Implementing Kinematic Wave Theory to Reconstruct Vehicle Trajectories from Fixed and Probe Sensor Data

Babak Mehrana,*, Masao Kuwaharaa, Farhana Naznin
aInstitute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505 Japan

Abstract

A data fusion framework is examined to reproduce vehicle trajectories on urban arterials by combining probe and fixed sensor data, and signal timing parameters. The methodology is based on the kinematic wave theory and employs the variational theory for the solution method. However, the original methodology cannot deal with the vehicles coming in and out in the middle of the study section despite the frequent existence of such vehicles in the real world. Therefore, the methodology is extended so as to incorporate the vehicles coming in and out. The proposed method is then applied to real world data and its robustness is confirmed by changing the input data characteristics.

Keywords: Data fusion, Variational theory, Kinematic wave theory, Probe data, Vehicle trajectory, Travel time

1. Introduction

Travel time is the most intuitive measure of traffic conditions on urban arterials. The efficiency of congestion management strategies can be evaluated directly through observation of the travel times. Travel time on urban streets can be measured directly using probe sensors (e.g. probe vehicles) or can be estimated indirectly from fixed sensor data (e.g. loop detectors). Probe vehicles provide spatial traffic information and direct measurements of travel time. However, their frequency is restricted and sometimes there are GPS reception errors included in the data. On the other hand, fixed sensors record traffic data continuously but only at fixed locations. Considering different accuracies and limitations of traffic data from various sources, data fusion techniques are applied to extend the spatial and temporal coverage of data in order to provide continuous and reliable travel time information.

Existing data fusion techniques for travel time estimation can be summarized in three categories (EL Faouzi, 2004): i) Statistical based, ii) Probabilistic based and, iii) Artificial cognition based. Among statistical techniques, a weighted combination of travel times from different sources is the most common approach. Weights are generally derived from variance-covariance estimation errors by applying methods such as the “voting technique” (Choi & Chung, 2002). Probabilistic approaches such as Bayesian approach (EL Faouzi, 2006) and Dempster-Shafer...
inference (EL Faouzi et al., 2009) have been used to tackle the problem of data fusion for the purpose of travel time estimation. El Faouzi (2006) describes application of the Bayesian approach to combine travel times estimated from conventional loop detectors and probe vehicles on urban routes. A mathematical basis for applying the Dempster-Shafer inference to improve travel time estimation by fusing estimated travel times from toll collection stations and conventional loop detectors is demonstrated by El Faouzi et al. (2009). Ivan et al. (1995) applied artificial neural networks (ANN) to develop arterial incident detection models that fuse probe and fixed detector data.

The majority of existing fusion techniques to estimate travel time are based on statistical methods without considering traffic engineering concepts. However, it is arguably more advantageous to combine multi-sensor traffic data according to traffic engineering principals prior to using statistical models. Such a procedure allows effective evaluation of shockwaves, queue propagation, and speed variations, and therefore culminates in the identification of the traffic state in time-space.

Berkow et al. (2009) proposed a fusion method to combine the data derived from loop detectors and probe buses to improve travel time estimations on urban routes. They developed an algorithm to identify congested time periods using detector data which were used to reproduce bus trajectories. However, they did not consider the delay at signalized intersections within their study area. Coifman (2002) presented a simple method for estimating vehicle trajectories on freeway links using traffic data from an individual loop detector. Coifman’s method employs the basic traffic flow theory to extrapolate local traffic conditions to an extended freeway link. Yet, estimated trajectories are not accurate during transition periods when the traffic condition changes from uncongested to congested, and vice versa. In addition, the methodology is not applicable to signalized highways.

Another flaw in traditional data fusion methods is the way probe information is implemented. Probe data contain much richer information of vehicle trajectories which represent traffic conditions in time-space. Nevertheless, probe trajectories are used only to extract travel times. Probe vehicles are sampled from the vehicle population and their travel time might not be the typical travel time for the whole population. Moreover, travel times estimated from probe vehicles (e.g. taxis or buses) might vary according to the vehicle types, though the location and the waiting times of the probe vehicles stopping in a queue do not depend on the vehicle types. Such information could be combined more effectively with fixed sensor data to reproduce the trajectories of all vehicles.

The objective of this research is to examine a data fusion framework to reconstruct vehicle trajectories on signalized urban arterials by combining traffic data from fixed and probe sensors according to traffic engineering principles. The methodology is based on the kinematic wave theory and implements the solution proposed by Daganzo (2005a; 2005b) based on variational theory. The original solution by Daganzo cannot deal with vehicles coming in and out in the middle of the study section despite the frequent existence of such vehicles in the real world. Therefore, the methodology is extended further to incorporate the vehicles coming in and out. The proposed methodology is then applied to real world data and its robustness is confirmed by changing input data characteristics such as the frequency of probe data and the aggregation interval of fixed sensor data. Once vehicle trajectories are estimated, they can be used for several purposes, including travel time estimation and prediction, signal coordination and emission monitoring.

2. Methodology

The methodology to estimate vehicle trajectories is based on the kinematic wave theory originally developed by Lighthill, Whitham and Richards (LWR) in the 1950s (Lighthill & Whitham, 1955; Richards, 1956). They described a theory of one-dimensional wave motion which could be applied to certain types of fluid motion or to highway traffic flow.

The key feature of the LWR theory was that there is some functional relation between the flow $q$ and the density $k$ which might vary with location $x$ and time $t$. Assuming no entering or exiting traffic, the conservation equation implies:

$$\frac{\partial k(x,t)}{\partial t} + \frac{\partial q(x,t)}{\partial x} = 0$$  \hspace{1cm} (1)

Equation (1) can be written as:
\[
\frac{1}{w} \frac{\partial q(x,t)}{\partial t} + \frac{\partial q(x,t)}{\partial x} = 0
\]  

(2)

where \( w(q, x) = \frac{\partial q}{\partial k} \) is called the “wave speed”.

Newell (1993a, 1993b, 1993c) combined the concept of cumulative curves with LWR theory and extended it to the 3D kinematic wave theory. Given \( N(x, t) \) as the cumulative number of vehicles at location \( x \) by time \( t \), Newell suggested evaluating \( N(x, t) \) rather than \( q(x, t) \). Assuming a one dimensional traffic movement where the first in-first out (FIFO) condition holds, variations in the cumulative number of vehicles over space can be addressed through (3).

\[
dN = (\partial N / \partial x)dx + (\partial N / \partial t)dt = -kdx + qdt = (-k + q/w)dx
\]  

(3)

With a piece-wise linear fundamental diagram, \( dN \) takes only two values: 0 for the forward wave and \( k_j dx \) for the backward wave, where \( k_j \) is the jam density.

The 3D kinematic wave theory has been used for several applications such as dynamic network assignment (Kuwahara & Akamatsu, 2001), traffic simulation (Daganzo, 1994), and so on. Recently Daganzo (2005a, 2005b) proposed an efficient calculation method based on variational theory of the 3D kinematic wave theory to evaluate forward and backward waves within the traffic flow. We employed Daganzo’s method to estimate vehicle trajectories and validated the proposed methodology using real world data.

Given \( N(x_0, t_0) \) and \( q(x_0, t_0) \) at some boundary point, variational theory provides a solution to estimate \( N(x, t) \) by treating the problem as a capacity constrained optimization problem. Flow at any point in time-space is bounded from above by \( q_{\text{max}} \), the capacity. A similar capacity constraint should also hold if the road is viewed from a rigid frame of reference that moves with speed \( v \). In this case the capacity relative to the frame (the “relative capacity”) is the maximum rate at which traffic can pass an observer attached to the frame. An observer that moves with speed \( v \) next to a traffic stream with density \( k \) and flow \( q \) is passed by traffic at a rate \( q-vk \). In the simplified variational theory, the fundamental diagram is triangular, and the “relative capacity” function \( R \) (the “cost function”) is linear. In this case the maximum cost \( (r) \) is:

\[
r = R(-w) = (1 + w/u)q_{\text{max}}
\]  

(4)

where \( u \) and \( w \) are forward wave speed, and backward wave speed, respectively.

In variational theory the problem is solved approximately by overlaying a dense but discrete, network with short straight valid paths as links, and the following two properties: i) the slopes of links branching from each node which represent the range of wave speeds with sufficient resolution; and ii) link costs (allowable change in the cumulative number of vehicles along each link).

Daganzo (2006) showed that the solution of a variational theory problem with bottlenecks exists for well-posed problems. Furthermore, the requirements for well-posedness were proven to be mild and met by practical applications including multi-lane streams and networks with a combination of fixed and moving bottlenecks (Daganzo, 2006).

The solution domain in time-space can be modeled with a lopsided-network in which the mesh resembles triangular fundamental diagrams with identical steps in the time-space plane. The nodes are on a rectangular lattice with space separation \( ststep \) and time separation \( tstep \). There are sets of links pointing to any node with slopes \( u \) (forward wave) and \( -w \) (backward wave). Each node in the network represents the height of the three dimensional cumulative surface at a specific point.

Coordinate system \( F^* \) is defined with \( i \)-coordinate aligned with the backward wave, and \( j \)-coordinate aligned with the forward wave. The cumulative number of the vehicles at node \((i, j)\), is presented by \( N(i, j) \).

With the piece-wise linear fundamental diagram, the link cost along the forward wave is equal to zero, which means there is no change in the cumulative number of the vehicles along the forward wave. On the other hand, the link cost along the backward wave, which is the maximum allowable change in the cumulative number of the vehicles, is estimated from \( k_j \) times \( ststep \). If there are any red intervals in the solution domain, the cost along the links corresponding to the red intervals becomes equal to zero. In this case, red intervals create shortcuts in the network. In other words, during the red intervals, it would be possible to move from node \((i-1, j-1)\) to node \((i, j)\).
without incurring any cost. Since the positions of the signalized intersections are known, considering the signal timing data, those links corresponding to the red intervals can be distinguished easily.

### 3. Calculation steps

Figure 1 shows the necessary steps to estimate vehicle trajectories. The process starts with defining a discrete network to address time-space plane. Considering a triangular fundamental diagram, given the forward wave speed $u$, jam density $k_j$ and the maximum flow rate $q_{max}$, the backward wave speed $w$ is estimated from (5). The horizontal distance between the nodes of the same row, $t_{step}$, is an input variable. Considering computational straightforwardness, a time step of 1 second is recommended. The vertical distance between network nodes, $s_{step}$, is estimated from (6).

$$ w = \frac{q_{max}}{k_j - \frac{q_{max}}{u}} $$ (5)

$$ s_{step} = \frac{u \times w \times t_{step}}{u + w} $$ (6)

Each node in the network represents the height of the cumulative surface. To set the initial conditions on the network boundaries, the height along the first column in the network is assumed to be 1. Considering the passing times of the vehicles recorded by detectors, cumulative traffic counts at the upstream and the downstream are assigned as the heights to the nodes along the lower and upper boundaries of the network.

When probe trajectories are used as a reference to reproduce other trajectories, some additional treatments are necessary. A probe trajectory can be interpreted as a contour on a three dimensional surface of cumulative curves. Therefore a constant height should be assigned to the network nodes in the vicinity of the probe trajectory. Since the cumulative vehicle counts at the upstream are known, a constant height can be estimated once the intersection point of the probe trajectory with the lower boundary of the network is known.

Finally, optimization is performed to find the height of each node in the network. According to the variational theory (Daganzo, 2005a; Daganzo, 2005b), calculation of the cumulative height for each of the network nodes is reduced to the shortest path calculation with the link costs explained above and considering the cumulative heights at the network boundaries. In general, the height of node $(i, j)$ that is $N (i, j)$ is estimated from (7). However, if node $(i, j)$ is located on a link which represents a red interval in time-space, considering the shortcut effect of the red intervals, $N (i, j)$ is estimated from (8).
Once the cumulative heights are calculated, a trajectory of a vehicle is obtained as a contour line of the cumulative heights.

4. Study area

As shown in Figure 2, a 1.15km stretch of Komazawa Street in Tokyo, Japan, was considered for the purpose of this study. The selected section is a single–lane facility and includes six signalized intersections. The analysis period was chosen to be part of a morning peak from 8:15 to 9:15 on 1 September 2006.

A pair of AVI cameras was installed at the entrance and exit points of the study area to record the number plate of the vehicles entering or leaving the section. Number plates were matched afterwards in order to estimate travel time during the analysis period. In addition to travel time, AVI data provided passing times of individual vehicles at the upstream and the downstream of the study area. Probe data were collected using GPS equipped probe taxis on the study area. Probe data included the time and the travelled distance (position) at 1–second intervals. Required AVI and probe data were provided from the COSE (Consortium for Software Engineering) project.

Traffic signal data including phasing, cycle length and signal timings were provided by the Tokyo Metropolitan Police Department for all signalized intersections except for the third signalized intersection as indicated in Figure 2, where such data were not available. The locations of the intersections, signal timing patterns and probe trajectories are shown in Figure 2.

\[
N(i, j) = \text{Min} \left\{ N(i, j-1), N(i-1, j) + k \cdot sstep \right\}
\]

\[
N(i, j) = \text{Min} \left\{ N(i, j-1), N(i-1, j-1), N(i-1, j) + k \cdot sstep \right\}
\]
At this stage it is assumed that there are no vehicles entering or leaving the study area from midblock intersections. Required parameters to estimate vehicle trajectories are summarized in Table 1. The forward wave speed was estimated from probe data during the free-flow conditions. The saturation flow rate was estimated using AVI data during the green intervals and an assumed value was considered for the jam density on the study section.

5. Qualitative analysis

Figure 3 shows estimated trajectories for the entire study area without using any of the probe trajectories as reference. Although shock waves and queues are emerging, estimated trajectories do not match with corresponding probe trajectories, particularly during the congested periods. As an example, the estimated trajectory corresponding to the 11th probe trajectory is highlighted in Figure 3 according to the entry times. While the 11th probe trajectory shows considerable shockwaves due to the congested traffic conditions, its corresponding estimated trajectory represents traveling at free-flow conditions between intersections. When probe trajectories are not used as reference, the study area is considered as an empty highway segment while setting the initial conditions at the network boundaries. As a result, the impact of congested traffic conditions and the corresponding backward waves from the earlier time intervals are not reflected in the estimated trajectories.

Figure 4 represents estimated trajectories when the 10th probe trajectory is used as a reference. When probe data are used as reference, even though the signal timings of the 3rd intersection is not considered, the agreement between the estimated trajectory and the corresponding probe trajectory improves significantly. The impact of the backward

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward wave speed (km/h)</td>
<td>32.4</td>
</tr>
<tr>
<td>Saturation flow rate (veh/h)</td>
<td>1750</td>
</tr>
<tr>
<td>Jam density (Veh/km)</td>
<td>165</td>
</tr>
</tbody>
</table>
waves from the earlier time intervals is conspicuous in Figure 4.

The agreement between the probe trajectories and their corresponding estimated trajectories is evaluated in Figure 5 by comparing the accordance between the trajectory paths for each set of the probe and its corresponding estimated trajectory. Beginning from the first probe trajectory, pairs of adjacent probe trajectories are considered. For each pair, the first probe trajectory is used as a reference and the trajectories are estimated up to the second probe trajectory. Then the second probe trajectory is compared with its corresponding estimated trajectory. The agreement is evaluated as a function of the difference between the entry times of the reference probe trajectory and
the estimated trajectory corresponding to the subsequent probe trajectory. Suppose the solution network consists of \( n \) rows in time-space. For each pair of the probe and its estimated trajectory, \( T_x(y) \) represents the measured time at space \( y \) for a probe vehicle leaving the upstream boundary at time \( x \) while \( t_x(y) \) refers to the estimated time at space \( y \) for a vehicle leaving the upstream boundary at the same time \( x \). For each set of probe and its corresponding trajectory which leave the upstream at time \( x \), equation (9) is used to estimate the Mean Absolute Error (\( MAE_x \)). Afterwards, the relative error (\( RE_x \)) for each set of the probe and the estimated trajectories is calculated from (10).

\[
MAE_x = \frac{1}{n} \sum_{y=1}^{n} |T_x(y) - t_x(y)|
\]

\[
RE_x = \frac{MAE_x}{TT_x} \times 100
\]

where \( TT_x \) represents the total travel time from the upstream to the downstream, measured from the probe trajectory leaving the upstream boundary at time \( x \).

According to Figure 5, generally when the difference between the entry time of the reference probe trajectory and the estimated trajectory increase, the relative error increases. The results imply that the estimated trajectories are more accurate over shorter intervals. Those trajectories which are closer to a reference probe trajectory are strongly influenced by the shape of the reference trajectory. However, the shape of the farther trajectories is mostly governed by the solution of network characteristics. Since the proposed methodology is deterministic, stochastic aspects of the traffic flow and the driving behavior factors are not considered. In addition, the impact of incoming/outgoing vehicles from midblock intersections are not considered which result in some discrepancy between the probes and the corresponding estimated trajectories.

Performance of the proposed method regarding travel time estimation is evaluated in Figure 6. It compares estimated travel times derived from the estimated trajectories with the corresponding measured travel times from AVI data. Figure 6a) shows the result without using probe trajectories as reference, while Figure 6b) represents travel time comparison when all probe trajectories are used as reference. There were 68 matched AVI records available for travel time measurement during the analysis period. AVI records were matched with estimated trajectories based on the entry times at the upstream. While the proposed method underestimates the travel times without using probe trajectories as reference, when probe trajectories are used as reference, estimated travel times significantly improve and the Root Mean Square Error (\( RMSE \)) becomes smaller. When probe trajectories are not used as reference, the impact of congested traffic conditions from earlier times cannot be considered properly. As a
result, travel times are shorter and less variable. It should be noted that in the absence of incoming/outgoing vehicles, the travel time from the upstream to the downstream derived from estimated trajectories depends only on the ordered correspondence of the vehicles at the upstream and the downstream boundaries, but are independent of the solution of network characteristics. Assuming there are no incoming/outgoing vehicles from midblock intersections, the ordered correspondence of the vehicles at the upstream and the downstream can be determined by each probe trajectory, since the cumulative number of the vehicles along each probe trajectory is supposed to be a constant value in the solution network.

Discrepancies between estimated trajectories and probe trajectories as well as travel time estimation errors are partly due to the fact that vehicles coming in or going out from midblock intersections are not considered. As can be seen in Figure 4, neglecting incoming vehicles may result in unrealistic empty areas between the estimated trajectories. On the other hand, disregarding outgoing vehicles may result in unrealistic backward waves in the congested regions. In addition to incoming/outgoing vehicles, the shape of the fundamental diagram and stochastic characteristics of traffic flow which are not yet considered contribute in the discrepancy between estimated and measured trajectories.

6. Sensitivity analysis

6.1. Sampling frequency of probe data

The sampling frequency of probe data used in this research is 1 second. However, in practice the probe data are usually collected at longer intervals (e.g. every 2 minutes). When sampling frequency of the probe data decreases, the shape of the resulting probe trajectory changes from a continuous curve to a piece-wise linear curve. Variations in the shape of a reference probe trajectory due to different sampling frequencies may affect the shape of the estimated trajectories as well.

Figure 7 demonstrates the estimated trajectories using the 10th probe trajectory as a reference when probe data are assumed to be sampled only every 5 minutes. The reference probe trajectory in this case consists of several straight lines which connect the representative points in the time-space plane. The shape of the estimated trajectory which corresponds to the 11th probe trajectory is clearly affected by the shape of the reference probe trajectory.

Figure 8 shows the impact of different sampling frequencies of the probe data on the estimated trajectories when the 10th probe trajectory is used as a reference. The sampling frequencies of 1 sec, 2 minutes and 5 minutes are considered for the 10th probe trajectory data. An extreme case in which the 10th probe trajectory is presented by only a direct line connecting the entry and the exit points at the upstream and the downstream is also considered. For each scenario, vehicle trajectories are estimated considering different shapes of the 10th probe trajectory. The agreement between the rest of the probe trajectories and their corresponding estimated trajectory is evaluated and the relative error ($RE_e$) is estimated in each case from (9) and (10).

Figure 8 shows that different sampling frequencies of the probe data only affect the trajectories which are close to the reference probe trajectory where the relative error increases as the sampling frequency decreases. Regardless of the shape of the reference probe trajectory, the relative error is almost constant after the 12th probe trajectory. The results imply that the shape of the reference probe trajectory affects only closer estimated trajectories. The shape of the farther trajectories are mostly determined by the shape of the fundamental diagram, the solution network of characteristics and the signal timings, rather than the shape of the reference probe trajectory.

Estimation of the trajectories using less frequently sampled probe data is one of the future issues in order to make the method robust for practical use. For example, given only the entry and the exit points of a probe vehicle at the upstream and the downstream, one could connect these points considering the free-flow-speed and signal timings to create a reasonable reference trajectory. For practical applications, it is necessary to evaluate the accuracy of the estimated trajectories when such simple treatments are implemented to define the initial conditions.

6.2. Aggregation of AVI data

In this study AVI data are used to identify the entry and the exit times of individual vehicles. To investigate the impact of the different aggregation times on the estimated trajectories, traffic volumes at the upstream and the downstream are aggregated over different intervals of 1 minute, 3 minutes and 5 minutes.
Figure 9 demonstrates the estimated trajectories using the 10th probe trajectory as a reference when traffic volumes are aggregated at 3 minute intervals. Since the passing times of individual vehicles are not used, to avoid biased results, aggregated traffic volumes are distributed only during the green intervals at the nearest intersections from the upstream and the downstream ends.

Using each probe trajectory as a reference, trajectories of the proceeding vehicles were estimated for each scenario of different aggregation intervals. For each case, the estimated trajectory corresponding to the next probe trajectory was considered and its travel time was compared with the measured value from the probe trajectory.
Figure 10 shows the travel time estimation error (RMSE) for different aggregation intervals. Once aggregated traffic volumes are implemented, travel time estimation errors increase as the real entry or exit times of the vehicles are not used at the upstream and the downstream boundaries. However, for different aggregation intervals, estimation error is almost constant.

When aggregated traffic volumes are used to set the initial conditions, since the green intervals are generally shorter than the aggregation intervals, the distribution of the vehicles at the upstream and the downstream boundaries do not change significantly for different aggregation intervals. As a result, estimation errors are stable.
7. Considerations for midblock incoming/outgoing vehicles

The basic assumption in the kinematic wave theory as well as in the variational theory (Daganzo, 2005a; Daganzo, 2005b) is that there are no vehicles entering or leaving the study area from midblock intersections. As discussed in section 5, disregarding such incoming/outgoing vehicles may drastically affect the shape of the estimated trajectories.

As mentioned earlier, the estimated trajectories when a probe trajectory is used as a reference to set the initial conditions at network boundaries may not fully agree with a subsequent probe trajectory. Therefore, it is necessary to modify the estimates so as to improve the agreement between the estimated trajectory and the corresponding probe trajectory. There are several reasons for disagreement. On-street parked vehicles, pedestrians crossing a street and bicycle and motorcycle movements have influence on the shape of the fundamental diagram. On the other hand, the incoming/outgoing vehicles from the midblock intersections as well as those joining or leaving the study section from the shops and minor roads, directly affect the movement of other vehicles and the shape of the estimated trajectories. Unfortunately, we normally cannot clearly figure out the main causes of the disagreement solely from traffic detector information. However, since incoming/outgoing vehicles in the midblock of the study section would substantially affect the motions of other vehicles; let us discuss further the modification of the estimated trajectories by considering incoming/outgoing vehicles.

In this section, first a methodology is proposed to consider incoming/outgoing vehicles by adjusting the cumulative heights in the solution network. The methodology is applicable whenever the locations of the incoming/outgoing vehicles are known in time-space. However, in practice, the traffic data regarding the turning volumes at an intersection are rarely available. Consequently, supplementary discussions are provided to define suitable locations to allocate incoming/outgoing vehicles so as to improve the agreement between the estimated and measured trajectories.

7.1. Influential area by an incoming/outgoing vehicle

To proceed further, coordinate system $F^*$ is implemented in Figure 11 with the origin set at A which divides the time-space plane into four distinct regions. Considering the time-space plane in Figure 11, let us discuss the influential area on the cumulative heights by an incoming vehicle at location A in the midblock of the study section. Suppose that cumulative heights $N(i, j)$ for all $i$ and $j$ have been calculated without considering the incoming/outgoing vehicles in the midblock section. Considering an incoming vehicle at A, since the wave speed is bounded within the range of $[-w, u]$, the followings are true:

(i) The cumulative height at A influences only the cumulative heights in Region I

\[ (11) \]
(ii) The cumulative height at A is influenced only by the cumulative heights in Region III.

From (11) it is concluded that the cumulative height at A due to an incoming vehicle influences the cumulative heights only in Region I; that is, \( N(i, j) \) already calculated are not necessarily changed for \((i, j)\) in Regions II, III and IV.

7.2. Proposed treatment to reconstruct the trajectory of an incoming/outgoing vehicle at \((i, j)\)

7.2.1. Sequential calculations

Therefore, given the cumulative heights in Regions II, III and IV with a vehicle added at A as in Figure 11, the heights along \(i\)-coordinate, \(j\)-coordinate and the thick solid and dashed lines at the upstream and the downstream boundaries can be determined. These heights should be employed as the new initial condition to estimate the cumulative heights in Region I. If a vehicle is added at A, the cumulative heights along \(i\)-coordinate and D-E must be increased by 1. The reason is that vehicles passing \(i\)-coordinate and D-E must be those running after the vehicle was added and hence the cumulative heights along \(i\)-coordinate and D-E must be increased by 1. When an outgoing vehicle leaves the study section at A, the cumulative heights along \(i\)-coordinate and D-E must be decreased by 1 instead.

\[
N(i, j) = N(i, j) + \delta \quad \text{for} \quad (i, j) \quad \text{along} \quad i\text{-coordinate and D-E in Figure 11}
\]

\[
\delta = \begin{cases} 
+1, & \text{if a vehicle is added} \\
-1, & \text{if a vehicle is subtracted}
\end{cases}
\]  

(12)

On the other hand, given an incoming vehicle at A, the heights along \(j\)-coordinate and B-C would remain the same. Since the slope along \(j\)-coordinate is the free-flow speed, all the vehicles passing \(j\)-coordinate must be the vehicles running ahead of the vehicle coming in at A. Because an incoming vehicle at A increases the cumulative heights of the vehicles only after itself, the cumulative heights along \(j\)-coordinate would not change. For boundary B-C, since the vehicle counts at the downstream end must already include the incoming vehicles, the cumulative heights along B-C do not have to be changed, too. When an outgoing vehicle leaves the study section at A, based on above discussion, the cumulative heights along \(j\)-coordinate and B-C remain unchanged as well.

Given the new initial conditions at the boundaries, the cumulative heights for each node in Region I is found by calculating the shortest paths as discussed in section 3.

7.2.2. Simultaneous Calculation

The above procedure is sequentially repeating the setting of the new initial conditions at the boundaries and the calculation of the cumulative heights until the next incoming/outgoing vehicle. However, by changing some of the link costs in advance, we can estimate the cumulative heights at once. If we know there is an incoming vehicle at A, instead of defining new initial conditions at the boundaries, the costs of the links connected with \(i\)-coordinate are increased by 1 and the cumulative heights along D-E should be increased by 1 in advance as shown in Figure 11.

For an outgoing vehicle at A, the link costs along \(i\)-coordinate and D-E are decreased by 1. If we know the locations of all incoming/outgoing vehicles over time-space plane, these modifications of the link costs as well as the heights of the boundary nodes can be done in advance before start calculating the shortest paths. Therefore, with these treatments, we could estimate the cumulative heights of all nodes over the study area at once.

7.3. Determining appropriate locations to consider incoming/outgoing vehicles

Locations and volumes of the turning vehicles are not known often. This section suggests some treatments and provides supporting discussions in order to improve the agreement between estimated and measured trajectories by allocating incoming/outgoing vehicles in appropriate locations.

7.3.1. Required conditions to change an estimated trajectory by adding/subtracting a vehicle

As shown in Figure 12, suppose that the estimated trajectory (the solid curve) and the probe trajectory (the dashed curve) are completely agreed with each other by just before location P (or suppose these two trajectories
have been successfully modified by just before P). By drawing two hypothetical lines with slopes $v$ and $-w$ which intersect at P, four regions are denoted as in Figure 12. Given the two-dimensional time-space plane, there are many possible locations at which vehicles can be added or subtracted in order to modify the estimated cumulative height at P. In fact, according to (ii) in (11), an incoming/outgoing vehicle at any location in Region III may influence the cumulative height at P. Moreover, only the vehicles added/subtracted ahead of those passing P may change the estimated height at P. Therefore shape of the estimated trajectory at P may potentially be modified by adding/subtracting the vehicles only in the shaded region in Figure 12. However, in order to be able to modify the cumulative height at P by adding/subtracting a vehicle in the shaded area in Figure 12, particular conditions must exist between P and the location at which the vehicle addition/subtraction occurs which are discussed hereafter.

Given the estimated trajectory which passes through P in Figure 12; suppose a vehicle is added (or subtracted) in the shaded area in Region III. To discuss further, let us consider the situation as in Figure 13 where the location of the vehicle addition (or subtraction) is denoted as A. Coordinate system $F^*$ is implemented in Figure 13 with the origin set at A (0, 0). Once a vehicle is added (or subtracted) at A, the cumulative height at A is increased (or decreased) by 1. The valid paths from A to P in the solution domain are shown in Figure 13. Based on the discussion in section 7.2, in order to evaluate the change in the cumulative height at P, new initial conditions should be
considered along \(i\)-coordinate and \(j\)-coordinate in Figure 13. Given \(N(i, j)\) as the cumulative height at \((i, j)\), the minimum cost from \((i, j)\) to \((i', j')\) is defined as \(\Delta((i, j), (i', j'))\).

Recalling from section 2, the maximum allowable change in the cumulative height along the backward wave is bounded by the link cost:

\[
\Delta^*(((0,0),(i,0))) \geq N(i,0) - N(0,0) \quad 0 < i \leq i_p
\]  

(13)

On the other hand, the maximum allowable change in the cumulative height along the forward wave is equal to zero:

\[
N(0, j) \leq N(0,0) + \Delta^*((0,0),(0, j)) \quad 0 < j \leq j_p
\]  

(14)

Given the valid paths in the solution domain, the minimum cost from \((0,0)\) to any \((i, j)\) can be estimated either from (15) or (16):

\[
\Delta^*((0,0),(i,j)) = \Delta^*((0,0),(i,0)) + \Delta^*((i,0),(i,j)) \quad 0 < i \leq i_p, \quad 0 < j \leq j_p
\]  

(15)

\[
\Delta^*((0,0),(i,j)) = \Delta^*((0,0),(0,j)) + \Delta^*((0,j),(i,j)) \quad 0 < i \leq i_p, \quad 0 < j \leq j_p
\]  

(16)

By combining (13) with (15) and (14) with (16) respectively, the following equations can be derived:

\[
N(0,0) + \Delta^*((0,0),(i,j)) \geq N(i,0) + \Delta^*((i,0),(i,j)) \quad 0 < i \leq i_p, \quad 0 < j \leq j_p
\]  

(17)

\[
N(0,0) + \Delta^*((0,0),(i,j)) \geq N(0,j) + \Delta^*((0,j),(i,j)) \quad 0 < i \leq i_p, \quad 0 < j \leq j_p
\]  

(18)

As for location \(P\), (17) and (18) become as (19) and (20) respectively.

\[
N(0,0) + \Delta^*((0,0),(i_p,j_p)) \geq N(i_p,0) + \Delta^*((i_p,0),(i_p,j_p)) \quad i_p > 0, \quad j_p > 0
\]  

(19)

\[
N(0,0) + \Delta^*((0,0),(i_p,j_p)) \geq N(0,j_p) + \Delta^*((0,j_p),(i_p,j_p)) \quad i_p > 0, \quad j_p > 0
\]  

(20)

Equations (19) and (20) mean that \((i_p, j_p)\) is determined by the minimum cost along the path either through \(A_i (i_p, 0)\) or from \(A_j (0, j_p)\). Therefore, there are two cases; whether (19) holds or (20).

When RHS of (20) is smaller than that of (19) the following holds:

\[
N(0, j_p) + \Delta^*((0,j_p),(i_p,j_p)) \leq N(i_p,0) + \Delta^*((i_p,0),(i_p,j_p))
\]  

(21)

Since \(\Delta^*((i_p,0),(i_p,j_p))\) is equal to zero, the condition becomes as in (22).

\[
N(i_p,0) - N(0,j_p) \geq \Delta^*((0,j_p),(i_p,j_p))
\]  

(22)

The LHS of (22) is the number of vehicles passing through \(i\)-coordinate and \(j\)-coordinate in Figure 13 while the RHS of (22) is the minimum cost from \((0, j_p)\) to \((i_p, j_p)\) that is equal to the spatial distance between \(A_i (0, j_p)\) and \(P\) multiplied by the jam density.

If the inequality in (22) is not satisfied, the minimum cost along the path through \(A_i (i_p, 0)\) determines the height at \(P\). Since \(\Delta^*((i_p,0),(i_p,j_p))\) is equal to zero, the height of \(P\) is the same as \(N(i_p,0)\). If a vehicle is added at \(A_i (0, 0)\); \(N(i_p,0)\) will also increase by 1 as explained in section 7.2. Consequently \(N(i_p,j_p)\) will also increase by 1 due to the addition at \(A_i (i_p, 0)\). Then, \(N(i_p,j_p)\) is on the contour of the same height \(N(i_p,0)\) before and after the vehicle addition at \(A_i\). An example is provided in Figure 14 where the location of \(P\) does not change after adding a vehicle at \(A_i\).

On the other hand if (22) is satisfied, the minimum cost along the path through \(A_j (0, j_p)\) determines the height of \(P\). In this case even if a vehicle is added at \(A_j (0, j_p)\) does not change as explained in section 7.2 and hence \(N(i_p,j_p)\) would not change as well. On the other hand, \(N(i, 0)\) will increase by 1 due to the vehicle addition at \(A_j\) as discussed in section 7.2.2. Relative to \(P\), \(A\) must be located in Region III (see Figure 12) to influence the cumulative height at \(P\). Also, the vehicle trajectory which passes through \(P\) should cross \(i\)-coordinate at \((i, 0)\)
0 < i_s ≤ i_p. This is because only incoming/outgoing vehicles ahead of the vehicle passing P can influence the trajectory as shown in the shaded region in Figure 12. As a result, the height at the intercept of the estimated trajectory with i-coordinate; N (i_s, 0) is increased by 1. Since the height at P; N (i_p, j_p) has not been changed, the trajectory path is forced to change. Namely, before a vehicle being added at A, trajectory of a vehicle which is passing P; passes through (i_s, 0) as well. But after adding a vehicle at A, the cumulative height corresponding to that particular vehicle increases by 1 while N (i_p, j_p) does not change. As a result the new cumulative height corresponding to that particular vehicle cannot be found at (i_p, j_p) but at (i, j_p), i > i_p. The situation is illustrated in Figure 15 where adding a vehicle at A results in the relocation of P. Direction of the relocation is indicated with a small arrow.

Above discussion is similarly valid for the condition where a vehicle is subtracted at A.
When signal red intervals exist along the valid paths from A to P, considering the shortcut effect of the red intervals in time-space (the minimum cost along the horizontal links is equal to zero) the region to be considered would be narrower as in Figure 16. However, basically the same discussion as above is applicable.

From above discussion the following is concluded:

(i) Only when \((22)\) holds, vehicle addition or subtraction at \(A\) can shift the estimated trajectory at \(P\).

(ii) Practically, possible locations for adding or subtracting the vehicles would be the intersections. For instance a vehicle can be added from the crossing street during the red interval or can be subtracted as a turning vehicle to the crossing street during the green interval.

(iii) Location \(A\) where vehicles are supposed to be added or subtracted must be found so as to satisfy the consistency condition (Daganzo, 2006).

7.3.2. Modification method

Suppose we want to modify the estimated trajectory at point \(P\) in Figure 12 so as to effectively improve the agreement between the estimated trajectory and the probe trajectory. Considering the conclusions in (23), one possible but time consuming way is as follows:

(i) Based on (22), all the locations at which a vehicle addition/subtraction can effectively adjust the height at \(P\) are defined along all intersections in the time-space plane.

(ii) Modification of the estimated trajectory at \(P\) is made by adding/subtracting required number of vehicles in a suitable locations defined in step (i).

(iii) Step (i) and step (ii) are repeated for all \(P\) locations along the estimated trajectory untill reaching the downstream.

Although modification of the estimated trajectories based on above steps is theoretically sound, the calculation would be cumbersome and time consuming (impractical). Moreover, as mentioned earlier, in addition to incoming/outgoing vehicles, several other factors such as on-street parked vehicles, pedestrians, and bicycles are affecting the movement of the vehicles as well as the shape of the fundamental diagram. For practical applications it is necessary to develop more straightforward methods to allocate incoming/outgoing vehicles in suitable locations.

7.3.2.1. Method I

To simplify the analysis, the following assumptions are made while adding/subtracting the vehicles to improve the agreement:

(i) Vehicles are coming in only at the nearest intersection to \(P\) during the red intervals.
(ii) Vehicles are leaving out only at the nearest intersection to $P$ during the green intervals.

(iii) Vehicles are added/subtracted along the backward wave line (with slope $-w$).

(iv) The minimum number of the vehicles are added or subtracted.

Based on assumptions (i), (ii) and (iii) in (25), the locations at which the vehicles should be added or subtracted can be determined as $S$ or $Q$ as in Figure 12, respectively. From assumption (i) in (11), the vehicle addition/subtraction at $S/Q$ will not influence the estimated trajectory before $P$. As a result, if a vehicle is added or subtracted along the backward wave line, we can modify the height at $P$ without changing the estimated trajectory before $P$. Therefore, with assumption (iii) in (25), the estimated trajectory can be modified sequentially from the beginning. Also, from assumption (iv) in (25), we minimize the number of the vehicle additions or subtractions. For instance, when the deviation in the cumulative height at $P$ is +1, only 1 vehicle is added, not 2 additions and 1 subtraction, 3 additions and 2 subtractions, and so on. Since we do not have sufficient information on the actual locations of the incoming/outgoing vehicles, these assumptions are proposed purely for simplification.

Figure 17 and Figure 18 illustrate two examples before and after implementation of the proposed treatment. For each case the Mean Absolute Error is estimated from (9) by comparing the estimated trajectory with its corresponding probe trajectory. Figure 17 shows considerable improvements after vehicles being added or subtracted in appropriate locations. However improvements in Figure 18 are marginal and insignificant. It turns out that the applicability of the proposed treatment is limited and it cannot always guarantee the best possible agreement between the estimated trajectories and corresponding probe trajectories.

Considering that the vehicle addition/subtraction can be made only at intersections along the backward wave, there are limited locations at which vehicles can be added or subtracted. However, even with such limited locations, there are several combinations of incoming/outgoing vehicles which can produce identical results.

7.3.2.2. Method II

As it is suggested in the following, for practical applications when traffic data regarding the turning vehicles at the intersections are not available, simpler procedures can be devised to modify the trajectories in order to improve the agreement between the estimated trajectories and the corresponding probe trajectories:

(i) Compare the number of entering and exiting vehicles in the upstream and downstream between two adjacent probe trajectories, and estimate the algebraic sum of them assuming a positive value for entering vehicles and a negative value for exiting vehicles.

(ii) The difference estimated in step (i) indicates the algebraic sum of the vehicles that must be added or subtracted in the midblock intersections assuming $+1$ for an added vehicle and $-1$ for a subtracted vehicle.

(iii) Based on step (ii) and visual inspection of the shape of the estimated trajectory, appropriate number of vehicles are added during the red intervals or subtracted during the green intervals in suitable locations at the intersections so as to matching the estimated trajectory and probe trajectory at least at the entry and the exit points in the upstream and the downstream.

Following above steps the entry and the exit points of the estimated trajectory will virtually match with that of the probe trajectory. However, in the middle of the study area the trajectories may not still match perfectly. Based on practical considerations and field observations extra rules could be set further to eliminate potential locations for vehicle addition/subtraction in the time-space plane. For instance some intersections might accommodate considerable volumes of the turning vehicles while in other intersection the turning movements could be prohibited. In addition outgoing (incoming) vehicles may use only the last portion of the red (green) interval to turn.

The example demonstrated in Figure 18b) is represented in Figure 19 after incoming/outgoing vehicles were allocated based on (26). The algebraic sum of the number of the entering and the exiting vehicles in the upstream and the downstream in this example is $-1.6$. Based on steps (ii) and (iii) in (26), (+)6 vehicles were added and (-)22 vehicles were subtracted in appropriate locations considering the shape of the estimated trajectory. Moreover, incoming vehicles were permitted only at the second intersection. It can be seen that after modification the entry and the exit points of the estimated trajectory at the upstream and the downstream is identical with that of the probe trajectory. Unlike in Figure 18b), the estimated trajectory in Figure 19 shows significantly improved agreement with the corresponding probe trajectory.
It is worthy to note that since the signal timing data of the 3rd intersection were not available, this intersection was not considered for vehicle addition/subtraction. Otherwise it would have been possible to achieve more significant improvements in this example.
8. Conclusions and future directions

The majority of data fusion techniques to estimate travel time on urban corridors are based on statistical methods and do not adequately implement traffic engineering methods in the fusion process. In addition, probe data which
include rich trajectory information are not properly incorporated when fused with fixed sensor data to produce travel time information. Instead, probe information is used only to extract travel time information. This research suggested a novel approach towards fusion of fixed and probe sensor data in order to reconstruct trajectories of all vehicles considering signal timing parameters and traffic engineering concepts. The proposed method is based on kinematic wave theory and is capable of using probe trajectory information completely. Such features enable reasonable estimation of vehicle trajectories using minimum input data. Furthermore, particular techniques were introduced to incorporate vehicles coming in or going out in the middle of the study section. The proposed method was applied to real world data and its performance was evaluated under different conditions.

Availability of trajectory information paves the way towards several other applications for the proposed method. Signal timings can be adjusted to evaluate the progression quality along the study section and iterative procedures can be adopted to find optimum signal coordination. Furthermore, since driving modes can be retrieved from estimated trajectories, the methodology can be used for the purpose of emission monitoring. Different signal patterns can be examined to figure out the most environmental friendly signal settings for a particular corridor.

The objective of this research went beyond travel time estimation, since the proposed method can be used for traffic state estimation as well. Traditionally simulation models have been used for the purpose of continuous state estimation. However, simulation components including car following and driver behavior models require careful calibrations. In addition, the simulation process is time consuming even for a short study area. The proposed method provides an analytical approach for traffic state estimation which does not require detailed calibrations and preparation. It would be very useful to compare the performance of the proposed method with simulation models in terms of their accuracy and efficiency to reproduce vehicle trajectories. Features such as required input data, calibration procedures, calculation time and flexibility to use probe information would be among the points of interest.

The analytical approach presented in this research is deterministic. Characteristics of traffic flow are simply described with a triangular fundamental diagram in which the forward and backward wave speeds are constant. Such simplifications result in some discrepancy between estimated trajectories and probe trajectories which was found to be increasing with time. To avoid accumulation of estimation errors, as soon as new probe information is available, initial conditions at network boundaries should be reset and trajectories should be estimated based on the new initial conditions from that point on. Yet, if probe data are available less frequently, estimation errors are unavoidable over
longer intervals. Such discrepancies might be eliminated if stochastic characteristics of traffic flow were to be considered in the proposed methodology.

The proposed methodology still cannot be used to estimate vehicle trajectories in multi lane facilities where vehicles can overtake or pass each other as it was assumed the first in-first out (FIFO) condition holds throughout the study area. As a result, application of the methodology is currently restricted to lane by lane analysis with no passing permitted. On the other hand, although a probe vehicle whose trajectory is used as a reference to estimate other trajectories may change its lane several times, it is assumed that the reference probe trajectories represent one dimensional movement and lane changing maneuvers are not considered. A possible extension to the current research would be to develop it further for multilane conditions where lane changing maneuvers are allowed. Considering that the methodology to consider incoming/outgoing vehicles has already been addressed in this research, a vehicle changing its lane in multilane conditions can be treated as an incoming or outgoing vehicle in a single lane.

Tackling these issues require adjustment of the fundamental theory as well as introducing extensions to the proposed methodology which are considered to be the future directions of this research.

References


