Real-time data fusion on stabilizing camera pose estimation output for vision-based road navigation

Zhencheng Hu* and Keiichi Uchimura
Dept. of Computer Science, Kumamoto University, 2-39-1 Kurokami, Kumamoto, Japan 860-8555

ABSTRACT

This paper presents a novel framework of vision-based road navigation system, which superimposes virtual 3D navigation indicators and traffic signs onto the real road scene in an Augmented Reality (AR) space. To properly align objects in the real and virtual world, it is essential to keep tracking camera's exact 3D position and orientation, which is well known as the Registration Problem. Traditional vision based or inertial sensor based solutions are mostly designed for well-structured environment, which is however unavailable for outdoor uncontrolled road navigation applications. This paper proposed a hybrid system that combines vision, GPS and 3D inertial gyroscope technologies to stabilize the camera pose estimation output. The fusion approach is based on our PMM (parameterized model matching) algorithm, in which the road shape model is derived from the digital map referring to GPS absolute road position, and matches with road features extracted from the real image. Inertial data estimates the initial possible motion, and also serves as relative tolerance to stable the pose output. The algorithms proposed in this paper are validated with the experimental results of real road tests under different road conditions.

Keywords: Data fusion, Registration problem, On-road navigation, Parameterized model matching, Augmented reality

1. INTRODUCTION

Tracking a moving camera's three-dimensional (3D) position and orientation is essential to the so-called registration problem in an Augmented Reality Context. Augmented Reality (AR) supplements reality by interactively superimposing virtual objects upon the real world scene. In contrast, virtual reality immerses a completely computer-generated world. AR technology has been widely applied in many application areas including visualization in surgery navigation, visual reconstruction, human-computer interfaces and navigation [1,2]. In the literature of road navigation, the new generation of navigation systems [3] superimpose virtual direction indicators and traffic information bulletins upon the real road scene to give drivers efficient and direct visual navigation information.

The *Registration problem* is one of the basic problems currently limiting Augmented Reality (AR) applications. The objects in the real and virtual world must be properly aligned with respect to each other, which requires knowing the observer's exact 3D viewing pose (position and orientation) data. Especially when the observer (camera) is moving, accurate estimation of the 3D pose data and tracking the temporal coherence from successive images will absolutely affect the synthesizing accuracy and visual performance of virtual objects in the AR space.

To deal with this problem, many approaches have been proposed in recent years. Previous work in this area can be divided into three main categories: 1) solutions based on external tracking devices like inertial sensors, beacons or transponders, 2) solutions based on image processing technology that directly estimates camera pose from the same imagery observed by the viewer, 3) hybrid solutions attempt to overcome the drawbacks of any single sensing solution.

Inertial sensors are widely used for motion tracking [1]. With the characteristics of self-contained, source-less and high sampling rate, they are suitable for tracking the rapid motions like vehicle or aviation movement. However, since inertial sensors only measure the variation rate or accelerations, the output signals have to be integrated to obtain the position and orientation data. As a result, longer integrated time produces significant accumulated drift because of noise or bias.

The general concept of vision-based camera 3D pose estimation is to find the best set of camera position and orientation data (the six extrinsic parameters) to fit a known model in the target image. When camera intrinsic parameters are given or available from the initial calibration, it is also known as the *absolute orientation problem*. Unlike other sensing technologies, vision solutions directly estimate camera pose from the same imagery that is also used as the real world background, therefore vision solutions always offer the best visual perceived performance when the virtual objects are projected to the background. However, since vision solutions absolutely depend on image feature extraction and tracking result, they also suffer from the high computational cost, sensitive to noise and lack of robustness.

In the first category, Foxlin [4,5] developed an inertial plus compass based tracking system, in which the drift caused by the Gyro data integration can be corrected by the data from inclinometers and a compass. More recently, he employed ultrasonic range finder and a wireless transponder beacons for more accurate pose estimation.

Pure vision solutions can be traced to the affine model proposed by Horn and Weldon [6]. They described the brightness constraints derived from the linear equation of intensity derivatives to rigid body motion parameters. Their solution will encounter with huge blunders when the general depth information is not available. Zhuang and Haralick provided a simplified linear motion parameters estimation algorithm in their early work [7], and it was demonstrated to be very efficient to solve the problem when there is limited noise and no corresponding errors. The algorithm is also considered to be very sensitive to noise and matching errors. In their later work [8], iterative reweighed least squares methods were proposed as a robust solution for the affection of incorrect matching outliers, although its impractical computational cost makes it unavailable to most of the real time applications. Our previous work [14] simplifies the 3D-2D feature matching process into a 2D-2D parameterized model matching, which significantly reduces the computation of camera pose. However it still suffers from the feature-less road shape and the output becomes unstable when feature tracking is failed.

Hybrid solutions are widely applied in recent research works since different sensors can be used to compensate others limitation. Chai et al. [9] employs an adaptive pose estimator with vision and inertial sensors for overcoming the problems of inertial sensor drift and vision sensor slow measurement. The extended Kalman filter (EKF) is used for data fusion and error compensation. You et al. [10] also combined vision and inertial sensor with a two-channel complementary EKF, which can take advantage of the low-frequency stability of vision sensors and the high-frequency tracking of gyro sensors. However, most of these approaches are designed for well-structured environment. Especially for the vision sensors, predefined artificial markers are vital for feature tracking process, which is however unavailable in the outdoor uncontrolled road navigation environment.

In this paper, we extended our previous work of pure vision based solution to a hybrid solution that combines vision, GPS and 3D inertial gyroscope sensing technologies. We still restrict our aim at the registration problem for on-road navigation applications. The fusion approach is based on our PMM (parameterized model matching) algorithm, in which the road shape model is derived from the digital map referring to GPS absolute road position, and matches with road features extracted from the real image. Inertial data estimates the initial possible motion, and also serves as relative tolerance to stable the pose output.

The remainder of this paper is organized as follows. Section 2 quickly reviews the new concept of direct visual navigation system. Theoretical analysis and implementation details of this algorithm are described in Section 3. Road structure constraints and parameterized road shape models are also provided in this section to derive a simplified PMM algorithm. Section 4 provides experimental results on real roads.

2. REVIEW OF DIRECT VISUAL NAVIGATION SYSTEM

Since Pioneer ® introduced the world's first commercial GPS car navigation system in 1990, on-road navigation related research has become one of the most active areas of many new techniques. An on-road navigation system is generally defined as the integrated system that is mainly used to provide location and navigation information to help drivers drive on road. The first generation of on-road navigation systems was equipped with the basic functionalities of 2D map displaying and road positioning. With the development of voice guidance and dynamical traffic information exchange techniques, recent navigation systems will guide you with voice instructions well in advance of your next move along a planned route. However even with the voice guidance and 2D road map, driver still has to compare by himself the road scene with the digital map to determine which lane he should take, or, at which intersection he should turn. It is inconvenient and even dangerous in some cases, especially during the high-speed driving in dense traffic roads.

To overcome these indirect, unsafe and inconvenient navigation problems, a new concept of direct visual navigation and its prototype system – Vision-based Car Navigation System (VICNAS) was proposed [3]. VICNAS employs Augmented Reality technique to superimpose virtual direction indicators and traffic information bulletins into the real driver's view. Therefore it gives more efficient and direct visual guidance to the drivers. Figure 1 provides a prototype driver interface of VICNAS. There are four main components in the graphic overlay: virtual direction indicator and traffic sign, virtual traffic bulletin board and road paint (speed limits), simplified road map and important landmark with icons.

Since all the virtual indicators and overlay graphics have to be aligned properly with the real traffic scene from driver's view, the accuracy of navigation that VICNAS can provide absolutely depends on the accuracy of the estimated viewing pose, which means camera registration accuracy directly determines the visually-perceived performance of AR application.

There are several factors that have to be considered to solve the Registration problem for VICNAS. Road navigation is mainly for passenger vehicles moving in high speed, which gives a fast translation along vehicle's moving axis. No predefined square or circle markers can be put in the wide-open real road scene. Territory map, some landmarks and road shapes are the only features that can be used to determine camera pose. Even a little drift in camera pose estimation will lead a significant displacement of virtual objects on the projected image.

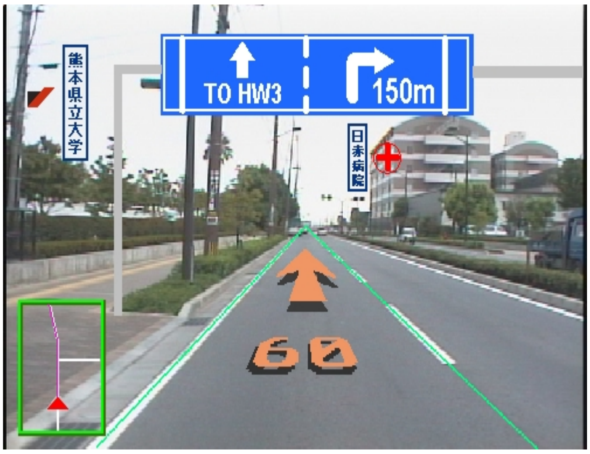


Figure 1. Prototype driver interface of VICNAS system

3. PARAMETERIZED MODEL MATCHING ALGORITHM

Our vision approach shares similar goals with the video-based model tracking solution described by Valinetti [11]. Valinetti introduced a scalar evaluation score based on the local image gradient along the projected model lines to evaluate the existence possibility of certain camera pose values. In our PMM algorithm, we choose road shape as the target model since it can be directly derived from the digital road map and is fairly easy to track in different lighting conditions.

Figure 2 shows the basic block diagram of PMM algorithm. Absolute road position is derived from the fusion of GPS and Gyro data. Road Modeling Block (RMB) uses digital road map data to generate a shape model of roads ahead from this position. It will match with the road features extracted from real image and output the estimation result. The angular rate data obtained from the Gyro sensor initializes the possible motion, and also serves as relative tolerance to stable the final output.

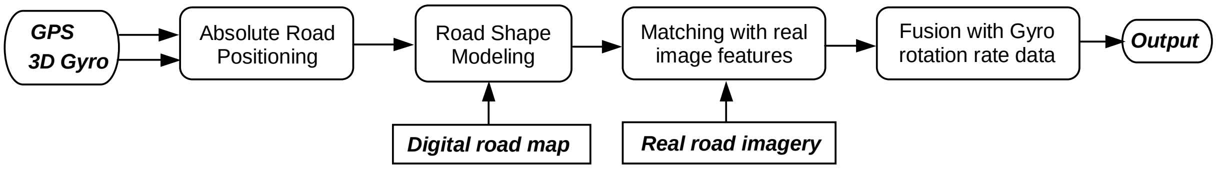


Figure 2. Block diagram of PMM algorithm

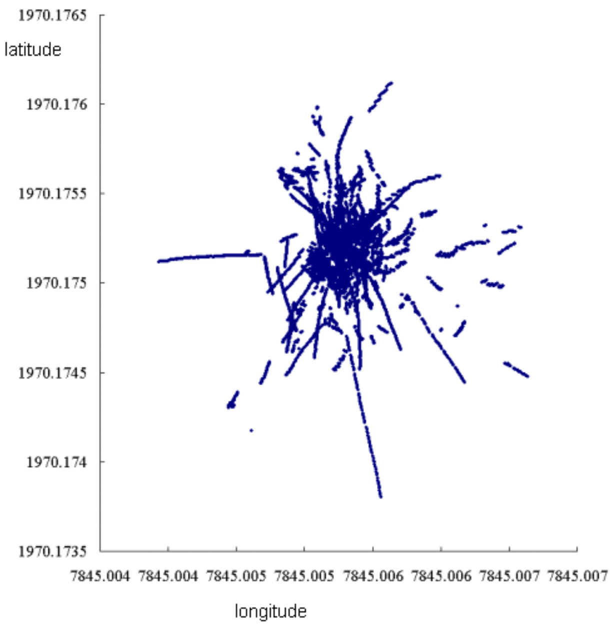


Figure 3. A static accuracy test result of DGPS sensor (GPS receiver: Trimble® AgGPS, duration time > 1 hour)

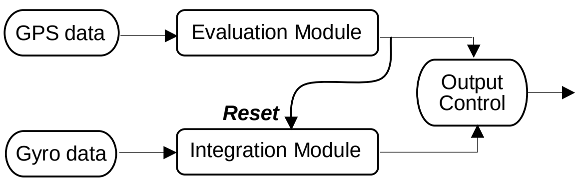


Figure 4. Fusion of GPS and Gyro data

3.1 Fusion of GPS and Gyro Data for Absolute Road Position

In an open, well-communicated environment, accuracy of differential GPS (DGPS) sensor can achieve 1.5m horizontally and 5m in altitude (Figure 3). In urban area, high buildings and signal random reflection (so-called multi-path) will significantly affect GPS accuracy. In this case, inertial sensors are employed to compensate the GPS data. Most commercial navigations systems will use map-matching algorithm to pull the absolute positioning data to the nearest possible road according to moving trace history.

Since GPS sampling rate (1Hz~10Hz) normally is lower than the inertial sensor sampling rate (10Hz~500Hz), the fusion of GPS and 3D gyro for absolute road position is based on a predictor-corrector control theory as shown in Figure 4. GPS data and gyro data are fed into evaluation module and integration module separately. After checking data integrity, captured satellites number and DOP value, every evaluated trustable GPS data will start a new loop and reset gyro's integrating module. The difference between new GPS position and integrated gyro's predication will be fed back into the gyro integration module as a dynamical correction factor. Assuming $P_g(t_i) = (X_{t_i}, Y_{t_i}, Z_{t_i})$ are evaluated trustable GPS position data, where $t_i = t_0, t_1, ..., t_n$, and $V_i(\tau) = (vx_\tau, vy_\tau, vz_\tau)$ are velocity data integrated from 3D gyro's acceleration output. Then the absolute road positioning output between two trustable GPS data $P_g(t_n)$ and $P_g(t_{n+1})$ can be calculated by:

$$P(t) = P_g(t_n) + \left[\int_{t_n}^{t} V_i(\tau) d\tau + \Delta P_{t_n}(t - t_n) \right] \quad (1)$$

where ΔP_{t_n} is the feedback adjustment factor to correct 3D gyro data.

$$\Delta P_{t_n} = \frac{1}{t_n - t_{n-1}} \left\{ P_g(t_n) - \left[\int_{t_{n-1}}^{t_n} V_i(\tau) d\tau + \Delta P_{t_{n-1}}(t_n - t_{n-1}) \right] \right\} \quad (2)$$

3.2 Reference Frames of Vision System

There are five coordinate systems involved in VICNAS: World Coordinate System (WCS), Vehicle Coordinate System (VCS), Camera Coordinate System (CCS), Inertial Coordinate System (GCS) and 2D Projected Image Coordinate System (ICS).

As shown in Figure 5, we assume that the origin of VCS (X_v, Y_v, Z_v) is located at the center of two rear wheels on the ground plane, Z_v is on the vehicle's central axis, X_v, Y_v are pointing left and up respectively. VCS can be treated as the relative WCS with an offset and heading angle on the ground plane. Mapping from CCS to ICS is a perspective projection. We will denote in this paper $\vec{P}_c = [X_c\ Y_c\ Z_c\ 1]^T$ as the homogeneous coordinates of a point P_c in CCS space. The corresponding point in ICS is $\vec{p}_i = [x_i\ y_i\ 1]^T$. A 3x4 matrix $\mathbf{\Gamma}$ represents the perspective projection:

$$\boldsymbol{\kappa}\vec{p}_i = \mathbf{\Gamma}\vec{P}_c \quad (3)$$

where $\boldsymbol{\kappa}$ is an arbitrary scale factor.

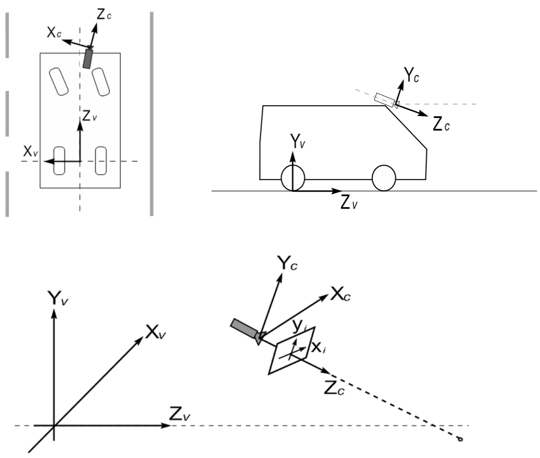


Figure 5. Reference frames of vision system

Matrix $\mathbf{\Gamma}$ can be decomposed into:

$$\mathbf{\Gamma} = K\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} = K[\mathbf{E}\,|\,\mathbf{0}], \tag{4}$$

where matrix K is called camera intrinsic parameters matrix, and normally can be extracted by scale factors (S_x, S_y), principle point (u_0, v_0) and screw factor S_θ:

$$K = \begin{pmatrix} S_x & S_\theta & u_0 \\ 0 & S_y & v_0 \\ 0 & 0 & 1 \end{pmatrix} \tag{5}$$

Relative orientation and position between CCS to VCS can be described as a rigid body translation and rotation. Eq. (6) shows the homogeneous transformation matrix.

$$\vec{P}_c = M\vec{P}_V = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix}\vec{P}_V \tag{6}$$

where R and T are the rotation matrix and translation vector respectively. Since inertial sensor is rigidly mounted to the vehicle, the transformation between frames ICS to VCS and CCS are pre-calibrated and constant. More details are given in Section 3.5.

Combining eqs. (3) to (6) then holds a linear constraint.

$$\kappa\vec{p}_i = \Gamma\,\vec{P}_v = K\,[\mathrm{E}\,|\,0]M\,\vec{P}_v = K\,[\mathrm{E}\,|\,0]\begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix}\vec{P}_v \tag{7}$$

Fixed camera intrinsic parameters can be easily obtained from the initial calibration. Therefore computing the 6 extrinsic parameters in the pose data matrix M becomes the major problem of camera pose estimation. If we collect the 6 camera pose parameters in one vector $\boldsymbol{\sigma}$, we can simply parameterize the perspective mapping relationship between the 2D image coordinates in ICS and the 3D world coordinates in VCS as follows:

$$\vec{p}_i = \mathbf{\Gamma}(\vec{P}_v; \boldsymbol{\sigma}). \tag{8}$$

In general, each matching pairs of ($\vec{p}_i, \vec{P}_v$) will contribute to the determination of camera pose vector $\boldsymbol{\sigma}$. However, in case of on-road navigation, there are no pre-defined markers in the view and it is almost impossible to obtain the environmental depth information with image data alone. To avoid directly matching between ICS and VCS, we proposed a parameterized model matching algorithm.

3.3 Road Shape Model

The general road information embedded in the 2D digital map consists of a list of road skeleton node positions (longitude and latitude values), road segments (between two adjacent road skeleton nodes), intersections information and the associated attributes to the road segments (road name, road construction level, direction information and lanes number in either direction). Lane width is generally fixed and can be determined by the information of road level.

The civil engineering regulations require the connections between road nodes have to be smooth and changing of curvature has to follow the road level and speed limitation. Clothoid [12] is a widely used road shape whose curvature c is proportional to its length l

$$c(l) = c_0 + c_1 l \tag{9}$$

where c_0 is the curvature at the beginning point and c_1 is the rate of change (derivation) for the curvature. Actual roads are – as environmental conditions allow – a set of successive clothoids (including straight lines and circles).

This geometric description of horizontal road shape results in a very compact parameterized multi-lane road shape model of road segment i on WCS (from node i to node $i+1$):

$$\Re_i = (c_{0i}, c_{1i}, n_{li}, n_{ri}, w_i, L_i)^T \text{ ,} \tag{10}$$

where n_{li} is the number of lanes on the node-descent side (from node $i+1$ to node i), n_{ri} is the number of lanes on the node-ascent side (from node i to node $i+1$) and w_i is the average lane width during this segment.

In order to obtain the road shape ahead at any driving location on the road, we have to transfer the road model $\Re_i$ at elapsed time i to a new model $\Re_i'$ in VCS according to the current offset and heading angle, since the road model's origin is based on the road central skeleton line. Assuming road surface in the nearby view is flat and the ground plane is at $Y_v = 0$, together with the perspective mapping eq. (8), the following perspective road model is obtained.

$$\vec{p}(\Re_i') = \mathbf{\Gamma}(\Re_i'; \boldsymbol{\sigma}) \tag{11}$$

3.4 Model Matching

To avoid direct matching between the 3D road model $\Re_i'$ and the 2D image data, we transfer the problem to the optimization of the matching between parameterized road model and the road shape data extracted from the image.

Gray scale correlation is not preferred in this application due to various types and colors of road lane markers, different lighting and weather conditions as well. To counteract this effect, a Road Shape Look-up table (RSL) is employed to give peak values at the position of lane marker, and lower values at their neighbors.

Calculation of RSL is based on the extracting result of road lane markers. Suppose point $p(x, y)$ belongs to the region of extracted lane markers, its RSL value is calculated through the convolution of neighbor pixels with a predefined kernel χ.

$$RSL :\rightarrow p(x, y) = \sum_i \sum_j (\lambda_{x-i, y-j} \chi_{i,j}) \tag{12}$$

where $\lambda_{x,y} = \begin{cases} 1, & p(x, y) \in \textit{lane mark} \\ 0, & \textit{otherwise} \end{cases}$

Kernel elements $\chi_{i,j}$ are set to 1 as a simple smooth filter. Kernel size is generally set to 5 or 7 in order to have a larger base of attempts for the optimum.

Therefore, a normalized camera pose estimation score function can be given as:

$$E(\sigma) = \frac{1}{|\eta_\sigma|} \sum_{p \in \eta_\sigma} \| RSL :\rightarrow p(x, y) \| = \frac{1}{|\eta_\sigma|} \sum_{p \in \eta_\sigma} \| RSL :\rightarrow \Gamma(\Re'; \sigma) \| \tag{13}$$

where $\eta_{\boldsymbol{\sigma}}$ represents the set of points belonging to the revised perspective road model $\Re'$. Every point with a non-zero RSL value will contribute to the score function. In other word, the maximum of estimation score will be reached at the perfect matching of projective road model to the road shapes on the image.

3.5 Searching for the best camera pose data

A direct search algorithm is adopted in the optimization searching operation[11], while the fusion of vision and gyro data gives out its initial state and searching range. Since the gyro data is defined in the inertial coordinate system, it is necessary to convert it to the world coordinate system. Let $\Omega(\theta, \phi, \psi)$ be the absolute rotation angle (Euler angle), and $W(\omega_x, \omega_y, \omega_z)$ represent the angular rate from inertial sensor output. According to [15], the conversion from inertial angular rate to the world will be

$$\Delta\Omega = \begin{bmatrix} \dot{\theta} \\ \dot{\phi} \\ \dot{\psi} \end{bmatrix} = \begin{bmatrix} \cos\phi & 0 & \sin\phi \\ \tan\theta \sin\phi & 1 & -\tan\theta\cos\phi \\ -\sin\phi / \cos\theta & 0 & \cos\phi / \cos\theta \end{bmatrix} \begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix} \tag{14}$$

Eq. (14) is employed to predict the motion of image features. It gives an inertial state in the optimization search of vision approach. Searching range is limited to a dynamical range depending on the angular rate. When the image feature tracking is failed, we will adopt gyro's prediction data as camera pose output.

4 Experimental Results

4.1 Initial Camera Calibration

An initial calibration test was carried out in order to calculate camera's intrinsic parameters as well as the position relationship between VCS and ICS.

Illustration of the calibration test is shown in Figure 6. We used 4x6 pre-defined grids on the ground as references. Vehicle rear wheels were aligned to the center of last row which gives the VCS coordinates of each grid points as:

$$(X_v, Y_v, Z_v)^T = ((1.5-i)d_{column}, 0, (5.0-j)d_{row})^T \tag{15}$$

where row number $i = 0 \sim 3$, column number $j = 0 \sim 5$, d_{row} and d_{column} are the interval distances of row and column respectively. ICS coordinates of each grid can be interactively or automatically obtained from the image.

We used Haralick's Modified Weights Method[6] to solve this typical absolute orientation problem. In order to simplify the computing, we assume the affection of lens distortion is neglectable and camera roll angle can be easily adjusted to 0 during the system setup. Image principle point is usually located in the center of image, which gives $u_0 = v_0 = 0$. Skew factor S_θ is also assumed to be 0. Calibration test result is shown in Table 1.

Table 1. Calibration results of intrinsic and extrinsic parameters

Notation	Description	Estimated values
f	Camera focal distance	467 pixels
T_X	Displacement on X_v axis	-12.24 cm
T_Y	Displacement on Y_v axis	197.02 cm
T_Z	Displacement on Z_v axis	213.22 cm
Ω_X	Displacement on pitch angle	0.262 rad
Ω_Y	Displacement on yaw angle	0.136 rad

4.2 Road Test Results

As shown in Figure 7, a finger-sized CCD analog video camera was mounted on the front of test platform minivan. Image sequences were recorded in NTSC format at the frame rate of 30fps. Differential GPS data (Trimble® AgGPS)

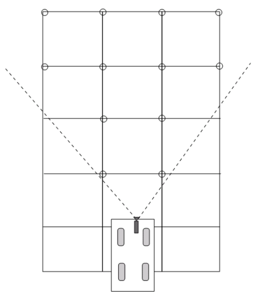


Fig. 6. Camera calibration

Fig.7. Camera and GPS device

and inertial data (DataTech® GU-3023) were sent to PC's serial port and recorded at the frequency of 16.7Hz. Zenrin®

Z-Map (Kumamoto region) was used as the 2D road map. As the Phase 1 of VICNAS project, our tests were based on the off-line processing.

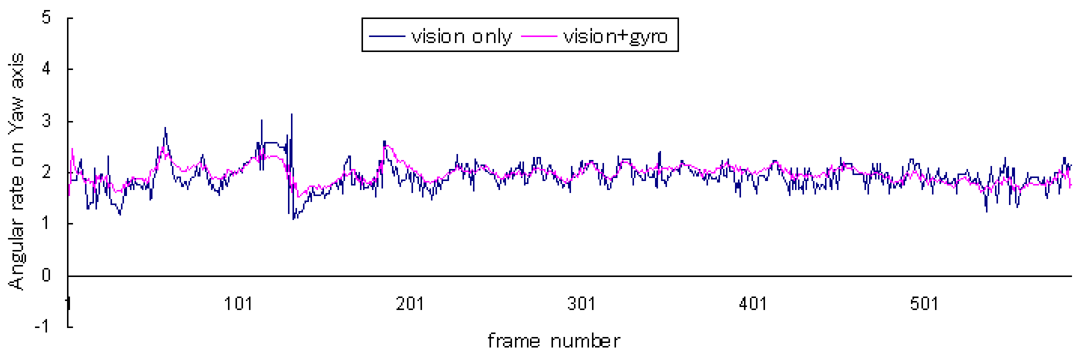
Road tests were carried out on different kinds of road (express toll-way, city highway, downtown street and countryside road), different lane structures (one-way or two-way, 1~6 lanes, with or without central separators) and shapes (straight, curve, S-curve).

Figure 9 shows part of the angular rate estimation results of vision only and hybrid approach from a 3500 frames image sequence (Figure 8). The results here has already been eliminated the original displacement between the camera and the vehicle reference frame. A very stable yaw rate with an average of 2.0 degree corresponds to fact of vehicle's heading angle was turning right, which can be also verified by the fact of road segments' right curve.

Figure 10 shows the result of integrated roll, pitch and yaw angle with respect to the world coordinate system.

Figure 8. Tested Road Sequence

Figure 7. Estimated result on Yaw rate by vision and gyro



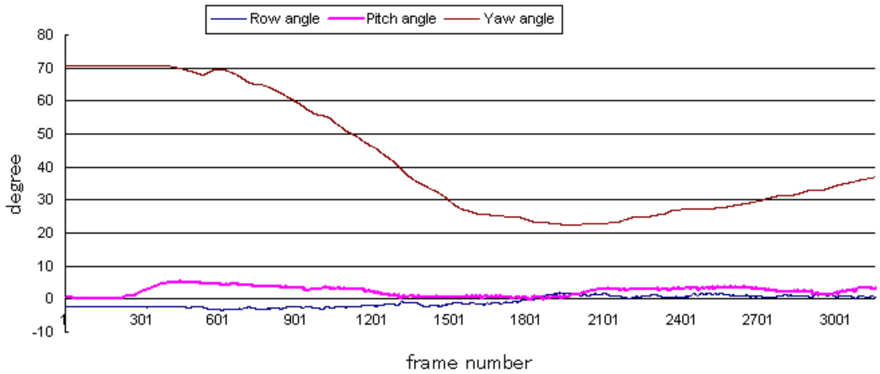


Figure 10. Integrated result by proposed hybrid approach

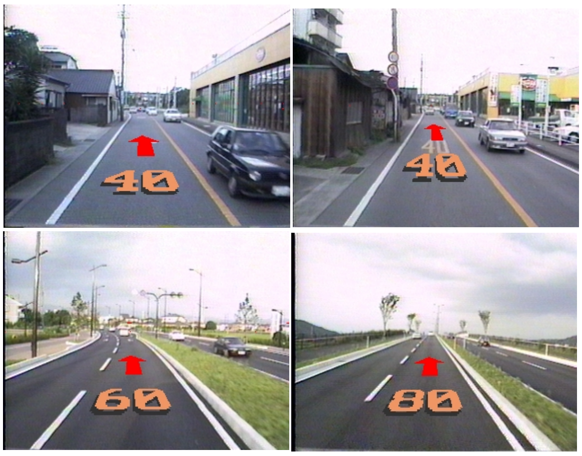


Figure 11. Visual perceived performance test on different roads

To verify the accuracy of our estimation algorithm, a virtual direction indicator (red arrow) and a virtual road bulletin (speed limitation painting) were generated and projected to the real road scene according to the estimated camera pose data. As shown in Figure 9, a very smooth and stable synthesized result in the AR space verified the effectiveness of our solution to the registration problem for on-road navigation.

5. Conclusion

This paper presents a novel framework of vision-based road navigation system, which superimposes virtual 3D navigation indicators and traffic signs onto the real road scene in an Augmented Reality (AR) space. To properly align the virtual object with real world, this paper proposed a hybrid camera pose tracking system that combines vision, GPS and 3D inertial gyroscope technologies. The fusion approach is based on our PMM (parameterized model matching) algorithm, in which the road shape model is derived from the digital map referring to GPS absolute road position, and matches with road features extracted from the real image. Inertial data estimates the initial possible motion, and also serves as relative tolerance to stable the pose output. The algorithms proposed in this paper are validated with the experimental results of real road tests under different conditions and types of road.

There are still some special road shape segments that are not covered by our algorithm of road model matching, such as intersection and diversion, which are also essential for the on-road navigation. 3D road shape is also another interesting topic and it will become more commercially valuable when the 3D digital map data is available in the near future. Our interests will be continuously focused on these topics as well as the real-time computation and implementations in AR world.

REFERENCES

[1] R.T. Azuma. "A survey of augmented reality," In *Presence: Teleoperators and Virtual Environments 6*, 4, pp.355-385 (1997).

[2] M. Bajura, H. Fuchs, and Ohbuchi. "Merging virtual objects with the real world: Seeing ultrasound imagery within the patient," *Computer Graphics*, pp.203-210, July 1992.

[3] Z. Hu and K. Uchimura. "Dynamical Road Modeling and Matching for Direct Visual Navigation", *IEEE workshop on Applications of Computer Vision*, Orlando, USA, pp.237-241 (2002).

[4] E. Foxlin, "Inertial Head-Tracker Sensor Fusion by a Complementary Separate-Bias Kalman Filter", *Proc. of IEEE Virtual Reality Annual International Symposium*, pp.184-194, 1996.

[5] E. Foxlin, M.Harrington, and G. Pfeifer, "Constellation: A Wide-Range Wireless Montion-Tracking System for Augmented Reality and Virtual Set Applications", *Prof. Of GRAPHICS 98*, 1998.

[6] B. Horn and E. Weldon. "Direct Mothods for recovering motion," Intl. Journal of Computer Vision, Vol.2, pp. 51-76 (1988)

[7] X. Zhuang, R. Haralick, and T. S. Huang, "Two-view motion analysis: A unified algorithm," *Opt. Soc. Am.*, 3(9), pp. 1492-1500, 1986.

[8] R.M. Haralick, H. Joo, R. C. Lee, X. Zhuang, V. G. Vaidya, and M. B. Kim. "Pose Estimation from Corresponding Point Data," *IEEE Transactions on Systems, Man and Cybernetics*, 19(6), pp.1426-1446, November-December 1989.

[9] L. Chai, K. Nguyen, B. Hoff, and T. Vincent, "An Adaptive Estimator for Registration in Augmented Reality", *Prof. of IEEE International Workshop on Augmented Reality*, pp. 23-32, 1999.

[10] S. You and U. Neumann. "Fusion of vision and gyro tracking for robust augmented reality registration", *Proc. of IEEE Conference on Virtual Reality*, pages 71-78, Japan, 2001

[11] A. Valinetti, A. Fusiello, V. Murino. "Model tracking for video-based virtual reality," *Proc. 11th Intl. Conf. On Image Analysis and Processing*, pp.372 –377 (2001)

[12] R. Gregor, M. Lutzeler, M. Pellkofer, K.-H. Siedersberger, and E. D. Dickmanns. "EMS-Vision: A Perceptual System for Autonomous Vehicles," *IEEE Trans. on Intelligent Transportation Systems*. 3(1), pp. 48-59, March 2002.

[13] R. Hooke and T. Jeeves. "Direct search solution of numerical and statistical problems," *Journal of the Association for Computing Machinery (ACM)*, pp.212-229 (1961).

[14] Z. Hu, K. Uchimura, "On-board Camera Pose Estimation in Augmented Reality for Direct Visual Navigation", *IS&T/SPIE's Electronic Imaging 2003*, Santa Clara, California, USA, pp.508-518 (2003)

[15] A. Kelly, "Essential Kinematics for Autonomous Vehicles", Technical Report of Carnegie Mellon University, CMU-RI-TR-94-14 (1994)