Weakly Interacting Object Tracking in Indoor Environments

Chieh-Chih Wang, Kao-Wei Wan and Tzu-Chien Lo

Abstract—Interactions between targets have been exploited to solve the occlusion problem in multitarget tracking but not to provide higher level scene understanding. In our previous work [1], a variable structure multiple model estimation framework with a scene interaction model and a neighboring object interaction model was proposed to accomplish these two tasks. The proposed approach was demonstrated in urban areas using a laser scanner. 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 is consisted 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 in the course of human everyday life such as navigation in crowded environments, surveillance, and human activity monitoring. A number of approaches have been proposed to tackle the data association problem of multitarget tracking in the robotics and computer vision literature. The classical approaches such as the multiple hypothesis tracking (MHT) algorithm [2] and the joint probabilistic data association (JPDA) approach [3] have been extensively applied in many applications. However, the observation and motion modeling issues of interactions among the tracked objects and the scene were not addressed.

Recently, a few papers addressed the related issues of interacting object tracking. To deal with the motion and observation modeling issues of interacting objects, Khan et al. [4] 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. Similar to [4], Smith et al. [5] adopt a simple interaction model to penalize object overlapping. Yu and Wu [6] presented a collaborative tracking approach in which the adjacent/interacting objects compete for the common observations. Qu et al. [7] proposed an interactive distributed multi-object tracking algorithm in which a magnetic repulsion model is used to model pairwise interactions. Sullivan and Carlsson [8] 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 our previous work [1], a variable structure multiple model estimation framework with a scene interaction model and a neighboring object interaction model was proposed to perform multiple interacting object tracking in urban areas using a laser scanner. 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. All these basic maneuver and interaction models are seamlessly intergraded through the variable-structure multiple-model estimation framework. Our approach not only solves the data association problem in multitarget tracking but also provides higher level scene understanding via interactions. In the existing approaches addressed above, interactions represent negative information. Conversely, interactions gain positive information in the interacting object tracking framework.

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 the scene interaction model and the neighboring object interaction model. As the scene interaction motion based on traffic behavior pattern recognition does not work in indoor environments, a place-driven scene interaction model is proposed to represent long-term interactions between the target and the static scene. As moving object perform more complicated short-term interactions in indoor than in urban areas, the neighboring object interaction model is consisted of three short interaction models, following, approaching and avoidance. The basic maneuver model, the stationary process

Chieh-Chih Wang is with the Department of Computer Science and Information Engineering and the Graduate Institute of Networking and Multimedia, National Taiwan University, Taipei, Taiwan. bobwang@ntu.edu.tw
Kao-Wei Wan is with the Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan. leo@robotics.csie.ntu.edu.tw
Tzu-Chien Lo was with the Department of Computer Science and Information Engineering, National Taiwan University. He is now serving in Taiwan’s military. bright@robotics.csie.ntu.edu.tw
model and these two interaction models are seamlessly fused via a digraph switching algorithm in the variable structure multiple model 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 remainder of the paper is organized as follows. Section II reviews the variable structure multiple model 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. The experimental results in Section V demonstrate that the proposed approaches are able to solve the difficult occlusion problem and to provide higher level scene understanding. Finally, conclusion and future work are in Section VI.

II. VARIABLE-STRUCTURE MULTIPLE MODEL ESTIMATION

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:

$$p(x_k,s_k | Z_k) \propto p(z_k | x_k,s_k)$$

$$\sum_{s_{k-1}} \int p(x_k,s_k | x_{k-1},s_{k-1}) p(x_{k-1},s_{k-1} | Z_{k-1}) dx_{k-1}$$

(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 $k$, $p(x_{k-1},s_{k-1} | Z_{k-1})$ is the posterior probability at time $k-1$, $p(x_k,s_k | Z_k)$ is the posterior probability at time $k$, $p(x_k,s_k | x_{k-1},s_{k-1})$ is the motion model and $p(z_k | x_k,s_k)$ 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. Li [9] provided a theoretical proof that even the optimal use of motion models does not guarantee better tracking performance. Use of a fixed set of models is not the only option for multiple model based tracking approaches. A variable structure can be used in multiple model approaches [10]. By selecting the most probable model subset, estimation performance can be improved. Not only motion but also other types of information or constraints can be selected and added to the model set. In [11], terrain conditions are used as constraint models and are added to the model set to improve performance of ground target tracking via a variable structure interacting multiple model (VS-IMM) algorithm. The details of the variable structure multiple model estimation and the related algorithms are available in [10].

In the variable structure multiple model estimation framework, it is assumed that the true motion mode $s_k$ is the set of all model states. In this paper, $s_k$ is consisted of the following motion models.

- The moving model ($m^{mv}$): the moving model is consisted of the constant velocity (CV) model and the constant acceleration (CA) model. These two models are fused using the interacting multiple model (IMM) approach [12].
- The stationary process model ($m^{st}$): 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 moving objects. The details of implementing $m^{ni}$ will be addressed in Section IV.

In our weakly interacting object tracking framework, four model sets are predetermined as:

$$D^{[1]} = \{m^{mv},m^{st}\}$$

$$D^{[2]} = \{m^{mv},m^{si},m^{st}\}$$

$$D^{[3]} = \{m^{sp},m^{si}\}$$

$$D^{[4]} = \{m^{sp},m^{si},m^{st}\}$$

(2)

$D^{[1]}$ is consisted of $m^{mv}$ and $m^{st}$. This model set 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 predetermined for the situation where the target is definitely moving and a moving object is nearby. $D^{[3]}$ is consisted of $m^{sp}$ and $m^{si}$. This model set is designed for the situations where the tracked object is stationary and no moving object is nearby. $D^{[4]}$ is designed for the situations that the target is stationary and a moving object is nearby.

To deal with move-stop-move maneuvers, the moving model and the stationary process model should not be mixed [1]. Therefore, a digraph switching algorithm [9] is applied to select one or two model sets for state estimation. Equation 3 show the predetermined rules to switch the model sets. Let $v$ be the minimum detection velocity and $\rho$ be the distance between the target and the neighboring and moving object. $t_1$ and $t_2$ are thresholds.

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

(3)

In the situations of $v \leq t_1$, 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. Algorithms 1 and 2 show two examples of the model set switching
progress.

\[
\text{if switch to } s_k = D^{[1]} \text{ then} \\
\quad \text{if } s_{k-1} = D^{[2]} \text{ then} \\
\qquad \text{Remove } m^{w}; \\
\qquad \text{Normalize } m^{w'} \text{ and } m^{s}; \\
\quad \text{end} \\
\text{if } s_{k-1} = \{D^{[1]} \& D^{[3]}\} \text{ then} \\
\quad \text{Remove } D^{[3]}; \\
\text{end} \\
\text{end} \\
\text{end}
\]

Algorithm 1: The digraph switching algorithm. The model
set is switched to \( s_k = D^{[1]} \).

\[
\text{if switch to } s_k = \{D^{[2]} \& D^{[4]}\} \text{ then} \\
\quad \text{if } s_{k-1} = D^{[1]} \text{ then} \\
\qquad D^{[2]} = D^{[1]} + m^{w}; \\
\qquad \text{Add } D^{[4]}; \\
\qquad p(D^{[2]}) = p(D^{[4]}) = 0.5; \\
\quad \text{end} \\
\quad \text{if } s_{k-1} = D^{[2]} \text{ then} \\
\qquad \text{Add } D^{[4]}; \\
\quad \quad p(D^{[2]}) = p(D^{[4]}) = 0.5; \\
\quad \text{end} \\
\quad \text{if } s_{k-1} = \{D^{[1]} \& D^{[3]}\} \text{ then} \\
\qquad D^{[2]} = D^{[1]} + m^{w}; \\
\qquad D^{[4]} = D^{[3]} + m^{w}; \\
\qquad p(D^{[2]}) = p(D^{[4]}) = 0.5; \\
\quad \text{end} \\
\text{end}
\]

Algorithm 2: The digraph switching algorithm. The model
set is switched to \( s_k = \{D^{[2]} \& D^{[4]}\} \).

Let \( \mu_k^{D[j]} \) and \( \Sigma_k^{D[j]} \) be the mean and covariance of the state estimate from a model set \( D^{[j]} \). \( \mu_k^{D[j]} \) and \( \Sigma_k^{D[j]} \) can be computed as:

\[
\mu_k^{D[j]} = \sum_{j} p(m^j) \mu_k^{m^j} \tag{4}
\]

\[
\Sigma_k^{D[j]} = \sum_{j} p(m^j) [\Sigma_k^{m^j} + (\mu_k^{m^j} - \mu_k^{D[j]})(\mu_k^{m^j} - \mu_k^{D[j]})^T] \tag{5}
\]

where \( N \) is the number of the models in the model set, \( p(m^j) \) is the probability of the model \( j \). \( \mu_k^{m^j} \) and \( \Sigma_k^{m^j} \) are the mean and covariance of the state estimate from the model \( j \).

III. SCENE INTERACTION MODEL

In this section, we briefly review the scene interaction
template for tracking in urban areas [1] and describe our
approaches to modify the scene interaction model to represent
long-term interactions between the target and its surrounding
and static objects in indoor environments.

A. Modeling

As temporal and spatial information of the dynamic scene
is embedded in the stationary and moving object maps built
by simultaneous localization, mapping and moving object
tracking (SLAMMOT) [13], the scene interaction model uses
the SLAMMOT maps to predict/constrain the possible future
motion and pose of the target. In addition to the occupancy
information, speeds and directions of moving objects are
embedded in the SLAMMOT maps for representing long-
term interactions.

In urban areas, the SLAMMOT maps are automatically
generated and maintained according to different behavior
patterns. The long-term interactions in urban areas are strong
as most moving entities obey the same traffic laws. The
behavior patterns of urban scenes could be easily classified
according to moving directions of all moving objects in the
scene. 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 in indoor environments
could be governed by places, and 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 NTU
CSIE building. In this paper, the visual images from the cam-
eras mounted on the second floor are only for visualization.
Place D is the main entrance of the building, Place C leads
to the second floor, Place B leads to the basement, and Places
A and E lead to the classrooms, restrooms and elevators.
According to these different places, the SLAMMOT maps
are generated and maintained separately using spatiotempo-
ral information of moving entities. Figure 1(b)–(f) depict
different long term interaction patterns between people and
the indoor environment and show that the place-driven scene
interaction model well represent the long-term interactions.

B. Prediction and Update

For interacting object tracking in urban areas, we predict
possible motions of a target using a sampling-based approach
with the use of the SLAMMOT maps and the scene be-
havior pattern recognition result. For accomplishing variable
structure multiple model estimation using the Kalman filter,
the mean and covariance of these weighted samples are
computed.

For interacting object tracking in indoor environments, the
place that the target came from is used to select the proper
SLAMMOT map, and the same sampling technique is used
to predict possible future motions. In addition, the k-means
clustering algorithm [14] is applied to find a couple possible
predictions. As the number of possible state predictions of
the scene interaction model is unknown, an iterative process
is applied to determine the number of clusters $k$. We select 5 as the initial value of $k$ and iteratively decreases $k$ value if the following criteria are met: (1) the number of samples belonging to a cluster is less than 10% of all samples; (2) the covariance’s determinant of a cluster is larger than a threshold, $0.4096 \, m^4$; (3) the distance of the means of two clusters is less than $0.25 \, m$. Figure 2 shows an example of motion predication of the scene interaction model using the proposed approaches in which three state predictions are generated. Only the best estimate is fused with other models. Note that the prediction from the scene interaction model only depends on the speed and location estimates of the tracked object but not on the previous estimates from the scene interaction model. The update of the scene interaction model is straightforward via the variable structure multiple model 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 our previous work [1], only the following interaction is modeled as it often occurs in urban areas. As neighboring object interactions in indoor environments could be more complicated and weaker than in urban areas, only dealing with the following interaction is insufficient.

In this paper, three short-term interactions, following, approaching and avoidance, are modeled. The following interaction is representing the situations that the target changes its moving direction to follow its neighboring object’s moving direction. The approaching interaction represents that two objects are moving toward to each other. The avoidance interaction is representing that the target performs avoidance maneuver to avoid collision with its neighboring object.

The spatial relationships between the target and its neighboring and moving object is computed to determine the directions of the three short-term 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 is the same as the moving direction of the target’s neighboring object. Eight canonical directions...
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 $2\sigma$ ellipses show the state predictions using the k-means algorithm. Empty grids are belonging to unoccupied or unobserved space. Black solid grids 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.

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 the approaching interaction. The speed estimates of these short-term interactions are set to equal to the speed estimate of the moving model. The covariance of these short-term interaction estimates is transformed accordingly.

In the situations that the distance between a target and its neighboring object is less than a threshold, these three short-term interaction models are initialized and used for tracking. 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 probabilities of these short-term interaction models successfully represent different short-term interaction patterns in indoor environments.

V. Occlusion and Higher Level Scene Understanding

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

A. Occlusion

As the classical multiple model estimation approaches only contain basic maneuver models, these 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 right columns of Figure 5 show the tracking results using the interacting multiple model (IMM) approaches. The IMM approaches fail in this case. The left columns show that the proposed approaches solve this challenging problem successfully.

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. In our previous work [1], a bicyclist disobeying the traffic law was successfully identified using the scene interaction model. Figure 6 shows an example that a person was wandering in the lobby and the probabilities of the place-driven scene interaction model quickly indicated this unusual or abnormal activity.
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, 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 8 shows the sum of all detected unusual activities of the five place-driven interaction patterns. Three new important places are identified. We verify these new identified places with the real world setting. There is a bulletin board at Location (-3,2). Location (-1,5) is close to the center of the lobby and there is a flat TV showing important 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 to figure out what these devices are. The identified places using the proposed approaches are consisted with the real world setting.

Fig. 8. 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 TV.
VI. 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 between the target and the dynamic scene. Based on our previous work on strongly interacting object tracking in urban areas, the main contribution of this work is to propose the place driven scene interaction model and 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. The experimental results using data from a laser scanner collected in the lobby of the NTU CSIE building demonstrate that the proposed approaches accomplish multitarget tracking and provide higher level scene understanding feasibly and effectively.

Future work will further analyze the computational complexity of the proposed algorithms. It would be of interest to explore the issues of simultaneous scene interaction modeling, unusual activity recognition and important place identification.

VII. ACKNOWLEDGMENTS

This work was partially supported by grants from Taiwan NSC (#95-2218-E-002-039, #96-2628-E-002-251-MY3, #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; and Intel.

REFERENCES


