Fluid-inspired field representation for risk assessment in road scenes

Xuanpeng Li  
School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China

Lifeng Zhu  
School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China

Qifan Xue  
School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China

Dong Wang  
School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China

Yongjie Jessica Zhang  
Department of Mechanical Engineering, CarnegieMellon University, Pittsburgh, PA 15213, USA

Follow this and additional works at: https://dc.tsinghuajournals.com/computational-visual-media

Recommended Citation  
Xuanpeng Li, Lifeng Zhu, Qifan Xue et al. Fluid-inspired field representation for risk assessment in road scenes. Computational Visual Media 2020, 6(4): 401-415.
Abstract  Prediction of the likely evolution of traffic scenes is a challenging task because of high uncertainties from sensing technology and the dynamic environment. It leads to failure of motion planning for intelligent agents like autonomous vehicles. In this paper, we propose a fluid-inspired model to estimate collision risk in road scenes. Multi-object states are detected and tracked, and then a stable fluid model is adopted to construct the risk field. Objects’ state spaces are used as the boundary conditions in the simulation of advection and diffusion processes. We have evaluated our approach on the public KITTI dataset; our model can provide predictions in the cases of misdetection and tracking error caused by occlusion. It proves a promising approach for collision risk assessment in road scenes.

Keywords  fluid-inspired risk field; multi-object tracking; road scenes

1 Introduction

Collision risk assessment is an essential task for vehicles driving on the road. It requires predicting the likely evolution of the current traffic situation, and assessing how dangerous the future situation might be. Risk can be intuitively understood as the likelihood and severity of the damage that a vehicle of interest may suffer in the future. Some quantitative risk indicators like time-to-collision, time-to-brake, and time-to-steer are widely used in advanced driver assistance systems like automatic emergency braking. But they are of limited use in complex traffic scenes: for example, they under-estimate collision risk at intersections with stopped cars, and over-estimate collision risk on a curved road [1].

Recently, risk assessment has mainly been based on prediction of future motion and estimation of collision occurrence from various sensors like cameras, LiDAR, and radar. The collision risk can be computed by integrating over all possible future trajectories and estimating collisions between each possible pair [2, 3]. Trajectory prediction plays an important role in risk assessment; approaches to modeling can be categorized into three types with an increasing degree of abstraction: physics-based, maneuver-based, and interaction-aware. Compared to the others, interaction-aware motion models can provide a reliable estimate of long-term motion and risk, but their computational requirements limit their applicability in real-time risk assessment.

Another way to present the surrounding environment is to use occupancy-based methods [4–6]. Spatial information concerning vehicles, cyclists, pedestrians, and other obstacles is quantified into a probabilistic distribution of occupancy of a grid. Dynamic grid occupancy is computed by recognizing surrounding objects using sensors mounted on the vehicle. However, interaction between multiple objects causes difficulties: occlusion can lead to failure of object detection, motion prediction, and occupancy grid generation. Risk assessment involving occluded objects is difficult to tackle because their state information cannot be acquired.

Humans use internal physics to predict how dynamical systems evolve [7]: capturing and utilizing the physics of a system is critical to improving the
performance of motion prediction and risk assessment. Accurate physical models are also essential for controlling intelligent vehicles which can interact with the traffic environment. For instance, field theory is widely adopted in modeling traffic behaviors. Yang et al. [8] employed an artificial potential field for modeling car following, where the host vehicle works as an individual charge in a potential field. Wang et al. [9] proposed to model driver–vehicle–road interactions in a unified way including a potential field, a kinetic field, and a behavior field. Their goal is to improve risk assessment under complex traffic environments. A recent line of work focuses on how to use physics to make inferences in a complex multi-agent environment. For example, robots can quickly learn manipulation skills when predicting the consequences of physical interactions [10]. Li et al. [11] proposed modeling interacting vehicle motions as an entire physical system instead of predicting a single motion.

However, for risk assessment for intelligent vehicles, practical applications of these models are limited due to the uncertainty of motion prediction and the high computational requirements. Hence, in this paper, we aim to integrate a fluid-inspired field model into approaches to collision risk assessment for real road scenes. The goal is to reveal evolving spatiotemporal information about object motions by considering multi-object interaction, and to improve the efficiency of path prediction and risk assessment, using a modest amount of computation.

Fluid dynamics is a common tool for simulating a variety of natural phenomena, e.g., water vapor forming into clouds and smoke curling from a glowing cigarette. We assume that fluid dynamics can be adapted to compute a risk map. If we imagine that each nearby vehicle is emitting virtual smoke which is animated by the motion of the vehicles, we presume that a vehicle of interest may sense the risk by observing the smoke. Using the above assumption, we design a method to construct a risk map inspired by fluid dynamics.

Classic fluid simulation involves an iterative process of advection and diffusion [12]. Advection is the process by which a fluid’s velocity transports itself and other quantities in the fluid, while diffusion is derived from the physics that viscous fluids have a certain resistance to flow. The influence of an object’s motion on the surrounding environment can be modeled inspired by this fluid model. If an object moves comparatively quickly, the velocity of advection is high, and its threat to other objects increases, and vice versa. By means of diffusion, the fluid model also helps to reveal hidden spatial information in the presence of missing data. In order to analyze interactive motion and estimate collision risk, we extract 3D object state information and modify the advection and diffusion processes based on the fusion of state space sequences. This new fluid-inspired approach to risk assessment provides an alternative risk representation for intelligent vehicles.

The remainder of this paper is organized as follows: in Section 2, related work on 3D object detection, object tracking, and risk assessment is presented. In Section 3, LiDAR-based 3D object detection is briefly introduced, and then a 3D Kalman filter is used to predict object motions and generate a continuous state space for multiple objects. The formulation of the fluid model and its application are covered as the key points of this work. In Section 4, the method is evaluated using the real-world public KITTI dataset. Its applications to several typical traffic scenarios are also considered. Section 5 provides a discussion and concludes this paper.

2 Related work

In this section, we briefly review recent studies on 3D object detection, object tracking, and risk assessment on road scenes. In recent years, with the development of deep neural networks, research into computer vision tasks, such as image classification, object detection, and semantic segmentation, has made significant progress. In the field of object detection, two kinds of network are used: two-stage networks, e.g., R-CNN, Faster R-CNN, and Mask R-CNN, and one-stage networks, e.g., YOLO and SSD. These methods show great improvements over traditional methods like SVM and AdaBoost. Some ideas are employed in 3D object detection using point clouds from LiDAR, RGB-D, and stereo cameras. Three categories of approach exist: (i) projection from 3D point clouds to 2D images, e.g., Complex-YOLO [13] and BirdNet [14], (ii) voxel feature based detection, e.g., 3DFCN [15], Vote3Deep [16], and VoxelNet [17], and (iii) multi-modal fusion approaches, e.g., x MV3D [18] and AVOD [19].
Research into multi-object tracking (MOT) has focused on the data association issue based on tracking-by-detection and learning-to-track methods. Both batch methods [20, 21] and online methods [22, 23] explore how to learn a similarity function for data association. More recent studies on MOT have integrated hierarchical features from deep convolution networks [24, 25] and correlation filters [26]. In addition, use of a reinforcement learning algorithm has been proposed to link data in online MOT: e.g., Markov decision processes (MDP) have proved suitable for dynamic environments [27]. Multiple objects can be modeled using multi-agents which have their own lifetime to perform certain tasks and maintain certain states.

Using history trajectory analysis based on multi-object tracking, object motion can be predicted in a reasonable way over a limited time interval. For example, Dueholem et al. [28] extracted trajectories of surrounding vehicles from panoramic camera arrays and then classified vehicle motions by a hidden Markov model into certain states like overtaking and lane change. According to a survey about trajectory prediction [2], physics-based models and maneuver-based models are widely studied and used [29, 30], while interaction-aware models form the focus of state-of-the-art research since they can provide longer-term predictions than the other models [31, 32]. Schulz et al. [31] proposed a dynamic Bayesian network to model the states of multi-agents for intelligent vehicles in urban scenarios. Generative adversarial networks have also been employed for the real-world task of vehicle behavior prediction, considering interactions between multiple entities as a whole system [32].

In risk assessment, field-based approaches have been proposed for automatic vehicle guidance over a long history. Electric field models are used to interpret the vehicle’s motion as an electron within an electric field, and the system is transformed into a risk field reflecting the risk at a certain position in the dynamic environment [33]. Wolf and Burdick [34] introduced a vehicle collision avoidance system in a full two-dimensional field with lane, road, car, and velocity potential function components. Recent work [35] assessed collision risks by risk potential modeling of predicted motion of the surrounding vehicles under various driving conditions. Wang et al. [36] proposed a hybrid field model to assess driving safety in a pre-collision warning system. We proposed a fluid-inspired risk map representation for surrounding vehicles from a forward-moving monocular camera [37], but it does not take the scale, orientation, or velocities of vehicles into account. In this work, we focus on collision risk assessment by building a fluid-inspired field from the state spaces of observed objects, giving an alternative method of risk representation in the traffic context. Moreover, it is a real-time risk assessment method, compatible with various model inputs.

3 Approach

In this paper, the key question is how to model interactions between multiple objects based on fluid dynamics. The system pipeline has two parts: (i) 3D object detection and tracking, and (ii) fluid-inspired risk assessment, as shown in Fig. 1. Since the risk assessment is independent from the 3D object detection and tracking part, this framework is compatible with various kinds of sensor input, such as LiDAR, radar and cameras. Here, we use point cloud data from LiDAR as input, as LiDAR is typically available on intelligent vehicles.

![Fig. 1](image-url) Our system has two main parts: 3D object detection, and tracking and fluid-inspired risk assessment.
3.1 3D object detection and tracking

In our work, a state-of-the-art 3D detector, PointRCNN [38], is employed to provide detection results using the KITTI pre-trained model. The whole PointRCNN network involves two stages: (i) bottom-up 3D proposal generation, and (ii) canonical 3D box refinement. The first stage directly generates 3D proposals while the second stage refines the proposals in canonical coordinates by fusion of semantic features and local spatial features. This 3D object detection module provides n 3D bounding boxes \( \mathcal{D}_t = \{ \mathbf{D}^1_t, \ldots, \mathbf{D}^n_t \} \) at time step \( t \), in terms of the 3D coordinates of each object’s center \( x^i, y^i, z^i \), size \( l^i, w^i, h^i \), and heading angle \( \theta^i \): \( \mathbf{D}^i_t = (x^i, y^i, z^i, \theta^i, l^i, w^i, h^i)^T \).

We then employ the Kalman filter to iteratively predict the entire state spaces of \( n \) object trajectories \( \mathcal{T}_t = \{ \mathbf{T}^1_t, \ldots, \mathbf{T}^n_t \} \) with a constant velocity model. The state space of each object trajectory is formulated as a 10-dimensional vector \( \mathbf{T}^i_t = (x^i, y^i, z^i, \theta^i, l^i, w^i, h^i, v^i_x, v^i_y, v^i_z)^T \). The motion model of object \( i \) at time step \( t \) is formulated as

\[
\mathbf{T}^i_t = \mathbf{F} \mathbf{T}^i_{t-1} + \mathbf{w}_{t-1} \tag{1}
\]

where

\[
\mathbf{F} = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
\end{pmatrix}
\]

and

\[
\mathbf{w}_{t-1} \sim \mathcal{N}(0, \mathbf{Q}_{t-1})
\]

Using the motion model, the trajectories of all observed objects can be predicted as follows: the trajectory at the prediction step is

\[
\hat{\mathbf{T}}^i_t = \mathbf{F} \hat{\mathbf{T}}^i_{t-1} \tag{2}
\]

and the covariance matrix at the prediction step is

\[
\hat{\mathbf{P}}^i_t = \mathbf{F} \hat{\mathbf{P}}^i_{t-1} \mathbf{F}^T + \mathbf{Q}_{t-1} \tag{3}
\]

Note that the initial covariance matrix \( \hat{\mathbf{P}} \) and the noise matrix \( \mathbf{Q} \) are given at the first step of the iteration.

Then, the data association module is employed by matching the current detection results \( \mathbf{D}^i_t \) with predicted trajectories \( \hat{\mathbf{T}}^i_t \) by using the metric of 3D IoU and the Hungarian algorithm [39]. The data association module generates matched pairs of detected and predicted trajectories \( (\mathbf{D}^i_t, \hat{\mathbf{T}}^i_t)_{\text{match}} \). Then, using the matched pairs in the correction step of the Kalman filter, the measurement model of the Kalman filter is involved, defined as

\[
\mathbf{D}^i_t = \mathbf{H} \hat{\mathbf{T}}^i_t + \mathbf{v}_t \tag{4}
\]

with

\[
\mathbf{H} = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
\end{pmatrix}
\]

and

\[
\mathbf{v}_t \sim \mathcal{N}(0, \mathbf{R}_t)
\]

During the correction step, the Kalman gain matrix is

\[
\mathbf{K}_t = \hat{\mathbf{P}}^i_t \mathbf{H}^T (\mathbf{H} \hat{\mathbf{P}}^i_t \mathbf{H}^T + \mathbf{R}_t)^{-1} \tag{5}
\]

the corrected trajectory is

\[
\tilde{\mathbf{T}}^i_t = \hat{\mathbf{T}}^i_t + \mathbf{K}_t (\mathbf{D}^i_t - \mathbf{H} \hat{\mathbf{T}}^i_t) \tag{6}
\]

and the correction covariance matrix is

\[
\tilde{\mathbf{P}}^i_t = (\mathbf{I} - \mathbf{K}_t \mathbf{H}) \hat{\mathbf{P}}^i_t \tag{7}
\]

The final output of the Kalman filter comprises the corrected trajectories of objects based on the matched pairs, denoted \( \tilde{\mathbf{T}}_t = \{ \tilde{\mathbf{T}}^1_t, \ldots, \tilde{\mathbf{T}}^n_t \} \) where \( \tilde{\mathbf{T}}^i_t = (\tilde{x}^i, \tilde{y}^i, \tilde{z}^i, \tilde{\theta}^i, \tilde{l}^i, \tilde{w}^i, \tilde{h}^i, \tilde{v}_x^i, \tilde{v}_y^i, \tilde{v}_z^i)^T \), reflecting the motions of surrounding objects relative to the ego-vehicle (the agent which perceives the surrounding environment by mounted sensors).

In order to manage the birth and death of trajectories, unmatched detections are considered as potential objects entering the perspective field and corresponding new trajectories are created. Conversely, unmatched trajectories are regarded as potential objects leaving the perspective field. After an unmatched trajectory has been present for several consecutive frames, it is removed from the state space \( \tilde{\mathbf{T}}_t \), terminating the corresponding object’s trajectory.

Next, we employ the fluid model to perform collision risk assessment using the trajectories \( \tilde{\mathbf{T}}_t \).

3.2 Fluid-inspired field representation

In our work, instead of employing long-term trajectory prediction, risk assessment in the traffic
environment is based on the fluid motion. Using the trajectories of tracked objects, we are able to compute the dynamic 2D risk field in a bird’s-eye view based on a fluid-inspired field representation. Our key assumption is that the fluid-like animation of a density field provides the risk map. Our model is that each vehicle on the road emits some invisible smoke, whose animation is driven by the motion of the vehicle. Our intuition is that high density of the virtual smoke indicates a closer distance to the nearby vehicle. In addition to the static distance, the density of the smoke also responds to vehicle motion, providing dynamic update of the risk. Therefore, by sensing the density of the smoke, the ego-vehicle may obtain a risk map for the driving environment.

We now describe how we compute the fluid-like risk field. Because risk is a virtual concept which cannot be perfectly modeled as a fluid, we do not need to strictly adhere to incompressible flow. Instead of implementing an accurate fluid simulation, we propose a computational pipeline inspired by the Navier–Stokes equations. In the field of fluid dynamics, a fluid whose density and temperature are nearly constant can be described by a velocity field \( u \) and a pressure field \( p \) [40]. Given the initial velocity and pressure at \( t = 0 \), the evolution of these quantities over time is given by the Navier–Stokes equations for incompressible flow [12]:

\[
\nabla \cdot u = 0
\]

\[
\frac{\partial u}{\partial t} = -(u \cdot \nabla)u - \frac{1}{\rho} \nabla p + \nu \nabla^2 u + f \tag{8}
\]

where \( \nu \) is the kinematic viscosity of the fluid, \( \rho \) is the (constant) fluid density, and \( f \) represents an external force that acts on the fluid.

Instead of solving the Navier–Stokes equation (8) with an accurate numerical solver, we only need an approximate solution that produces a fluid-like density field. We therefore solve the equations based on the stable fluids technique proposed in Ref. [40]. Following the standard Euler solver, the Navier–Stokes equations are split into three simpler equations: driving force \( \partial u / \partial t = f \), advection \( \partial u / \partial t = -u \cdot \nabla u \), diffusion \( \partial u / \partial t = \nu \nabla^2 u \) and projection \( \partial u / \partial t = -(1/\rho) \nabla p \), such that \( \nabla \cdot u = 0 \). We solve them separately for the velocity field, followed by an advection step for the density field \( R \):

\[
\frac{\partial R}{\partial t} = -(u \cdot \nabla) R + \nu \nabla^2 R \tag{9}
\]

where the velocity field \( u \) is found from Eq. (8) and \( \nu \) controls the diffusion rate of the density field. See Ref. [40] for implementation details.

In our model, \( f \) encapsulates acceleration due to external forces applied to the fluid. We assume that external forces come from the motions of objects in the fluid. This leads to the velocity of the fluid field changing. We set the external force to the relative velocities of the surrounding vehicles with respect to the ego-vehicle. It captures the relative motion between the vehicles and drives the risk map for the ego-vehicle. We have also tested our approach with only advection, only diffusion and with both advection and diffusion, as shown in Fig. 2. We do not recommend omitting the advection or diffusion steps, in order to produce the desired animation of the virtual smoke.

Unlike simulating an incompressible fluid, we model the fluid-like risk map by assuming that each surrounding vehicle emits virtual smoke. In other words, we need to add sources to the field, causing the velocity field to not be divergence-free. Therefore, before we animate the virtual smoke using the fluid-like model, we need to add sources at the locations of the surrounding vehicles. Specifically, each tracked vehicle is “injected” into the fluid as a source, which is formulated as

\[
\frac{\partial R}{\partial t} = S(x, l, w) \tag{10}
\]

with the boundary conditions \( R(x, 0) = w_0(x) \) and \( w_1(x) = R(x, \Delta t) \). The source \( S \) is added at the beginning of each iteration, and is set to be linearly correlated with the positions and sizes of objects. An approximation method can be used here to solve Eq. (10) since the “injected” vehicles have continuous trajectories, and do not vary considerably. This linear method may be written:

\[
R(x) = R_0(x) + \Delta t S(x, l, w) \tag{11}
\]

where \( R_0 \) is the density field in the previous frame.

If we only add sources to the density field, after a
short period of computation, the entire map will be blurred, with a high density. Therefore, we model dissipation of the virtual smoke to keep the density field visually clear. The dissipation term can be solved implicitly, as done for the diffusion term, and is formulated as

\[(I + \Delta t \alpha) R'(x) = R(x)\]  

(12)

where \(I\) is the identity operator, \(\Delta t\) is the time step, the coefficient \(\alpha\) controls the dissipation rate, and \(R'\) is the dissipated density field.

The original diffusion problem is solved by applying an isotropic Gaussian filter as shown in Fig. 4(a); thus, the distribution of the risk field around each object is isotropic. To better present the possible risk, we introduce anisotropy along the velocity of the tracked objects in the diffusion step, as illustrated in Fig. 4(b).

The diffusion stencil for isotropic diffusion is uniform around each grid cell, i.e., \(\gamma(i \pm 1, j \pm 1) = \lambda\). The stencil at neighboring grid cells is updated as

\[\gamma(i + \text{sgn}(v_x), j) = \lambda + \beta v_x\]
\[\gamma(i, j + \text{sgn}(v_y)) = \lambda + \beta v_y\]

(13)

where \(\beta\) is the parameter controlling anisotropy and the \(\text{sgn}\) function is formulated as

\[\text{sgn}(x) = \begin{cases} \frac{\text{d}}{\text{d}x} |x|, & \text{if } x \neq 0 \\ 0, & \text{otherwise} \end{cases}\]  

(14)

Note that more complex anisotropy could be modeled by using the Laplace operator \(\nabla^2\), but we adopt the above updating scheme because it is a simple and efficient encoding method for the short-term prediction of surrounding objects.

4 Results and discussion

4.1 Introduction

In the experiments, we first evaluate the risk assessment module on synthetic data. Then, the proposed method is tested on the KITTI dataset, which contains LiDAR data clips, calibration files, and image sequences (with 10 Hz sampling frequency).

We highlight the behavior of this method using a case of tracking failure, as shown in Fig. 3, where the 11th vehicle is lost in (b) compared with (a). Because it takes time to advect and diffuse the smoke, even if a vehicle disappears due to tracking failure, its evolving smoke from the previous frames persists and still leaves useful hints to the driving system. Due to time-coherence, the missing vehicle has high probability of reappearing and it makes sense to keep the information from it. In comparison, as shown in the upper row of Fig. 3, when a vehicle disappears and reappears, the distance field drastically changes and may misguide the decision-making. In comparison, the fluid field continues to predict the evolving system while considering uncertainty.

Next, we evaluated the whole pipeline of the fluid-based risk assessment method on the KITTI tracking test set, including several typical traffic scenarios, such as an expressway, an intersection, and a roundabout. We focus on vehicles as a unique category and ignore other objects like pedestrians and cyclists. The PointRCNN with model on the pre-trained KITTI object detection dataset was employed
to detect 3D objects in the LiDAR point cloud data. Then, the Kalman filter was adopted to track multiple objects and corrected trajectories were used to assess collision risk based on the fluid model. The process of fluid modeling is computed on a single 2.5 GHz CPU thread, taking about 25 ms for each iteration. The field was updated frame by frame at a resolution of $512 \times 512$, corresponding to an area of $80 \text{ m} \times 80 \text{ m}$, as illustrated in Fig. 5. The risk map covered an area of $-40 \text{ m}$ to $+40 \text{ m}$ in the lateral direction and $-20 \text{ m}$ to $+60 \text{ m}$ in the longitudinal direction. By default, we set the parameters $\alpha = 0.96$, $\beta = 5.0$, and $\mu = 1.0$. The region is selected arbitrarily in this experiment, and one could set different scales according to the requirements of risk assessment. Note that a higher resolution involves more computational cost. Our aim was to verify the effectiveness of the model, without considering various weather and illumination conditions such as a rainy day or night.

4.2 Expressway

The first case is on an expressway as illustrated in Fig. 6. Two vehicles are driving near the ego-vehicle: the green one is in front of it, and the red one is...
overtaking it. Figure 6(b) makes it easy to see that these two vehicles are moving faster than the ego-vehicle. At the beginning of this sequence, the green vehicle with a faster speed has been tracked for a while, its distance from the ego-vehicle increasing continuously. Thus, the risk field of the green vehicle is an arrow-like shape, with almost no probability of collision between the green vehicle and the ego-vehicle. In frame 70 (the next sample), the red vehicle appears in the lane to the left of the ego-vehicle. Since the ego-vehicle has just recognized the red one, it does not update the risk field. Then, the interaction between the red one and the ego-vehicle continuously changes risk field, and the red one shows a similar risk representation as the green one. This case qualitatively shows that the fluid-inspired risk representation matches a human assessment.

4.3 Intersection

The second case is at an intersection, and is much more complex than the expressway. There are two states for the ego-vehicle: waiting at a crossroads, and crossing the intersection. The former is shown in Fig. 7. In this case, several vehicles are driving in front of the ego-vehicle. Since the ego-vehicle is stopped, the influence of target vehicles on the risk field depends solely on their own motions. For example, when target vehicles turn left and approach the ego-vehicle (in a stopped state), the high risk area is located at the front-left side of these vehicles. It is consistent with the intuition that if an object approaches with a certain velocity, the collision risk increases as the relative distance decreases. Moreover, as certain vehicles are static in this sequence, their influence on risk is small since there is almost no relative movement between the ego-vehicle and these vehicles.

As the ego-vehicle turns and crosses the intersection (see Fig. 8), the surrounding objects make a reasonable risk field representation. Especially, at the beginning of the sequence, there is no risk because all vehicles are static including the ego-vehicle. Once the vehicle moves, the risk field changes. The last frame shows when the ego-vehicle moves, the stationary vehicles still contribute to a high risk region. This is because the target vehicles are dense in a small area.

Fig. 7 Risk assessment when the ego-vehicle is stopped at an intersection, from KITTI tracking test set 0010: (a) LiDAR point cloud annotated with 3D bounding boxes; (b) 3D object tracking results projected onto the synchronized image sequence; (c) risk representation based on the fluid field.
The relative motion changes the pressure in the fluid where substances tend to squish and slosh.

4.4 Roundabout

The third case is at a roundabout as shown in Fig. 9. This is a typical traffic scenario in which traffic is allowed to flow in one direction around a central island, with priority typically given to traffic already in the junction. It is still difficult for intelligent vehicles because the roundabout conditions are complex and in most cases no traffic light controls the traffic flow and provides traffic guidance.

The samples are selected from the KITTI tracking test set 0008, where the ego-vehicle follows another vehicles driving into a roundabout. Around the roundabout, each vehicle follows its lane until it drives out of the roundabout. Qualitative results are shown in Fig. 9(c). The headings of target vehicles obviously differ from the directions of their trajectories. The effects of their motion on the risk representation are calculated based on their historic trajectories. The areas of high risk are in the direction of target vehicles’ trajectories in the view of the ego-vehicle due to their motions. We find that when objects leave the ego-vehicle’s perceptive field, the substance of the risk field dissipates immediately: there is no risk from these vehicles.

When the ego-vehicle approaches another vehicle, the risk is computed by using their relative movement as shown in Fig. 10. In this case, 6 samples from frame 70 to frame 120 are used to illustrate the process of approach. Since the relative velocity decreases, the risk correspondingly declines to a low level. Since the ego-vehicle does not completely stop, there is a low level risk field around the vehicles.

As shown in the cases above, the fluid-inspired risk assessment method not only works for the ego-vehicle, but also provides a risk estimation for the local traffic situation. It could also be used for traffic situation analysis for road-side surveillance systems. As mentioned above, failure of detection and tracking due to various issues like occlusion is a potential threat to intelligent vehicles. Next, we consider risk assessment in the case of the detection and tracking failure.

4.5 Detection and tracking failure

In Fig. 11, consider the area marked by red circles in the risk map. They indicate changing risk when a vehicle becomes lost due to occlusion. In frames 122
Fig. 9  Risk assessment at a roundabout from KITTI tracking test set 0008: (a) LiDAR point cloud annotated with 3D bounding boxes; (b) 3D object tracking results projected on the synchronized image sequence; (c) risk representation based on the fluid field.

Fig. 10  Risk assessment when the ego-vehicle approaches other vehicles from KITTI tracking test set 0013: (a) LiDAR point cloud annotated with 3D bounding boxes; (b) the 3D object tracking results projected on the synchronized image sequence; (c) risk representation based on the fluid field.
Fluid-inspired field representation for risk assessment in road scenes

Fig. 11 The fluid model provides a field-based representation of an object for which tracking fails. Red circles indicate risk map changes when the vehicle becomes lost due to occlusion.

to 124, it is successfully detected and tracked (blue 3D bounding box), but then it is fully occluded by another vehicle and no longer tracked in frames 125 to 127. In the last two frames 128, 129, this vehicle is detected and tracked as a new object. In this sequence, the risk field still reflects the influence of the occluded vehicle through the processes of advection and diffusion, even though it is not detected and tracked.

4.6 Comparison

Lastly, we compare our model with the state-of-the-art predictive occupancy map (POM) [3], which represents risk as a reciprocal of time taken for surrounding vehicles to occupy a region, as shown in Fig. 13. The risk is computed based on the advanced time-to-occupancy (ATTO), an indicator similar to TTC.

We take the KITTI tracking test set 0006 as an example to show the difference between our model and POM; results are illustrated in Fig. 12. Compared to the POM, the fluid-inspired risk is calculated based on both the current motion of surrounding vehicles and their historical influence on the road. It highlights the effect of each object’s previous motion on the

Fig. 12 Comparison of the fluid-inspired and POM risk maps.
traffic scene. In addition, the fluid-inspired risk map further considers uncertainty of relative motion and observation between surrounding vehicles and the ego-vehicle, while the POM uses a simple and direct method, just taking the observations to calculate the risk. Once an object is no longer detected due to occlusion, it is difficult for the POM to handle the risk representation.

One should note that in real urban scenarios, there are not only vehicles but also other vulnerable road users, such as pedestrians, cyclists, and motorcyclists, with different attributes to vehicles. Especially, in complex road scenes like busy intersections, such objects should be carefully considered. Moreover, road geometric attributes are another important factor which should be studied in collision risk estimation. Geometric attributes like drivable road space and curvature can be acquired by state-of-the-art technologies like CNN-based semantic segmentation and lane marking detection, but are outside the scope of this paper. We hope to study the impact of road geometry and other road users in following work.

5 Conclusions

In this paper, we propose an approach of fluid-inspired risk assessment, which can be easily integrated with different modes of data, like point clouds and image sequences. We use a state-of-the-art 3D detector and a 3D Kalman filter for multi-object tracking to estimate the state space of surrounding objects (using vehicles as an example). Then, advection and diffusion processes handle the physically evolving risk field in order to build a dynamic risk map for intelligent vehicles. This method enables short-term prediction of the distribution along with the trajectory of each object, even when object detection and tracking fails and sensor data is imperfect. We have evaluated our proposed approach on the public KITTI dataset.

This method targets advanced driver assistance systems and autonomous driving systems. For planning and control of autonomous vehicles, our proposed risk estimation method provides constraints for behavioral decisions and local path planning. Once decisions and local paths have been determined, control of the vehicle would be executed based on methods like PID and model predictive control. Note that both driving assistance and autonomous driving are high-level tasks which require understanding of traffic scenarios. Our method can capture dynamic information about multi-object interactions at the conceptual level by solely using possibly uncertain estimated trajectories. In future we will explore how to combine other road elements like lane markings, geometric information, and static obstacles for further planning and decision-making. While in this work, although we do not focus on tracking occluded objects, we do believe stable tracking is beneficial for risk assessment. Therefore, a systematic field-based method to deal with the challenges of tracking surrounding vehicles is also part of our future work.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant No. 61906038, the Fundamental Research Funds for the Central Universities under Grant No. 2242019K40039, and the Zhishan Youth Scholar Program of Southeast University.

References

[1] Laugier, C.; Paromtchik, I. E.; Perrollaz, M.; Yong, M.; Yoder, J.; Tay, C.; Mekhnacha, K.; N`egre, A. Probabilistic analysis of dynamic scenes and collision risks assessment to improve driving safety. *IEEE Intelligent Transportation Systems Magazine* Vol. 3, No. 4, 4–19, 2011.

[2] Lefèvre, S.; Vasquez, D.; Laugier, C. A survey on motion prediction and risk assessment for intelligent vehicles. *ROBOMECH Journal* Vol. 1, No. 1, 1–14, 2014.

[3] Lee, M.; Sunwoo, M.; Jo, K. Collision risk assessment of occluded vehicle based on the motion predictions using the precise road map. *Robotics and Autonomous Systems* Vol. 106, 179–191, 2018.

[4] Coué, C.; Pradalier, C.; Laugier, C.; Fraichard, T.; Bessière, P. Bayesian occupancy filtering for multitarget tracking: An automotive application. *The International Journal of Robotics Research* Vol. 25, No. 1, 19–30, 2006.
[5] Nguyen, T. N.; Michaelis, B.; Al-Hamadi, A.; Tornow, M.; Meinecke, M. M. Stereo-camera-based urban environment perception using occupancy grid and object tracking. *IEEE Transactions on Intelligent Transportation Systems* Vol. 13, No. 1, 154–165, 2012.

[6] Lee, K.; Kum, D. Collision avoidance/mitigation system: Motion planning of autonomous vehicle via predictive occupancy map. *IEEE Access* Vol. 7, 52846–52857, 2019.

[7] Hamrick, J.; Battaglia, P.; Tenenbaum, J. B. Internal physics models guide probabilistic judgments about object dynamics. In: Proceedings of the 33rd Annual Conference of the Cognitive Science Society, Vol. 2, 2011.

[8] Yang, Z. S.; Yu, Y.; Yu, D. X.; Zhou, H. X.; Mo, X. L. APF-based car following behavior considering lateral distance. *Advances in Mechanical Engineering* Vol. 5, 207104, 2013.

[9] Wang, J. Q.; Wu, J.; Li, Y. The driving safety field based on driver–vehicle–road interactions. *IEEE Transactions on Intelligent Transportation Systems* Vol. 16, No. 4, 2203–2214, 2015.

[10] Villegas, R.; Yang, J.; Zou, Y.; Sohn, S.; Lin, X.; Lee, H. Learning to generate long-term future via hierarchical prediction. In: Proceedings of the IEEE International Conference on Machine Learning, 3560–3569, 2017.

[11] Li, J.; Ma, H.; Zhan, W.; Tomizuka, M. Generic probabilistic interactive situation recognition and prediction: From virtual to real. In: Proceedings of the IEEE International Conference on Intelligent Transportation Systems, 3218–3224, 2018.

[12] Chorin, A. J.; Marsden, J. E. *A Mathematical Introduction to Fluid Mechanics*. New York: Springer, 1990.

[13] Simon, M.; Milz, S.; Amende, K.; Gross, H. M. Complex-YOLO: An Euler-region-proposal for real-time 3D object detection on point clouds. In: *Computer Vision – ECCV 2018 Workshops. Lecture Notes in Computer Science*, Vol. 11129. Leal-Taixé, L.; Roth, S. Eds. Springer Cham, 197–209, 2019.

[14] Beltran, J.; Guindel, C.; Moreno, F. M.; Cruzado, D.; Garcia, F.; de La Escalera, A. BirdNet: A 3D object detection framework from LiDAR information. In: Proceedings of the 21st International Conference on Intelligent Transportation Systems, 3517–3523, 2018.

[15] Li, B. 3D fully convolutional network for vehicle detection in point cloud. In: Proceedings of the IEEE International Conference on Intelligent Robots and Systems, 1513–1518, 2017.

[16] Engelcke, M.; Rao, D.; Wang, D. Z.; Tong, C. H.; Posner, I. Voten3DDeep: Fast object detection in 3D point clouds using efficient convolutional neural networks. In: Proceedings of the IEEE International Conference on Robotics and Automation, 1355–1361, 2017.

[17] Zhou, Y.; Tuzel, O. VoxelNet: End-to-end learning for point cloud based 3D object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4490–4499, 2018.

[18] Chen, X.; Ma, H.; Wan, J.; Li, B.; Xia, T. Multi-view 3D object detection network for autonomous driving. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 6526–6534, 2017.

[19] Ku, J.; Mozifian, M.; Lee, J.; Harakeh, A.; Waslander, S. L. Joint 3D proposal generation and object detection from view aggregation. In: Proceedings of the IEEE International Conference on Intelligent Robots and Systems, 1–8, 2018.

[20] Butt A. A.; Collins, R. T. Multi-target tracking by Lagrangian relaxation to min-cost network flow. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1846–1853, 2013.

[21] Kuo, C.; Huang, C.; Nevatia, R. Multi-target tracking by on-line learned discriminative appearance models. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 685–692, 2010.

[22] Bae S.-H.; Yoon, K.-J. Robust online multi-object tracking based on tracklet confidence and online discriminative appearance learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1218–1225, 2014.

[23] Yoon, J. H.; Lee, C.; Yang, M.; Yoon, K. Online multi-object tracking via structural constraint event aggregation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1392–1400, 2016.

[24] Leal-Taixé, L.; Canton-Ferrer, C.; Schindler, K. Learning by tracking: Siamese CNN for robust target association. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 33–40, 2016.

[25] Tang, S. Y.; Andres, B.; Andriluka, M.; Schiele, B. Multi-person tracking by multicut and deep matching. In: *Computer Vision – ECCV 2016 Workshops. Lecture Notes in Computer Science*, Vol. 9914. Hua, G.; Jégou, H. Eds. Springer Cham, 100–111, 2016.

[26] Park, S.; Lee, K.; Yoon, K. Robust online multiple object tracking based on the confidence-based relative motion network and correlation filter. In: Proceedings of the IEEE International Conference on Image Processing, 3484–3488, 2016.

[27] Xiang, Y.; Alahi, A.; Savarese, S. Learning to track: Online multi-object tracking by decision making. In: Proceedings of the IEEE International Conference on Computer Vision, 4705–4713, 2015.
[28] Dueholm, J. V.; Kristoffersen, M. S.; Satzoda, R. K.; Moeslund, T. B.; Trivedi, M. M. Trajectories and maneuvers of surrounding vehicles with panoramic camera arrays. *IEEE Transactions on Intelligent Vehicles* Vol. 1, No. 2, 203–214, 2016.

[29] Xie, G. T.; Gao, H. B.; Qian, L. J.; Huang, B.; Li, K. Q.; Wang, J. Q. Vehicle trajectory prediction by integrating physics- and maneuver-based approaches using interactive multiple models. *IEEE Transactions on Industrial Electronics* Vol. 65, No. 7, 5999–6008, 2018.

[30] Deo, N.; Trivedi, M. M. Multi-modal trajectory prediction of surrounding vehicles with maneuver based LSTMs. In: Proceedings of the IEEE Intelligent Vehicles Symposium, 1179–1184, 2018.

[31] Schulz, J.; Hubmann, C.; Löchner, J.; Burschka, D. Interaction-aware probabilistic behavior prediction in urban environments. In: Proceedings of the IEEE International Conference on Intelligent Robots and Systems, 3999–4006, 2018.

[32] Li, J.; Ma, H.; Tomizuka, M. Interaction-aware multi-agent tracking and probabilistic behavior prediction via adversarial learning. In: Proceedings of the IEEE International Conference on Robotics and Automation, 6658–6664, 2019.

[33] Reichardt, D.; Shick, J. Collision avoidance in dynamic environments applied to autonomous vehicle guidance on the motorway. In: Proceedings of the IEEE Intelligent Vehicles Symposium, 74–78, 1994.

[34] Wolf, M. T.; Burdick, J. W. Artificial potential functions for highway driving with collision avoidance. In: Proceedings of the IEEE International Conference on Robotics and Automation, 3731–3736, 2008.

[35] Kim, K.; Kim, B.; Lee, K.; Ko, B.; Yi, K. Design of integrated risk management-based dynamic driving control of automated vehicles. *IEEE Intelligent Transportation Systems Magazine* Vol. 9, No. 1, 57–73, 2017.

[36] Wang, J. Q.; Wu, J.; Zheng, X. J.; Ni, D. H.; Li, K. Q. Driving safety field theory modeling and its application in pre-collision warning system. *Transportation Research Part C: Emerging Technologies* Vol. 72, 306–324, 2016.

[37] Zhu, L.; Li, X.; Lu, W.; Zhang, Y. J. A field-based representation of surrounding vehicle motion from a monocular camera. In: Proceedings of the IEEE Intelligent Vehicles Symposium, 1761–1766, 2018.

[38] Shi, S.; Wang, X.; Li, H. PointRCNN: 3D object proposal generation and detection from point cloud. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770–779, 2019.

[39] Kuhn, H. W. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly* Vol. 2, Nos. 1–2, 83–97, 1955.

[40] Stam, J. Stable fluids. In: Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques, 121–128, 1999.

Xuanpeng Li received his B.S. and M.S. degrees in instrument science and technology from Southeast University, China, in 2007 and 2010, and his Ph.D. degree in information technology from the Université de Technologie de Compiègne, France, in 2014. From 2014 to 2015, he was a post-doctor at LIVIC, IFSTTAR, and VEDECOM in France. Since 2015, he has been an assistant professor with the School of Instrument Science and Engineering. His research interests include causal perception, scene understanding, driving behavior analysis, and risk estimation for intelligent transportation systems.

Lifeng Zhu received his doctoral degree in computer science from Peking University in 2012. From 2012 to 2015, he was a post-doctor at the Universities of Tokyo and Pennsylvania. Since 2018, he has been an associate professor in the Department of Instrument Science and Technology in Southeast University. His research topics are visual computing and human computer interaction. His interests are in particular shape modeling, simulation, and visualization methods for medical science and intelligent transportation systems.

Qifan Xue is currently a Ph.D. student in Southeast University, China. He received his B.S. degree from Southeast University in 2017. His research interests include causal perception and scene understanding on the traffic scene.

Dong Wang received his B.S. degree from the School of Instrument Science and Engineering, Southeast University in 2011, and his Ph.D. degree from Southeast University in 2016. He is currently an assistant professor with the School of Instrument Science and Engineering, Southeast University. His research interests include intelligent transportation, sensor technology, and signal processing.
Yongjie Jessica Zhang is the George Tallman Ladd and Florence Barrett Ladd Professor of Mechanical Engineering at Carnegie Mellon University with a courtesy appointment in the Department of Biomedical Engineering. She received her B.Eng. degree in automotive engineering and her M.Eng. in engineering mechanics from Tsinghua University, China, and M.Eng. degrees in aerospace engineering and engineering mechanics, and her Ph.D. degree in computational Engineering and Sciences from the Institute for Computational Engineering and Sciences, the University of Texas at Austin. Her research interests include computational geometry, mesh generation, computer graphics, visualization, finite element methods, isogeometric analysis, and their application to biomedicine and engineering.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made.

The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Other papers from this open access journal are available free of charge from http://www.springer.com/journal/41095. To submit a manuscript, please go to https://www.editorialmanager.com/cvmj.