Robotica (2004) volume 22, pp. 51–59. © 2004 Cambridge University Press
DOI: 10.1017/S0263574703005381 Printed in the United Kingdom

A new approach using sensor data fusion for mobile robot navigation

Tae-Seok Jin*, Jang Myung Lee* and S. K. Tso†

(Received in Final Form: May 30, 2003)

SUMMARY

To fully utilize the information from the sensors, this paper proposes a new sensor-fusion technique where the data sets for the previous moments are properly transformed and fused into the current data sets to enable an accurate measurement. Exploration of an unknown environment is an important task for the new generation of mobile service robots. The mobile robots may navigate by means of a number of monitoring systems such as the sonar-sensing system or the visual-sensing system. Note that in the conventional fusion schemes, the measurement is dependent on the current data sets only. Therefore, more of sensors are required to measure a certain physical parameter or to improve the accuracy of the measurement. However, in this approach, instead of adding more sensors to the system, the temporal sequence of the data sets are stored and utilized for the accurate measurement. The theoretical basis is illustrated by examples and the effectiveness is proved through the simulations and experiments. The newly proposed, STSF (Space and Time Sensor Fusion) scheme is applied to the navigation of a mobile robot in an unstructured environment, as well as in structured environment, and the experimental results show the performance of the system.

KEYWORDS: Multi-sensor fusion, Mobile robot, Measurement, Image processing, Navigation.

1. INTRODUCTION

Much research has been done on the spatial fusion technique, i.e. multiple sensor data are utilized either for the purpose of providing complementary or redundant data for measuring physical parameters; all of the current data from the sensors are integrated and fused to obtain a correct set of measurement.[1] In recent years interest has been growing in the synergistic use of multiple sensors to increase the capabilities of intelligent machines and systems. For these systems to use multiple sensors effectively, a strategy to integrate the information provided by these sensors into the operations of the system is necessary. While in many multi-sensor systems the information from each sensor serves as a separate input to the system, the actual combination or fusion of information prior to its use in the system has been a particularly active area of research. Typical applications that can benefit from the use of multiple sensors are industrial tasks like an assembly, a military command and control for battlefield management, mobile robot navigation, multi-target tracking, and aircraft navigation. In all of these applications, the system needs to intelligently interact with the human and operates in an unstructured environment with the human operator's assistant.

Fusing temporal information recursively is crucial to many applications, such as navigation, robotics, target identification, and multi-target tracking. There are algorithms that can be used to integrate temporal information.[2] Among them, the distributed Kalman filtering[2] and the Bayesian approach[3] are adopted for many applications. However, these algorithms require substantial prior information, such as initial values and initial covariance matrices for the distributed Kalman filtering and prior probabilities for the Bayesian approach. In many cases, the prior information is either not available or not known precisely. Theoretically, some estimated values could be used as the prior information if the models are correct and consistent measurements are provided. However, it takes some time for the systems to converge to the correct values. On the contrary, the Dempster-Shafer technique has the strong capability of handling the information uncertainty, at the cost of more expensive computation.[4,5] Note that the technique has been employed mostly for spatial information fusion[6,7] and for temporal information fusion when the data structure is disjoint.[8]

In this paper, as a general approach of sensor fusion, a STSF (Space and Time Sensor Fusion) scheme is proposed for the joint and disjoint data structure and applied to the landmark identification and a mobile robot's navigation problem. This newly proposed STSF is inevitable for the complementary case where unless there is the sensor fusion, the measurement cannot be completed. Therefore the effectiveness is very clear and the utilization method will be determined by the sensory data structure. However, for the redundant case where some sensor data are not essential for the measurement, it is required to define that how to fuse the previous data sets to the current data set. The minimum square solution is generally adopted for the redundant case fusion without considering the error variance in the measurement.[9,10]

The paper is organized as follows: Section II first presents basic concepts of the conventional fusion, and the concept of STSF is derived. Section III represents a typical example of STSF. And the application of STSF to the mobile robot navigation is illustrated, and the experimental results are shown in Section IV. Finally, Section V summarizes the current research and proposes further topics.

* Department of Electronics Engineering, Pusan National University, Pusan, 609-735 (Korea).
Fax: 82-51-515-5190; E-mail: jmlee@pusan.ac.kr
† City University of Hong Kong.

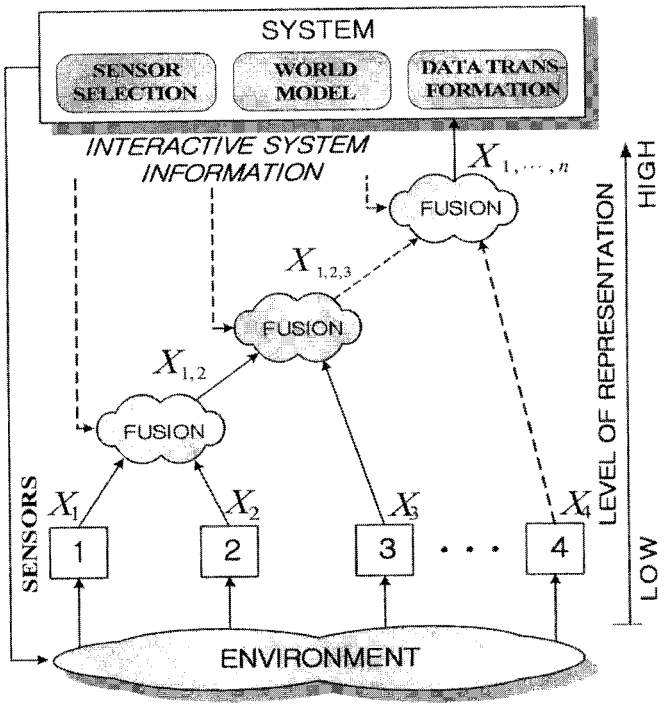


Fig. 1. General pattern of the multi-sensor integration and fusion.

2. CONCEPT OF STSF

2.1 A general pattern of sensor fusion

Multi-sensor fusion refers to any stage in the integration process where there is an actual combination (or fusion) of different sources of sensory information into one representational format. Figure 1 shows a general pattern of multi-sensor integration and fusion in a system. While the fusion of information tasks place at the nodes in the figure, the entire network structure together with the integration functions, shown as part of the system, are parts of the multi-sensor integration process. In the figure, n sensors are integrated to provide the measurement to the system. The output X_1 and X_2 from the first two sensors are fused at the lower left-hand node into a new representation $X_{1,2}$. The output X_3 from the third sensor could then be fused with $X_{1,2}$ at the next node, resulting in the representation $X_{1,2,3}$ that might then be fused at a higher node in the structure. In a similar manner, the output from all n sensors could be integrated into the overall network structure. The dashed lines from the system to each node represent any of the possible signals sent from the integration functions within the system.

One of the simplest and most intuitive sensor fusion is to take a weighted average of the measurements provided by a group of sensors and to use this as the fused value. While this method allows a real-time processing of dynamic low-level data in the most cases, a higher-level sensor fusion is required to achieve a reliable and accurate measurement in the unstructured environment. For the case, the Kalman filter is preferred because it provides a method that does not require heavy computation compared to the weighted average and results in the estimated fused measurements that are optimal in a statistical sense. A weighted average has been used for the multi-sensor fusion in the mobile robot navigation by HILARE,[11] where the sensory information is preprocessed by the threshold operation to eliminate spurious measurements.

Let us define the kth moment data provided by ith sensor as, $z_i(k)$, which is transformed into a measurement, $x_i(k)$. Then the conventional sensor fusion, that is, a spatial fusion technique provides the measurement as

$$\hat{x}(k) = \sum_{i=1}^{n} W_i x_i(k) \tag{1}$$

where $\Sigma_{i=1}^{n} W_i = 1$, $x_i(k) = H_i z_i(k) \in R^m$, n is the number of sensors, H_i represents a transformation from the sensory data to the m dimensional measurement vector, and $W_i \in R^{m \times m}$ represents the weight for ith sensor.

Note that in obtaining $z_i(k)$, the low-level fusion might be applied with multiple sets of data with known statistics.[12] The transformation H_i is purely dependent on the sensor type and the decision of W_i can be done through the sensor fusion process. Later these data are provided to the linear model of the control/measurement system as current state vector, $x(k)$.

2.2 STSF (Space and Time Sensor Fusion)

The STSF (Space and Time Sensor Fusion) scheme combines the sensory information acquired at different instants from different sensors to determine the measurement. It may expend its applicability to the systems where the states at each instant can be predicted, as shown in Figure 2.

Estimation of parameter block may provide the measurement vector at each sampling moment. The blocks of verify the significance and adjust weight are pre-processing stages for the sensor fusion. After these steps, the previous data set will be fused with the current data set, which provides a reliable and accurate data set as the result of multi-sensor temporal fusion. In the figure, the significance implies that how much the previous data set is related to the current data.

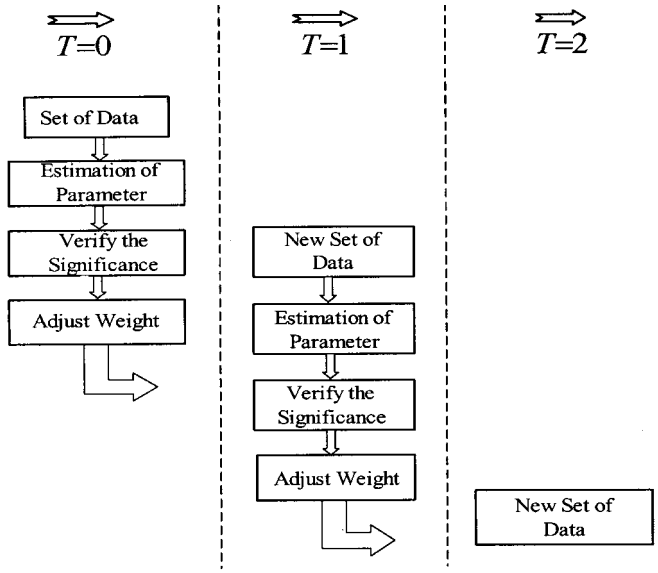


Fig. 2. Data processing for STSF.

https://doi.org/10.1017/S0263574703005381 Published online by Cambridge University Press

An arbitrary value of significance may cause the problem to be complex. Therefore, some may consider whether it corresponds to the same data or not, that is, 1 or 0. When the significance is 0, the weight can be adjusted simply to 0. However, when the significance equals 1, the adjustment of weight should be properly performed to provide reliable and accurate data. In the next sub-section, we will introduce a simple methodology for the weight adjustment and significance decision. The STSF can be represented mathematically as follows:

$$\hat{x}(k) = \sum_{i=1}^{n} W_i \left\{ \sum_{j=1}^{k} P_j TS_i(j) \right\} \tag{2}$$

where $\Sigma_{j=1}^{k} P_j = 1$.

Note that when each of sensor information can provide the measurement vector, that is, the redundant case $TS_i(j)$ can be expanded as

$$TS_i(j) = T_{ij} + H_i z_i(j) \tag{3}$$

where T_{ij} represents the homogeneous transformation from the location of the jth measurement to the ith measurement.

However, when the multi-sensors are utilized in the complementary mode, the transformation relationship cannot be defined uniquely; instead, it will be defined depending on the data constructing algorithm from the measurements. For an example, a single image frame captured by a camera on a mobile robot cannot provide the distance to an object until the corresponding object image is provided from a different location.

2.3 Use of Spectral Estimation Techniques (Auto-correlation)

Each previous data set is transformed to the kth (current) sampling location, and represented by the measurement vector, $TS_i(j)$. Now how can we fuse the k data sets into a reliable and accurate data set? In the Equation (2), W_i can be determined by the geometrical relationship among sensors, in other words, by the spatial sensor fusion considering the noise. While the image frame is tracking a feature, the sensor generates a stream of measurements. In analyzing the measurements from the sensor, the grey values g and image f over time can be treated as random processes. If there is no relative motion between the feature being tracked and the sensor, then the random processes can be viewed as stationary. When there is a relative motion between the feature and the sensor, the processes then cease to be stationary. As an illustration of low-level sensor fusion, we shall only consider the stationary case. In other words, we assume that there is no motion between the sensor and the feature being tracked in the image. The interest in the proceeding analysis lies only in determining whether the process noise is white or not.

Determination of the significance P_j is the final step for the temporal sensor fusion. Note that this expands the dimension of sensor fusion from one to two. As one of solid candidate, we propose here to use the auto-correlation as an index for the significance adjustment and have the form,

$$\Psi_j = \sum_{k=-\infty}^{\infty} TS_i(k) TS_i(j+k). \tag{4}$$

Depending on the correlation, P_j will be determined as:

$$P_j = \frac{\Psi_j}{\Sigma_{j=1}^{k} \Psi_j}. \tag{5}$$

2.4 Related Fusion Methods

Most of the similar approaches among the general multi-sensor fusion methods to the STSF are surveyed in this subsection. Most methods of multi-sensor fusion make explicit assumptions concerning the nature of the sensory information. The most common assumptions include the use of a measurement model for each sensor that includes a statistically independent additive Gaussian error or noise term (i.e. location data) and an assumption of statistical independence among the error terms for each sensor. The difference in the fusion methods exits in their techniques of calibration and threshold to transform the raw sensory data into a particular form. Also the above assumptions are utilized reasonably in achieving the mathematically tractable fusion methods. The conceptually inherent issues in any fusion method that is based on these common assumptions have been reviewed completely by Richardson and Marsh.[13] Their paper provides a proof that the inclusion of additional redundant sensory information almost always improves the performance of any fusion method based on the optimal estimation.

Among the works similar to the STSF approach, there is a research result by M. Rombaut,[14] which is associated within the framework of the European project, *Prometheus*. As a typical sensor for navigation, an active camera is placed on an automobile. There exist blind zones for the camera such as the leftmost, rightmost sides, and the behind. However, the blind zones do not cause any trouble since the role of automobile is following the road mostly. The STSF concept is proved to be very suitable for predicting the curved road by matching the road images with the road image database. A. Nifle[15] also has utilized the STSF concept in classifying the missiles by analyzing their dynamic behaviors. However, the recognition is based on the observation of known events.

3. A TYPICAL STSF

In this STSF approach, the data obtained by the sensors are utilized until they do not have any efficiency for the measurement decision. The data set can be either redundant to improve the accuracy or complementary for the complete measurement. For the latter case, this STSF scheme is essential for the measurement. For example, a single camera system is not capable of measuring the distance to an object. However, if we utilize this STSF scheme, the distance measurement is possible by only a single camera. Note that human keeps the image only about 1/15 seconds on his retina and he cannot measure the distance by the single eye. However, the robot is capable of keeping the image for a

https://doi.org/10.1017/S0263574703005381 Published online by Cambridge University Press

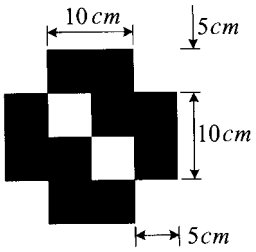


Fig. 3. Landmark pattern and size used by *IRL-2001*.

couple of hours. This enables a single camera to take over the function of stereo camera system that measures the distance using the images captured by the two cameras separated by d Cm.

3.1 Landmark Design for Template Matching

The proposed landmark is made of square patterns with symmetric and repetitive arrangement of black and white patches, as shown in Figure 3. Such arrangement of black and white patterns in the landmark is robust against geometric distortions in indoor environments. Even if the landmark rotates and changes its scale in the image frame, the grey histogram and its projection characteristics do not change very much.

To recognize the landmarks, this paper describes a system for landmark tracking by a template matching approach. An adaptive template matching approach based on a Bayesian decision technique[3] is used for self-localization.

As a measure, how well an arbitrary pattern of grey-values, a *template* $g(m,n)$, matches to a given image $f(i,j)$, the distance function, d, is defined as,

$$d=\int_G (f-g)^2 \text{ or } \max_G |f-g|. \tag{6}$$

The minima of these measures imply the best match. In the discrete case, this takes the form as,

$$M(i,j)=\sum_m \sum_n f(i+m, j+n)\cdot g(m,n), \tag{7}$$

where the maximum of $M(i,j)$ corresponds to the best match. This "cross-correlation" yields a result only if the integral is computed over the whole area G.

To detect the landmark in the cluttered scene robustly, the landmark should show invariant characteristics under some distortions. Invariance under geometric distortions is related to the shape and gray pattern of the landmark, and invariance under the photometric distortions is related to the size and shape in which processing is preformed. Shape and gray pattern of the landmark is also important for robust detection and tracking.

3.2 Fuse of Two Image Frames

Let us consider a short scenario for the mobile robot localization. If a camera captures and stores an image for an object, it moves d Cm to the right, and captures another image for the same object. That is, two image frames are captured for an object by a single camera. This is the same as capturing two image frames by two cameras, d Cm apart at the same time.

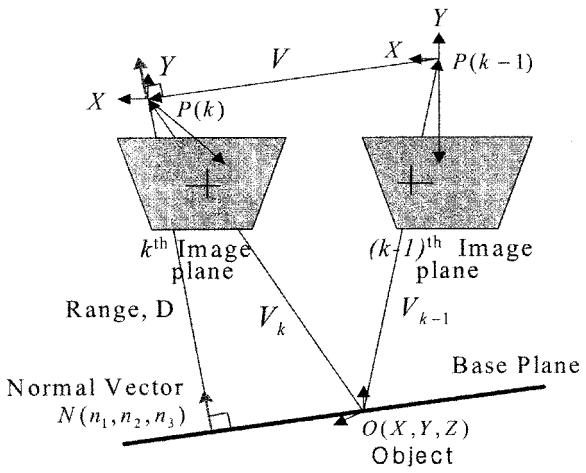


Fig. 4. Transformation of camera coordinates.

However in this scenario, the control of camera motion needs to be accurate enough. If there exist uncertainties in the position control, this will directly affect the measurement error. Even though this error can be minimized through the multi-sensor fusion process, this is a drawback for this method. All the other drawbacks, such as consuming a lot of memory space and requiring a lot of computation in real time, can be released by the rapid improvement of computer technologies.

As a typical geometrical model for a camera, a pinhole model is widely used in vision application fields as shown in Figure 4. At the kth sampling moment, a scene point O(X,Y,Z) is captured by a camera on the mobile robot. The vectors from the object point to the kth and $(k-1)$th camera lens center are represented by V_k and V_{k-1} respectively. The motion of mobile robot from $(k-1)$th moment to kth moment is represented by $\boldsymbol{V}$. Now we can write the vector relationship as:

$$V_{k-1}=V_k-V. \tag{8}$$

This can be represented as a matrix form,

$$\alpha\begin{bmatrix} x_{k-1} \\ y_{k-1} \\ -f \end{bmatrix}=\beta\begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}\begin{bmatrix} x_k \\ y_k \\ -f \end{bmatrix}-\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \tag{9}$$

where $(x_k,y_k,-f)$ and $(x_{k-1},y_{k-1},-f)$ represent the projection of the object point onto the camera image planes, $V(v_1, v_2, v_3)$ represents the translational motion of the mobile robot, r_{ij} is an element of the rotation matrix, R represents the relative rotation between the two camera frames, and α and β are scaling constants.

Now consider the reference base plane passing through the scene point P with a normal vector $N(n_1, n_2, n_3)$, then the range value, D, can be represented as:

$$D=V_k\cdot N. \tag{10}$$

This can be represented again as:

$$D=\beta(n_1x_k+n_2y_k-n_3f). \tag{11}$$

https://doi.org/10.1017/S0263574703005381 Published online by Cambridge University Press

Now, Equation (8) is reformulated as:

$$\alpha\begin{bmatrix} x_{k-1} \\ y_{k-1} \\ -f \end{bmatrix} = \beta\begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}\begin{bmatrix} x_k \\ y_k \\ -f \end{bmatrix} - \frac{\beta}{D}\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}[n_1 \;\; n_2 \;\; n_3]\begin{bmatrix} x_k \\ y_k \\ -f \end{bmatrix} \tag{12}$$

$$(\alpha/\beta)\begin{bmatrix} x_{k-1} \\ y_{k-1} \\ -f \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}\begin{bmatrix} x_k \\ y_k \\ -f \end{bmatrix} \tag{13}$$

where $a_{ij}=r_{ij}-(v_i \cdot n_j/D)$.

In Equation (11), expanding the matrices and dividing rows one and two by row three gives:

$$x_{k-1}=-f\frac{a_{11}x_k+a_{12}y_k-a_{13}f}{a_{31}x_k+a_{32}y_k-a_{33}f} \tag{14}$$

$$y_{k-1}=-f\frac{a_{21}x_k+a_{22}y_k-a_{23}f}{a_{31}x_k+a_{32}y_k-a_{33}f} \tag{15}$$

Equations (12) and (13) can be written more compactly as:

$$D(R_3x_{k-1}+R_1f)=C_3x_{k-1}+C_1f \tag{16}$$

$$D(R_3y_{k-1}+R_2f)=C_3y_{k-1}+C_2f \tag{17}$$

where $R_i=r_{i1}x_k+r_{i2}y_k-r_{i3}f$ and $C_i=v_i(n_1x_k+n_2y_k-n_3f)$.

In matrix form, these equations can be expressed as:

$$AD=B \tag{18}$$

where $A^T=[a\ b]$, $B^T=[c\ d]$, $a=R_3x_{k-1}+R_1f$, $b=R_3y_{k-1}+R_2f$, $c=C_3x_{k-1}+C_1f$, and $d=C_3y_{k-1}+C_2f$.

The pseudo-inverse matrix enables computation of the range value, D that is associated with image point (x_k, y_k), and is written as:

$$D=(A^TA)^{-1}A^TB. \tag{19}$$

The calculation result is:

$$D=\frac{(ac+bd)}{a^2+b^2}. \tag{20}$$

So far, we have shown that using the consecutive two image frames, the distance information of the scene point can be obtained as using the stereo images at a certain moment. This is an example of the complementary sensor fusion.

Since the camera captures multiple image frames for the same object consecutively, the extra image frames from the third can be utilized for the better localization of the mobile robot by the STSF.

3.3 STSF Filter

We represent a single observation of an object as a two dimensional Gaussian distribution in Figure 5. The center, or mean of the distribution is the estimated location of the object and the standard deviations along the maximum and minimum axes of the distribution correspond to the estimates of the uncertainty (or noise) in the observation along the corresponding axis. The value of the distribution at any point corresponds to the probability that the object is actually in that location.

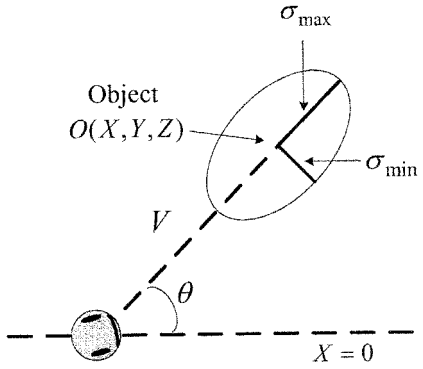


Fig. 5. Gaussian distribution parameter definition: mean (X, Y), standard deviations along maximum and minimum axes σ_{max} and σ_{min}, and distance to the object, V.

For the given observations, we need to determine the mean, standard deviations, and angle of the merged distribution to estimate object position and characterize the quality of the estimate. We compute the mean, standard deviations, and angle of measurement distributions from sensor readings (mean and angle) and models of sensor error (deviations). Thus, we require a method of determining combined parameters from those of individual distributions.

The matrix form of Kalman filtering adopted by Smith and Cheeseman makes this computation relatively simple.[16] Because the mean standard deviations and orientation of the major axis are independent of scaling, they can be extracted from the merged covariance matrices without considering scaling factors.

The canonical form of the two-dimensional Gaussian distribution depends on standard deviation, σ, a covariance matrix, $\underline{C}$, and the mean.[16] The covariance matrix of an observation, $\underline{C}$, is initially determined from the major and minor axis standard deviations as:

$$C\begin{bmatrix} \sigma_{max}^2 & 0 \\ 0 & \sigma_{min}^2 \end{bmatrix}. \tag{21}$$

Since the observation may be oriented arbitrarily with respect to the global coordinate frame, it must first be rotated to align with this frame:

$$C=R(-\theta)^T\underline{C}R(-\theta) \tag{22}$$

where θ is the angle of the distribution's principal axis with respect to the global x-axis. This rotation accomplishes the transformation from observation parameters to the canonical form. Once the observations are in the canonical form, we continue to merge the observations into one.

The covariance matrices of two distributions, C_1 and C_2, can be combined into a single covariance matrix, C, as:

$$C=C_1-C_1[C_1+C_2]^{-1}C_1. \tag{23}$$

https://doi.org/10.1017/S0263574703005381 Published online by Cambridge University Press

Now the mean of the resulting merged distribution, X, is computed from the individual distribution means and covariance matrices as follows:

$$\hat{X}\text{-}\hat{X}_1 + C_1[C_1 + C_2]^{-1}(\hat{X}_2 - \hat{X}_1). \quad (24)$$

4. EXPERIMENTS

4.1 Experimental Environment

The mobile robot used in the experiment is *IRL-2001* developed in the IRL (Intelligent Robot Laboratory), PNU. Initially, it was designed as an intelligent service robot. This robot is shown in Figure 6 along with some of its sensory components. Its main controller is made of Pentium IV Processor that has the system clock 1.0 GHz. The sensors, 16-ultrasonic and a robust odometer system, are installed on the mobile robot. Ultrasonic sensors and infrared sensors in eight sides (25°) detect the obstacles, and the main controller processes this information. For the visual information, a two d.o.f CCD camera is mounted on the top of the mobile robot in order to sense obstacles or landmarks near the mobile robot. And DC servomotors are used for steering and driving of the *IRL-2001* robot.

Experiments were performed in the corridor shown in Figure 7. The conventional sensor fusion and STSF have been tested and compared by experiments to show the usefulness of STSF. The walls in the artificial environment are drawn based on the real map. There are eight landmarks

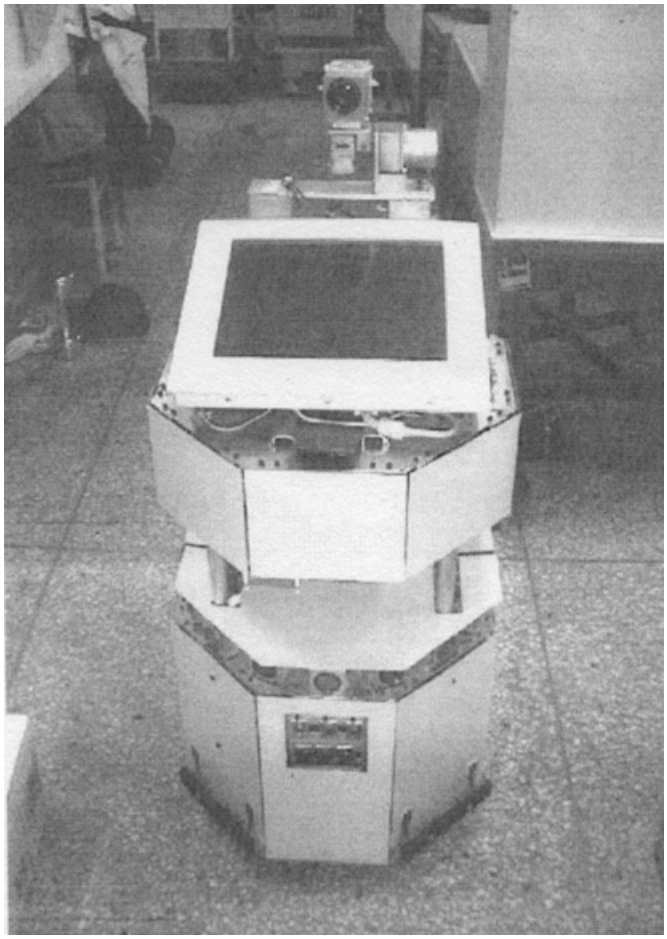


Fig. 6. *IRL-2001* robot.

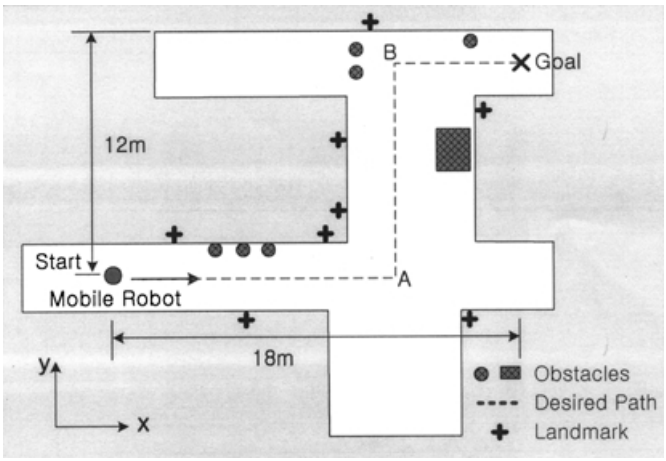


Fig. 7. Corridor environment of the IRL.

on the wall along the navigation path so that at any time the CCD camera on the mobile robot can capture the image of one landmark during the navigation.

4.2 Preliminary STSF Experiments

In the middle of the corridor path, there is a landmark on the wall, which is captured by the CCD camera on the robot.

To begin with, the 2-D landmark used by *IRL-2001* experiments is shown in Figure 3. The primary patterns of landmark are 5cm white square blocks on the black background. The major reasons for choosing the square blocks are provides as follows:[17,18]

- The projection of a square block in the image plane can always be approximated by an ellipsoid, which makes it easy to recognize the landmark using the elliptical Hough transformation technique.
- A square pattern is more robust to noise and occlusoon than circular/polygonal patterns during template matching process, even though all these patterns can be detected by using Hough transformation technique.

Figure 8 shows six different images for the same landmark captured by the moving service robot. Using the consecutive two image frames, the position/orientation of the camera is calculated based on the Equation (20). Each camera position/orientation is represented by a pyramid in Figure 9 where the position is indicated by the vertex and the orientation is shown by the direction of the pyramid.

In obtaining the relative distance of the landmark from the mobile robot based on the Equation (20), the predefined values of the landmark are given as follows: the origin of coordinates is equal to the origin of mobile robot, a Y-axis is fit to face the mobile robot and an X-axis is perpendicular to the Y-axis.

This experiment investigates and compares the performances of the conventional sensor fusion and STSF with respect to the number of image planes (more precisely, the number of landmark images captured by the mobile robot). When the mobile robot finds the landmark, it may capture the landmark images consecutively. In this experiment, at six different locations, the CCD camera captured the landmark images as shown in Figure 8 (1)–(6). For each location, ten image frames are captured and matched to the

https://doi.org/10.1017/S0263574703005381 Published online by Cambridge University Press

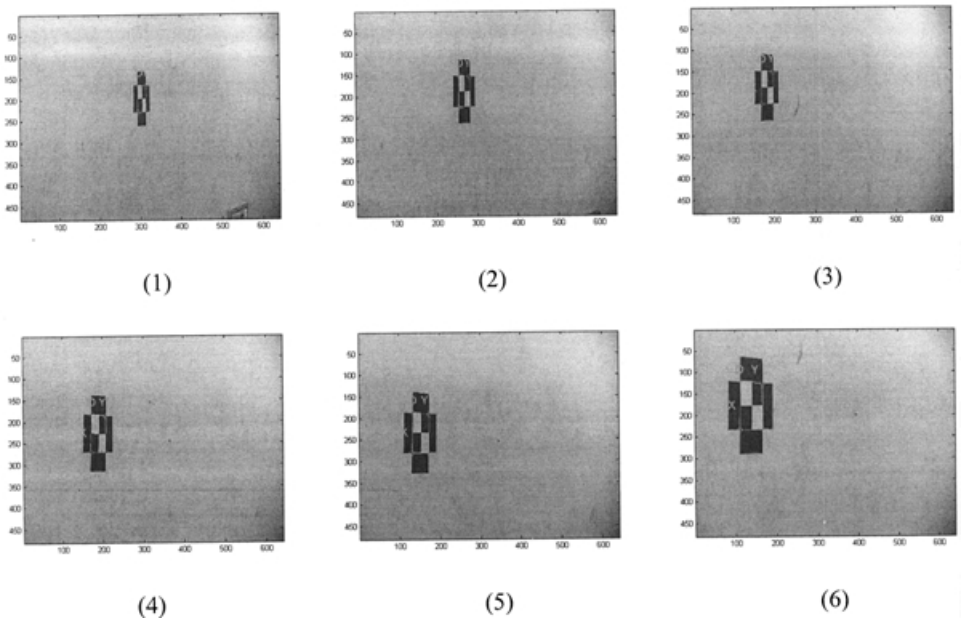


Fig. 8. Landmark locations detected by the camera.

reference images to measure the relative position and orientation. In the extracting the feature points of the landmark, the noise characteristics were estimated that the mean was zero and the standard deviation was 0.5 pixels in the image frame.

Table I compares the distance measurement errors in the conventional sensor fusion and STSF. For the conventional sensor fusion, two consecutive image frames are utilized to calculate the distance, while for the STSF, all of the previous image frames are utilized. Therefore, it is clear that the distance error in the STSF decreases by the number of frames that are utilized as shown in Table I.

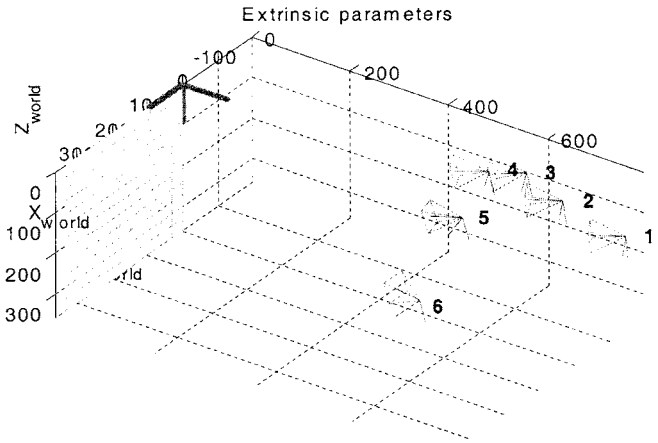


Fig. 9. Mobile robot position and orientation.

4.3 Effectiveness of STSF Using Images

The effectiveness of using images is verified by the experiments shown in Figure 10. The circles indicate the robot trajectory when it utilizes only the odometer data, while the triangles represent the robot trajectory when it utilizes the STSF scheme with images. For both cases of experiment, the mobile robot moves along the X-axis with the velocity of 12.2 cm/sec.

The tracking by only odometer data has an approximately 40cm deflection error after 10 m navigation along the Y-axis. It took 84 seconds. The robot position error comes from wheel slippage, a rough surface, and sensor error. The further the robot moves from the start point, the greater the error becomes in real experiments. The smallest error in the X-axis was 0.2 cm and in the Y-axis was 0.13 cm at the start point. And the biggest error in the X-axis and Y-axis were 21.2cm and 40.53cm, respectively.

To overcome the navigation proportional error problem, STSF scheme is used, which utilizes the landmark to recognize the current position of the mobile robot. In the

Table I. Result of the distance measurement.

		Conventional sensor fusion		STSF	
Frame	World Coordinate Distance (m)	Image Coordinate Distance (m)	Error (m)	Image Coordinate Distance (m)	Error (m)
1	7.81	8.24	0.43	8.24	0.43
2	7.02	7.36	0.34	7.30	0.28
3	6.28	6.53	0.25	6.48	0.20
4	5.06	4.89	0.17	4.92	0.14
5	5.52	5.39	0.13	5.63	0.11
6	6.32	6.46	0.14	6.43	0.11

https://doi.org/10.1017/S0263574703005381 Published online by Cambridge University Press

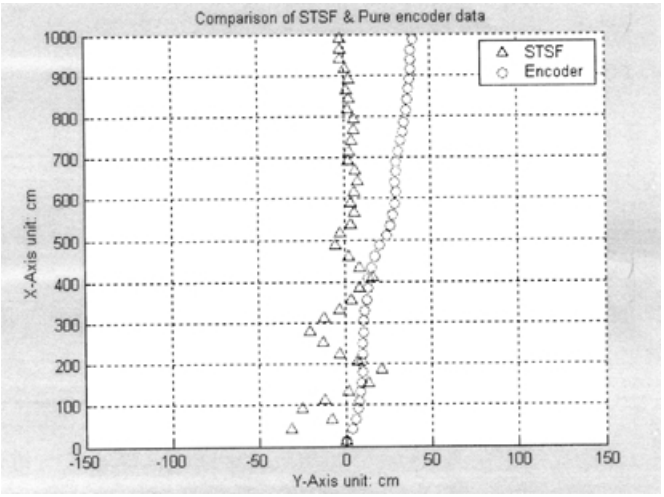
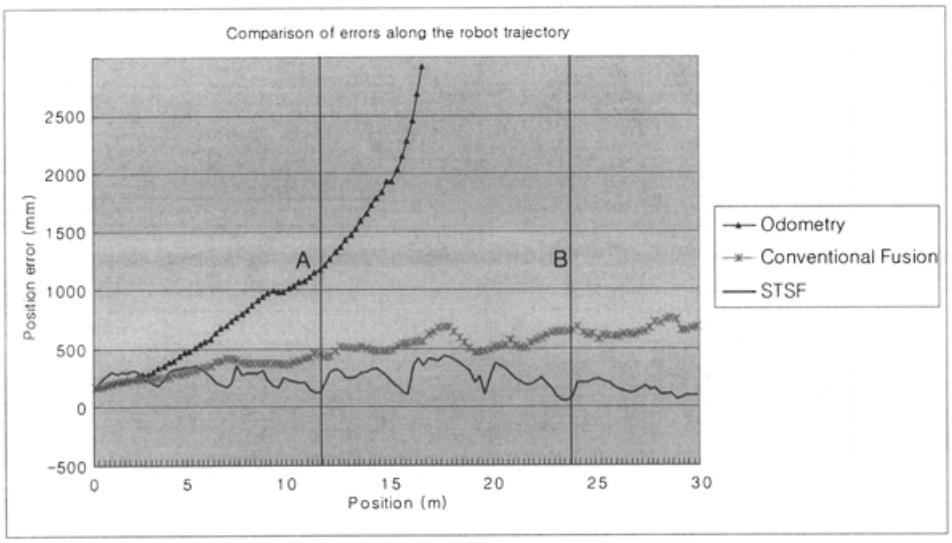


Fig. 10. Trajectories of the mobile robot by the encoder data and by the STSF.

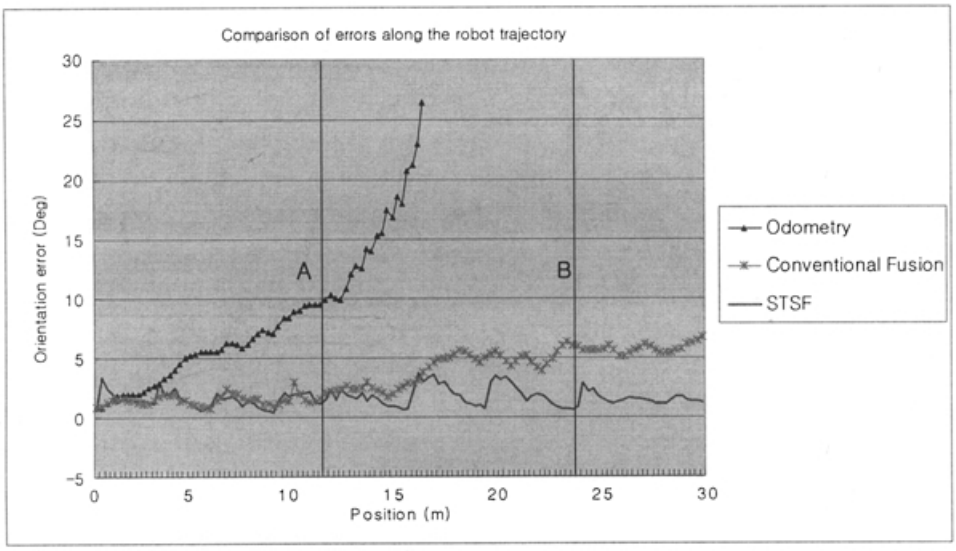
STSF experiments, the average time spent to judge the existence of a landmark was 0.127 sec and to localize was 0.332 sec. The experimental data with the STSF have 20–30 cm drift errors after the 4 m navigation. However the navigation error is reduced to 10 cm at the 5 m navigation and kept below the value later on, which explicitly shows the effectiveness of the STSF. The navigation took 92 seconds that is 8 seconds longer than the previous case. The reason is not because of image processing time but because of the frequent orientation change of the mobile robot to keep the desired path.

4.4 Navigation Experiments

The navigation is planned to follow the path from the start point to the goal using the *IRL-2001* mobile robot, as shown in Figure 6. Three different localization schemes are utilized for the estimation of the mobile robot position during the navigation: the odometer based estimation, the conventional sensor fusion, and the STSF.



(a) Position error.



(b) Orientation error.

Fig. 11. Position/orientation errors in the navigations.

https://doi.org/10.1017/S0263574703005381 Published online by Cambridge University Press

The navigation characteristics are illustrated by the position/orientation error in Figure 11. In the experiment with the odometer based estimation, the mobile robot stopped at 17 m location by colliding to the wall as the result of the accumulated error in reading the encoder. For the latter two schemes, the navigations were successfully completed to the goal utilizing the landmarks to localize the mobile robot. However, in the navigation with the conventional sensor fusion, the position/orientation error increases gradually along with the traveling distance since the uncertainties in measuring the relative distance becomes bigger and bigger.[19,20] The STSF has the best performance, especially in the latter part of the path. The matching rate of landmarks is about 22% higher for the STSF than for the conventional sensor fusion, which provides the better localization performance.

5. CONCLUSIONS

In this paper, a new sensor fusion concept, STSF (Space and Time Sensor Fusion), was introduced. The effectiveness of STSF was demonstrated through the preliminary and navigation experiments. To follow the navigation trajectories without *a priori* information on the environment, not only the data from the sensors located at different places but also the previous sensor data are inevitably necessary. This scheme may require more memory space and computing power in the navigation system. However, it becomes nonpoisonous with the rapid price drop of I.C.s. Sonar and vision systems can be cooperatively utilized for collision avoidance based upon STSF such that a mobile robot successfully navigates in an unstructured environment, as well as in a structured environment. Although we have used an active CCD camera system for landmark recognition and navigation in indoor environment, the results from the experiments clearly show that by utilizing the mobile robot and by applying the active sensing to adapt to different situations, a high level of competent collision avoidance behavior can be achieved by the STSF.

Based on these results, further experiments will aim at applying the STSF to the control of a mobile robot in an unstructured environment with various sensors.

References

1. R.C. Luo, C.C. Yih and K.L. Su, "Multisensor fusion and integration: approaches, applications, and future research directions," *Sensors Journal, IEEE* **2**, Issue 2, 107–119 (2002).
2. L. Hong and A. Lynch, "Recursive temporal-spatial information fusion with applications to target identification," *Aerospace and Electronic Systems, IEEE Transactions on*. **29**, Issue 2, 435–445 (1993).
3. S. Thrun, "A Bayesian approach to landmark discovery and active perception for mobile robot navigation," (*Tech. rep. CMU-CS-96-122*). Pittsburgh, PA: Carnegie Mellon University, Department of Computer Science (1996).
4. A.P. Dempster, N.M. Laird and D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Statist. Soc*. **39**, 1–38 (1977).
5. W. Pieczynski, "Unsupervised Dempster-Shafer fusion of dependent sensors," *Proc. IEEE Southwest Symp. Image Analysis and Interpretation* (SSIAI'2000), Austin, TX, (2000) pp. 247–251.
6. R.C. Ruo and K.L. Su, "A Review of High-level Multisensor Fusion: Approaches and applications," *Proc. of IEEE Int'l. Conf. On Multisensor Fusion and Integration for Intelligent Systems*, Taipei, Taiwan, (1999) pp. 25–31.
7. P. Weckesser and R. Dillman, "Navigating a Mobile Service-Robot in a Natural Environment Using Sensor-Fusion Techniques," *Proc. of IROS*, (1997) pp. 1423–1428.
8. L.A. Zadeh, "Outline of a New Approach to the Analysis of Complex Systems and Decision Processes," *IEEE Transactions on Systems, Man, and Cybernetics* **3**(1), 28–44 (1973).
9. V. Ayala, J.B. Hayet, F. Lerasle and M. Devy, "Visual localization of a mobile robot in indoor environments using planar landmarks," *Intelligent Robots and Systems, (IROS 2000). Proceedings. IEEE/RSJ International Conference on*. (2000) **Vol. 1**, pp. 275–280.
10. A. Ohya, A. Kosaka and A. Kak, "Vision-Based Navigation by a Mobile Robot with Obstacle Avoidance Using Single-Camera Vision and Ultrasonic Sensing," *IEEE Transactions on Robotics and Automation* **14**(6), 969–978 (1998).
11. R. Chatila, J. Laumond, "Position referencing and consistent world modeling for mobile robots," *Robotics and Automation. Proceedings. 1985 IEEE International Conference on* (1985) **Vol. 2**, pp. 138–145 .
12. J.M. Lee, B.H. Kim, M.H. Lee, M.C. Lee, J.W. Choi and S.H. Han, "Fine Active Calibration of Camera Position/Orientation through Pattern Recognition," *Proc. of IEEE Int'l. Symp. on Industrial Electronics*, Slovenia, (1999) pp. 100–105.
13. J.M. Richardson and K.A. Marsh. "Fusion of multisensor data," *International Journal of Robotics Research* **7**(6), 78–96 (1988).
14. M. Rombaut and D. Meizel, "Dynamic data temporal multisensor fusion in the Prometheus Prolab2 demonstrator," *IEEE Int. Conference on Robotics and Automation*, San Diego, (1993), pp. 36–76.
15. A. Nifle and R. Reynaud, "Behavior classification based on events occurrence in possibility theory," (in French) *Traitement du signal* **Special 14**(5), 423–434 (1997).
16. R.C. Smith and P. Cheeseman, "On the Representation and Estimation of Spatial Uncertainty." *The Int. J. Robotics Research* **5**(4), 56–68 (1986).
17. S. Bentalba, A. ElHajjaji and A. Tachid, "Fuzzy Control of a Mobile Robot: a New Approach," *Proc. IEEE International Conference on Control Applications*, (1997), pp. 69–72.
18. H.R. Beom and H.S. Cho, "A Sensor-Based Navigation for a Mobile Robot Using Fuzzy Logic and Reinforcement Learning", *IEEE Trans. on System, Man, and Cybernetics* **25**(3), 464–477 (1995).
19. M. Kam, X. Zhu and P. Kalata, "Sensor Fusion for Mobile Robot Navigation," *Proc. of the IEEE* **85**(1), 108–119 (1997).
20. F. Delmotte and P. Borne, "Modeling of Reliability with Possibility Theory," *IEEE Trans. On Systems, Man, and Cybernetics* **28**(1), 78–88 (1998).

https://doi.org/10.1017/S0263574703005381 Published online by Cambridge University Press