rotation and 3-D translation), if at least a point in the set is an outlier, the resulting 3-D motion will be significantly different. Hence, the RANSAC robust estimator is applied to choose the group with the highest number of features that gives the same motion estimate. The advantage of this method is that the resulting motion can be used directly in our application as the incremental ego-motion estimate for the DR. Furthermore, the minimum sample size for RANSAC is three rather than five. This means it is considerably less computationally expensive than the method based on 3-D homography.

A widely used error uncertainty estimate of the reconstructed 3-D point is a *spherical* error covariance [30]. This means that uncertainty is a constant in every direction at any point in 3-D space and mutually uncorrelated. In contrast, it is more accurate to use an approximate Gaussian distribution *ellipsoidal* error covariance model reported in [31], which results in different uncertainty dependent on the direction and the position of the point in the 3-D space. In addition, the errors between directions are no longer uncorrelated. The ellipsoidal model and the correlated error mean that a standard RANSAC distance function cannot be assumed to be a chi-square distribution, which is necessary for discriminating inlier and outlier under RANSAC algorithm [16]. Therefore, we introduce the Mahalanobis distance h_m^2, which brings about the chi-square distribution. As a consequence, the inlier and outlier discrimination for the m^{th} testing tracked 3-D pairs $\mathbf{p}_m(k)$ and $\mathbf{p}_m(k+1)$ in frame k and $k+1$ with distance threshold τ becomes

$$\begin{matrix} h_m^2 < \tau^2; \text{ inlier} \\ h_m^2 \geq \tau^2; \text{ outlier} \end{matrix}, \ \tau^2 = F_d^{-1}(\alpha) \tag{11}$$

Fig. 4. Landmark (upper left corner of printer) is shown by dot and circle.

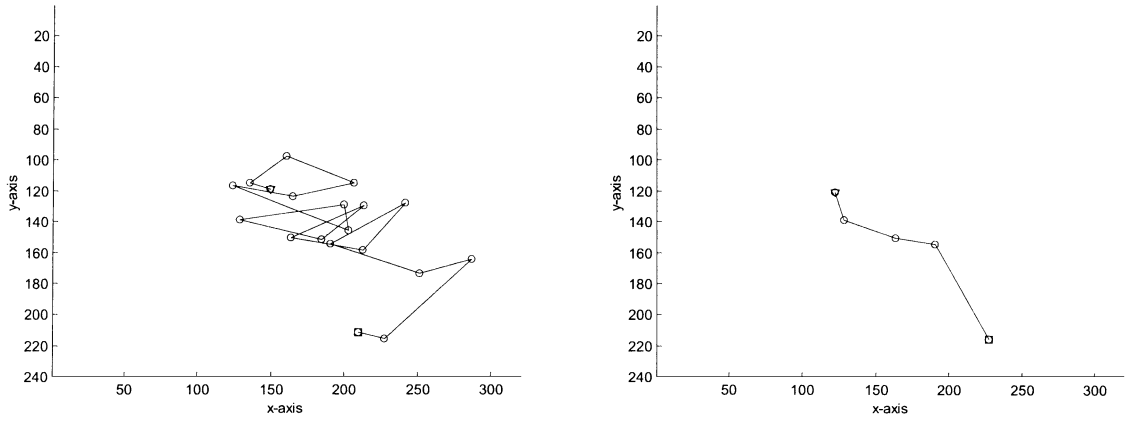
$$h_m^2 = \mathbf{e}_m^T \mathbf{C}_m^{-1} \mathbf{e}_m \tag{12}$$

$$\mathbf{e}_m = \mathbf{p}_m(k+1) - (\mathbf{R}(\boldsymbol{\theta}_n)\mathbf{p}_m(k) + \mathbf{t}_n) \tag{13}$$

$$\mathbf{C}_m = E\{\mathbf{e}_m \mathbf{e}_m^T\} = \boldsymbol{\Sigma}_m(k+1) + \mathbf{R}^T(\boldsymbol{\theta}_n)\boldsymbol{\Sigma}_m(k)\mathbf{R}(\boldsymbol{\theta}_n) \tag{14}$$

where $F_d^{-1}(\alpha)$ is the inverse cumulative chi-square distribution with co-dimension d and the probability α that the chosen pair is an inlier. The probability α is typically set as 0.95 and the co-dimension for estimating the motion is three. $\Sigma_m(k)$ and $\Sigma_m(k+1)$ are the ellipsoidal error covariance of $\mathbf{p}_m(k)$ and $\mathbf{p}_m(k+1)$, respectively. The total number of testing 3-D pairs M is equal to $L-3$, where L is the number of tracked pairs. $\boldsymbol{\theta}_n = [\theta_x\ \theta_y\ \theta_z]_n^T$ and $\mathbf{t}_n = [t_x\ t_y\ t_z]_n^T$ are the orientation and translation vectors calculated from the n^{th} sampled set of size three using least square solution [30]. $\mathbf{R}(\boldsymbol{\theta})$ is a 3×3 rotation matrix function of $\boldsymbol{\theta}$. The number of sampled sets N can be calculated adaptively [16]. The inliers are chosen from the sampled set that gives the highest number of them.

After that, the ego motion can be estimated from these inliers. The ellipsoidal error covariance is used for the estimation. As a result of the ellipsoidal error covariance, a least square solution for estimating motion is not applicable and a maximum likelihood (ML) solution [31] is therefore used. As shown in (15),



(a) (b)

Fig. 5. The trajectory of a landmark. (a) The trajectory of a landmark re-sampled at 600 ms and (b) the trajectory of the same landmark at every even trough.

the likelihood l is the joint conditional probability density of the error vectors

$$l = f(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_Q | \mathbf{R}(\boldsymbol{\theta}), \mathbf{t}) = \prod_{q=1}^{Q} \frac{1}{\sqrt{|2\pi \mathbf{C}_q|}} e^{-\frac{1}{2}\mathbf{e}_q^T \mathbf{C}_q^{-1} \mathbf{e}_q}. \tag{15}$$

It can be seen that the likelihood is directly proportional to $-(1/2)\mathbf{e}_q^T \mathbf{C}_q^{-1} \mathbf{e}_q$. Therefore, maximizing the likelihood is similar to minimizing the cost function c_{ML} defined in (17)

$$\arg\max_{\boldsymbol{\theta},\mathbf{t}} l \leftrightarrow \arg\min_{\boldsymbol{\theta},\mathbf{t}} c_{\mathrm{ML}} \tag{16}$$

$$c_{\mathrm{ML}} = \sum_{q=1}^{Q} \mathbf{e}_q^T \mathbf{C}_q^{-1} \mathbf{e}_q \tag{17}$$

$$\mathbf{e}_q = \mathbf{p}_q(k+1) - (\mathbf{R}(\boldsymbol{\theta})\mathbf{p}_q(k) + \mathbf{t}) \tag{18}$$

$$\mathbf{C}_q = E\{\mathbf{e}_q \mathbf{e}_q^T\} = \boldsymbol{\Sigma}_q(k+1) + \mathbf{R}(\boldsymbol{\theta})\boldsymbol{\Sigma}_q(k)\mathbf{R}^T(\boldsymbol{\theta}) \tag{19}$$

where $\Sigma_q(k)$ and $\Sigma_q(k+1)$ are the ellipsoidal error covariance of the q^{th} 3-D inlier pair $\mathbf{p}_q(k)$ and $\mathbf{p}_q(k+1)$, respectively, and Q is the number of inliers. A numerical optimization algorithm is employed to find the solution of the optimization problem. (*Levenberg-Marquardt* algorithm [32] was used to obtain the result.) After the new $\boldsymbol{\theta}$ and $\mathbf{t}$ are calculated, the inlier and outlier discrimination are repeated until the maximum number of inliers converges. This normally takes only two iterations.

D. Image Capturing

Since the stereo cameras are attached to the user, three-dimensional rotations of the user during walking result in noisy and winding trajectories of the image features [33], which can easily cause a failure of the tracker. It is possible to solve this problem by including the parameters of the sinusoidal-like trajectories in Kalman filter state variables. However, since the rotations, which cause the winding trajectories, are not ideally sinusoidal, the pure sinusoidal models do not match such underlying motion. An example of nonsinusoidal rotation pattern can be seen from in Fig. 3 [34].

On the other hand, with an additional sensor that can provide rotation measurements, it is possible to remove the winding trajectories by a motion-stabilized platform [35]. Ideally, the resulting stabilized trajectories should have pure translations that can be easily tracked. Using such platform, it is necessary to have the prior knowledge of transformation between the sensor-centered coordinate and the camera coordinate systems. This means that a static calibration of these two systems has to be done before using the visual odometer. Moreover, in the case that the additional sensor is an inertial sensor, a dynamic registration, which is an online calibration, has to be implemented. This is because inertial drift and error are accumulated and their distributions are difficult to model for analytical correction [36]. Similarly, for a compass that relies on earth magnetic field, the compass bias and errors caused by magnetic field disturbance are also difficult to model; therefore, the transformation also has to be calibrated online.

In this work, 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 as shown in Fig. 3, the images should be captured only at almost the same stage of the cycle. Hence, the stereo cameras are set to capture images at every other trough of the pitch signal measured by the orientation sensor. The sensor in this study is a 3-DM sensor module from MicroStrain. In this module, there are an orthogonal array of accelerometers and an orthogonal array of magnetometers. The array of accelerometers forms two sets of inclinometers. The sensor is placed around the waist on the back of the user for the pitch detection, Fig. 2. It should be noted that we could do this because our application does not need high frame rate. 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 track by the Kalman filter-based tracker without those transformation calibrations for stabilizing the images.

III. Results

A. Effect of Winding Trajectory

To demonstrate the less winding trajectory, we captured a 320×240 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. 4. The winding trajectory of the

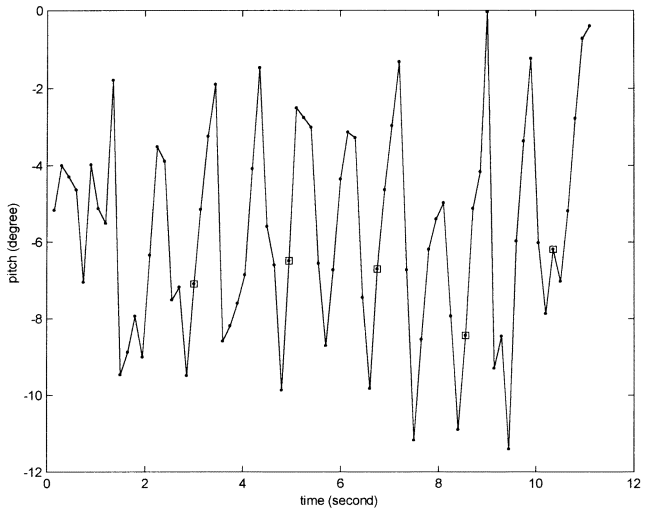


Fig. 6. Pitch sampled at 150 ms and even detected trough (square).

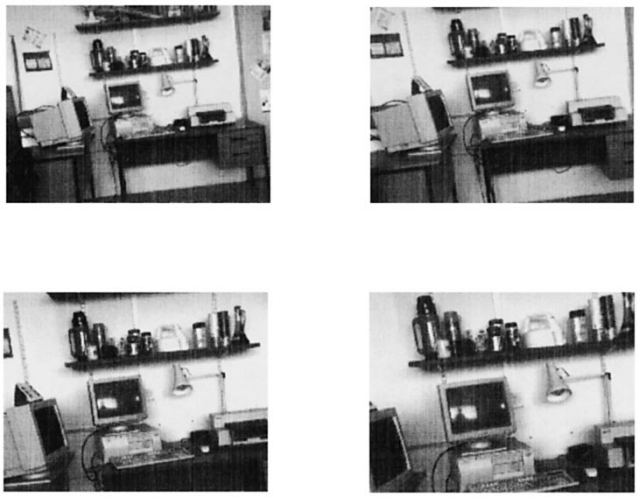


Fig. 7. Left images captured at every other trough. Upper left, upper right, lower left, and lower right sub-figures show left images of frame number 1, 2, 3, and 4, respectively.

landmark in the same sequence re-sampled at 600 ms, which is practical in capturing a pair of 640 × 480 image sequence, is shown in Fig. 5(a). In comparison, in Fig. 5(b), the trajectory of the same landmark re-sampled at every other trough Fig. 6 is considerably less winding and therefore easier to track.

B. Three-Dimensional Tracking

This section shows the result of tracking on 640 × 480 stereo images captured at every other trough of pitch signal. Fig. 7 shows the left images at frame number 1–4. In Fig. 8, it can be seen that although the 3-D 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 were completely removed. Consequently, in the next frame, these incorrect tracks could not steal measurements of other tracks.

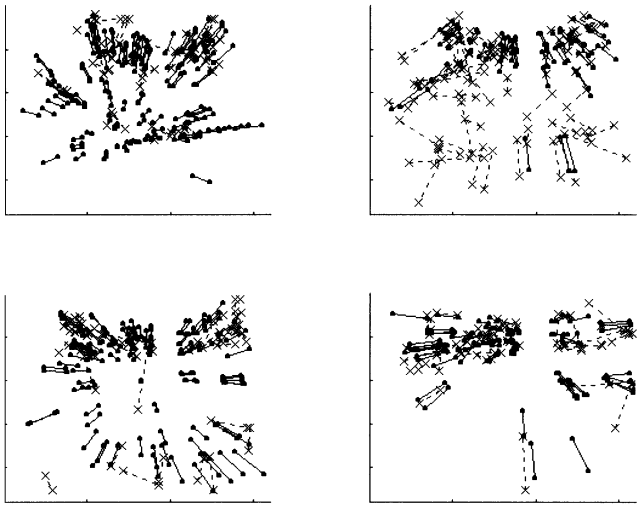


Fig. 8. Result of tracking of image sequence captured at every other trough. Dot-solid lines show tracks with inliers. Cross-dash lines show tracks with outliers, which are removed. Upper left, upper right, lower left, and lower right sub-figures show tracks of frame 1&2, frame 2&3, frame 3&4, and frame 4&5, respectively.

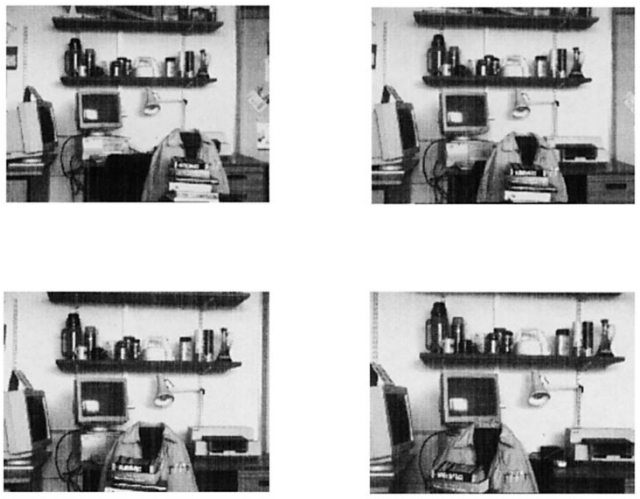


Fig. 9. Upper left, upper right, lower left, and lower right sub-figures show left images of frame number 5, 6, 7, and 8, respectively.

C. Accuracy Test With Independent Moving Object

To test the accuracy, we need to know the true rotation and translation of the stereo system. An experiment was set up on a stereo camera system on a tripod in order to have only translation. The tripod was moved 30 cm forward for eight frames. The left images of frame number 5–8 are shown in Fig. 9. The independent moving objects are on a wheeled chair moving 9 cm per frame from right to left. An example of the result of tracking of frame number 5 and 6 is shown in Fig. 10. It can be seen that all outliers including the independent moving objects can be detected and completely removed.

The *L2 norms* of rotation and translation error vectors are shown in Fig. 11(a) and (b), respectively. The least square ego-motion estimation (LS) on all points including outliers results in the highest error. The maximum likelihood solution on all points (ML) provides a more accurate result for both rotation

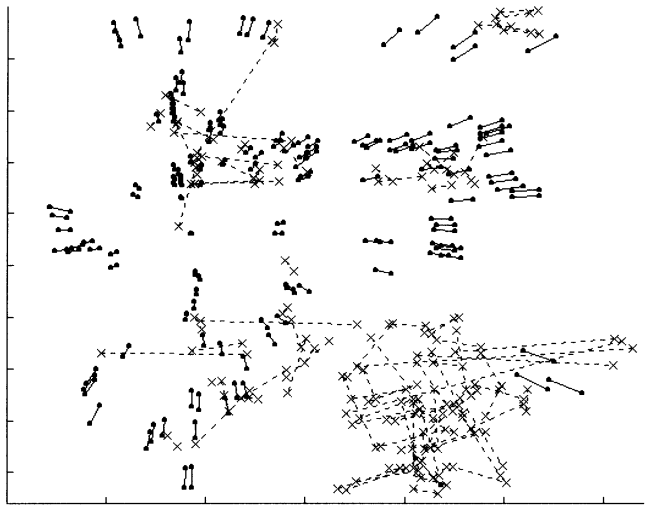


Fig. 10. Example of tracking result of frame number 5 and 6. Dot-solid lines show tracks with inliers. Cross-dash lines show tracks with outliers including tracks of independent moving features, which are removed.

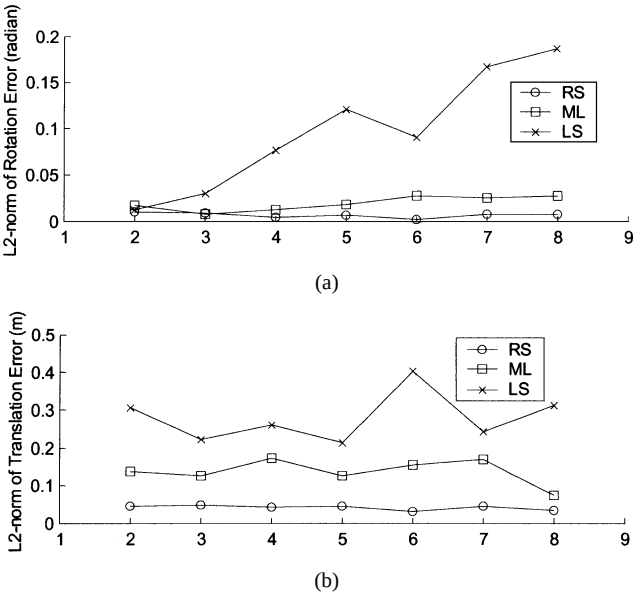


Fig. 11. (a) L-2 norm of rotation error vector and (b) L-2 norm of translation error vectors. The number i on x-axis refers to the results in frame number (i-1)-th and i-th.

and translation. By applying our robust ego-motion estimation algorithm (RS), the resulting rotation and translation errors are minimized.

IV. 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 an independently moving object, we introduced our robust ego-motion estimation algorithm to solve this problem. The results showed that not only the tracks with independently moving objects but also those with stereo mismatched and incorrectly assigned measurements were removed. This leads to more accurate ego-motion estimation.

V. Future Work

Further improvement may be done by modeling the residual winding trajectories in Kalman filter state variables. Since the residual is considerably smaller, a less complex model is required. However, the computational complexity of the Kalman filter will be increased exponentially with the number of state variables.

It should be noted that our stereo baseline, which is the distance between the left and right cameras, is 9 cm. This narrow baseline results in the limitation of the ego-motion estimation accuracy. If the baseline is wider, for instance, about 30 cm, such that the left and right cameras are placed around the shoulders of the pedestrian, the accuracy of the ego-motion estimation can be improved.

Acknowledgment

The authors would like to thank M. Pilu for his SVD stereo matching code.

References

[1] V. Garaj, F. Cecelja, and W. Balachandran, "The Brunel navigation system for blind," in *ION GPS*, Salt Lake City, UT, 2000.
[2] V. Garaj, "The Brunel navigation system for blind: determination of the appropriate position to mount the external GPS antenna on the user's body," in *ION GPS*, Salt Lake City, UT, 2000.
[3] P. Ptasinski, R. Jirawimut, F. Cecelja, and W. Balachandran, "Assessment of DGPS performance in personal navigation and location systems," in *ION GPS*, Salt Lake City, UT, 2000.
[4] P. Ptasinski, M. A. Shah, F. Cecelja, and W. Balachandran, "Use of a mobile telephone as a communication link for correction data transmission in a DGPS system," in *GNSS Global Navigation Satellite Systems Conference*, Edinburgh, U.K., 2000.
[5] P. Ptasinski, R. Jirawimut, F. Cecelja, and W. Balachandran, "Brunel inverse DGPS positioning system," *NAVIGATION, J. Inst. Navigation*.
[6] R. Jirawimut, M. A. Shah, P. Ptasinski, F. Cecelja, and W. Balachandran, "Integrated DGPS and dead reckoning for a pedestrian navigation system in signal blocked environments," in *ION GPS*, Salt Lake City, UT, 2000.
[7] R. Jirawimut, P. Ptasinski, V. Garaj, F. Cecelja, and W. Balachandran, "A method for dead reckoning parameter correction in pedestrian navigation system," in *IEEE Instrum. Meas. Technol. Conf.*, Budapest, Hungary, 2001.
[8] C. F. Olson, L. H. Matthies, H. Schoppers, and M. W. Maimone, "Stereo ego-motion improvements for robust rover navigation," in *ICRA*, 2001.
[9] N. Molton, S. Se, J. M. Brady, D. Lee, and P. Probert, "A stereo vision-based aid for the visually impaired," *Image Vis. Comput.*, no. 4, pp. 251–263, 1998.
[10] S. Se and M. Brady, "Vision-based detection of kerbs and steps," in *Proc 8th British Machine Vis. Conf.*, Colchester, U.K., 1997.
[11] S. Se, "Zebra-crossing detection for the partially sighted," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 2, 2000.
[12] C. G. Harris and M. Stephens, "A combined corner and edge detector," in *Proce. 4th Alvey Vis. Conf.*, 1988.
[13] M. Pilu, "A direct method for stereo correspondence based on singular value decomposition," in *Proc. 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1997.
[14] R. Klette, K. Schluns, and A. Koschan, *Computer Vision: Three-Dimensional Data From Images*, New York: Springer-Verlag, 1998.
[15] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*. Boston, MA: PWS-Kent, 1998.
[16] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2000, p. 607.

[17] I. J. Cox and S. L. Hingorani, "An efficient implementation of Reid's multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, pp. 138–150, Feb. 1996.
[18] C.Cj. Veenman, E. A. Hendriks, and M. J. T. Reinders, "A fast and robust point tracking algorithm," in *Proc. Int. Conf. Image Processing*, vol. 3, 1998.
[19] B. Galvin, B. McCane, and K. Novins, "Robust feature tracking," in *5th Int./Nat. Biennial Conf. Digital Image Computing, Techniques, and Applications (DICTA)* Perth, Australia, 1999.
[20] P. Tissainayagam and D. Suter, "Visual tracking with automatic motion model switching," *Pattern Recognit.*, vol. 34, no. 3, pp. 641–660, 2001.
[21] Y. P. Hung, C. Y. Tang, S. W. Shin, Z. Chen, and W. S. Lin, *A 3D Feature-Based Tracker for Tracking Multiple Moving Objects With a Controlled Binocular Head*. Taipei, Taiwan: Academia Sinica Institute of Information Science, 1995, p. 35.
[22] J. W. Yi and J. H. Oh, "Recursive resolving algorithm for multiple stereo and motion matches," *Image Vis. Comput.*, vol. 15, no. 3, pp. 181–196, 1997.
[23] MAY. Bar-Shalom, *Multitarget-Multisensor Tracking: Advanced Applications*. Norwood: Artech House, 1990.
[24] MAS. Blackman and P. Robert, *Design and Analysis of Modern Tracking Systems*. Norwood: Artech House, 1999.
[25] A.Aa. Argyros, M. I. A. Lourakis, P. E. Trahanias, and S. C. Orphanoudakis, "Qualitative detection of 3D motion discontinuities," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)* , vol. 3, 1996.
[26] J. Shi and J. Malik, "Motion segmentation and tracking using normalized cuts," in *6th Int. Conf. Computer Vision*: Narosa Publishing House, 1998.
[27] P. H. S. Torr and D. W. Murray, "Statistical detection of independent movement from a moving camera," *Image Vis. Comput.*, vol. 11, no. 4, pp. 180–187, 1993.
[28] D. Demirdjian and R. Horaud, "Motion-egomotion discrimination and motion segmentation from image-pair streams," *Comput. Vis. Image Understanding*, vol. 78, no. 1, pp. 53–68, 2000.
[29] Z. Zhang and O. D. Faugeras, "Three-dimensional motion computation and object segmentation in a long sequence of stereo frames," *Int. J. Comput. Vis.*, vol. 7, no. 3, pp. 211–241, 1992.
[30] R. M. Haralick, H. Joo, C. Lee, X. Zhuang, V. G. Vaidya, and M. B. Kim, "Pose estimation from corresponding point data," *IEEE Trans. Syst., Man, Cybern.*, vol. 19, pp. 1426–1446, Dec. 1989.
[31] L. Matthies and S. A. Shafer, "Error modeling in stereo navigation," *IEEE J. Robot. Automat.*, vol. RA-3, pp. 239–248, June 1987.
[32] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 1992.
[33] N. D.Nicholas David Molton, "Computer vision as an aid for the visually impaired," in *Robotics Research Group, Department of Engineering Science*. Oxford, U.K.: Oxford Univ., 1998, p. 212.
[34] V. T. Inman, H. J. Ralston, and F. Todd, *Human Walking*, 2nd ed. Baltimore, MD: Williams & Wilkins, 1994, p. 263.
[35] R. Azuma, B. Hoff, H. Neely III, and R. Sarfaty, "A motion-stabilized outdoor augmented reality system," in *Proc. IEEE Virtual Reality*, 1999.
[36] S. You, U. Neumann, and R. Azuma, "Hybrid inertial and vision tracking for augmented reality registration," in *Proc. IEEE Virtual Reality*, 1999.

Rommanee Jirawimut received the B.Eng. degree from King Mongkut's University of Technology, Thonburi, Thailand, in 1998 and the M.Sc. degree in communications and signal processing from Imperial College of Science, Technology, and Medicine, University of London, London, U.K., in 1999. She is currently pursuing the Ph.D. degree at the Department of Systems Engineering, Brunel University, Middlesex, U.K. She

Simant Prakoonwit (M'00) received the B.Eng. degree in electrical engineering from Chulalongkorn University, Thailand, in 1989 and the M.Sc. degree in communications and signal processing and the Ph.D. degree in computer vision from Imperial College of Science, Technology, and Medicine, University of London, London, U.K., in 1990 and 1996, respectively.

He is currently a lecturer with the Department of Systems Engineering, Brunel University, Middlesex, U.K. Prior to coming to Brunel University, he was a Postdoctoral Research Assistant with the Department of Biological and Medical Systems (now the Department of Bioengineering), Imperial College. His research interests are in computer vision, evolutionary computation, and artificial intelligence.

Dr. Prakoonwit is a member of the IEE.

Franjo Cecelja was born in Krizevci, Croatia, on February 21, 1955. He received the Dipl. Eng. degree in electronics from the University in Zagreb, Croatia, the M.Sc. degree in control from the Cranfield Institute of Technology, Cranfield, U.K., and the Ph.D degree in electrical engineering from Brunel University, Middlesex, U.K.

From 1978 to 1992, he had pursued research in control and sensor technology. In 1992, he joined the University of Surrey, Surrey, U.K., where he was engaged in research on optical sensors and EM interaction with biological systems. He is currently a Lecturer at Brunel University, with research interests in EM radiation, near-field measurements, and general signal processing, which resulted in more than 80 scientific papers.

Wamadeva Balachandran (M'91–SM'96) received the B.Sc. degree from the University of Ceylon, Colombo, Sri Lanka, in 1970 and the M.Sc. and PhD degrees from the University of Bradford, Bradford, U.K., in 1975 and 1979, respectively.

He is a Professor of electronic systems and Head of the Department of Systems Engineering, Brunel University, Middlesex, U.K. He is actively pursuing research in electromagnetic sensors, measurement systems, medical electronics, DGPS navigation systems, electrohydrodynamics, and charge particle dynamics. He has published over 200 papers in these fields and has filed eight patent applications.

Prof. Balachandran is a Fellow of the IEE, Institute of Measurement and Control, Institute of Physics, and the Royal Society of Arts, U.K. He is an editorial board member of the *International Journal of Atomization and Sprays* and the *International Journal of Particle Science and Technology*. He is the past Chairman of the Static Electrification Committee of the Institute of Physics, London, and the International President of the Institute of Liquid Atomization and Spray Systems.