Paper

A Combination of Monocular CCD Camera and Inertial-sensor as Hybrid Vision Inertial-Navigation For Range Estimation

Salah Derrouich* Non-member
Kiichiro Izumida* Member
Kazuhisa Shiiya* Member
Okihiko Ishizuka** Non-member

The problem under consideration centers on building a three-dimensional description of the unprepared environment of an autonomous mobile robot. In an image sequence, tracking is to be performed after image rectification. This intermediate process minimizes the token relative displacements between two frames and simplifies the tracking phase, because it reduces the disparity between two relative tokens and thus simplifies the matching process. In this paper we present a new hybrid approach to range estimation that combine inertial and visual based technologies; this allows us to calculate the image-space distance between the robotic head and 3D tokens.

Two frames from the image sequence obtained from passive-target-tracking system, moving CCD video camera, will represent a set of data with the output of the inertial-tracking system, that report the relative changes of orientations and accelerations between the two frames. By integrating these data in our algorithm the image-space distances of different 3D points was estimated theoretically and experimentally.

keywords : Robotic vision, Vertical Rectification, Hybrid inertial and vision, Sequence of images, 3D reconstruction

1. Introduction

Artificial visual systems face difficulties in tasks like navigation on uneven terrain or detecting other movements when they are moving themselves. Paradoxically, these are tasks that biological systems like insects with very simple brains can accomplish very easily. It appears that this is not only a matter of computational power but also a question of representation and sensor design. Several projects have addressed the method that simultaneously estimate 3D form and depth recovery[(1) (2)]. Tagawa[(3)] and Ohata[(4)] proposed the use of optical flow in place of light and darkness vector method; Ueshiba[(5)] and Sugimoto[(6)] proposed 3D reconstruction methods based on epipolar equation. It is therefore surprising that most artificial visual systems use only one kind of sensor, a passive-target-sensor (CCD camera). We know that the ability to both track-pose and manage residual errors is unique to vision. However, vision suffers from a notorious lack of robustness, including subject to signal degradation such as poor lightning, and high computational expense, that it is no longer desirable especially if the requirements of many potential applications in advanced robotics are to be met.

To compensate for this shortcoming, a number of researches have recently been done; most of them involved hybrid techniques based on visual and inertial sensors[(7)~(10)]. Pose-tracking methods often compute 2D-image motion, since these motions are often due to rotation, inertial gyro sensor can aid the vision system in tracking these motions and increase the robustness and computing efficiency of a vision system by providing a frame-to-frame prediction of camera orientation and position. On the other hand, vision can correct the long-term-drift of inertial sensors.

Inertial sensors measure acceleration or motion rates, so their signals must be integrated to produce position or orientation. Because the choice of sensors has to be realistic, and the cost of an inertial system on a robot cannot be the same as the very expensive ones such as those with high accuracy and optimum long-term-drift used on airplanes, noise calibration error and the gravity field impart errors on these signals, producing accumulated position and orientation drift. Position requires double integration of linear acceleration, so the accumulation of position drift grows as the rate of elapsed time squares. Orientation only requires a single integration rate, so the drift accumulates linearly with elapsed time.

In robotic applications the origin of inertial forces may be grouped into two categories: expected forces and unexpected forces. Expected forces are the intrinsic displacements due to the action of the robot's effectors; these displacements are then partially known. Unexpected forces that cause a 'sudden' change of movement can be either extrinsic displacement (displacements of the robot not due to its effectors) or unwanted vibrations due to actuators and mechanical transmissions. In order to solve this sudden change of pose (position and orientation), we propose the combination of visual and inertial sensors to develop a new algorithm for robotic tasks such as navigation and trajectory-generation among others. This algorithm is independent from any pose changes that could happen between two positions in the trajectory of a mobile robot, and afford the same kind of information as that provided by passive navigation algorithms using artificial vision, but with a different dynamic range and precision [(11)].

* Information Communication Engineering Lab., Miyazaki University, 1-1 Gakuen kibanadai Nishi, Miyazaki-shi 889-2192
** Center for Integration Technology Lab., Miyazaki University, 1-1 Gakuen kibanadai Nishi, Miyazaki-shi 889-2192

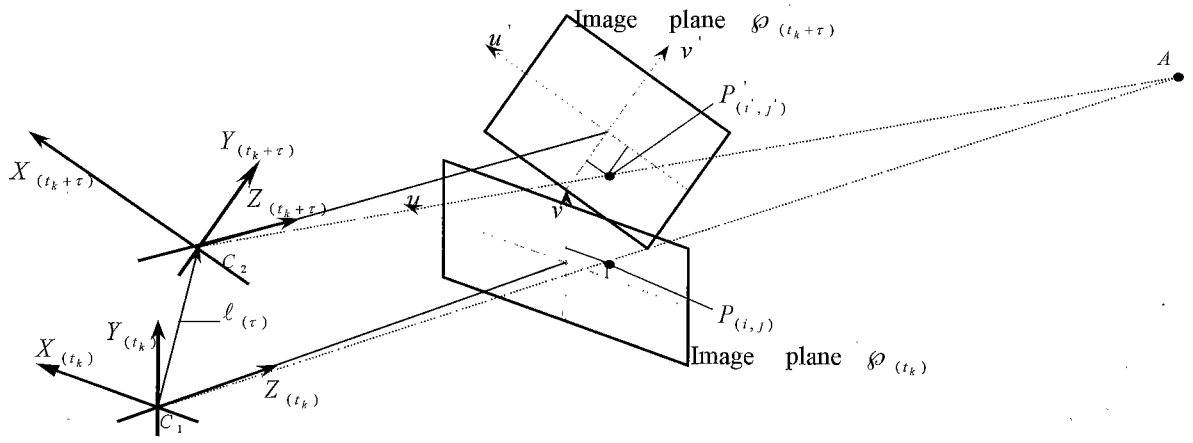


Fig. 1. Geometric constraint of monocular stereo vision.

The following sections describe a shape-from-motion method based on epipolar geometry, and vertical-rectification method to minimize the token relative displacements between two frames and estimate the image-space distance. Experimental results are also shown.

2. Epipolar geometry for range estimation

Using a calibrated camera, the geometric quality of the CCD sensor legitimates the use of the pinhole model for the camera, with a unit focal distance. Every quantity is referred to in the camera intrinsic coordinates *(X(tk), Y(tk), Z(tk))*. The origin of this frame of reference C_1 is the optical center of the camera, the Z-axis corresponds to the optical axis of the camera, and Y-axis is aligned with the vertical in the image. The image plane $\wp$ has as axes *(v)* parallel with *(Y)* and *(u)* parallel with *(X)* (Fig.1)

Let's consider a 3D point $A=(x,y,z)^T$ in homogenous coordinates, its projection P on the image plane $\wp(t_k)$ at time t_k is represented by i as ordinate and j as abscissa. The projection of A on image plane $\wp(t_k+\tau)$ after τ seconds of movement P', is represented by i' and j' ordinate and abscissa respectively. Next we decompose the 3D model into two 2D projections on planes (X, Z) and (Y, Z). Fig.2 shows the projection of Fig.1 on plane (X, Z) where we suppose:

* $\wp(t_k)$ and $\wp(t_k+\tau)$ are parallel. It is possible to project the image in a way that there is no more gyro-rotation while the vertical columns are aligned with the absolute vertical, and both images become parallel.

* $P_{(i,j)}$ and $P'_{(i',j')}$ are correspondent.

Therefore, the distance L'_x between C_{2x} (camera $(t_k+\tau)$) *(C_{2x} is the projection of C_2 on the plane (X,Z))* and point A_x *(the projection of A on the plane (X,Z))* is given by (1):

$$L'_x = \ell_{xy}\sin(\beta_x)/\sin(\theta_{x3}) \quad \cdots\cdots(1)$$

With $\beta_x = \theta_{x1} - \alpha_x$ given θ_{x1} as the angle between Z and C_1A_x. θ_{x3} is the angle between A_xC_1 and A_xC_{2x}, from Fig.2 one can see that $\theta_{x3} = \theta_{x2} - \theta_{x1}$, where θ_{x2} is the angle between Z and $C_{2x}A_x$

$$\ell_{xy} = \sqrt{l_x^2 + l_z^2} \quad \cdots\cdots(2)$$

$$\alpha_x = \arctan(l_x/l_z) \quad \cdots\cdots(3)$$

ℓ_{xy} is the moving-camera's position after τ second, and α_x is the angle between Z and C_1C_{2x}.

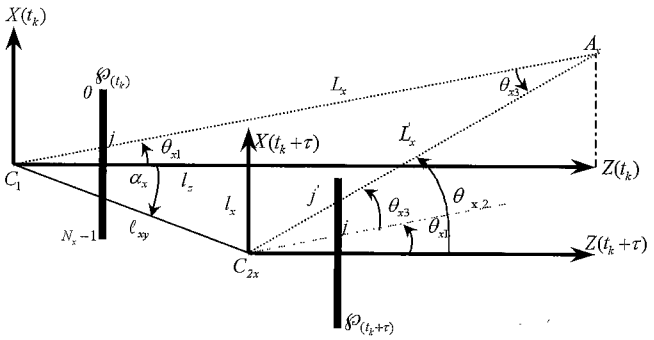


Fig. 2. Projection of Fig.1 on plane (X,Z).

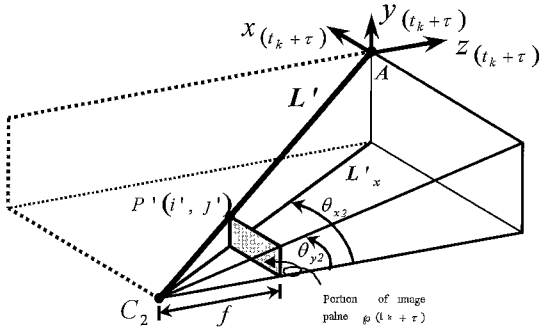


Fig. 3. Calculation of 3-D distance L' from 2-D distance L'_x.

Hence (1) can be rewritten as:

$$L'_x = \sqrt{l_x^2 + l_z^2}\,\frac{\sin(\theta_{x1} - \arctan(l_x/l_z))}{\sin(\theta_{x2} - \theta_{x1})} \quad \cdots\cdots(4)$$

In this equation we have lx, lz, θ_{x1} and θ_{x2} where lx and lz are obtained from inertial-sensor (see appendix A). θ_{x1} and θ_{x2} are known by fixing the corresponding points $P_{(i,j)}$ and $P'_{(i',j')}$ (Fig.2)

$$\theta_{x1} = \arctan\left(\frac{j-(N_x-1)/2}{f}\right) \quad \cdots\cdots(5)$$

$$\theta_{x2} = \arctan\left(\frac{j'-(N_x-1)/2}{f}\right) \quad \cdots\cdots(6)$$

with (N_x,N_y) represent the horizontal (width in pixels) and vertical (height in pixels) image resolution (see Fig.B-1 Appendix B), f is the focal length. From 2D distance L'_x we get the 3D image-space distance L' as expressed by (7) see Fig.3

$$L' = L'_x\sqrt{1+cos^2(\theta_{x2})tan^2(\theta_{y2})} \quad \cdots\cdots(7)$$

from (4) and (7) we get the final expression of L' :

$$L' = \sqrt{l_x^2+l_z^2}\,\frac{\sin(\theta_{x1}-\arctan(l_x/l_z))}{\sin(\theta_{x2}-\theta_{x1})}\sqrt{1+\cos^2(\theta_{x2})\tan^2(\theta_{y2})} \quad \cdots\cdots(8)$$

$\theta_{y2} = \arctan\left((i'-((N_y-1)/2))/f\right)$ is the angle between Z and $C_{2y}A_y$ (where $A_y(C_{2y})$ is the projection of $A(C_2)$ on plane (Y,Z) and $j(j')$ is the abscissa of $P(P')$ in image plane $\wp(t_k)(\wp(t_k+\tau))$)

3. The Gyro Rotation and the Vertical Rectification

However, in typical cases we face two problems

- Correcting the rotations of camera(t_k) and camera$(t_k+\tau)$, and having them vertically parallel.
- Looking for the corresponding point of $P_{(i,j)}$.

In both cases, mobile robot and manipulator, the displacement is bounded and several realistic physical hypotheses can be made.

- Inertial forces can be calculated using classical law of kinematics and dynamics.

- The gravity field is a constant, homogeneous and isotropic field of acceleration, not varying during robot displacements.

- Astronomic movements such as earth rotation have no influence as generators of force, on the robot.

Considering these hypotheses, we can separate gravity from linear acceleration. From this mechanism we can obtain an estimation of the vertical, combine it with a visual estimation of the vertical and reanalyze visual information considering this important 3D cue. This can be summarized in (9) (see Appendix B)

$$\begin{pmatrix} i_c \\ j_c \end{pmatrix} = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} f\tan\left[\arctan\left(\dfrac{i-(N_y-1)/2}{f}\right)+\theta\right] \\ f\tan\left[\arctan\left(\dfrac{j-(N_x-1)/2}{f}\right)-\psi\right] \end{pmatrix} + \begin{pmatrix} \dfrac{N_y-1}{2} \\ \dfrac{N_x-1}{2} \end{pmatrix} \quad \cdots\cdots(9)$$

Applying this relation to point $P_{(i,j)}$ we get its correspondent $P'(i_c,j_c)$ in case of pure rotation around the 3 axes: X (by θ), Y (by ψ), and Z (by ϕ). (Refer to Fig.4 for notations).

The gyroscopes determine the values of θ, ψ and ϕ. Therefore, by applying (9) to image plane $\wp(t_k)$ and $\wp(t_k+\tau)$, we will have new vertically parallel image-planes with less displacement, ready to be matched to extract the disparity caused by the translation along X, Y and Z.

Fig.5 (b) shows the result of the application of (9) to the image in Fig 5(a) with θ=-9.75°, ψ=-4.78° and ϕ=-1.66°

Moreover, performing picture rectification, we can reduce the disparity between tokens in two consecutive frames, thus improving matching method. Matching these vertically parallel pictures afford us the corresponding points in both images. Many methods have been proposed for the matching purpose but are not discussed here (Further explanations can be found in references (12) to (15)). In this paper, the correspondence process was done manually for high accuracy purpose.

By fixing the correspondents $P_{(i,j)}$ and $P'(i',j')$, the image-space distance L' is fixed as presented in second section.

The following are the experiments' results of our method applied on a simple target in the first experiment, to check the efficacy of the epipolar geometry. The second experiment's objective is range estimation check.

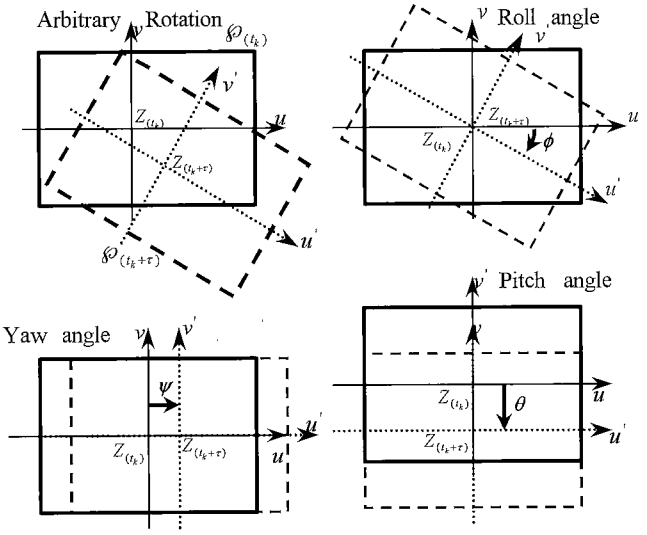
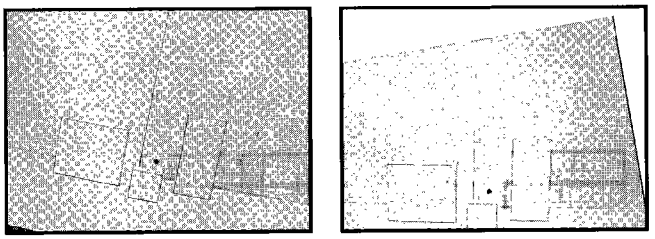


Fig. 4. Illustration of $\wp(t_k)$ and $\wp(t_k+\tau)$ after the three rotations.



(a) Raw picture **(b)** Rectified picture

Fig. 5. Result of application of (9).

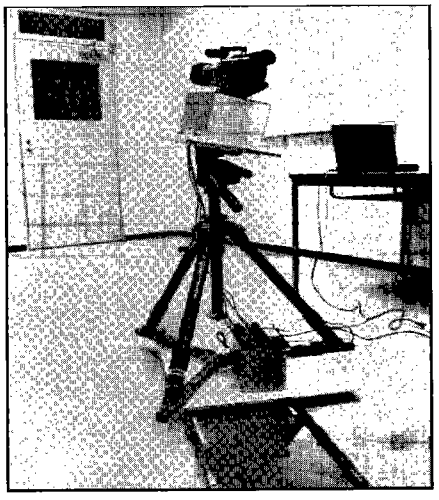


Fig. 6. Picture of the moving sensor-system: CCD camera and inertial-sensor on a moving carrier.

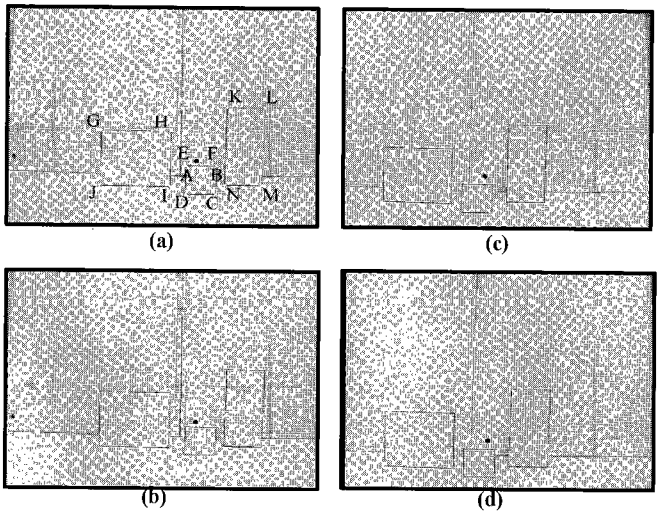


Fig. 7. Epipolar Geometry Experiment **(a)** Raw picture at t_k, **(b)** Vertically rectified picture of (a), **(c)** Raw picture at time $t_k+\tau$, **(d)** Vertically rectified picture of (c).

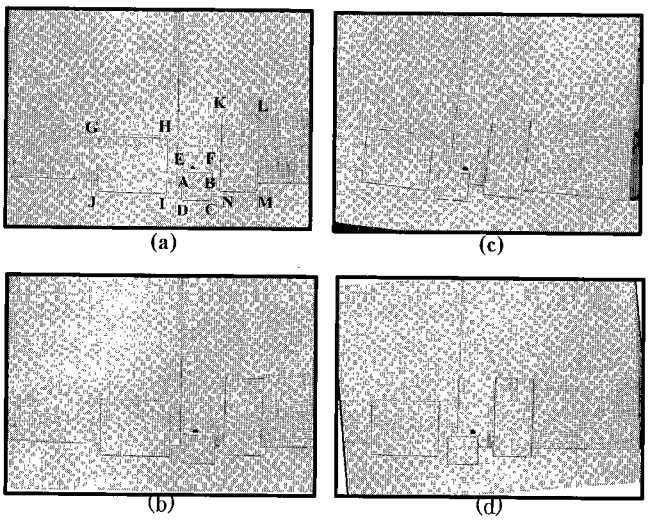


Fig. 9. Range Estimation experiment **(a)** Raw picture at t_k, **(b)** Vertically rectified picture of (a), **(c)** Raw picture at time $t_k+\tau$, **(d)** Vertically rectified picture of (c).

Table 1. Epipolar Geometry Method Applied on 14 points in Fig.7.

P(i,j)	L'c	L'm	Error(cm)	Error(%)
A	122.10	122.00	0.10	0.08
B	122.74	122.00	0.74	0.61
C	126.75	124.00	2.75	2.22
D	125.21	124.00	1.21	0.98
E	128.46	127.80	0.66	0.52
F	129.24	127.80	1.44	1.12
G	139.41	136.40	3.01	2.21
H	130.91	132.20	-1.29	0.98
I	136.60	135.80	0.80	0.59
J	143.84	139.00	4.84	3.48
K	138.49	133.80	4.69	3.51
L	132.18	135.00	-2.82	2.09
M	135.96	139.50	-3.54	2.54
N	142.22	138.50	3.72	2.69

Table 2. Range Estimation Method Applied on 14 points in Fig.9.

P(i,j)	L'c	L'm	Error(cm)	Error(%)
A	121.64	120.50	1.14	0.95
B	120.98	120.50	0.48	0.40
C	124.37	122.40	1.97	1.61
D	123.23	122.00	1.23	1.01
E	127.97	126.40	1.57	1.24
F	128.32	126.50	1.82	1.44
G	142.29	136.50	5.79	4.24
H	133.65	131.40	2.25	1.71
I	135.92	134.00	1.92	1.44
J	143.32	138.60	4.72	3.41
K	134.58	133.00	1.58	1.19
L	136.60	133.60	3.00	2.24
M	140.01	136.80	3.21	2.35
N	136.84	136.00	0.84	0.62

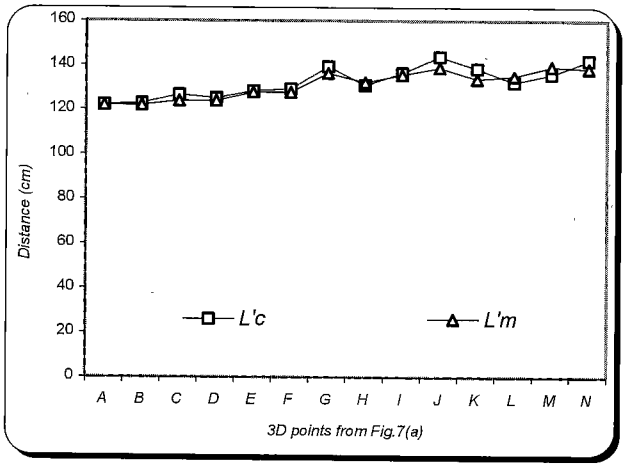


Fig. 8. Graphic representation of the calculated distance $L'c$ and the measured distance $L'm$ from Table 1.

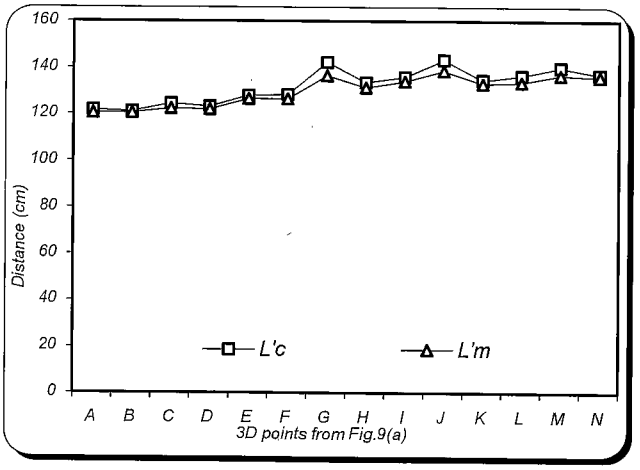


Fig. 10. Graphic representation of the calculated distance $L'c$ and the measured distance $L'm$ from Table 2.

4. Experimental Results

4.1 System Configuration To check the efficacy of this method we conducted some experiments using a digital camera *(SONY 3CCD DCR-VX1000)* to take two pictures separated by τ seconds of movement. The acceleration and angular changes are obtained from the inertial sensor (TAGAWA TA 7511N3XX0).

To conduct a pure translation, pure rotation, or both, while tracking the movement of the sensors-system, we set the camera and the inertial sensor on a platform that can perform all the 3 rotations (Pitch, Yaw, and Roll), fixed on an elevator for vertical translation. For horizontal translation, we placed the whole system on another platform that can slide on parallel rails (Fig.6).

4.2 Results To check our algorithm, we first tested the epipolar geometry part by an experiment with 2DOF, (a simple translation), and then with an experiment that covers the whole theory.

4.2.1 Epipolar Gometry Experiment The scene is composed of 3 boxes. The sensors-system performs a simple translation along X-axis l_x=-18.00 cm and Y-axis l_y=6.5 cm (Fig.7).

Because of the movement, the gyroscopes register some rotations around their respective axes that we took in consideration during the analysis. (θ,ψ,ϕ)=(0.17, -0.38, 0.01) are the Pitch, Yaw, and Roll of the first picture Fig.7(a) and (θ,ψ,ϕ)=(0.07, -0.09, 0.28) are the Pitch, Yaw, and Roll of the second picture Fig.7(c). The goal, as mentioned before, was to extract the distance between the camera and the scene. We focused on 14 points Fig.7(a). Table.1 shows the 14 points A~N in the first camera, the calculated distance L'_c (using our algorithm), the measured distance L'_m (manually measured), the difference between L'_c and L'_m in (cm) and the Error in %. Fig.8 represents graphically both the measured distance and the calculated distance.

4.2.2 Check of Range Estimation Theory In this experiment the sensors-system performs one translation and one rotation; $\ell(l_x,l_y,l_z)$=(0, 21.28, -0.12) is the distance between the 2 positions; the orientation of the picture taken at time (t_k) is (θ,ψ,ϕ)=(0.14, 0.58, 0.03) (Fig.9(a)) and the orientation of the picture taken at time $(t_k+\tau)$ is (θ,ψ,ϕ)=(-5.20, 0.58, 0.19) (Fig.9(c)).

The results of the 14points in Fig.9(a): L'_c calculated distance, L'_m measured distance, the difference between L'_c and L'_m in (cm), the percentage of the error are presented in Table.2. The graphic presentation of the results is plotted in Fig.10.

4.3 Noise Sensitivity and Expected Errors In this section we analyzed the noise sensitivity of sensors-system and the expected errors. Our algorithm can be divided into two major parts: first the epipolar-geometry and second the vertical-rectification. The epipolar geometry is where $\ell(lx,ly,lz)$ appears. This quantity is the double integration of linear acceleration measured by an inertial sensor. Noise, calibration error and the gravity field impart errors on these signals provided accumulated position drift. This accumulation of position drift grows as the square of the elapsed time to make errors originating from inertial sensors inevitable; to overcome this problem, we reduced the elapsed time to τ seconds, enough time for taking a pair of picture from two different positions, then we reset the system for the next pair of data.

Let's consider the case where the linear acceleration $\boldsymbol{a}_m$ in equation A-5 (Appendix A), is affected by error $\boldsymbol{e}$. we have the output of accelerometers $\boldsymbol{a'}_m=\boldsymbol{a}_m+\boldsymbol{e}$, and equation A-5 become:

$$\ell'(\tau)=\ell(\tau)+(\tau/N)^2 e\sum_{m=2}^{N}(N-m+1) \quad \cdots\cdots (10)$$

By calling the accumulated position drift: $\Delta\ell$ we have

$$\Delta\ell(\tau)=\ell'(\tau)-\ell(\tau)=(\tau/N)^2 eN(N-1)/2 \quad \cdots\cdots (11)$$

or $$\Delta\ell(N)=eT^2N(N-1)/2 \quad \cdots\cdots (12)$$

Where $T=\tau/N=1/f$ is the period of the accelerometers' output (in this equation f is the frequency).

In our actual experiment the accuracy of the accelerometers was e=±0.01 and T=20ms, if we give to the mobile sensor 5 periods of time (N=5) we will have τ=0.1s and an error of $\Delta\ell(5)=\pm0.03924$cm . To get a close qualitative-look to the error, let's consider the case where the system is running in straight-line with a constant speed of 7.2 km/h (close the human's speed), in a lapse of τ=*0.1s* the system will cross 20 cm, by considering $\Delta\ell(5)=\pm0.03924$cm we will be facing and error of *0.19%*.

As we can see from this example, the accumulated position drift $\Delta\ell$ grows as the square of the elapsed time τ. Thus the determination of τ depends on the kind of application of the robot; and its decision has to take in consideration: the range of allowed accumulated drift $\Delta\ell$, and the movement of the robot (translation and rotation) to keep the target inside the vision field of the tracking system.

The second major part of the algorithm is vertical rectification. Where 3 orientation measurements (Pitch, Yaw, and Roll) appear, the error expected relies on the inertial sensor precision just like the first part. To demonstrate this we map the specified error into the 2D image domain. Let (f_x,f_y) be the effective horizontal and vertical focal lengths of a video camera (in pixels), (N_x, N_y) represent the horizontal and vertical image resolution and (D_x, D_y) be the field-of-view of the camera, respectively. If pixels sample the rotation angles uniformly (Pitch and Yaw), the ratio of image pixel motion to the rotation angles (pixel/degree) is

$$N_X/D_X = N_X/2\arctan(N_X/2f_x)$$

$$N_y/D_y = N_y/2\arctan(N_y/2f_y)$$

To illustrate a concrete example of this relationship, we consider our video camera with F1:1.6,5.9~59mm lens. Through calibration, we determine the effective horizontal and vertical focal lengths as f_x=776.74 pixels and f_y=800.81 pixels, with a 640x480 image resolution, The ratios are N_x/D_x=14.29pixel/degree and N_y/D_y=*14.38* pixel/degree. This indicates that each degree of orientation angle error results in about 14pixels of alignment error in the image plane [8]. In our actual experiment, the accuracy of the gyroscopes was: ±0.3, which is about 4 pixels that could appear as noise.

The main objective of our research was the calculation of image-distance L' (equation (8)), inthis equation we see that L' depends implicitly on i' and j' (explicitly on θ_{x2} and θ_{y2}) the coordinates of P' the correspondent of P. to check the influence of the matching method to the total performance in our experiments we plotted the percentage of changes in the value of L', for 14 points A~N in Fig.9(a), in function of the accumulated errors in

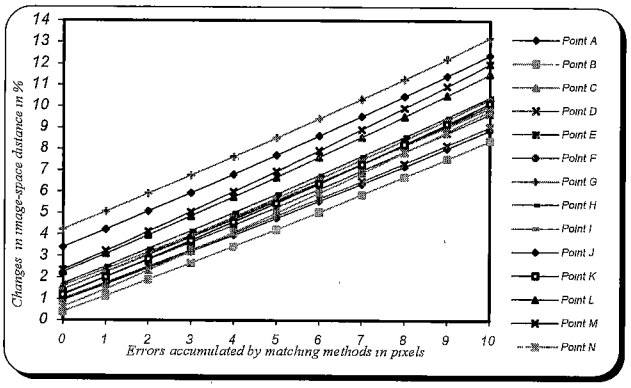


Fig. 11. Representation of the percentage of changes in image-space distance of 14 points *A~N* (Fig.9(a)) and accumulated errors by matching method.

the decision of $P'_{(i',j')}$ (the 2D position on the screen) (see Fig.11). From this representation we can see how does the errors accumulated by the correspondence process affect proportionally the changes in the image-space distance. In our experiments the correspondence between points was manual and the errors was around ± 1pixel.

The experiments results show the strength and the precision of this method even if the system was built from low cost equipment and many components were not precisely engineered.

5. Conclusion

We have presented a new hybrid monocular stereo vision technique based on moving camera and inertial sensor data.
Inertial tracking has advantages of robustness, range, and a system that is passive and self-contained. In our experiment and theoretical approach we overcame its major disadvantage that has been its lack of accuracy and drift over time. Vision tracking is accurate over long periods, but it suffers from occlusion and computation expense. We exploit the complementary nature of these two tracking technologies to compensate for the weakness in each separate component.

Furthermore, based on this method, the online range estimation and detection of moving target will be the subject of future studies.

(Manuscript received May 20, 2002, revised Jan. 6, 2003)

References

(1) Go Juo and S. Tengi . "3D Vision", Kyoritsu, Tokyo Apr (1998) (in Japanese)

(2) J Sato . "Computer Vision", Korona, May (1999)

(3) N. Tagawa, T. Torio, T. Endoh, T. Tanaka, and T. Moriya : "Estimation of 3-D Shape and Motion from Optical Flow Based on Depth Model", *IEICE*, Vol.81-D-II, No.8 pp.1727-1735 (1998) (in Japanese)

(4) N. Ohta : "Structure from Motion with Confidence Measure and Its Application for Moving Object Detection", *IEICE*, Vol j76-D-II, No.8, pp.1562-1571 (1993) (in Japanese)

(5) T. Ueshiba and F Tomita . "A Factorization Method for Multiple Perspective Views via Iterative depth Estimation", *IEICE*, Vol.j81-D-II, No.8, pp.1718-1726 (1998) (in Japanese)

(6) N Sugimoto and G. Xu : "A Simple Linear Algorithm for Motion from Three Weak Perspective Images Using Euler Angles", *IEICE*, Vol.j81-D-II, No.4, pp 681-688 (1998) (in Japanese)

(7) K. Izumida, K. Shiiya, H. Takahashi, and S. Derrouich . "Moving Object Detection from Traveling Monocular Camera Image", *Trans. of IEEJ*, Vol. 122-C, No 3, pp.498~505 (2002) (in Japanese)

(8) S. You, U. Neumann, and R. Azuma · "Hybrid Inertial and Vision Tracking for Augmented Reality Registration", *IEEE*, pp.260-266 (1999)

(9) U Neumann, S You, Y. Cho, j. Lee, and J Park : "Augmented Reality Tracking in Natural Environments", Proc. Int. Symposium on Mixed Realities, (1999)

(10) T Mukai and Ohnishi : "The Recovery of Object Shape and Camera Motion using a Sensing System with a video Camera and a Gyro Sensor", *Proc. Int. Conf. On Computer Vision (ICCV'99)*, Vol.1, pp.411-417 (1999)

(11) T. Vieville : "A Few Steps Towards 3D Active Vision", Springer-Verlag, Berlin (1997)

(12) Barrett, E B , Brill, M.H., Haag, N.N. and Payton, P.M. : "Linear Resection, Intersection, and Perspective-independent Model Matching in Photogrammetry· Theory", *Appli. Digital Image Processing XIV, editor Tescher, A., Proc. SPIE* 1567, p.142-169 (1991)

(13) S.T. Barnard, "Stereo matching by hierarchical micro canonical annealing", *in Proc. 10th IJCAI, Milano, Italy*, pp.832-835 Aug (1987)

(14) N. Ayache and B. Faverjon · "Efficient registration of stereo images by matching graph descriptions of edge segments", *Int. J. Comput. Vision*, Vol. 1, No.2, Apr. (1987)

(15) Z. Zhang, R. Deriche, O Faugeras, and Q.-T. Luong · "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry", Artificial Intelligence Journal, 78·87-119 (1995)

Appendix A

Derivation of l_x l_y l_z from Gyro's Data

Let N be the number of sets of raw data that we get from inertial sensors during τ seconds, those data sets are: $(\theta_m, \psi_m, \phi_m)$(rad) angular position vector (pitch, Yaw and Roll), $(\omega_{xm}, \omega_{ym}, \omega_{zm})$(rad/s) angular velocity vector and $(\alpha_{xm}, \alpha_{ym}, \alpha_{zm})$(m/s^2) linear acceleration vector, $1 \le m \le N$. with a frequency of *50* Hz.

By using the transfer matrix $[{}_kC_m]$ we get the transformed acceleration (a_{xm}, a_{ym}, a_{zm}) for each $(\alpha_{xm}, \alpha_{ym}, \alpha_{zm})$.

$$\begin{pmatrix} a_{xm} & a_{ym} & a_{zm} \end{pmatrix}^T = [{}_kC_m] \begin{pmatrix} \alpha_{xm} & \alpha_{ym} & \alpha_{zm} \end{pmatrix}^T \quad \cdots \text{(A-1)}$$

With

$$[{}_kC_m] = \begin{bmatrix} \cos\varphi_m \cos\theta_m & \sin\varphi_m \cos\theta_m & -\cos\psi_m \sin\theta_m \\ -\sin\varphi_m \sin\psi_m \sin\theta_m & +\cos\varphi_m \sin\psi_m \sin\theta_m & \\ \sin\varphi_m \cos\psi_m & \cos\varphi_m \cos\psi_m & \sin\psi_m \\ \cos\varphi_m \sin\theta_m & \sin\varphi_m \sin\theta_m & \cos\psi_m \cos\theta_m \\ +\sin\varphi_m \sin\psi_m \cos\theta_m & -\cos\varphi_m \sin\psi_m \cos\theta_m & \end{bmatrix} \quad \cdots \text{(A-2)}$$

We use $[{}_kC_m]$ to represent all the accelerometers' data (vectors) in the first local-coordinate system at time t_k (knowing that the accelerometers refer to the previous coordinate system each 20ms). The relation between $\boldsymbol{a_m}=(a_{xm}, a_{ym}, a_{zm})^T$ and $\ell_m=(l_{xm}, l_{ym}, l_{zm})^T$ the position vector is

$$\boldsymbol{a_m} = (N/\tau)^2(\ell_m - 2\ell_{m-1} + \ell_{m-2}) \quad \cdots \text{(A-3)}$$

Multiplying each side of (A-3) by $(N-m+1)$ and sum both sides for

m=2, 3, 4...N, we get:

$$\sum_{m=2}^{N}(N-m+1)a_m=(N/\tau)^2(\ell_N-N\ell_1+(N-1)\ell_0) \quad \cdots\cdots (A\text{-}4)$$

Considering that $\ell_0=0$, $\ell_N=\ell(\tau)=(l_x, l_y, l_z)^T$ and ignoring ℓ_1 (the 1st position of the gyro after 20 ms!) we get:

$$\ell(\tau)\cong(\tau/N)^2\sum_{m=2}^{N}(N-m+1)a_m \quad \cdots\cdots (A\text{-}5)$$

Appendix-B

Derivation of Equation (9)

Let's consider the image plane $\wp(t_k)$ at time t_k in Fig.B-1. the relation between the coordinates (i,j) in pixels and (x,y) of point P is:

$$\begin{cases} y=i-(N_y-1)/2 \\ x=j-(N_x-1)/2 \end{cases} \quad \cdots\cdots (B\text{-}1)$$

same in case of point P' in image plane $\wp(t_k+\tau)$ at time $t_k+\tau$:

$$\begin{cases} y'=i_c-(N_y-1)/2 \\ x'=j_c-(N_x-1)/2 \end{cases} \quad \cdots\cdots (B\text{-}2)$$

The relation between $P_{(x,y)}$ and $P'_{(x',y')}$ (its correspondent in case of pure rotation around Z axis with angle ϕ), see Fig.B-2, is :

$$\begin{pmatrix} y' \\ x' \end{pmatrix}=\begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}\begin{pmatrix} y \\ x \end{pmatrix} \quad \cdots\cdots (B\text{-}3)$$

In case of full rotation including pitch and yaw, additional factors Δx and Δy that represent the rotation distortion around Y and X axes respectively, will appear as shown in equation B-4:

$$\begin{pmatrix} y' \\ x' \end{pmatrix}=\begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}\begin{pmatrix} y+\Delta y \\ x+\Delta x \end{pmatrix} \quad \cdots\cdots (B\text{-}4)$$

From Fig.B-3 and Fig.B-4 we have:

$$\begin{cases} \Delta y=y'-y \\ \Delta x=x'-x \end{cases} \quad \cdots\cdots (B\text{-}5)$$

$$\begin{cases} y'=f\tan(\theta_{y2}) \\ x'=f\tan(\theta_{x2}) \end{cases} \quad \cdots\cdots (B\text{-}6)$$

and $$\begin{cases} \theta_{y2}=\theta_{y1}+\theta \\ \theta_{x2}=\theta_{x1}-\psi \end{cases} \quad \cdots\cdots (B\text{-}7)$$

From B-5, B-6 and B-7 we get B-8:

$$\begin{cases} y+\Delta y=f\tan[\arctan(y/f)+\theta] \\ x+\Delta x=f\tan[\arctan(x/f)-\psi] \end{cases} \quad \cdots\cdots (B\text{-}8)$$

Finally, from B-4 B-8 B-1 and B-2 we have B-9 the final expression of equation (9):

$$\begin{pmatrix} i_c \\ j_c \end{pmatrix}=\begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}\begin{pmatrix} f\tan\left[\arctan\left(\left(i-\frac{N_y-1}{2}\right)\Big/f\right)+\theta\right] \\ f\tan\left[\arctan\left(\left(j-\frac{N_x-1}{2}\right)\Big/f\right)-\psi\right] \end{pmatrix}+\begin{pmatrix} \frac{N_y-1}{2} \\ \frac{N_x-1}{2} \end{pmatrix} \quad \cdots\cdots (B\text{-}9)$$

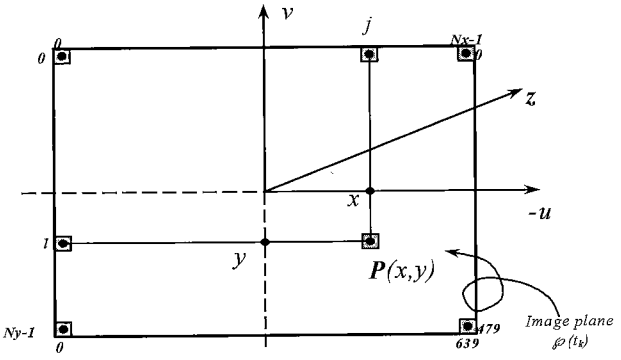


Fig. B-1. Representation of point $P_{(x,y)}$ with its coordinates (i,j) in image plane.

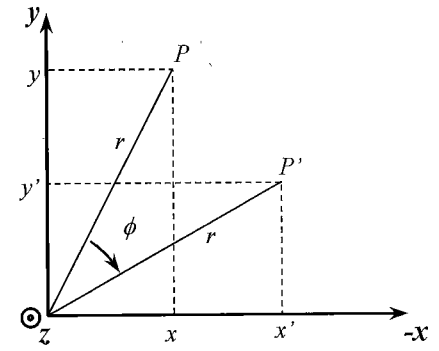


Fig. B-2. Rotation of base (x,y,z) around Z axis by ϕ.

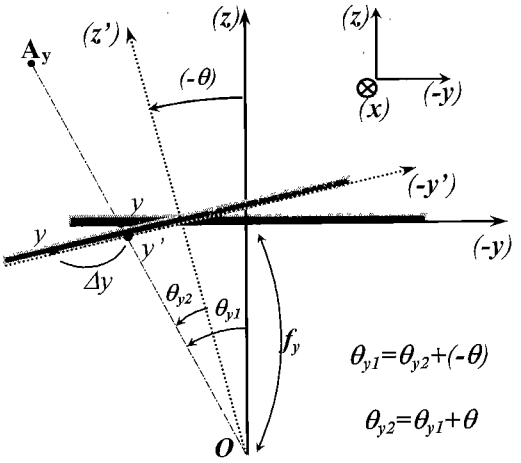


Fig. B-3. Rotation of the base (x,y,z) around X axis by $(-\theta)$.

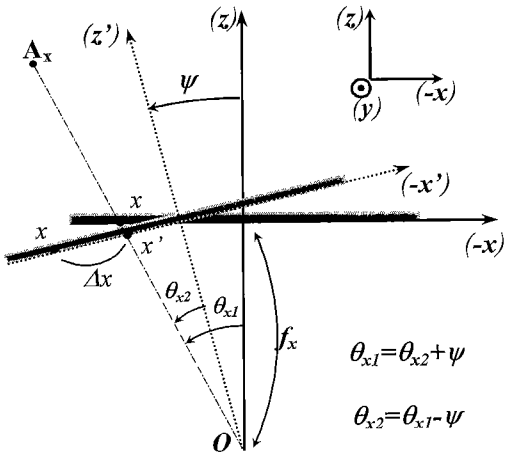


Fig. B-4. Rotation of the base (x,y,z) around Y axis by (ψ).

Salah Derrouich (Non-member) was born in Bouanan, Morocco; in 1974 He received B.S. in Electronic Engineering from Mohammed I University Oujda Morocco in 1997, and his M.S. degree in Information Communication Engineering laboratory, Miyazaki University, Japan. He is currently working towards his Ph.D. degree in the same laboratory. His research interests include image processing, range sensing and moving target detection.

Kiichiro Izumida (Member) graduated from Kumamoto University in 1968 and then joined Mitsubishi Electric Corp. In 1986 he was dispatched to Mitsubishi Space Software Co. Since 1995, he has been a professor in Department of Electrical and Electronic Engineering, Miyazaki University. He holds a D.Eng degree (Tohoku University). He has engineered in research on system estimation/control and digital signal processing concerning vision.

Kazuhisa Shiiya (Member) graduated from Miyazaki University in 1994 and became a Technical Assistant there. His research interests are image processing and digital signal processing.

Okihiko Ishizuka (Non-member) was born on September 7 1939. He received a B.E. degree in 1964 from Kagoshima University, and a M.E. degree in 1966 and a Ph.D. degree in 1978 from Kyushu University. He is a Professor of Electrical & Electronic Engineering at Miyazaki University, Miyazaki, Japan. His research interests are network synthesis, circuit analysis on multiple-valued logic, threshold logic and fuzzy logic. Dr. Ishizuka is a member of IEEE and the Executive Subcommittee of the IEEE computer Society Technical Committee on Multiple-Valued Logic.