Weakly Interacting Object Tracking in Indoor Environments

Kao-Wei Wan[†] and Chieh-Chih Wang[†‡]
[†]Department of Computer Science and Information Engineering
[‡]Graduate Institute of Networking and Multimedia
National Taiwan University, Taipei, Taiwan
Email: leo@robotics.csie.ntu.edu.tw, bobwang@ntu.edu.tw

Tu T. Ton
National Services, Transport Planning
Parsons Brinckerhoff Australia Pty Limited
Sydney, NSW, Australia
Email: tton@pb.com.au

Abstract— **Interactions between targets have been exploited to solve the occlusion problem in multitarget tracking but not to provide higher level scene understanding. As indoor environments are relatively unconstrained than urban areas, interactions in indoor environments are weaker and have more variants. Weak interactions make scene interaction modeling and neighboring object interaction modeling challenging. In this paper, a place-driven scene interaction model is proposed to represent long-term interactions in indoor environments. To deal with complicated short-term interactions, the neighboring object interaction model consists of three short-term interaction models, following, approaching and avoidance. The moving model, the stationary process model and these two interaction models are integrated to accomplish weakly interacting object tracking. In addition, higher level scene understanding such as unusual activity recognition and important place identification is accomplished straightforwardly. The experimental results using data from a laser scanner demonstrate the feasibility and robustness of the proposed approaches.**

I. Introduction

Multiple moving object tracking or multitarget tracking is a key prerequisite for automating many useful robotics applications. The classical approaches such as the multiple hypothesis tracking (MHT) algorithm [1] and the joint probabilistic data association (JPDA) approach [2] have been extensively applied in many applications. However, only a few works addressed the observation and motion modeling issues of interactions among the tracked objects and the scene. Khan *et al.* [3] proposed a Markov chain Monte Carlo (MCMC)-based particle filter to track interacting ants in which interactions are modeled through a Markov random field motion prior. Their interaction potential is only based on static poses which cannot provide higher level scene understanding. Smith *et al.* [4] adopt a simple interaction model to penalize object overlapping. Sullivan and Carlsson [5] proposed to construct an interaction graph and then apply a two-stage clustering scheme to label the identity of the target. Instead of modeling or understanding interactions explicitly, these studies use the term, interaction, to describe the situations that the target and adjacent objects share the common measurements and cannot be correctly labeled. In these existing approaches, interactions represent negative information.

Wang *et al.* [6] proposed a variable structure multiple model (VSMM) estimation framework[7] with a scene interaction model and a neighboring object interaction model to perform multiple interacting object tracking in urban areas using a laser scanner. In this framework, interactions gain positive information. The scene interaction model and the neighboring object interaction model respectively take the long-term and short-term interactions between the tracked object and its surroundings into account. This approach not only solves the data association problem but also provides higher level scene understanding.

As moving objects in urban areas always obey the strict traffic rules, the interactions in these urban areas are stronger than in indoor environments. *Weaker* interactions make scene interaction and neighboring object interaction modeling more challenging as objects have more freedom to move and the interactions could have more variants. In this paper, we propose to accomplish weakly interacting object tracking by exploiting a place-driven scene interaction model and a neighboring object interaction model consisting of three short interaction models. The basic maneuver model, the stationary process model and these two interaction models are seamlessly fused via a digraph switching algorithm in the VSMM estimation framework. In addition, higher level scene understanding such as unusual activity recognition and important place identification is accomplished straightforwardly through the proposed interacting object tracking framework. The performance of the proposed approaches is evaluated with manually labeled ground truth data.

The remainder of the paper is organized as follows. Section II reviews the VSMM estimation framework, and describes our approaches to integrate the basic maneuver and interaction models. The scene interaction model and the neighboring object model are described in Sections III and IV, respectively. In Section V, we demonstrate that the proposed approaches are able to solve the difficult occlusion problem and to provide higher level scene understanding. The experimental results and performance evaluation are in Section VI. Finally, conclusion and future work are in Section VII.

II. Variable-Structure Multiple Model Estimation

In this section, we review the theoretical foundations of the variable-structure multiple model (VSMM) estimation framework briefly, and describe our approaches to integrate the moving models, the stationary model, the scene interaction model and the neighboring object interaction model in detail.

A. Theory

The tracking problem can be solved with the mechanism of Bayesian approaches such as the Kalman filter and the particle filter. As the true motion mode is often unavailable in many applications, online motion modeling is needed. Moving object tracking can be formalized in the probabilistic form as:

$$p(x_k, s_k \mid Z_k) \propto p(z_k \mid x_k, s_k) \sum_{s_{k-1}} \int p(x_k, s_k \mid x_{k-1}, s_{k-1}) p(x_{k-1}, s_{k-1} \mid Z_{k-1}) dx_{k-1} \tag{1}$$

where x_k is the true state of the moving object at time k, s_k is the *true motion mode* of the moving object at time k, $Z_k = \{z_1, z_2, \cdots, z_k\}$ is the perception measurement set leading up to time k, $p(x_{k-1}, s_{k-1} \mid Z_{k-1})$ is the posterior probability at time $k-1$, $p(x_k, s_k \mid Z_k)$ is the posterior probability at time k, $p(x_k, s_k \mid x_{k-1}, s_{k-1})$ is the motion model and $p(z_k \mid x_k, s_k)$ is the measurement or observation model.

For online motion modeling, using more models is not necessarily the optimal solution. Additionally, it increases computational complexity considerably. Use of a fixed set of models is not the only option for multiple model based tracking approaches. The details of the VSMM estimation and the related algorithms are available in [8].

In the VSMM estimation framework, it is assumed that the true motion mode s_k is the set of all model states. Thus the equation can be further expanded as:

$$\begin{aligned} & P(x_k, s_k|Z_k) \\ \propto & \sum_{m_k \in s_k} P(z_k|x_k, m_k) \sum_{s_{k-1}} \sum_{m_{k-1} \in s_{k-1}} \\ & \int_{x_{k-1}} P(x_k, m_k|x_{k-1}, m_{k-1}) \cdot P(x_{k-1}, m_{k-1}|Z_{k-1}) \end{aligned} \tag{2}$$

where $P(x_{k-1}, m_{k-1}|Z_{k-1})$ is the posterior probability of tracked object pose and motion models at time $k-1$. $P(x_k, m_k|x_{k-1}, m_{k-1})$ is the motion model including model transitions.

B. Weakly Interacting Object Tracking Framework

In our weakly interacting object tracking framework, s_k consists of the following motion models.

- The moving model (m^{mv}): the moving model consists of the constant velocity (CV) model and the constant acceleration (CA) model. These two models are fused using the interacting multiple model (IMM) approach [9].
- The stationary process model (m^{sp}): the stationary process model is assumed to be properly described by a second order stationary series. Because of limited data and time in practice, the mean and the covariance of the series are used to decide if the series is a stationary process.
- The scene interaction model (m^{si}): this model is designed to represent the long term interactions between the target and the static scene. The details of implementing m^{si} will be described in Section III.
- The neighboring object interaction model (m^{ni}): this model is designed to represent the short-term or immediate interactions between the target and its neighboring and moving objects. The details of implementing m^{ni} will be addressed in Section IV.

All these models are seamlessly intergraded through the VSMM framework.

We further predetermine three model sets as:

$$\begin{aligned} D^{[1]} &= \{m^{mv}, m^{si}\} \\ D^{[2]} &= \{m^{mv}, m^{si}, m^{ni}\} \\ D^{[3]} &= \{m^{sp}, m^{si}\} \end{aligned} \tag{3}$$

$D^{[1]}$ is designed for the situations where the speed of the target is higher than a minimum detection velocity and no moving object is nearby. $D^{[2]}$ is designed for the situation where the target is moving and a moving object is nearby. $D^{[3]}$ is designed for the situations where the tracked object is stationary.

To deal with move-stop-move maneuvers, the moving model and the stationary process model should not be mixed [6]. Therefore, a digraph switching algorithm [7] is applied to select one or two model sets for state estimation. Equation 4 show the rules to switch the model sets. Let υ be the minimum detection velocity and ρ be the distance between the target and the neighboring and moving object. t_1 and t_2 are thresholds.

$$s_k = \begin{cases} D^{[1]} & \upsilon > t_1,\ \rho > t_2 \\ D^{[2]} & \upsilon > t_1,\ \rho \leqq t_2 \\ D^{[1]} \& D^{[3]} & \upsilon \leqq t_1,\ \rho > t_2 \\ D^{[2]} \& D^{[3]} & \upsilon \leqq t_1,\ \rho \leqq t_2 \end{cases} \tag{4}$$

In the situations of $\upsilon \leqq t_1$, the model sets associated with the moving model and the stationary process model are evaluated separately without making any hard decision. While switching between the predetermined model sets, some models are added or removed and the probabilities of the model sets and the motion models are normalized or initialized.

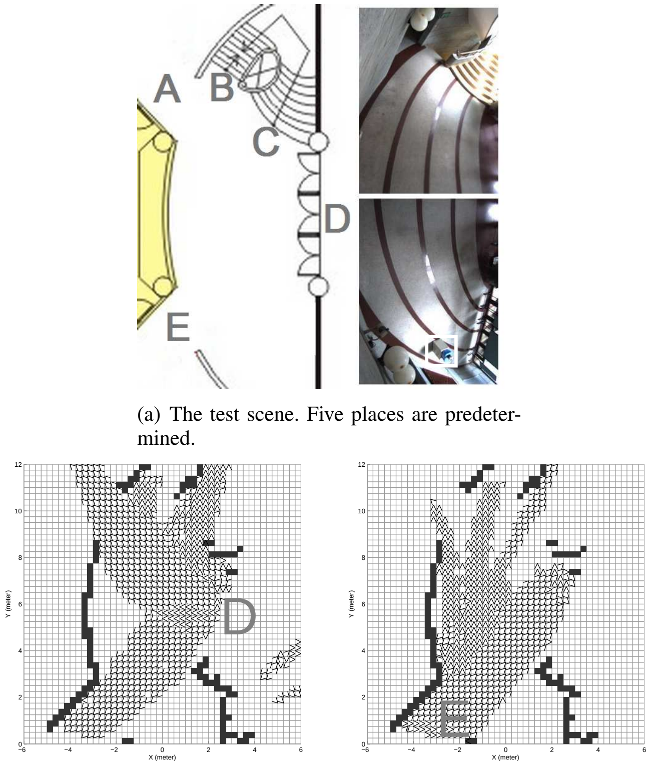
III. Scene Interaction Model

In this section, we briefly review the scene interaction model for tracking in urban areas [6] and describe our approaches to modify the scene interaction model for indoor environments.

A. Modeling

In urban areas, the simultaneous localization, mapping and moving object tracking (SLAMMOT) [10] maps are automatically generated and maintained according to different behavior patterns. The long-term interactions with surrounding environments in urban areas are strong as most moving entities obey the traffic laws. The behavior patterns of urban scenes could be easily classified according to moving directions of all moving objects in the scene[6]. Unfortunately, this approach does not work in indoor scenes because of no traffic control.

To deal with this issue, we follow an observation that the weak and long term interactions with dynamic environments in indoor scenes could be governed by *places*, and



(a) The test scene. Five places are predetermined.

(b) Place D-driven Scene Interaction Pattern.

(c) Place E-driven Scene Interaction Pattern.

Fig. 1. The place-driven scene interaction model. In the visual images of (a), a SICK LMS 291 laser scanner is located at the place indicated by a white rectangle. (b) and (c) show the place-driven patterns of place D and E respectively. Black solid grids are belonging to stationary objects. White grids are unobserved or unoccupied areas. Regarding the remainder grids, only the most observed motion direction of a grid is shown by an arrow inside the grid. The laser scanner is located at the origin $(0, 0)$ of the SLAMMOT maps.

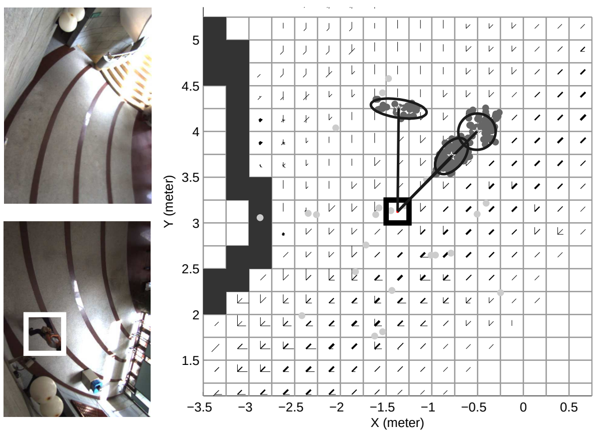


Fig. 2. The prediction stage of the scene interaction model using the sampling and k-means approaches. Solid gray circles are predicted samples and three 2σ ellipses show the state predictions using the k-means algorithm. Empty grids are belonging to unoccupied or unobserved space. Black solid girds are belonging to stationary objects. Inside the grids belonging to the moving object map, the distributions of eight canonical moving directions are represented by lines. The length of the line indicates the speed distribution. The width of the line shows the distribution of the occurrence times. It is more likely for an object to follow a direction if the width of the line in the direction is bigger. See [6] for more details.

propose a *place-driven scene interaction model* in which the SLAMMOT maps are generated and maintained according to predetermined or online recognized places such as entrances and exits. To illustrate the fundamental principle of the place-driven scene interaction model, Figure 1 shows an example in which five important places are predetermined in the lobby of the a building. In this paper, visual images from the cameras mounted on the second floor are only for visualization. Figures 1(b)∼(c) depict different long term interaction patterns between people and the indoor environment and show that the place-driven scene interaction model well represents the long-term interactions.

B. Prediction and Update

For the place-driven scene interaction model, the place that the target came from is used to select the proper SLAMMOT map, and the sampling technique is used to predict possible future motions. In addition, the k-means clustering algorithm [11] is applied to find a couple possible predictions. An iterative process is applied to determine the appropriate number of clusters k. Figure 2 shows an example of motion predication of the scene interaction model using the proposed approaches in which three state predictions are generated. All the estimates are weighted with probabilities which are proportional to cluster sizes. Note that the prediction from the scene interaction model only depends on the location of the tracked object in the SLAMMOT map which contains the statistical information but not on the previous estimates from the scene interaction model. The update of the scene interaction model is straightforward via the VSMM state estimation framework.

IV. Neighboring Object Interaction Model

The neighboring object interaction model is designed to represent the short-term or immediate interactions between the target and its neighboring and moving objects. In this paper, three short-term interactions, following, approaching and avoidance, are modeled for indoor environments. The following interaction represents the situations that the target changes its moving direction to follow its neighboring object. The approaching interaction represents that the object moves toward to its neighboring object. The avoidance interaction represents that the target performs avoidance maneuver to avoid collision with its neighboring object.

It is assumed that interactions change the moving direction of the tracked object. The prediction steps of these models first determine the moving direction of the tracked object and then estimate the next pose of the tracked object with this new direction. The spatial relationships between the target and its neighboring and moving object are computed to determine the moving directions of the three interactions as shown in Figure 3. The direction of the shortest distance between the target and its neighboring object is the direction of the approaching interaction model. The direction of the following interaction

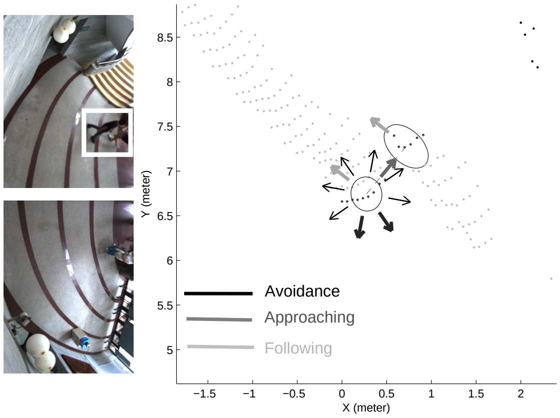


Fig. 3. Moving direction determination of the approaching, following and avoidance interaction models. See the text for more details.

is the same as the moving direction of the target's neighboring object. Eight canonical directions are considered as the moving direction candidates of the avoidance model. The directions of the avoidance interaction are chosen by selecting two canonical directions which are the most closest to the moving direction of the target, but not close to the moving direction of its neighboring object. The speed estimates of these models are set to equal to the speed estimate of the moving model. The covariance of these short-term interaction estimates is transformed accordingly.

Note that these three interaction models are not mixed during tracking. Figure 4 shows the tracking results using the proposed neighboring object interaction model in which the probability history of these short-term interaction models successfully represent different short-term interaction patterns in indoor environments. The probability can be computed by the following equation:

$$P(m) \propto exp[-d_m^2/2]/((2\pi)^M|S_m|)^{0.5}$$
$$where\ d_m^2 = y'S_m^{-1}y \qquad (5)$$

where $p(m)$ is the probability of the model m, S_m is the predicted observation covariance and y_m is the difference between the actual and expected observation with model m.

V. Occlusion and Higher Level Scene Understanding

In this section, we will demonstrate that the proposed framework is capable of solving the challenging occlusion problem and is able to provide higher level scene understanding such as unusual activity recognition and important place identification.

A. Occlusion

The classical approaches can not deal with the situations where the target is occluded and abruptly changes its motion due to short-term interactions. Figure 5 shows an example of this challenging occlusion problem. The IMM approach fails in this case. The proposed approaches solve this challenging problem successfully.

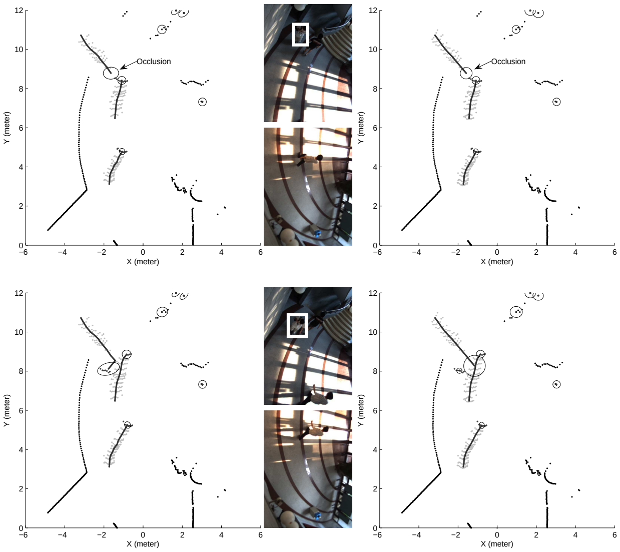


Fig. 5. Dealing with occlusion. Left columns show the tracking results of the proposed algorithm. Right columns show the tracking results of the IMM approach. Middle column are visual images. See the text for more details.

B. Unusual Activity Recognition

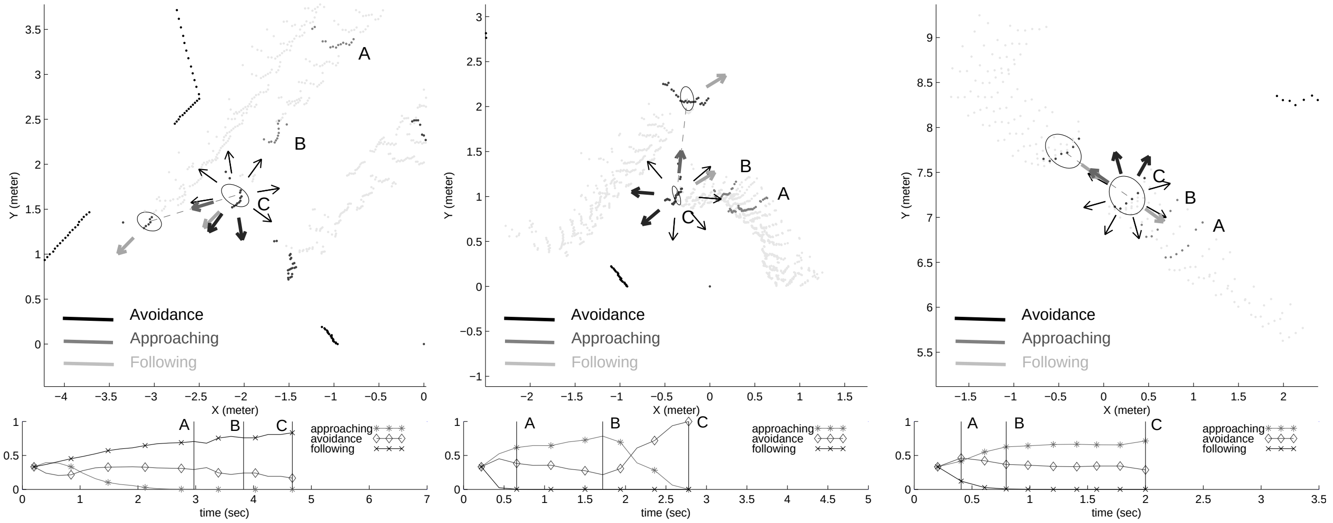
Usual or normal activities are embedded in the scene interaction model and the neighboring object interaction model. Low probabilities of these interaction models could indicate unusual or abnormal activities. Figure 6 shows an example in which a person was wandering in the lobby. Our approach quickly shows that the probabilities of the place-driven scene interaction model is very low. This event is highly likely to be unusual or abnormal.

Figure 7 shows the unusual activity recognition results under different patterns. Note that it is also possible to recognize unusual or abnormal activities using the neighboring object interaction model.

C. Important Place Identification

As the scene interaction model is place-driven, important place determination is critical. Although we demonstrated the feasibility of the scene interaction model with predetermined places, it is feasible to online recognize new important places and build the SLAMMOT map accordingly by accumulating and analyzing the results of unusual activity recognition. In Figure 7(a)∼(c), the locations at which the targets performed unusual activities and their speeds were less than the minimum detection velocity are indicated. These places could be important. Figure 7(d) shows the sum of all detected unusual activities of the five place-driven interaction patterns. Three new important places are identified. These new identified places are consistent with the real world setting.

There is a bulletin board at Location (-3,2). People stopped at location (-1,5) to watch a flat screen TV showing information at the location (-3, 5). Interestingly, Location (0,0) is identified as important simply because our experiment equipments were located there and people stopped by to figure



(a) The following interaction.

(b) The avoidance interaction.

(c) The approaching interaction.

Fig. 4. Interacting object tracking using the neighboring object interaction model. Two ellipses indicate the target and its neighboring object. The laser scan points collected at different times are shown and data collected at time A, B, C are highlighted. The predicted directions of the three interaction models are shown. The probabilities successfully indicate the real situations.

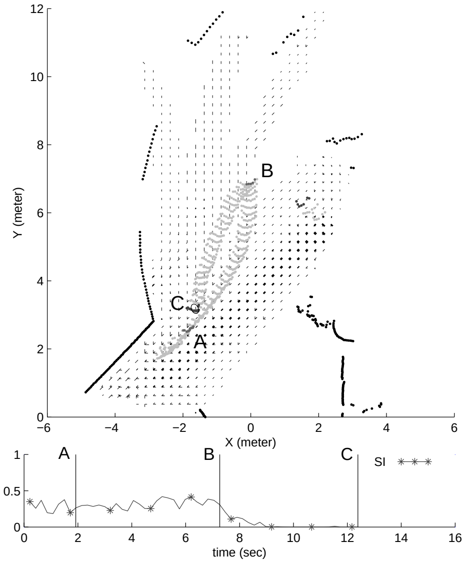
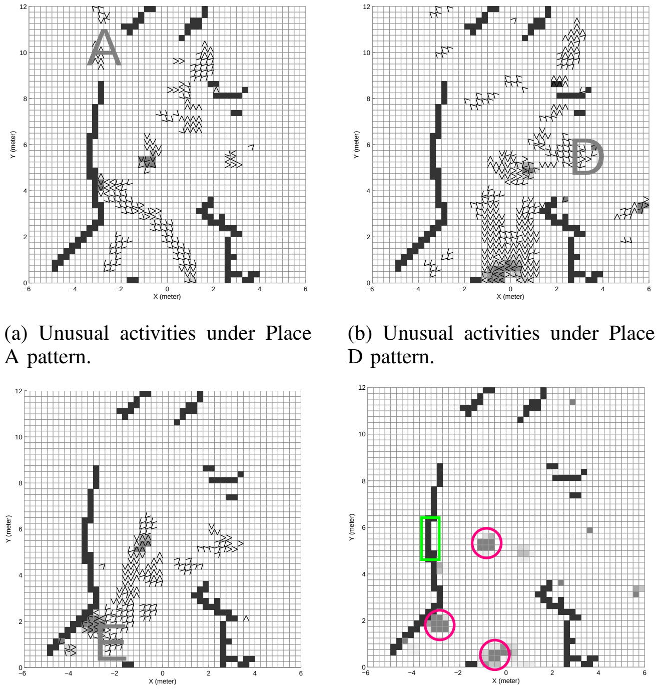


Fig. 6. Unusual or abnormal activity recognition using the scene interaction model. The probabilities of the scene interaction model indicate unusual events.



(a) Unusual activities under Place A pattern.

(b) Unusual activities under Place D pattern.

(c) Unusual activities under Place E pattern.

(d) Important place identification.

Fig. 7. Unusual activity recognition and important place identification. In (a)∼(c), arrows inside the grids show the moving directions of the detected unusual activities. Gray grids with arrows inside indicate the locations where the speeds of the targets are less than the minimum detection velocity. The darker the grid, the higher the occurrence times of these events. (d) shows important place identification using the sum of all unusual activities under five place-driven interaction patterns. Circles are the identified important places. A rectangle indicates the location of a flat screen TV.

out what these devices are. The identified places using the proposed approaches consist with the real world setting.

VI. Experiments and Results

The proposed algorithm is evaluated using data from a SICK LMS 291 laser scanner in the lobby of our department. The ground truth of the testing data were labeled manually. The performance of our approaches is shown in terms of tracking accuracy, success rate and interaction classification.

TABLE I
TRACKING PERFORMANCE EVALUATION

	No Moving Object Nearby		Other Moving Object Nearby	
	Error (μ,σ)	Success rate	Error (μ,σ)	Success rate
IMM	(2.37m,1.14m)	92%	(2.89m,1.31m)	81%
WI	(2.43m,1.50m)	95%	(2.62m,1.54m)	93%

TABLE II
DETECTION OF INTERACTIONS

	GroundTruth			
	approaching	avoidance	following	no interaction
approaching	90%	4%	7%	6%
avoidance	1%	87%	27%	33%
following	5%	6%	51%	4%
no interaction	4%	3%	15%	57%
total	100%	100%	100%	100%

A. Tracking Performance

The tracking performance of the proposed approach is compared with the IMM tracker. The results and comparison are shown in Table I. In the situations that no other moving object is nearby, the performance of the proposed approach is similar the IMM approach. The scene interaction is suitable for detecting abnormal activities but not for improving tracking accuracy. In the situations that a moving objet is near the tracked object, our approach outperforms the IMM tracker in terms of tracking accuracy and occlusion handling. As the scene interaction model and the neighboring object interaction model provide a good prediction while the IMM tracker only predicts using the current motion information, the success rate of tracking using our approaches is better.

B. Interaction Classification

Here we show the correctness of short-term interaction classification. As the neighboring object interaction models are not fused during tracking, the probabilities of these models are maintained and used to classify the short-term interaction between the tracked object and its neighboring and moving object. If the probability of some short-term interaction model is greater than other models, we regard this model as the interaction mode of the tracked object with its neighboring and moving object. Table II shows the classification results. The approaching and avoidance interactions are classified accurately. However, the following interaction and the situation of no interaction are easily confused. We would argue that the results are reasonable as it may not be easy for human beings to differentiate these two situation in indoor environments.

VII. CONCLUSION AND FUTURE WORK

Our interacting object tracking framework not only deals with the challenging data association problem in multitarget tracking but also provides a means to understand higher level interactions and activities. Based on [6], the main contributions of this work are to propose the place driven scene interaction model and to apply three key short term interaction models to accomplish weakly interacting object tracking in indoor environments. We also contribute a simple yet effective approach to accomplish unusual activity recognition and important place identification via the interacting object tracking framework.

Future work will apply the proposed algorithms in both indoor and outdoor environments to test their feasibility. The computational complexity of the proposed algorithms will be analyzed. It would be of interest to explore the issues of simultaneous scene interaction modeling, unusual activity recognition and important place identification.

ACKNOWLEDGMENTS

This work was partially supported by grants from Taiwan NSC (#96-2628-E-002-251-MY3, #96-2218-E-002-035, #96-2218-E-002-008); Excellent Research Projects of National Taiwan University (#95R0062-AE00-05); Taiwan DOIT TDPA Program (#95-EC-17-A-04-S1-054); Taiwan ITRI, III; and Intel.

REFERENCES

[1] I. J. Cox and S. L. Hingorani, “An efficient implemenation of Reid’s multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking,” *IEEE Trans. on Pattern Analysis and Machine intelligence*, vol. 18, no. 2, pp. 138–150, Feb. 1996.

[2] T. Fortmann, Y. Bar-Shalom, and M. Scheffe, “Sonar tracking of multiple targets using joint probabilistic data association,” *IEEE Journal of Oceanic Engineering*, vol. OE-8, pp. 173–184, July 1983.

[3] Z. Khan, T. Balch, and F. Dellaert, “Mcmc-based particle filtering for tracking a variable number of interacting targets,” *IEEE Transactions on Pattern Analysis and Machine Intelligence,*, vol. 27, no. 11, pp. 1805–1918, November 2005.

[4] K. Smith, D. Gatica-Perez, and J.-M. Odobez, “Using particles to track varying numbers of interacting people,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 2005, pp. 962–969.

[5] J. Sullivan and S. Carlsson, “Tracking and labelling of interacting multiple targets,” in *9th European Conference on Computer Vision (ECCV)*, 2006.

[6] C.-C. Wang, T.-C. Lo, and S.-W. Yang, “Interacting object tracking in crowded urban areas,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Roma, Italy, April 2007.

[7] X.-R. Li and Y. Bar-Shalom, “Multiple-model estimation with variable structure,” *IEEE Transactions on Automatic Control*, vol. 41, no. 4, pp. 478–493, April 1996.

[8] X. R. Li and V. P. Jilkov, “Survey of maneuvering target tracking. part v. multiple-model methods,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 41, no. 4, pp. 1255– 1321, October 2005.

[9] H. A. P. Blom and Y. Bar-Shalom, “The interacting multiple model algorithm for systems with markovian switching coefficients,” *IEEE Trans. On Automatic Control*, vol. 33, no. 8, pp. 780–783, Aug. 1988.

[10] C.-C. Wang, C. Thorpe, S. Thrun, M. Hebert, and H. Durrant-Whyte, “Simultaneous localization, mapping and moving object tracking,” *The International Journal of Robotics Research*, vol. 26, no. 9, pp. 889–916, September 2007.

[11] J. MacQueen, “Some methods for classification and analysis of multivariate observations,” in *5-th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1. University of California Press, 1967, pp. 281–297.