Statistical Inference of Motion in the Invisible

Haroon Idrees, Imran Saleemi, and Mubarak Shah
{haroon,imran,shah}@eecs.ucf.edu

Computer Vision Lab, University of Central Florida, Orlando, USA

Abstract. This paper focuses on the unexplored problem of inferring motion of objects that are invisible to all cameras in a multiple camera setup. As opposed to methods for learning relationships between disjoint cameras, we take the next step to actually infer the exact spatiotemporal behavior of objects while they are invisible. Given object trajectories within disjoint cameras' FOVs (field-of-view), we introduce constraints on the behavior of objects as they travel through the unobservable areas that lie in between. These constraints include vehicle following (the trajectories of vehicles adjacent to each other at entry and exit are time-shifted relative to each other), collision avoidance (no two trajectories pass through the same location at the same time) and temporal smoothness (restricts the allowable movements of vehicles based on physical limits). The constraints are embedded in a generalized, global cost function for the entire scene, incorporating influences of all objects, followed by a bounded minimization using an interior point algorithm, to obtain trajectory representations of objects that define their exact dynamics and behavior while invisible. Finally, a statistical representation of motion in the entire scene is estimated to obtain a probabilistic distribution representing individual behaviors, such as turns, constant velocity motion, deceleration to a stop, and acceleration from rest for evaluation and visualization. Experiments are reported on real world videos from multiple disjoint cameras in NGSIM data set, and qualitative as well as quantitative analysis confirms the validity of our approach.

1 Introduction

The proliferation of large camera networks in recent past has ushered research in multiple camera analysis, and several methods have been proposed to address the problems of calibration, tracking and activity analysis with some degree of reliability [1,2,3,4,5,6,7]. However, despite significant efforts in this area, the majority of literature has been confined to solution of problems like object correspondence and activity correlation between visible objects, while estimation and inference of object behaviors in *unobservable regions* between disjoint cameras has mainly remained unexplored. Such invisible regions between disjoint cameras are always present as visual sensor networks have an inherent inability to provide exhaustive coverage of all areas of interest, while failure of a sensor is always a possibility which can result in loss of coverage of a particular area that was previously being observed. Besides the issue of coverage, there are several other applications that justify research into such inference: improving object correspondences across cameras; estimating patterns of motion and scene structure; aiding

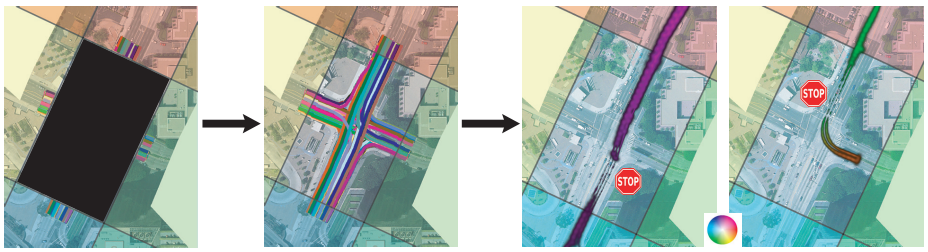

Fig. 1: The first image depicts the input to our method - correspondences across multiple disjoint cameras. In this case, there are five cameras, the FOV of cameras are shown with different colors whereas invisible region is represented by black. Given the input, we reconstruct individual trajectories using constraints introduced in this paper. Next, reconstructed trajectories are used to infer expected behavior at each location in the scene, shown as thick color regions, where the direction of motion is shown by HSV color wheel. We also infer different behaviors such as stopping and turning from the reconstructed trajectories.

expensive operations for PTZ camera focusing by precise localization of unobservable objects; and generalized scene understanding, etc.

In this paper, we pose the question of what information can be inferred about objects while they are not observable in any camera, given tracking and object correspondence information in a multiple camera setup? This is an under-constrained problem, for instance, the correspondence provides two sets of constraints (position, velocity), but if the object is invisible for a hundred time steps, then we have to solve for a hundred variables. For instance, in the scenario shown in Fig. 1, the knowledge that an object exits a camera's FOV on the top, and enters another's on the right is of little use in guessing what its behavior was while invisible. The best we can do it to assume that object moved with the constant velocity through the invisible region. But, this is a rather strong assumption, since the object may have stopped at some unknown locations for unknown amounts of time, or may have taken an indirect path between exit and re-entry into the camera network. Such behavior is influenced by scene structure (such as allowable paths), obstacle avoidance, collision avoidance with other objects, and if the object is a vehicle, it may further be influenced by dynamic aspects of the invisible scene such as traffic signals. Besides being an assumption that is not always true, constant velocity does not provide with any useful information about motion of object or the invisible region itself. The question then becomes, can we do better than constant velocity? If we assume absolutely no information about the invisible region and treat objects independently, the answer is no. But, if we have correspondences for multiple objects available, then the fact that motion of an object is dependent on proximal objects can be used to constrain its movement to a certain degree. The idea of using object-specific contextual constraints has been used as social force models for tracking [8] and simulation, and for describing vehicular motion in transportation theory [9,10,11]. But, these models differ in application from the problem addressed in his paper, in that they assume object positions are known with certainty, we on the other hand, use these constraints as costs

which we minimize to obtain positions at each time instant. This is further complicated by the fact that, each object affects motion of its nearby objects and thus, circular interdependencies exist between objects. This requires some form of precedence mechanism where some objects receive priority in the solution over others.

In the proposed approach, given object trajectories within disjoint cameras' FOVs, we introduce constraints on the behavior of objects as they travel through the unobservable areas that lie in between. The first of these constraints is *vehicle following*, which is based on the observation that trajectories of proximal vehicles that exit a camera and enter another camera are time-shifted versions of each other. The second constraint is *collision avoidance*, which ensures that vehicles do not collide with each other. The next constraint *temporal smoothness* restricts the allowable movements of vehicles based on physical limits. The last constraint *behavior localization* bounds the cost introduced by collision avoidance ensuring that solution is non-trivial. The constraints are embedded in a generalized, global cost function for the entire scene, incorporating influences of all objects, followed by a minimization to obtain trajectory representations of objects that define their exact dynamics and behavior while invisible. Finally, a statistical representation of motion in the entire scene is estimated to obtain a probabilistic distribution representing individual behaviors, such as turns, constant velocity motion, deceleration to a stop, and acceleration from rest, for evaluation and visualization.

To the best of our knowledge, there are currently no methods in literature that attempt to infer any salient properties for unobservable areas (trajectories, motion patterns, static and dynamic behavior of objects). In addition to the applications listed earlier, such inference will allow economically viable deployment of large sensor networks without the need to cover all possible regions of interest, and learning of patterns of activity for an invisible region will allow detection of anomalous behavior without directly observing it. To summarize our contributions, the proposed framework has the ability to infer the following about an *invisible* scene without any observations:

- Estimation of object trajectories in (x, y, t) as they travel through the unobservable area,
- Inference of static and dynamic aspects of the scene such as positions where objects generally stop and traffic lights without direct observation, and
- Completely unsupervised discovery and statistical representation of salient object patterns of motion for the entire scene including large invisible regions.

The organization of the rest of the paper follows. We briefly review relevant literature in §2, followed by detailed problem formulation and solution in §3, inference of static and dynamic scene structure in §4 and presentation of experimental evaluation in §5.

2 Related Work

The literature involving multi-camera scenarios contains techniques proposed for object tracking as well as scene and activity modeling and understanding [12,13,14]. In particular, many methods have been proposed specifically for analysis of disjoint, or non-overlapping multiple cameras [4,6,5,15]. The objective of modeling camera relationships is to use these models to solve for hand-off between cameras, but the pat-

terns are modeled only for motion within the field of views of cameras. Dockstader and Tekalp [16] use Bayesian networks for tracking and occlusion reasoning across calibrated overlapping cameras while in a series of papers [17,18,19], authors employ Kalman Consensus Filter for tracking in multiple camera networks.

In terms of inference of topological relationships between multiple disjoint cameras, Makris et al. [3] determined the topology of a camera network by linking entry and exit zones using co-occurrence relationships between them, while Tieu et al. [20] avoid explicit feature matching by computing the statistical dependence between observations in different cameras to infer network topology. Stauffer [21] proposed an improved linking method to handle cases where exit-entry events may be correlated, but the correlation is not due to valid object transitions. Another interesting area of research is the focus of work by Loy et al. [7], where instead of assuming feature correspondences across cameras, or availability of any tracking data, regions of locally homogenous motion are first discovered in all camera views using correlation of object dynamics. A canonical cross correlation analysis is then performed to discover and quantify the time delayed correlations of regional activities observed within and across multiple cameras in a common reference space.

For our purpose of behavior inference in unobservable regions, avoidance of object collision in structured scenes is one of the most important cues. In this respect, research in transportation theory has attempted to perform collision prediction and detection in the visual domain. Atev et al. [22] propose a collision prediction algorithm using object trajectories, and van den Berg et al. [23] solve the problem of collision avoidance for independently moving robots that can observe each other. In our proposed framework, however, it is essentially assumed that no collisions took place in the unobservable region, and the goal is to infer unobserved object behavior given this assumption. Notice that although some of the proposed constraints bear similarity to 'motion planning' algorithms in robotics [24], some of the significant differences include the fact that for path planning the obstacles are directly observable and the length of time taken to reach the destination is unconstrained. Our method essentially deals with the reverse problem of path planning, i.e., inferring the path that has already been traversed. We therefore propose a solution to a previously unexplored problem. In the following, we formally define the problem and discuss our approach to solve it.

3 Problem Formulation and Solution

Given a set of trajectories (for only observable areas) that have been transformed to a global frame-of-reference, we focus on the difficult and interesting scenario when the unobservable region contains traffic intersections even though the solution we propose in this paper can handle simpler situations such as straight roads as well. The input variables of the problem are the correspondences, i.e., a vehicle's position, velocity, and time when it enters and exits the invisible region (or equivalently exits a camera's field of view and enters another's).

Let p_i^t, v_i^t, and a_i^t denote the position, velocity and acceleration respectively, of the i^{th} vehicle at time t while traveling through the invisible region and η_i and χ_i be the time instants it enters and exits the invisible region. Thus, given the pair of triplets for

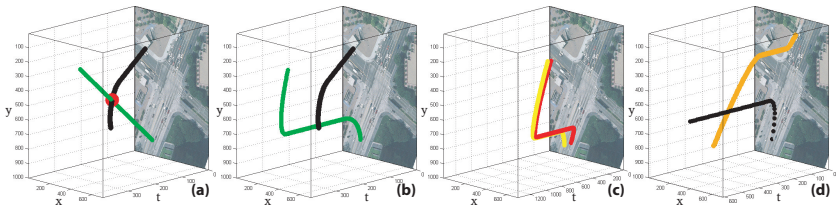

Fig. 2: Depiction of constraints using vehicle trajectories in (x, y, t): (a) The point of collision between green and black trajectories shown with a red sphere, whereas collision is avoided in (b). (c) shows an example of vehicle-following behavior where vehicle in yellow trajectory follows the one in red. In (d) the orange trajectory does not violate smoothness constraint but the one in black does (abrupt deceleration).

entry (p^η, v^η, η) and exit (p^χ, v^χ, χ), our goal is to find p_i^t for all $t \in [\eta_i, \chi_i]$, for each vehicle which correspondingly determines v_i^t and a_i^t.

A path $\mathcal{P}_i$ is a set of 2d locations traversed by a vehicle i and is obtained by connecting p^η with p^χ such that derivative of $\mathcal{P}_i$ is computable at all points i.e. there are no sudden turns or bends. The path so obtained does not contain any information about time. Associating each location in $\mathcal{P}_i$ with time gives us the trajectory $\{p_i^t\}$. Two vehicles i and j have the same path i.e., $\mathcal{P}_i \equiv \mathcal{P}_j$, if $\|p_i^{\eta_i} - p_j^{\eta_j}\|$ and $\|p_i^{\chi_i} - p_j^{\chi_j}\|$ are less than threshold T, and they have temporal overlap $\mathcal{O}(i,j) = 1$, if $\eta_i < \chi_j \wedge \eta_j < \chi_i$. Moreover, their paths intersect, i.e., $i \perp j$, if $\mathcal{P}_i$ obtained by joining $p_i^{\eta_i}$ to $p_i^{\chi_i}$, intersects with $\mathcal{P}_j$.

Since inference of motion in invisible regions in a severely under-constrained problem, we impose some priors over the motion of vehicles as they travel through the region. These priors in §3.1-3.4 below, are used as constraints that will later allow us to reconstruct complete trajectories in the invisible region. The first two constraints (collision avoidance and vehicle following) essentially capture the context of spatially and temporally proximal vehicles while third constraint (smoothness of trajectories) establishes physical limits on the mobility of vehicles as they travel through the region. Using these constraints, we propose an algorithm to reconstruct trajectories in §3.6.

3.1 Collision Avoidance

The first prior we exploit is the fact that vehicles are driven by intelligent drivers who tend to avoid collisions with each other. The probability that a vehicle will occupy a location at particular time becomes low if the same location is occupied by another vehicle at that same time. This effectively reduces the possible space of the solution, leaving only those solutions that have low probability of collisions. Consider the two vehicle trajectories shown in Fig. 2(a) where black trajectory shows a vehicle making a left-turn while vehicle with green trajectory moves straight. The corresponding 2d paths, $\mathcal{P}$ intersect at the point marked with a red sphere. A collision implies that a single point in (x, y, t) is occupied by more than one object. Collision avoidance, thus, enforces that no two trajectories pass though the same (x, y, t) point. In Fig. 2(b), the collision is avoided by a change in shape of the green trajectory. Note that, collision

avoidance doesn't necessarily mean that vehicles change paths in space, but that they don't occupy the same spatial location at the same time. Formally, let τ be the time when vehicles with intersecting paths are closest to each other in (x, y, t), then the collision cost for vehicle i given by:

$$C_i^{\alpha} = \sum_j \exp\left(\omega^{\alpha} \cdot \frac{v_i^{\tau}.v_j^{\tau}}{\|p_i^{\tau} - p_j^{\tau}\|}\right), \text{ where } \tau = \underset{t}{\operatorname{argmin}} \left\|p_i^t - p_j^t\right\|, \tag{1}$$

$\forall j | i \perp j \wedge \mathcal{O}(i,j) = 1$, ω^{α} being the weight. The above equation captures the cost due to motion at the point of closest approach from all vehicles with respect to the vehicle under consideration. The exponentiation softens the impact of collision to nearby points in (x, y, t), thus forcing vehicles to not only avoid the same point but avoid close proximity as well. Two proximal vehicles both moving with a high velocity will have a high cost, however, if at least one vehicle is stationary, this cost will be low.

3.2 Vehicle Following

Like collision avoidance, this constraint reduces the solution space by making sure that relative positions of adjacent and nearby vehicles remain consistent throughout their travel in the invisible region. It is inspired from transportation theory, where vehicle following models describe the relationship between vehicles as they move on the roadway [9,10,11]. Many of them are sophisticated functions of distance, relative velocity and acceleration of vehicles and have several parameters such as desired velocity based on speed limit, desired spacing between vehicles and comfortable braking distance.

We, on the other hand, use vehicle following to define a spatial constraint between leading and following vehicles. The leader $\mathfrak{l}$ and follower $\mathfrak{f}$ are given by the pair:

$$\begin{aligned}(\mathfrak{l}, \mathfrak{f}) = \big\{(i,j) | \mathcal{P}_i = \mathcal{P}_j, \mathcal{O}(i,j) = 1 \wedge \eta_i < \eta_j, \\ \chi_i < \chi_j \wedge \nexists k | \eta_i < \eta_k < \eta_j \vee \chi_i < \chi_k < \chi_j \big\}.\end{aligned} \tag{2}$$

We use the relationship between leader and follower to constrain the possible movement of follower by forcing it to remain behind its leader throughout its travel through the invisible region. This also caters for the correct stopping position of follower since it must stop behind the leader and not occupy the same spot, an event which is highly likely if we only take into account cost from collision. The vehicle-following cost for follower given the leader is written as:

$$C_i^{\beta} = exp\left(\omega^{\beta} \sum_t \|p_j^t - p_i^t\|_+\right), \tag{3}$$

where $\|p_j^t - p_i^t\|_+ = \|p_j^t - p_i^t\|$, if $\|p_j^t - p_j^{\chi_j}\| < \|p_i^t - p_j^{\chi_j}\|$ and 0 otherwise; ω^{β} is the weight associated to this cost.

Vehicle-following constraint enforces the condition that trajectories of vehicles adjacent to each other following the same path are time-shifted versions of each other, as can be seen in Fig. 2(c) where red and yellow trajectories belong to the leading and following vehicles respectively.

3.3 Smoothness Of Trajectory

The smoothness constraint restricts the allowable movements of vehicles based on physical limits as it happens in real life. It prevents the solution from having abrupt acceleration or deceleration as well as sudden stops. This is an object-centric constraint and is computed as:

$$C_i^{\gamma} = exp\left(\omega^{\gamma} \sum_t \left(1 - \sqrt{\frac{\pi}{2}} \mathcal{N}(\frac{v_i^t}{v_i^{t-1}}; 1, \sigma^{\gamma})\right)\right), \tag{4}$$

where ω^{γ} is the weight, and $\mathcal{N}$ is the normal distribution with $\sigma^{\gamma} = 0.25$ variance.

The above equation ensures that distance in space-time volume between any two adjacent points in a single trajectory is a small multiple of the other. In Fig. 2(d), the orange trajectory is has low smoothness cost whereas black trajectory has higher cost due to abrupt deceleration in the beginning.

3.4 Stopping Behavior Localization

The above three constraints do not completely specify the solution because trivial solutions with high values of acceleration and deceleration can exist. This is possible when a vehicle is made to stop with high deceleration i.e. near $p_i^{\eta_i}$, stays there as long as possible before leaving the invisible region with high acceleration while satisfying the smoothness constraint. This *afraid-of-collision* solution for a vehicle is not only incorrect, it also will result in wrong results for all of the following vehicles. The following additional cost will rectify this problem avoiding such solution,

$$C_i^{\delta} = exp\left(\omega^{\delta} \left\| x_{i,j} - p_i^{t_{\varphi}} \right\|\right), j | i \perp j \wedge \mathcal{O}(i,j) = 1, \tag{5}$$

where ω^{δ} is the associated weight, $x_{i,j}$ is the spatial point where vehicle paths intersect and t_{φ} is the time when vehicle stops. This constraint dictates that stopping point for a vehicle cannot be arbitrarily away from the possible collision locations, essentially localizing the move-stop-move events in space and time.

3.5 Trajectory Parametrization

As mentioned earlier, given entry and exit locations, each trajectory is represented by a 2D path, $\mathcal{P}$ by joining the two locations in a way that allows for bends as they happen in the case of turns. Parameterizing the curve by placing $\chi - \eta$ equidistant points for a vehicle gives us a trajectory that represents motion with constant velocity. Parameterizing in this way reduces the number of variables to one-half (from 2D to 1D). Our goal then becomes to find the temporal parametrization of the trajectory that minimizes cost from the constraints. But, given that we might be dealing with thousands of vehicles, each invisible for hundreds of time units, the extremely large variable space ($\sim 10^6$) makes the problem intractable.

In order to make the problem tractable, we reduce the parameters defining a trajectory to three: deceleration (ϕ), duration of stopping time (φ) and acceleration (ψ). The

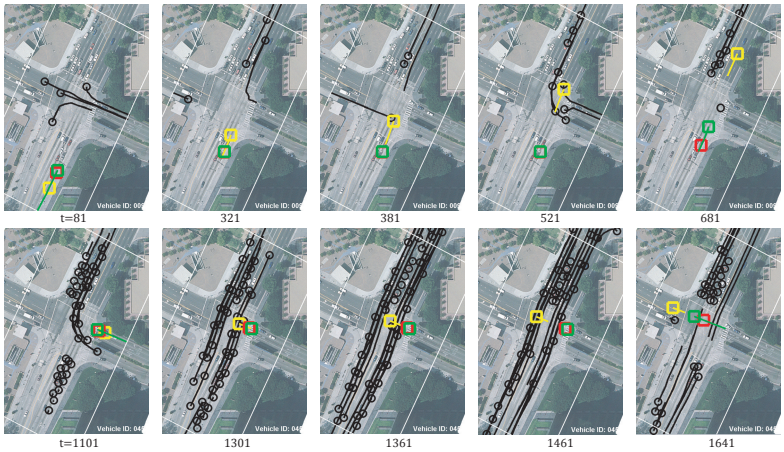

Fig. 3: Each row is an example of trajectory reconstruction. Vehicle under consideration is shown with squares, yellow depicts constant velocity, red is from proposed method and green square marks the ground truth. Rest of the vehicles are shown in black. In first row, reconstruction with constant velocity causes collisions at $t = 381$ and 521, and in the second row, between $t = 1200$ and 1500. On the other hand, proposed method and ground truth allow the vehicles to pass without any collision.

constant velocity case corresponds to $\phi = \varphi = \psi = 0$. Thus, we can model all cases between constant velocity to complete stopping by varying values of these variables. However, since the exact duration of invisibility and end-point velocities are known, we have only two-degrees of freedom making one of the variables dependent on the other two. We choose ϕ and φ to represent the trajectory, while ψ is determined based on time and velocity constraints. Thus, the parametrization results in constant or zero acceleration and deceleration while satisfying entry and exit velocities.

3.6 Optimization for Motion Inference

Considering each vehicle individually, given the time it enters and exits the region, the best estimate for its motion is constant velocity. However, since an invisible region (in our case, an intersection) may involve vehicles traveling in from all directions, each vehicle influences the motion of other vehicles. A constant velocity prediction for each vehicle will result in collisions even though it may satisfy the constraints of vehicle-following and smoothness.

Given entry and exit triplets of position, velocity and time for each vehicle, our goal is to find the parameters ϕ and φ for all vehicles that pass though the invisible region. Since each vehicle's position at every time instant is conditionally dependent on all other vehicles that also pass through the invisible region, this information is exploited in the form of four constraints. The proposed solution iteratively minimizes the local cost of each vehicle by making sure it satisfies the constraints, fixing its parameters and then moving onto other vehicles.

Algorithm 1 Algorithm to infer motion of vehicles given $[p^{\eta}, v^{\eta}, \eta]$ and $[p^{\chi}, v^{\chi}, \chi]$ for vehicles $1:n$

procedure INVISIBLEINFERENCE
  Prioritize vehicles using Eq. 6
  for all $i \leftarrow 1, n$ as per Θ_i **do**
    Identify $j|\mathcal{O}(i,j) = 1$ and $i \perp j$
    $\phi_i, \varphi_i \leftarrow \underset{\phi,\varphi}{\operatorname{argmin}}\ C_i^{\alpha} + C_i^{\beta} + C_i^{\gamma} + C_i^{\delta}$
    Parameterize trajectory i according to ϕ_i, φ_i
  end for
end procedure

We impose a prioritizing function on the trajectories, which is a linear function of entry and exit times:

$$\Theta_i = \omega^{\chi}\chi_i + \omega^{\eta}\eta_i. \tag{6}$$

If we set $\omega^{\chi} = 1$ and $\omega^{\eta} = -1$, the criteria becomes *shortest duration first*, on the other hand, if $\omega^{\chi} = 1$ and $\omega^{\eta} = 0$, the criteria becomes *earlier exit first*. The former biases the solution towards high priority vehicles putting very strong constraints on vehicles that spend longer times in the invisible region. We used the latter which makes more intuitive sense also since vehicles will yield way to the a vehicle that exits before them.

The cost for each vehicle is the sum of costs due to collision, vehicle-following, and smoothness including penalty for trivial solution. The parameters are bounded so that -20 ft/s^2 $< \phi < 0$, and $0 < \varphi < \chi - \eta$, and the cost is minimized through an Interior Point Algorithm with initialization provided by uniform grid search over the parameter space. The summary for this simple algorithm is provided in Alg. 1 while Fig. 3 shows the results on real examples.

4 Inference of Scene Structure

After obtaining trajectories using the method and constraints described in the previous section, we now propose methods to statistically represent motion in the invisible region (§4.1), followed by extracting some key features of the scene such as locations of stopping points (static) and timings of traffic signals (dynamic) in §4.2.

4.1 Statistical Representation of Motion

Given the inferred trajectory representations for objects in invisible region, we compute the features, $(x_i^t, y_i^t, u_i^t, v_i^t, t)$, for each point on the i^{th} trajectory derived from $\mathcal{P}_i$, ϕ_i, φ_i, and ψ_i, where $u^t = x^t - x^{t-1}$, and $v^t = y^t - y^{t-1}$. To obtain a probabilistic distribution that represent individual behaviors, such as turns, constant velocity motion, deceleration to a stop, and acceleration from rest, we first cluster feature points using the

k-means algorithm, and then treat the clusters as components of a 4d Gaussian mixture model, B for each leg of traffic. A feature point $\mathbf{x}$ induced by B is given as,

$$\mathbf{x} \sim \sum_{k=1}^{N_B} \omega_k \mathcal{N}\left(\cdot \middle| \mu_k, \mathbf{\Sigma}_k\right), \tag{7}$$

where the mixture contains N_B Gaussian components, each with parameters μ, $\mathbf{\Sigma}$, and mixing proportion ω. This representation serves as a generative model that summarizes the spatiotemporal dynamics of objects induced by each behavior, and is potentially useful for scene understanding, anomaly detection, tracking and data association, as well as visualization of results of the proposed inference framework. For visualization, we compute the per-pixel expected motion vectors conditioned on the pixel location along each leg, i.e., $\mathrm{E}_B\left[\sqrt{u^2+v^2}, \tan^{-1}(v/u)|x,y\right]$, to obtain expected magnitude and orientation of motion at each pixel for each behavior, and depict them using the HSV colormap as is done for motion patterns [25].

4.2 Scene Structure & Status Inference

Given the exact motion and behavior of objects in the invisible regions, we propose to estimate some key aspects of the scene structure and status to, show the importance and usefulness of our framework, and allow evaluation. We briefly explain our methods for finding the locations where vehicles exhibit the stopping behavior (equivalent to stopping positions), and the times at which such behaviors occur (equivalent to status of traffic signals) for each path in the invisible region.

We use the locations, $p_i^{t_\varphi}$, for $i|v_i < T$, to vote for regions corresponding to stop positions. Specifically, given n vehicles, we compute the following kernel density estimate of the 2d surface, $\mathit{\Gamma}$, representing probability of a pixel, p, belonging to stopping location:

$$\mathit{\Gamma}(\mathrm{p}) = \frac{1}{n}|\mathbf{H}|^{-\frac{1}{2}} \sum_{i=1}^{n} K\left(\mathbf{H}^{-\frac{1}{2}}\left(\mathrm{p} - p_i^{t_\varphi}\right)\right), \tag{8}$$

where K is a 2d Gaussian kernel with a symmetric, positive, diagonal bandwidth matrix, $\mathbf{H}$, of fixed horizontal and vertical variances, set to 10 pixels. Fig. 8(a,b) shows an example of the distribution reflecting probabilities of pixels being stopping positions. The proposed framework therefore estimates salient scene structure in a statistical manner, without making a single observation within the scene.

Secondly, given the representation of behaviors learned earlier which divides the scene into possibly overlapping segments corresponding to traffic intersection legs, by thresholding $\int\int \mathrm{Pr}_B dudv$, we can effectively estimate the signal status for each leg. We use the following simple process: at a given time t, the inferred status (red, green) of a traffic signal for a leg, l, is $\sum_i \|p_i^{t+1} - p_i^t\|$, if i belongs to l and $t > t_\varphi$. Therefore, if any vehicle traveling on leg l has a non-negligible velocity at its stopping positions, it votes for the green signal for that leg at that time. The results of signal status and transitions for all legs of traffic (blue), compared to the results obtained by applying the same process to ground truth trajectories (black), are shown in Fig.8(c).

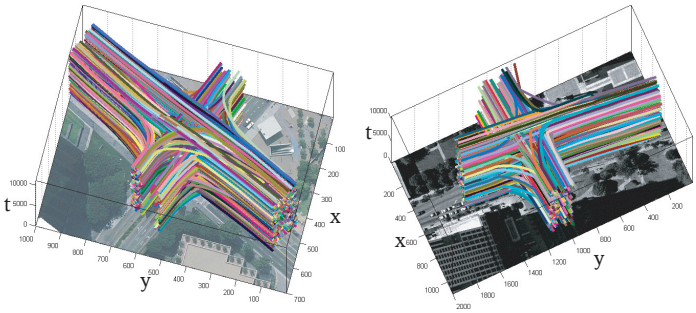

Fig. 4: All trajectories inferred for each dataset shown in 3D. Left and right images are inferred trajectories from Lankershim and Peachtree datasets respectively.

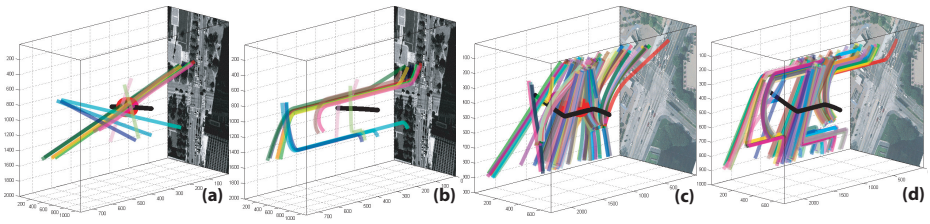

Fig. 5: Trajectories in (a) and (c) represent constant velocity while (b) and (d) show output of proposed method. Collisions due to constant velocity prediction are marked with red spheres in (a) and (c), but this does not occur in (b) and (d), which are the results of proposed trajectory inference.

5 Experiments

We ran our experiments on two datasets from NGSIM (see [26] for details). The first invisible region was from Lankershim 8:30am - 8:45am located at the intersection of Lankershim/Universal Hollywood Dr. (LA) with a total of 1211 vehicles passing through the region. The second invisible region was from Peachtree 4:00pm to 4:15pm located at the intersection of Peachtree/10th Street NE (Atlanta) with 657 vehicles passing through the region. Both intersections were typical four-legged with three possible paths that could be taken by a vehicle entering a particular leg, thus, resulting in 12 total paths. Fig. 4 shows the trajectories that were output by Alg. 1 for both the datasets.

We next analyze the performance of motion inference employing the different constraints, followed by results for motion behaviors and scene structure. Figure 5 provides qualitative results for motion inference where the (a,b) is from Peachtree and (c,d) from Lankershim dataset. The black trajectory corresponds to the vehicle under consideration while proximal vehicles which it could possibly collide with are shown in colors. In both (a) and (c), the trajectories are drawn assuming constant velocity for each vehicle. In (a), the vehicle collides with one of the vehicles whereas in (c), vehicle under consideration collides with six different vehicles. The locations of collision are shown with red spheres partially invisible due to other vehicles. Notice the change in shape in

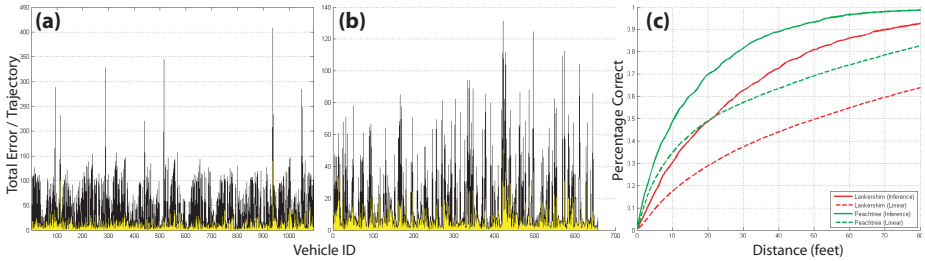

Fig. 6: (a,b) Error profile for our method (yellow) vs. constant velocity (black) for both datasets. As can be seen, our method has lower error (it has smaller magnitude), thus provides more accurate inference. (c) ROC curves for our method (solid) vs. constant velocity (dashed) for the Lankershim (red) and Peachtree (green). The x-axis is the distance threshold in feet while y-axis gives the percentage of points that lie within that threshold distance of the ground-truth.

(b) and (d) after inferring motion for all trajectories with the outcome that none of the trajectories collides with the black trajectory. Both vehicle-following and smoothness constraints are also visibly in effect in both the examples.

Fig. 6(a,b) gives a per-trajectory comparison of error with and without motion inference. These graphs for Lankershim and Peachtree respectively were obtained by computing total error (in feet) for each trajectory by computing Euclidean distance of each point to the groundtruth. The yellow bars correspond to motion inference whereas black bars represent the case of using constant velocity only. Fig. 6(c) gives the ROC curves for the two datasets. On the x-axis is the threshold distance in feet, on y-axis are the percentage of points in all invisible trajectories that lie within that threshold. Using inference, we get an improvement of at least 20% over the baseline in both datasets. After obtaining the inferred trajectories, we statistically represented the motion in the invisible region using the method described in §4.1. Fig. 7 shows MoG for three different legs where three columns represent constant velocity, proposed method and ground truth.

Figures 8 give results for some of the salient features of the invisible region using §4.2. Fig. 8(a) shows the probability map superimposed on the image of invisible region for locations where vehicles stop using only the inferred trajectories from Lankershim whereas Fig. 8(b) shows the same probability map for Peachtree. It can be seen that all of the locations are correct, just before the intersection due to collision avoidance and extend beyond due to vehicle following constraint when vehicles queue up at intersection. Figure 8(c) gives the probability of which traffic light was green at each time instant using the proposed method (blue) and the results are also compared against groundtruth (black). In this figure, we show traffic light behavior over time for 8 of the 12 paths as right turns do not get subjected to signals. Below each blue graph which is obtained using inferred trajectories, is the black graph showing probability of that light being green using groundtruth. The results show little difference, validating the performance and quality of inference.

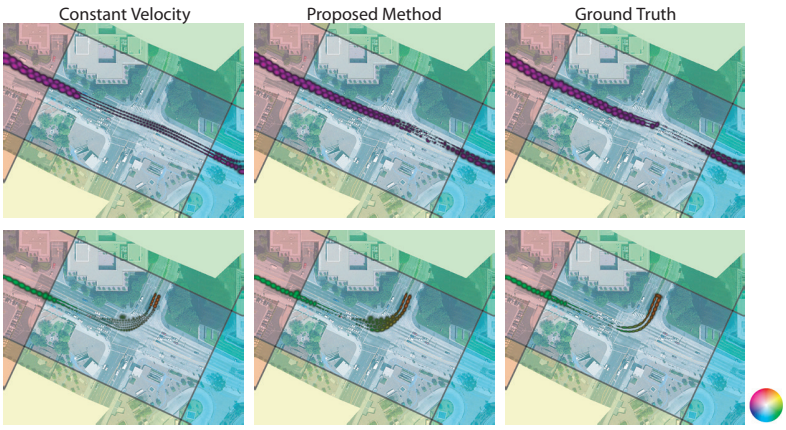

Fig. 7: Each row is the Mixture of Gaussians representation for a particular path using constant velocity, proposed method and ground truth. The patterns in the second and third column are similar and capture acceleration, deceleration, start and stop behaviors whereas in first column, all Gaussians have the same variance due to constant velocity.

6 Conclusion and Future Work

We presented the novel idea of understanding motion behavior of objects while they are in the invisible region of multiple proximal cameras. The solution used three constraints, two of which employ contextual information of neighboring objects to infer correct motion of object under consideration. Though, an interesting proposal from the perspective of scene and motion understanding, the idea has several potential applications in video surveillance. Possible extensions include handling situations where correspondences are missing or incorrect in some cases and to humans where social force models can be leveraged in addition to current constraints.

References

1. Stauffer, C., Tieu, K.: Automated multi-camera planar tracking correspondence modeling. In: CVPR. (2003)
2. Khan, S., Shah, M.: Consistent labeling of tracked objects in multiple cameras with overlapping fields of view. IEEE PAMI **25** (2003)
3. Makris, D., Ellis, T., Black, J.: Bridging the gaps between cameras. In: CVPR. (2004)
4. Gheissari, N., Sebastian, T., Hartley, R.: Person reidentification using spatiotemporal appearance. In: CVPR. Volume 2. (2006)
5. Hu, W., Hu, M., Zhou, X., Tan, T., Lou, J., Maybank, S.: Principal axis-based correspondence between multiple cameras for people tracking. IEEE PAMI **28**(4) (2006)
6. Gray, D., Tao, H.: Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: ECCV. (2008)
7. Loy, C.C., Xiang, T., Gong, S.: Multi-camera activity correlation analysis. In: CVPR. (2009)
8. Pellegrini, S., Ess, A., Schindler, K., van Gool, L.: You'll never walk alone: Modeling social behavior for multi-target tracking. In: ICCV. (2009)

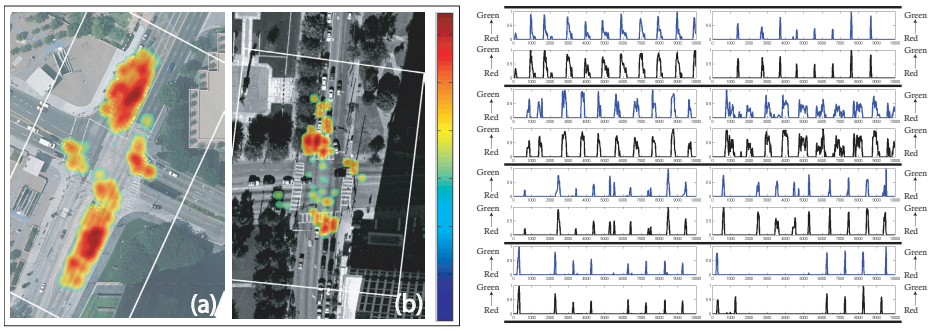

Fig. 8: Left (a,b): The probability map for stopping positions as inferred from Eq. 8 for both datasets which are correct as vehicles in reality stop and queue before the signal. Right: Probability of green signal for each of eight possible legs. The x-axis is time and y-axis in each graph is the *probability* from our method (blue) and groundtruth (black), which are evidently, perfectly aligned in time.

9. Kikuchi, S., Chakroborty, P.: Car following model based on fuzzy inference system. Transport. Res. Record (1992)
10. Nagel, K., Schreckenberg, M.: A cellular automaton model for freeway traffic. J. Phys. I France **2**(12) (1992)
11. Newell, G.: Nonlinear effects in the dynamics of car following. Ops. Res. **9**(2) (1961)
12. Javed, O., Shafique, K., Rasheed, Z., Shah, M.: "Modeling inter-camera space-time and appearance relationships for tracking across non-overlapping views". CVIU **109** (2008)
13. Huang, T., Russell, S.: "Object identification in a bayesian context". In: IJCAI,. (1997)
14. Kettnaker, V., Zabih, R.: "Bayesian multi-camera surveillance". In: CVPR. (1999)
15. Prosser, B., Gong, S., Xiang, T.: Multi-camera matching using bi-directional cumulative brightness transfer functions. In: BMVC. (2008)
16. Dockstader, S., Tekalp, A.: Multiple camera fusion for multi-object tracking. In: IEEE WMOT. (2001)
17. Soto, C., Song, B., Roy-Chowdhury, A.: Distributed multi-target tracking in a self-configuring camera network. In: CVPR. (2009)
18. Song, B., Kamal, A., Soto, C., Ding, C., Farrell, J., Roy-Chowdhury, A.: Tracking and activity recognition through consensus in distributed camera networks. In: IEEE TIP. (2010)
19. Kamal, A., Ding, C., Song, B., Farrell, J.A., Roy-Chowdhury, A.: A generalized kalman consensus filter for wide area video networks. In: IEEE CDC. (2011)
20. Tieu, K., Dalley, G., Grimson, W.: Inference of non-overlapping camera network topology by measuring statistical dependence. In: ICCV. (2005)
21. Stauffer, C.: "Learning to track objects through unobserved regions". In: WMVC. (2005)
22. Atev, S., Arumugam, H., Masoud, O., Janardan, R., Papanikolopoulos, N.: A vision-based approach to collision prediction at traffic intersections. IEEE TIT Systems **6**(4) (2005)
23. van den Berg, J., Guy, S., Lin, M., Manocha, D.: Reciprocal *n*-body collision avoidance. **70** (2011) 3–19
24. LaValle, S.M.: Planning Algorithms. Cambridge University Press (2006)
25. Saleemi, I., Hartung, L., Shah, M.: Scene understanding by statistical modeling of motion patterns. In: CVPR. (2010)
26. Next Generation Simulation (NGSIM) dataset. Available at http://www.ngsim.fhwa.dot.gov/.