A Hybrid People Tracker on Moving Platforms†

YING-YING CHAO, CHUNG-HSIEN HUANG*, MING-YU SHIH AND BWO-CHAU FU
Advanced Technology Center, Information & Communications Research Laboratories
Industrial Technology Research Institute, Taiwan

ABSTRACT

A method to detect and track moving objects with a Laser Range Finder (LRF) and a camera on a mobile platform is proposed. A LRF provides high-precision 2-D scene structure information, however, the small amount of measurement variation combined with the biases in odometry estimation makes it extremely difficult for pure laser-based algorithms to provide robust detection and tracking, especially when moving objects are close to a static background structure. In the proposed method, a hybrid tracker is developed which takes full advantage of the high-precision of a LRF-based Kalman-filter tracker and the flexibility of vision-based mean-shift tracker. With scan matching and background subtraction on laser-scanning data, moving objects are most-of-the-time tracked by a LRF-based Kalman filter. In circumstances where a vision-based mean-shift tracker provides better results, it takes over to maintain tracking until the LRF-based tracker is stabilized. Experimental results show the proposed method provides robust moving people detection and tracking on a mobile platform.

Key words: mobile platform, Kalman filter, mean-shift, scan matching, background subtraction.

1. INTRODUCTION

Moving object detection and tracking are two of the most critical issues within the field of computer vision study. They are extremely important in applications such as video indexing, human-computer interaction, autonomous surveillance, traffic monitoring and vehicle navigation, for the detection and tracking of interesting objects between frames for the purpose of performing advanced scene analysis.

Conventional moving object detection techniques use fixed cameras (or other sensors) to monitor interesting scenes and build their appearance models for background subtraction. These techniques provide moving object detection, in most cases using fixed cameras. In some circumstances, however, such as intruder-following with a security robot, human interaction with a companion robot, navigation on an automatic vehicle or obstacle-avoiding on an intelligent wheelchair, cameras are no longer fixed in one location. The motion of moving platforms makes background modeling extremely difficult. Therefore, the increasing demand of mobile platform applications generates tougher issues for moving object detection and tracking.

Many researchers have worked on the study of moving object detection and tracking for applications on mobile platforms. In 1991, Zhu (Zhu, 1991) was the

† This work is a partial result of Project 6353C41300 conducted by the Industrial Technology Research Institute under sponsorship from the Ministry of Economic Affairs, R.O.C, Taiwan.
* Corresponding author. E-mail: DavidCHHuang@itri.org.tw

first to model a moving object's motion, by the Hidden Markov Model, to avoid collision between a moving platform and moving objects. However, this algorithm was only examined in a computer-simulated environment, without actual deployment in real space. The rapid advances in laser sensor technologies encouraged researchers to have more and more mobile platforms taking advantage of a laser's high-precision depth estimation to obtain spatial information of the outer environment. This led Kluge (Kluge, Koehler & Prassler, 2001) to build a wheelchair equipped with a laser range finder (LRF) to collect environment data. The data was then further analyzed by using a computer-graphic based method. Since Kluge's method did not model the behavior of moving objects, it could not keep tracking moving objects while occlusions occurred. Fod (Fod, Howard & Matarić, 2002) utilized a Kalman filter while Lindström (Lindström & Eklundh, 2001) adopted a Gaussian hypotheses to track moving objects. Their proposed methods solved object occlusion problems for a short period occlusion. Montemerlo (Montemerlo, Thun & Whittaker, 2002) proposed the conditional particle filter, a probability-based algorithm which tracks a large distribution of object locations conditioned upon a smaller distribution of robot positions over time. Since this algorithm used different particle filters to simultaneously estimate the positions of a robot and moving objects, it could localize their positions at the same time. However, when the number of moving objects is increased, mismatched tracking arose frequently because their methods used a nearest neighbor (NN) method for data association. To solve this problem, Schulz (Schulz, Burgard, Fox & Cremers, 2003) applied a Sample-based Joint Probabilistic Data Association Filter (SJPDAF) to replace the NN-based method; Almeida (Almeida, Almeida & Araújo, 2005) combined SJPDAF with a particle filter to estimate the non-linear and non-Gaussian behavior of moving objects. The state-of-the-art research may be referred to Wang's work (Wang, Thorpe & Thrun, 2003). Wang integrated the study of Detection And Tracking of Moving Object (DATMO) and Simultaneous Localization And Mapping (SLAM) by using a probability-based method which estimates the uncertainty of the data obtained from a LRF and odometer. This framework seamlessly integrated robot localization, map reconstruction, and moving object detection while providing a robust result. In addition, Lee (Lee, Tsubouchi, Yamamoto & Egawa, 2006) considered different situations within the scan range of a LRF, such as entrance, exit, and occlusion, and modeled the human gait to track people by using an Extended Kalman filter. MacLachlan (MacLachlan & Mertz, 2006) mounted a LRF on a bus to detect cars and pedestrians. Moreover, radar and infrared sensors are frequently used on mobile platforms, but these devices are limited in depth estimation and spatial resolution thus providing only limited support for moving object detection and tracking in dynamic scenes.

In this paper, we focus on the detection and tracking of moving objects using a camera and a LRF on a mobile platform. Pure vision-based approaches (Irani, Rousso & Peleg, 1997; Talukder & Matthies, 2004; Talukder, Goldberg, Matthies & Ansar, 2003) provide useful detection and tracking results in certain circumstances, but these approaches are sensitive to illumination variations, shadows, and parallax effects. Moreover, the time-consuming nature of these

approaches makes them difficult to operate in an unconstrained environment. On the other hand, a LRF is insensitive to the light effects and provides accurate depth information which forms discrete points on a 2-D horizontal plane in 3-D world space. A significant feature of a laser sensor is the instability of the emitting direction of laser rays. As shown in Figure 1, even though the robot is not moving, the scanning points circulated are unstable during different scans. A conventional way to cope with this problem is to give each scanning point an error-tolerant range. Since the LRF emits in a radial manner, it has a higher sampling rate when the object is close to the LRF. Therefore, in this situation a lower error-tolerant range should be set; contrariwise, the farther the object is, the higher the error-tolerant range. However, this strategy deducts the sensitivity advantage of LRF sensors. For example, when a moving object is close to a static background within the tolerable range, the moving object is regarded as a part of the background, resulting in failed object detection. Consequentially, we tried to make a camera and a LRF mutually beneficial to keep both their advantages while compensating for the limits of each. A framework that combines a vision-based approach with the use of a high-precision LRF to provide a robust moving object detection and tracking for mobile platform applications is therefore proposed.

The proposed framework is described as follows: With scan matching and background subtraction on laser-scanning data, moving objects can be detected and then tracked by a LRF-based Kalman filter. However, LRF tracking frequently failed when the moving objects are too close to the background or stay for such a length of time as to be regarded as background. Although a LRF takes advantage of high-precision measurement in depth, it covers only a 2-D plane in a 3-D world space. In contrast, cameras capture the color information of the complete object with the absence of depth information. To keep both the advantages of a LRF and a camera while compensating for the limits of each, a vision-based mean-shift tracker (Comaniciu, Ramesh & Meer, 2000; 2003) is applied. The mean-shift tracker is widely applied in computer vision research; however, it requires an image template of the tracked object and its initial position, and only returns a 2-D position of the moving object on an imaging plane. Consequently, by integrating the information from both a Kalman-filter tracker in the laser-domain and a mean-shift tracker in the image-domain, a hybrid tracker is proposed.

This paper is organized as follows: The problem definition and the proposed framework are described in Section 2. Experimental results are revealed and discussed in Section 3. Finally, a conclusion is presented in Section 4.

2. MATERIALS AND METHODS

The proposed framework uses a mobile platform with a LRF and a camera to detect and track moving objects. The system flowchart is illustrated in Figure 2. The observed data $O_t=\{S_t, I_t\}$ was obtained from the LFR and camera, where S_t denotes the calibrated laser scanning data and I_t is the image captured at time t.

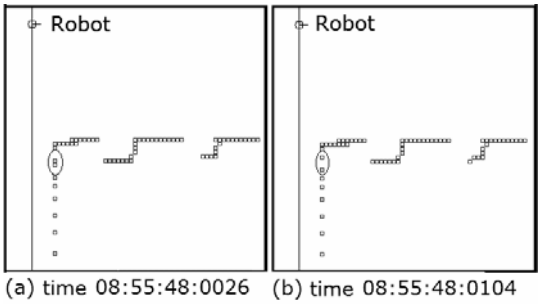


Figure 1. The laser readings are unstable due to the small amount of measurement variations from LRF devices. The effects of these small variations are enlarged on mobile platforms while biases in odometry estimation occur.

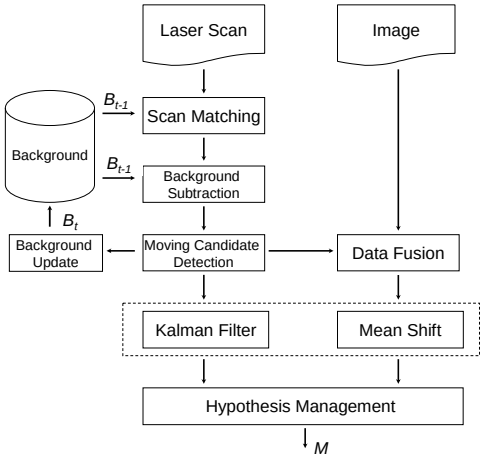


Figure 2. Flowchart of the proposed framework.

The coordinate of the raw data obtained by the LRF is computed by bearing α and range γ of a scan. According to the platform's position and heading obtained from the platform's odometry, the scan data can be transformed from the laser's coordinate system to the world coordinate system. However, considering the measurement error of the odometry and the drifting effect caused by the platform's wheels, the current scan data is aligned to the background model by using an iterative closest point algorithm (ICP) (Besl & McKay, 1992), a scan matching algorithm. While current scan and background model are aligned, candidate moving objects can be detected with background subtraction.

Furthermore, since the camera and the LRF are calibrated, candidate moving objects detected by LRF can be projected onto the image plane. After data fusion on laser-scanning data and image, each point of the laser scan is projected to its corresponding position on the image, as shown in Figure 3. Meanwhile, a

background model B_{t-1} is maintained using the laser-scanning data from time 0 to time t-1 by a local background modeling method. A Kalman filter is applied to provide moving object tracking on laser data.

The measurement error of a LRF makes it difficult to distinguish moving objects close to static background, such as walls. In such cases, a vision-based object tracking methods such as a mean-sift algorithm is integrated to track moving objects on an image plane. Finally, a hypothesis management module which integrates a Kalman-filter and mean-shift is applied to manage hypothetical trajectories and track moving objects (Gong, 2005).

The detail of each module is presented in the following subsections.

2.1 Calibration and Fusion of An LRF and Camera

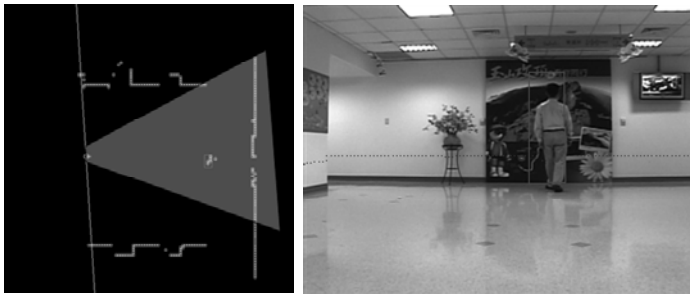
The LRF and camera are rigidly attached to the mobile platform. With corresponding transformations T^L and T^C, the LRF and camera readings are transformed to the platform's coordinate system for data fusion. As shown in Figure 4, every scan of the LRF consists of hundreds of laser rays and each laser ray is represented as a bearing parameter α and a range parameter γ. A laser point *p,* detected by a laser ray in 3-D workspace, is first represented using the platform's coordinate system according to Equation (1) and then transformed to the camera's coordinate system as p^c by Equation (2):

$$p = T^L \begin{pmatrix} d\cos\alpha \\ d\sin\alpha \\ 1 \end{pmatrix}, \tag{1}$$

$$p^C = \begin{pmatrix} x^c \\ y^c \\ z^c \end{pmatrix} = (T^C)^{-1} p\,. \tag{2}$$

Using a pinhole camera model, after camera calibration, the camera's intrinsic parameters, focal length, principal point, and distortion can be obtained and results in a projection matrix T^P (Hartley & Zisserman, 2003). With this projection matrix, a laser point p^c on 3D space could be projected onto its corresponding 2D image coordinates *(u, v)* through the following equation:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = T^P p^C = T^P (T^C)^{-1} T^L \begin{pmatrix} d\cos\alpha \\ d\sin\alpha \\ 1 \end{pmatrix}. \tag{3}$$



(a) (b)

Figure 3. Data fusion of laser data and image. (a) The laser data represented by an occupancy-grid model and the field of view of the camera. (b) Fusing the laser data on the corresponding image frame.

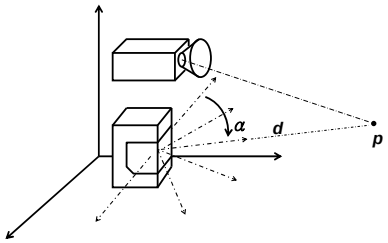


Figure 4. LRF and camera calibration.

2.2 Local Background Modeling

The occupancy-grid model proposed by Moravec and Elfes (Moravec & Elfes, 1985) is the most widely used world-representation model in robotics. As illustrated in Figure 5, an occupancy grid describes a robot's world as a two dimensional array of probability values. The occupancy-grid model is usually constructed by integrating the scan readings from LRFs, sonar, or infrared sensors installed on mobile platforms. In this model, each grid represents a user-defined discrete location marked as occupied or empty according to *a priori* observations. However, since the scan readings are discrete, readings on long-distance objects are much sparser than readings on close-by objects. This made it extremely difficult to maintain an accurate background model, especially for a far away scene structure.

Centered on the LRF as origin, each grid *(i, j)* at time *t* is associated with an occupancy confidence c_t which is updated along time with the following equation:

$$c_t(i,j) = w \times c_{t-1}(i,j) + (1-w) \times o_t^L(i,j) \tag{4}$$

where o_t^L stands for the current observation from LRF and *w* is the adaptation rate. When an obstacle is observed at grid *(i, j)*, it is marked as $o_t^L(i,j)=1$ for occupied. Otherwise, $o_t^L(i,j)=0$ for empty. In cases having higher confidence on historical observations, *w* is given a higher value to represent higher confidence than on the current observation. In our case, *w* is empirically set to 0.9.

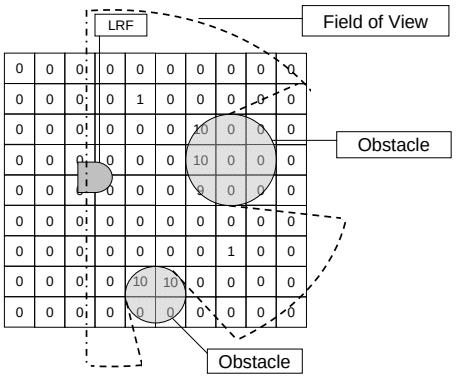


Figure 5. Occupancy grid after 10 scans.

In many cases, a patrolling robot moves freely in a wide area. It is unwise to maintain a huge or unlimited background model in the system memory. Applications such as moving object detection need only short-term information to obtain accurate results. Therefore, a local background model is implemented.

Let the initial position of the robot (x_G, y_G) be the origin of the global coordinate system. A W×H-unit memory space is allocated to maintain the local background model. In each unit, the occupancy confidence c_t is stored, and the correspondence between the global coordinate system and the local background model is estimated through the coordinate of the left-top corner (x_L, y_L) of the background model. To update the background model when the robot moves to (x_R, y_R), Equation (5) and Equation (6) are utilized, and the content of each unit is updated according to its spatial relationship to the left-top corner of the background model.

$$x_L = x_R - (W/2) + x_G\,, \tag{5}$$

$$y_L = y_R - (H/2) + y_G\,. \tag{6}$$

2.3 Scan Matching

Most mobile platforms rely on on-board odometers to estimate 2-D platform positions. However, the odometry measurement error and the drifting effect caused by platform wheels make odometry readings unreliable, especially in rotation

movement. Scan matching is a process that is frequently used to compensate for unreliable odometry readings and to align laser scans obtained at different times.

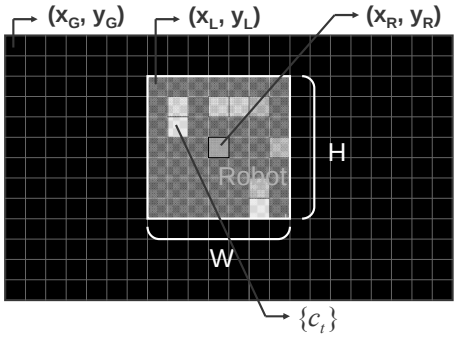


Figure 6. Illustration of the local background model.

The most popular scan-matching method is the iterative closest point algorithm (ICP) (Besl & McKay, 1992). ICP was originally a 3-D point registration algorithm applied widely in the computer vision community. After the seminal work by Lu (Lu & Mulios, 1997), ICP is utilized nowadays to align range scans and estimate robot motion. Figure 7 illustrates the procedure of ICP. First, as shown in Figure 7(a), the closest point from one data set to the other is identified. Second, as shown in Figure 7(b), the rigid transformation relating to the corresponding points is computed by a least-mean-square method. The algorithm then re-determines the closest point set and repeats until it converges to a minimum match between the two scans, as shown in Figures 7(c) and (d).

Suppose the point set of a background model is denoted as $A=\{a_1, a_2,\ldots, a_m\}$, and the point set of a current scan is denoted as $B=\{b_1, b_2,\ldots, b_n\}$ while the robot is currently at c (x_R, y_R) where $a, b, c \in R^2$. ICP registers both background and current data sets by finding the best least-mean-square matching. Detailed ICP iterations are summarized as the following four steps.

Step 1. For each point a_k from the data set A, find its closest neighbor point b_k from the data set B., i.e. $d(a_k, b_k) = \min\{d(a_k, B)\}$.

Step 2. Compute a transform T which comprises one rotation parameter and two translation parameters by using the least-mean-squares method.

Step 3. Calculate the following objective function:

$$f(T) = \frac{1}{m}\sum_{i=1}^{m}\left\|a_i - T(b_i)\right\|^2 . \qquad (7)$$

Step 4. Apply the transform T to the data set B. If the stopping conditions are satisfied such as that the cost function is smaller than a given threshold or the iteration number reaches a given number, then the algorithm goes to

step 5, otherwise it goes back to the first step to rematch the closest points of A from the updated B.

Step 5. Since ICP corrects the measurement error of odometry, the robot's position is then updated to $c' = T(c)$.

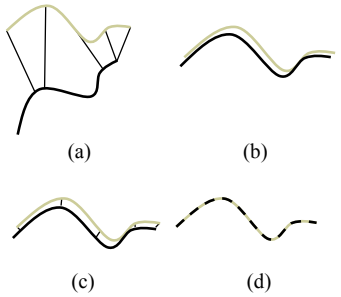


Figure 7. Illustration of ICP. (a) Find the closest point from one data set to the other. (b) Calculate the transformation from the relation in (a). (c) Re-determine the closest point set. (d) The convergence.

2.4 Hypothesis Management and the Hybrid Tracker

After the scan matching adjustment, the differences between the background model and the most up-to-date transformed scan are marked as occupied units. A connected component algorithm is then applied to group those connecting units as candidates for moving objects. Each candidate moving object is hereby called a hypothesis for clarity (Gong, 2005).

At the start of the algorithm we maintain a list of hypotheses in the memory. If a current hypothesis is similar enough to an element in the hypothesis list, this hypothesis will be associated with this element by updating it into the hypothesis list. Otherwise, this hypothesis is identified as a new element to be inserted into the hypothesis list.

In this study, a hybrid tracker is developed to handle issues of hypothesis updating, as shown in Figure 8. First, a hypothesis $\mathbf{c}_t^i \in \mathbf{C}_t$, the i-th connected component at frame t, is compared with all elements in the hypothesis list $\mathbf{M} = \{\mathbf{H}_{t-1}^j \mid j = 1,2,...,n\}$, which maintains n possible elements while each element $\mathbf{H}_{t-1}^j$ records the history of the associated objects from time 0 to t-1, i.e. $\mathbf{H}_{t-1}^j = \{\mathbf{h}_0^j, \mathbf{h}_1^j, ..., \mathbf{h}_{t-1}^j\}$ where $\mathbf{h} \in R^2$. Each element of hypothesis list $\mathbf{H}_{t-1}^j$ has a status flag $S_{t-1}^j = \{L \mid I\}$ indicating the tracking domain. The status L means that the object is tracked in the laser domain, while I represents a tracking in the image domain. Meanwhile, each element $\mathbf{H}_{t-1}^j$ maintains an image template $\mathbf{T}_{t-1}^j$ for the possible need of a vision-based mean-shift tracking.

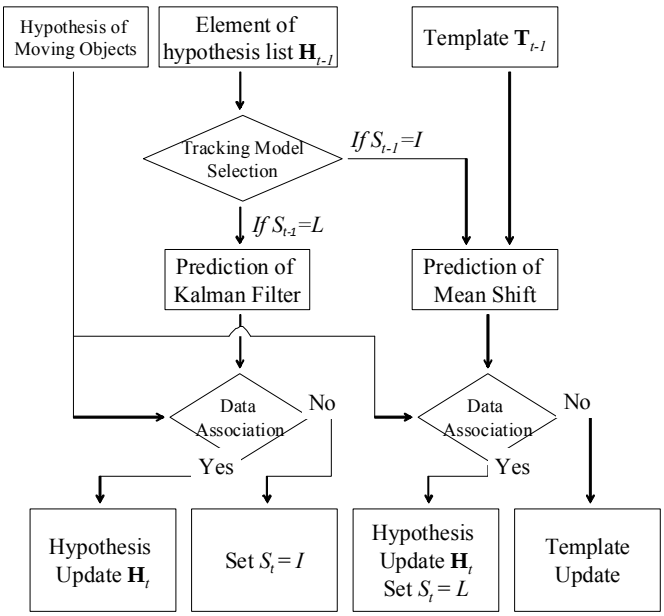


Figure 8. The flowchart of the hypothesis update.

The image template $\mathbf{T}_{t-1}^j$ is a patch extracted from the image according to the projected laser-scanning data $\mathbf{h}_{t-1}^j$. The rectangle in Figure 3(b) is an example of the image template. When we track the moving object in the laser domain, a Kalman filter is utilized to predict the future position of each element of the hypothesis list $\mathbf{H}_{t-1}^j$ at time t as $\tilde{\mathbf{h}}_t^j = KF(\mathbf{H}_{t-1}^j)$. If $\mathbf{c}_t^i$ is close enough to the prediction $\tilde{\mathbf{h}}_t^j$ of an existing element $\mathbf{H}_{t-1}^j$, that is, $(\tilde{\mathbf{h}}_t^j - \mathbf{c}_t^i) < \varepsilon_1$, $\mathbf{c}_t^i$ is then regarded as $\mathbf{h}_t^j$, updating $\mathbf{H}_{t-1}^j$ to $\mathbf{H}_t^j$.

If an element of the hypothesis list is not associated with any hypothesis of a moving object, it means that the object of the element may be out of the field of view, merged to the background, or not detected at the background subtraction stage, $S_t^j = I$ is set and the mean-shift algorithm is evoked at the next frame. The mean-shift keeps tracking the moving object and updates the template $\mathbf{T}_t^j$ by its prediction $\tilde{\mathbf{T}}_{t-1}^j$. If a patch corresponding to a new hypothesis $\mathbf{c}_t^i$ is close to $\tilde{\mathbf{T}}_{t-1}^j$ when a new frame comes, it means that the moving object is robustly detected in the laser domain and laser-based detection and tracking with Kalman filter is applied again before the updating of the element in the hypothesis list $\mathbf{H}_t^j$. The pseudo-code of the hypothesis management is listed in Algorithm 1.

Algorithm 1: Hypothesis Management

Input: $\mathbf{M}_{t-1}$, $\mathbf{T}_{t-1}$, $\mathbf{S}_{t-1}$, $\mathbf{C}_t$
Output: $\mathbf{M}_t$, $\mathbf{T}_t$, $\mathbf{S}_t$,

for each element j, **do**
  if $S_{t-1}^j = L$, **then**
    $\widetilde{\mathbf{h}}_t^j = KF(\mathbf{H}_{t-1}^j)$
    if $(\widetilde{\mathbf{h}}_t^j - \mathbf{c}_t^i) < \varepsilon_1$, **then**
      $S_t^j = S_{t-1}^j$, $\mathbf{H}_t^j = \mathbf{H}_{t-1}^j \cup \mathbf{c}_t^i$
      $\mathbf{T}_t^j = GetTemplate(\mathbf{c}_t^i)$
      remove $\mathbf{c}_t^i$ from $\mathbf{C}_t$
    else
      $S_t^j = I$, $\mathbf{H}_t^j = \mathbf{H}_{t-1}^j$, $\mathbf{T}_t^j = \mathbf{T}_{t-1}^j$
    end if
  else
    $\widetilde{\mathbf{T}}_t^j = MS(\mathbf{T}_{t-1}^j)$
    if $(\widetilde{\mathbf{T}}_t^j - GetTemplate(\mathbf{c}_t^i)) < \varepsilon_2$, **then**
      $S_t^j = L$, $\mathbf{H}_t^j = \mathbf{H}_{t-1}^j \cup \mathbf{c}_t^i$
      $\mathbf{T}_t^j = GetTemplate(\mathbf{c}_t^i)$
      remove $\mathbf{c}_t^i$ from $\mathbf{C}_t$
    else
      $S_t^j = S_{t-1}^j$, $\mathbf{H}_t^j = \mathbf{H}_{t-1}^j$, $\mathbf{T}_t^j = \widetilde{\mathbf{T}}_t^j$
    end if
  end if
end for
for each $\mathbf{c}_t^i$ **in** $\mathbf{C}_t$, **do**
  $\mathbf{M}_t \leftarrow Insert(InitHypo(\mathbf{c}_t^i))$
end for

3. EXPERIMENTAL RESULT

To verify the proposed algorithm, a preliminary experiment is conducted by setting a JAI CV-s3200 CCD camera and a SICK laser on a Pioneer DX-3 mobile platform, as shown in Figure 9. The camera produces consecutive images with a 320x240 resolution at 22 frames/per second, while the LRF provided 24 laser readings per second. A laptop is installed on the DX-3 to command its motion and connects to the camera and the LRF to perform the proposed algorithm. The results of the proposed hybrid tracker are demonstrated as follows.

A rectangular lobby where three elevators are located separately on both sides and a person who walks zigzag are adopted as our experimental scenario, depicted in Figure 10. The circle is the initial position of the mobile platform. The square

grids are the laser scanning points, and the dotted line represents the person's path starting at (a) and going to (l). In the experiment, the robot keeps moving forward and approaches the moving person. The positions (a)-(l) on the white dotted line corresponded to the images (a) to (l) of Figure 11.

The white rectangles in Figures 11(a), (d), (f), (h), (i) and (j) are the results of laser-based Kalman filter tracking. When the walking person stands very close to the wall, the measurement error of LRF makes it difficult to distinguish this person from the wall with a pure laser-based algorithm. To solve this issue, a vision-based mean-shift tracker is integrated to form a hybrid tracker. The black rectangles in Figures 11(b) and (e) are the initial regions applied to the mean-shift tracker when the confidence value of laser-based tracking is below a certain threshold. The gray rectangles in Figures 11(b), (c), (e), (g) and (k) are the consecutive results of the mean-shift tracking. Figure 11(h) shows the hand-shaking process from mean-shift-tracking (gray rectangle) to laser-based (white rectangle) Kalman filter tracking. The mean-shift tracker matches its hypothesis to the nearest hypothesis maintained by the laser-based Kalman filter. Afterwards, the Kalman filter keeps tracking the walking person without the use of the mean-shift tracker until the walking person is too close to the wall again.

To compare the performance of the proposed method, we tested the same video sequence by a pure-vision mean-shift-based tracker. Since an initial position and an image template should be provided to the mean-shift tracker, we contoured a rectangle manually on the frame when the moving person appeared, as shown in Figure 12(a). Figures 12(b) to (f) show some sample images in a short period when the person walks from position (a) to position (d) in Figure 10. Obviously, the tracker starts getting lost at frame #37 and diverges from the object to the background. The mean-shift tracker maintains an image template and updates it continuously. However, when the camera moves, noise caused by an unstable background misleads the mean toward wrong directions. As a result, the tracker loses the object. Another strategy of mean-shift is to use the same image template, the initial one, but this can not cope with appearance variations of the object under tracking. The mean-shift tracker only works well when the camera is fixed or when the camera is moving in a short period.

4. CONCLUSION

In this study, we propose a framework for people-tracking on moving platforms using a hybrid tracker. The hybrid tracker brings image information to solve the conventional laser-based object tracking problem. If the laser scanning data are reliable, the moving objects are tracked by Kalman filters. However, when the moving object can not be distinguished from the static background in the laser domain, a mean-shift algorithm comes into play to track the object in the image domain. As a result, different trackers are automatically applied in their most suitable circumstances to provide robust tracking results on moving platforms.

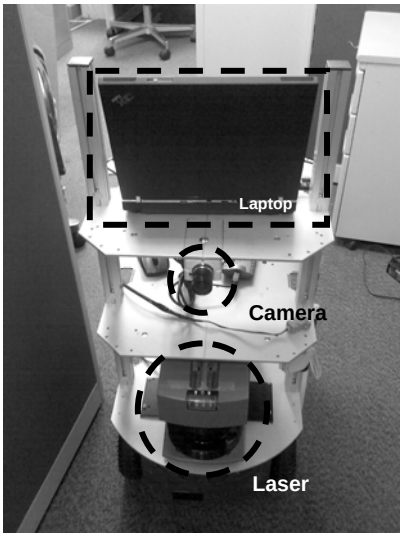


Figure 9. The testing platform with a CCD camera and a SICK LRF. A laptop connects all the devices to control the mobile platform and perform the proposed algorithm.

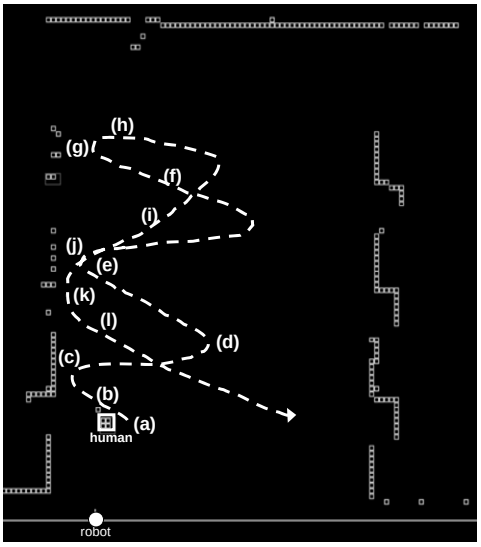


Figure 10. An illustration of the experimental scenario. A person walks zigzag and the moving platform keeps moving forward to approach the moving person. The white dotted line represents the human path starting from (a) to (l) and the corresponding images are shown from (a) to (l) of Figure 11.

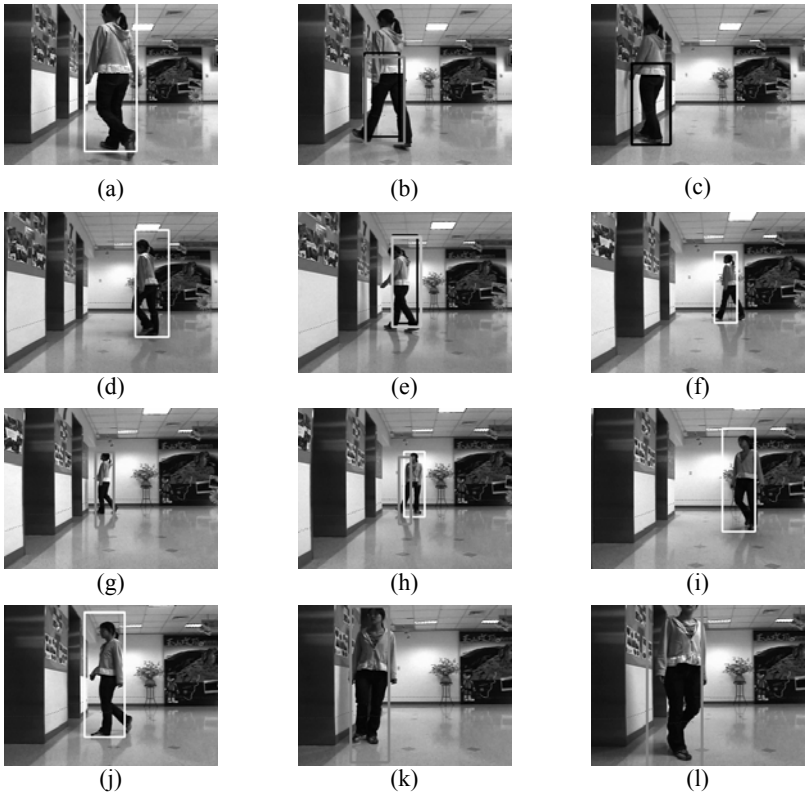


Figure 11. The tracking results of the proposed method shown in the image domain. The figures from (a) to (l) correspond to the related positions from (a) to (l) of the walking person in Figure 10. The white rectangles are the results of laser-based Kalman filter tracking. The black rectangles are the initial regions applied to the mean-shift tracker. The gray rectangles are the consecutive results of mean-shift tracking.

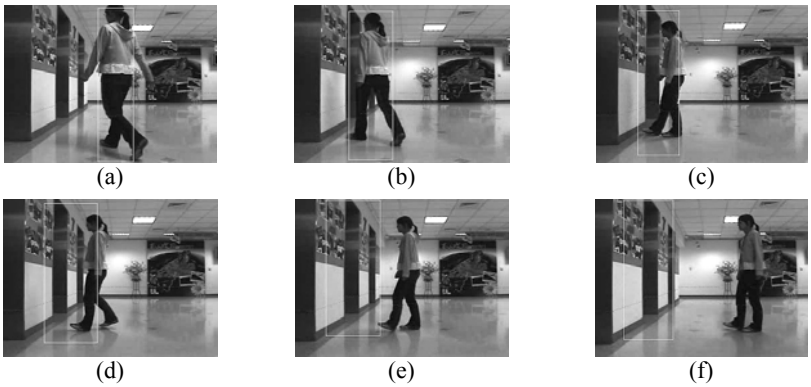


Figure 12. The mean-shift-based tracking results shown in image domain. (a) Frame #1. (b) Frame #11. (c) Frame #37. (d) Frame #46. (e) Frame #58. (f) Frame #77.

ACKNOWLEDGEMENT

This work is a partial result of Project 7353C41300 conducted by the Industrial Technology Research Institute under the sponsorship of the Ministry of Economic Affairs, R.O.C., Taiwan.

REFERENCES

Almeida, J., Almeida, A., & Araújo, R. (2005). Tracking Multiple Moving Objects for Mobile Robotics Navigation. *Proceedings of the IEEE Conference on Emerging Technologies and Factory Automation (ETFA)*, 203-210.

Besl, P. J., & McKay, N. D. (1992). A Method for Registration of 3-D Shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *14*, 239-256.

Comaniciu, D., Ramesh V., & Meer, P. (2000). Real-Time Tracking of Non-Rigid Objects using Mean Shift. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, *2*, 142-149.

Comaniciu, D., Ramesh V., & Meer, P. (2003). Kernel-Based Object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *25*(5), 564-577.

Fod, A., Howard , A., & Matarić, M. J. (2002). A Laser-based People Tracker. *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, *3*, 11-15.

Gong, Y. (2005). Integrated Object Detection and Tracking by Multiple Hypothesis Analysis. *NEC Journal of Advanced Technology*, *2*(1), 13-18.

Hartley, R., & Zisserman, A. (2003). Multiple View Geometry in Computer Vision. (2nd ed.). New York, USA: Cambridge University Press.

Irani, M., Rousso, B., & Peleg, S. (1997). Recovery of Ego-Motion Using Region Alignment. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, *19*(3), 268–272.

Kluge, B., Koehler, C., & Prassler, E. (2001). Fast and Robust Tracking of Multiple Moving Objects with a Laser Range Finder. *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, *2*, 1683-1688.

Lee, J. H., Tsubouchi, T., Yamamoto, K., & Egawa, S. (2006). People Tracking Using a Robot in Motion with Laser Range Finder. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2936-2942.

Lindström, M., & Eklundh, J.-O. (2001). Detecting and Tracking Moving Objects from a Mobile Platform Using a Laser Range Scanner. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, *3*, 1364-1369.

Lu, F., & Mulios, E. (1997). Robot Pose Estimation in Unknown Environments by Matching 2D Range Scans. *Intelligent and Robotic Systems*, *18*, 249-275.

MacLachlan, R. A., & Mertz, C. (2006). Tracking of Moving Objects from a Moving Vehicle Using a Scanning Laser Rangefinder. *Proceedings of the IEEE Intelligent Transportation Systems Conference*, 301-306.

Montemerlo, M., Thun, S., & Whittaker, W. (2002). Conditional Particle Filters for Simultaneous Mobile Robot Localization and People-Tracking. *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, *1*, 695-701.

Moravec, H., & Elfes, A. (1985). High-resolution maps from wide-angle sonar. *Proceedings of IEEE International Conference on Robotics and Automation*, 116-121.

Schulz, D., Burgard, W., Fox, D., & Cremers, A. B. (2003). People Tracking with a Mobile Robot Using Sample-Based Joint Probabilistic Data Association Filters. *International Journal of Robotics Research*, *22*(2), 99-116.

Talukder, A., Goldberg, S., Matthies, L. & Ansar, A. (2003). Real-time detection of moving objects in a dynamic scene from moving robotic vehicles. *Proceedings of the IEEE Conference on Intelligent Robots and Systems*, *2*, 1308-1313.

Talukder, A., & Matthies, L. (2004). Real-time detection of moving objects from moving vehicles using dense stereo and optical flow. *Proceedings of the IEEE Conference on Intelligent Robots and Systems (IROS)*, *4*, 3718-3725.

Wang, C.-C., Thorpe, C., & Thrun, S. (2003). Online Simultaneous Localization and Mapping with Detection and Tracking of Moving Objects: Theory and Results from a Ground Vehicle in Crowded Urban Areas. *Proceedings of the IEEE International Conference on Robotics and Automation*, *1*, 842-849.

Zhu, Q. (1991). Hidden Markov Model for Dynamic Obstacle Avoidance of Mobile Robot Navigation. *IEEE Transactions on Robotics and Automation*, *7*(3), 390-397.

Ying-Ying Chao received her Master's degree in Computer Science from National Chiao Tung University in 2006. She is an associate researcher in Information & Communications Research Laboratories at the Industrial Technology Research Institute. Her research is related to computer vision and image processing.

Chung-Hsien Huang received his B. E. E, M. S. and Ph. D. degrees in **Electrical** Engineering in 1998, 2000 and 2007 respectively from Chang Gung University, Taiwan. He is now a researcher working with the Information & Communications Research Laboratories at the Industrial Technology Research Institute. His research interests include medical image processing, computer vision and pattern recognition.

Ming-Yu Shih received his Master's and Ph. D. degrees in Computer Science from the National Central University in 1997 and 2005, respectively. Currently he is a researcher in the Information & Communications Research Laboratories at the Industrial Technology Research Institute. His research interests include image processing, computer vision and pattern recognition.

Bwo-Chau Fu received his Master's degree in Computer Science at the National Chiao Tung University, Taiwan. After graduation, he worked as a researcher at the Information & Communications Research Laboratories in the Industrial Technology Research Institute from 2003 to 2007. His research interests include computer vision and pattern recognition. He is now an MBA candidate at the Johns Graduate School of Management in Cornell University.