



Heterogeneous urban traffic data and their integration through kernel-based interpolation

Journal:	<i>Journal of Facilities Management</i>
Manuscript ID:	Draft
Manuscript Type:	Research Paper
Keywords:	Data fusion, Big Data, Transport network , Automatic Number Plate Recognition , GPS , urban loop detectors

SCHOLARONE™
Manuscripts

Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Heterogeneous urban traffic data and their integration through kernel-based interpolation

August 2015

ABSTRACT

This paper presents collection and analysis of heterogeneous urban traffic data, and integration of them through a kernel-based approach. The recent development in sensing and information technology opens up opportunities for researching the use of this vast amount of new urban traffic data. In this paper, the data fusion algorithm is developed by using a kernel-based interpolation approach. Our objective is to reconstruct the underlying urban traffic pattern with fine spatial and temporal granularity through processing and integrating data from different sources. The fusion algorithm can work with data collected in different space-time resolution, with different level of accuracy, and from different kinds of sensors. The properties and performance of the fusion algorithm is evaluated by using a virtual test-bed produced by VISSIM microscopic simulation. The methodology is demonstrated through a real-world application in Central London. This paper contributes to analysis and management of urban transport facilities.

Keywords: Data fusion, Big Data analytics, transport network, Automatic Number Plate Recognition (ANPR), GPS, urban loop detectors

1. Introduction

A detailed and reliable picture of spatio-temporal variations of traffic is essential for understanding and managing congestion (Tsapakis et al., 2012; Chow et al., 2014). Much previous research on traffic data has been focusing on freeways where Kwon and Varaiya (2005) provide a review on relevant studies. Compared with freeways or motorways, we see relatively less research work done on urban networks. It is due to the lack of required data and the complexity of problem. Recently, the increasing availability of urban traffic data provides new research opportunities.

Urban traffic data varies greatly in terms of spatio-temporal granularity, latency and accuracy. Typical sources of traffic flow data in urban environment include:

- Fixed sensors – such as loop detectors and automatic traffic counters - provide information of traffic volume, composition of traffic (e.g. proportion of buses, heavy good vehicles, etc), concentration and speed.
- Global Positioning System (GPS) devices – such as smart phones, personal navigators, etc - are attached to vehicles or persons. The GPS devices report location and speed of the attached objects regularly (typically every second). Various information such as trip lengths and journey times can also be derived from GPS data.
- Automatic Vehicle Identification (AVI) – With vehicle (re)-identification techniques, on-road or roadside sensors (e.g. cameras) can provide information including journey times and trip lengths.

Integrating heterogeneous traffic data in a consistent way is always a challenging problem, where Ou (2011) provides a review of different kinds of data fusion approaches. One the most popular approach for integrating traffic data is through the model-based Kalman Filter (KF, Kalman, 1960) and its variants such as Extended KF, Unscented KF, and Particle Filter (PF) (see examples: Wang and Papageorgiou, 2005; Mihaylova et al., 2007; Herrera and Bayen, 2010; Ngoduy, 2011). Under the KF and PF (and their variants) framework, traffic flow estimation is produced by comparing and combining the model estimates and actual measurements. The filtering framework consists of two parts: a state equation to model and predict the traffic state evolution over time, and an observation equation to relate the measurements to the underlying traffic state. Both state and observation equations consider explicitly the associated errors involved. An estimate is then generated through minimizing the associated expected pooled error in state predictions and observations. Such filtering framework is used widely in data fusion algorithms due to a number of its desirable properties. The explicit state and observation models allow data from different kinds of sensors to be incorporated. Moreover, the well-defined statistical measures of uncertainty make it possible to quantify the value of each data source.

Nevertheless, the filtering approach is model based, which mean it has to operate based upon an assumed traffic model. Typical examples of traffic model for this purpose include Cell Transmission Model (Daganzo, 1994), METANET (Papageorgiou et al., 1990), and a recently proposed two-regime traffic model by Balijepalli et al. (2013). As noted in Ou (2011), a major difficulty for this model-based approach is on the selection and calibration of the underlying traffic model. Choosing and calibrating a suitable traffic model is not

99 straightforward, and inappropriate choice will lead to inconsistency and instability of the
100 fusion framework.

101
102 This paper presents a kernel-based data fusion algorithm that can integrate heterogeneous
103 urban traffic data with different characteristics. The objective is to reconstruct the urban
104 traffic pattern with fine spatial and temporal granularity through processing and integrating
105 data from different sources. The fusion algorithm does not require assumption of any
106 underlying model and it can work with data collected in different spatio-temporal granularity,
107 with different level of accuracy, and from different kinds of sensors. We need to note the
108 proposed method will be limited to offline application due to the lack of an underlying traffic
109 model as a state estimator. Nevertheless, the fusion algorithm in this paper will be a valuable
110 and cost-effective tool for offline transport planning and policy evaluation through
111 integrating existing sources of data. The performance and properties of the fusion algorithm
112 is evaluated by using a synthetic scenario generated by VISSIM micro-simulation. The
113 algorithm will also be used to integrate actual traffic data collected from a road section in
114 Central London (UK) as an illustration of real-world application.

115
116 This paper starts with Section 2 which discusses the characteristics of different traffic data on
117 urban road networks. We use Central London in the UK as an illustration. Section 2 will also
118 highlight the difference between different data in terms of their spatial and temporal
119 granularity. Section 3 presents the data fusion algorithm that we adopt. Section 4 presents an
120 application of the fusion algorithm to Central London road network. The fusion algorithm is
121 also analysed by using a virtual test-bed generated by VISSIM micro-simulation. Finally,
122 Section 5 gives some concluding remarks.

123
124

125 **2. Urban traffic data – London, UK**

126

127 Journey times are one of the most important performance indicators for urban road networks.
128 As an illustration, journey times in London are measured by using the Automatic Number
129 Plate Recognition (ANPR) technique. In London, there are about 500 cameras for enforcing
130 various policies such as congestion charging and low emission zones. When a vehicle passes
131 by a camera, its plate number will be recognized and recorded along with the associated time.
132 The journey time of the vehicle between two camera sites can then be derived by matching
133 the plate number. The derived journey times are further processed and stored in 5-min
134 averages in the database.

135

136 It is noted that various errors may arise in matching the license plate numbers due to various
137 reasons such as misreading of license plates, vehicles stopping en-route, and vehicles taking
138 unusual long route between the two camera locations, data loss due to road closure, and
139 failure of hardware system. Consequently, a set of data filtering and processing rules is
140 adopted to improve the journey time estimation (Robinson and Polak, 2006). Some patching
141 or imputation algorithms may be used for imputing missing data. The associated ANPR
142 journey time data is flagged with a code referring to the type of patching mechanism which is
143 applied. This code is ranging from 0 (best: no patching applied) to 3 (worst: patched by
144 typical profile).

145

146 Figure 1 shows a 1-km stretch of Waterloo Road (A301) in Central London, and Figure 2
147 shows the associated speeds (i.e. reciprocals of journey times) measured along the road in

148 April 2010. In the figure, the position and width of the bar at each 5-min interval reflects the
 149 average and dispersion of the journey times in the month at that particular interval.
 150

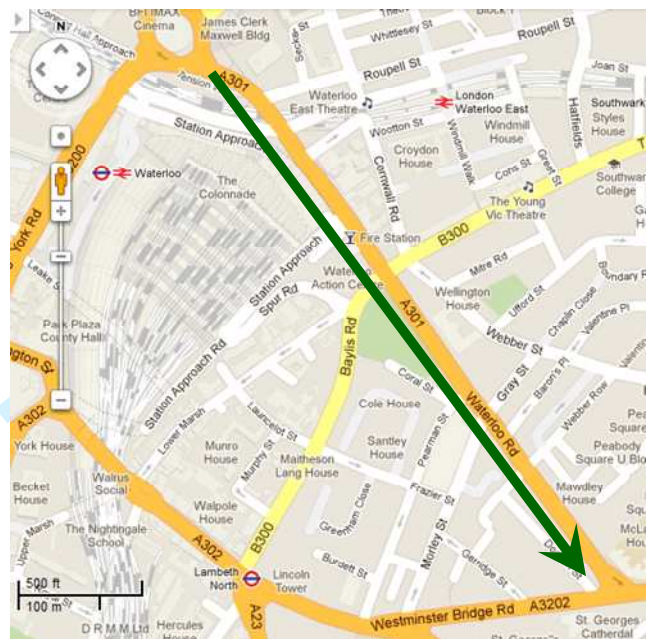


Figure 1 Waterloo Road, London (UK)

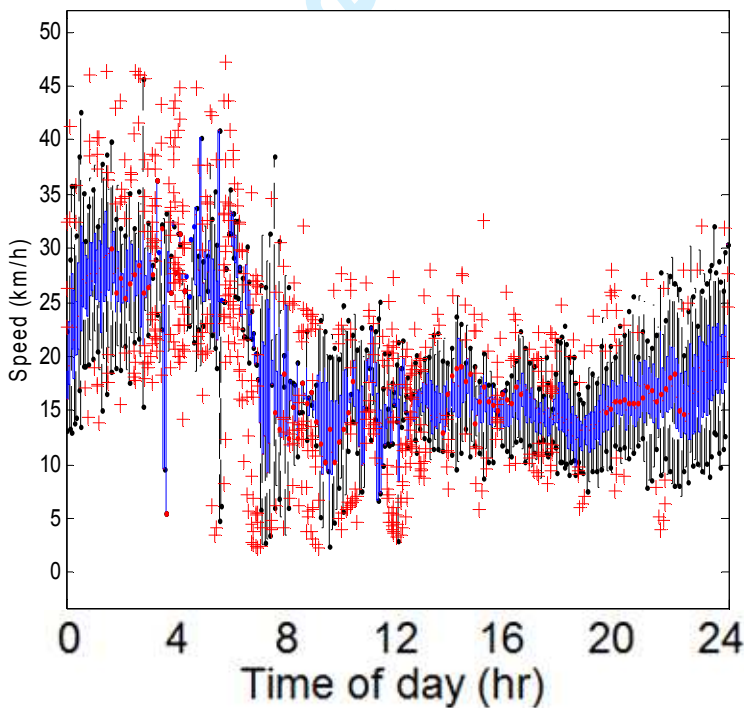


Figure 2 Variations of traffic speeds in April 2010 along Waterloo Road, London (UK)

Figure 3 shows the average speed field over time and space produced by the ANPR journey times. The colour scale on the space-time grid represents the average speed of traffic. As can

158 be seen from Figure 3, a major weakness with the ANPR journey times is that they do not
 159 capture much spatial feature of traffic. It is because the distance between a pair of ANPR
 160 camera sites is typically far apart (in the range of kilometres), which implies a lot of spatial
 161 variations are missed along the route.

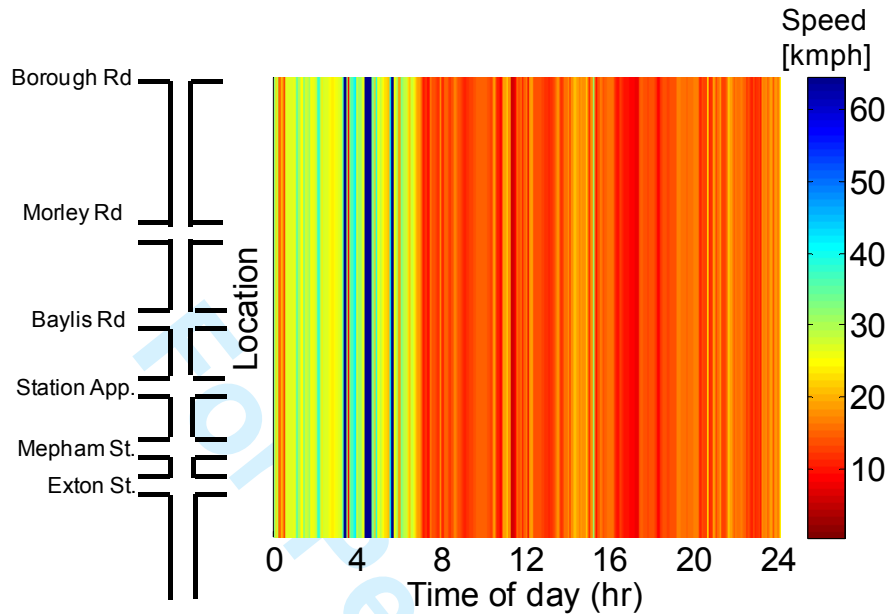


Figure 3 ANPR-Speed field along Waterloo Road, London (UK)

162
 163
 164
 165 To extract detailed spatial traffic variations, we use floating car data provided by
 166 Trafficmaster*. Some vehicles on the road are equipped with Trafficmaster GPS (Global
 167 Positioning System) devices. The GPS devices on these vehicles report the locations of the
 168 vehicles on a regular basis (~8-10 seconds). Figure 4 shows the corresponding speed field
 169 generated by these Trafficmaster data, which can reveal much more spatial feature compared
 170 with the ANPR data. Nevertheless, there are only very limited samples of Trafficmaster
 171 vehicles on the road (about 1,500 such vehicles in Greater London Area). With such small
 172 sample size, Trafficmaster can only reveal limited temporal characteristics of traffic.
 173

* <http://www.trafficmaster.co.uk/>

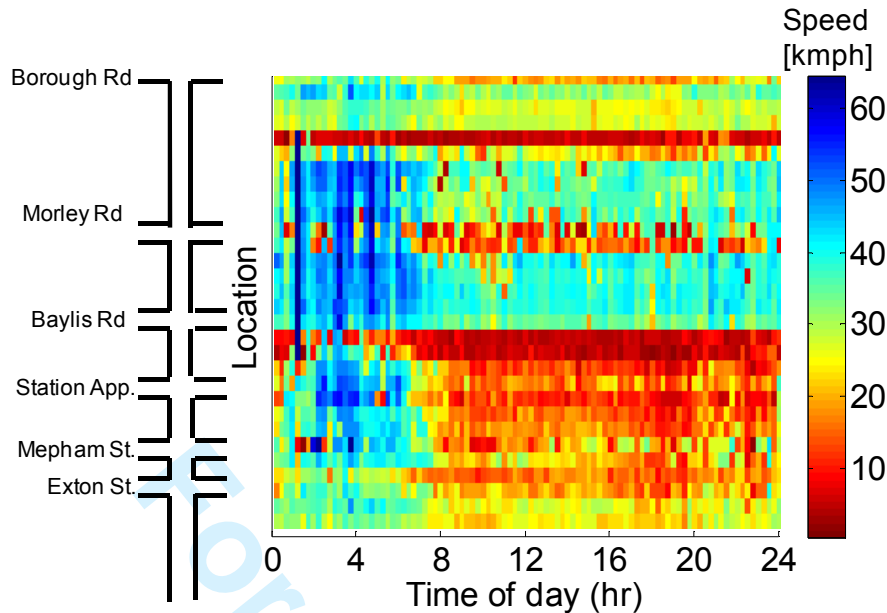


Figure 4 Trafficmaster-Speed contour along Waterloo Road, London (UK)

3. Data fusion

The study presents a kernel-based interpolation method to reconstruct the underlying traffic pattern with fine spatial and temporal details through integrating data from heterogeneous sources. It should be emphasized that the algorithm presented herein is generic and it can be applicable in other scenarios with different data sources from what we present herein. The data fusion algorithm consists of two steps: smoothing and integration.

3.1. Data smoothing

Different traffic data often come in different spatio-temporal granularity. Before integration, it is necessary to first process and reconstruct the data on a common space-time grid. Here we adopt the kernel-based interpolation method which is a popular approach used in data fusion literature (Lanckriet, et al., 2004; Camps-Valls et al., 2006). This kernel-based interpolation can also be regarded as a kind of fuzzy regression approach (Tanaka et al., 1982; Choi and Chung, 2002). Moreover, this kernel-based method is also adopted by Treiber and Helbing (2002 a, b) for fusing freeway traffic data as will be discussed in the next section.

Consider a set of traffic measurements u_i from a source taken as location x_i and time t_i , where $i = 1, 2, \dots, n$ and n is the total number of measurements. With this set of data, Treiber and Helbing's algorithm reconstructs the traffic state on a space-time domain with user-defined spatial interval δx and temporal interval δt .

We define (x, t) be the space-time coordinate on this new space-time domain, the associated traffic state $u(x, t)$ is estimated by

$$u(x,t) = \frac{1}{\Phi(x,t)} \sum_{i=1}^{M(x,t)} \phi_i(x-x_i, t-t_i) u_i(x_i, t_i), \quad (1)$$

where $\Phi(x,t) = \sum_{i=1}^{M(x,t)} \phi_i(x-x_i, t-t_i)$ is a normalizing factor. The notation ϕ_i denotes the value of a kernel smoothing or shape function at $(x-x_i, t-t_i)$, in which $(x-x_i)$ and $(t-t_i)$ are the respectively the spatial and temporal lags between the space-time of interest (x,t) , and space-time of the data source (x_i, t_i) . The kernel function is added here to capture the ‘fuzziness’ in the raw data and to smooth out high frequency noise and fluctuations. It is usual the function ϕ_i is symmetric or isotropic in (x, t) in which it depends on the quantities $\left[\frac{(x-x_i)}{\sigma}, \frac{(t-t_i)}{\tau} \right]$, where σ and τ represent respectively the spreads of the spatial and temporal influence regions. Model (1) converges to an ordinary regression model without fuzziness as σ and τ tend to zero, while the model become a simple arithmetic mean of all data points, regardless of their space-time location, as σ and τ tend to infinity.

Treiber and Helbing (2002) and Treiber et al. (2009) adopt the following exponential kernel function:

$$\phi_i(x-x_i, t-t_i) = \exp \left[- \left(\frac{|x-x_i|}{\sigma} + \frac{|t-t_i|}{\tau} \right) \right],$$

as the shape function in their freeway applications. Moreover, Treiber and Helbing (2002) suggests that σ and τ should be taken as halves of the associated spatial and temporal granularities. For example, if the spacing and sampling interval of detection are 500-m and 5-min respectively, then σ should be 250-m and τ should be 2.5 min.

Furthermore, in expression (1), $M(x,t)$ is number of data points that we consider when calculating $u(x,t)$. In extreme case, we can consider all data points in which $M(x,t) = n$. The choice of $M(x,t)$ is a trade-off between computational speed and accuracy: the larger $M(x,t)$, the more accurate the estimates while the heavier computational burden.

3.1.1. Anisotropic filter

It is known in traffic flow theory that the propagation speeds of traffic characteristics are different in free-flow and congested conditions. In free flow, traffic characteristics propagate *along* with the direction of traffic with a ‘free-flow’ speed v_f ; in congestion, traffic characteristics travels *against* the direction of traffic with a speed w . Empirical findings show the propagation speed w is generally less than speed v_f , which suggests the traffic influence is *anisotropic* with respect to direction of influence.

To capture the anisotropic feature, Treiber and Helbing (2002 a,b) propose different formulae of the kernel function for free-flow and congested traffic conditions respectively as

245

$$u_{free}(x, t) = \frac{1}{\Phi(x, t)} \sum_{i=1}^{M(x, t)} \phi_i \left(x - x_i, t - t_i - \frac{x - x_i}{v_f} \right) u_i$$

247

248 and

$$u_{cong}(x, t) = \frac{1}{\Phi(x, t)} \sum_{i=1}^{M(x, t)} \phi_i \left(x - x_i, t - t_i - \frac{x - x_i}{w} \right) u_i,$$

250

(2)

251

252 in which $\Phi(x, t)$ is the corresponding normalizing factor for both cases. In the case of urban
 253 streets, we take the forward propagation speed v_f as 25 kmph as revealed from the statistics
 254 shown in Figure 2, while w is -8 kmph. Note that w is negative which represents that it is
 255 travelling against the direction of traffic.

256

257 The smoothed speed at each (x, t) is then determined as

258

$$u(x, t) = \gamma(x, t)u_{cong}(x, t) + [1 - \gamma(x, t)]u_{free}(x, t),$$

259

(3)

260

261

262 where $\gamma(x, t)$ is weighting factor which manipulates the superposition of the free-flow and
 263 congested speeds. The weighting factor is expected to be approximately zero in free-flow (or
 264 high speeds), and approximately one at low speeds.

265

266 Treiber and Helbing (2002) and Treiber et al. (2009) adopt the “s-shaped” hyperbolic tangent
 267 function:

268

$$\gamma(x, t) = \frac{1}{2} \left[1 + \tanh \left(\frac{u_{crit} - u^*}{\Delta u} \right) \right],$$

269

(4)

270

271

272 where $u^* = \min[u_{free}(x, t), u_{cong}(x, t)]$; u_{crit} is a speed threshold distinguishing free-flow and
 273 congested traffic; and Δu is the transition window width adopted in the weighting function.
 274 Following our observations in Figure 2, we set the u_{crit} be 25 kmph and Δu be 5 kmph for
 275 classifying free-flow and congested data.

276

277

278 3.1.2 Processing journey time data

279

280 The data smoother above takes fixed data points, while journey times are measurements over
 281 a section. The journey time data will first have to be converted into equivalent series of point
 282 measurements before they can be used.

283

284 Consider a road section with a length L and an estimated journey time ω from a sample of
 285 vehicles within a time window Δt (say, Δt is 5 minutes as in the ANPR data). The average
 286 speed of this set of vehicles along the section is calculated as $\bar{v} = L / \omega$.

The following two assumptions are made:

- all sampled vehicles travel steadily (i.e. no change in speed) with this speed within the time window;
- the sampled vehicles enter the road section uniformly with a common time headway $\Delta t/(n-1)$, where n is the number of sampled vehicles in Δt .

We can then construct a set of ‘virtual trajectories’ of the vehicles as shown in Figure 5. The dotted lines in the figure refer to the ‘true’ trajectories of the sampled vehicles, which are unknown. The solid lines refer to the ‘virtual’ trajectories constructed. We consider that the vehicles enter the link at times s_0, s_1, \dots, s_{n-1} , where $s_i = s_{i-1} + \Delta t/(n-1)$ for all $i = 1, 2, \dots, (n-1)$. All vehicles travel through the link with a constant speed \bar{v} , and exits the link at $s_i + \omega$, where $i = 0, 1, 2, \dots, (n-1)$. Denote the location ‘0’ and ‘L’ be the starting and ending points of the link respectively. For a vehicle enters the link at s_i , data points are sampled every Δs (say, 1 min) at the following space-time coordinates: $(0, s_i), (\bar{v}\Delta s, s_i + \Delta s), (\bar{v}(2\Delta s), s_i + 2\Delta s), \dots, (L, \omega)$. At each of these time and location, it is regarded that the sampled point will report a speed \bar{v} , and hence we convert the sectional journey time measurement into a series of point measurements.

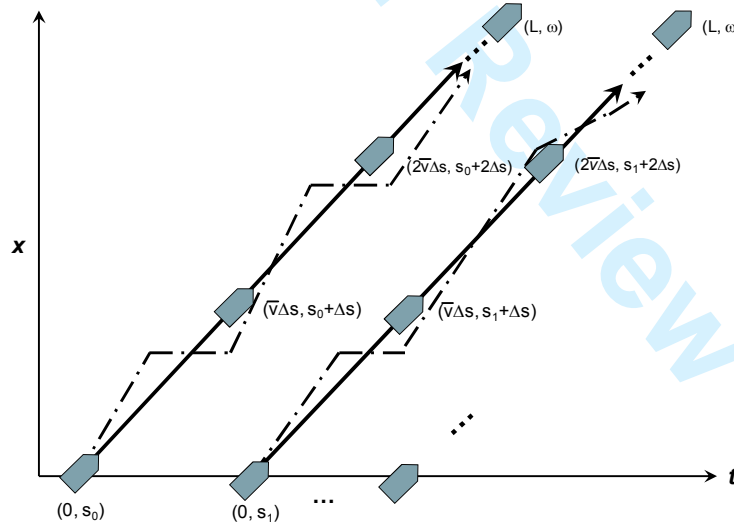


Figure 5 Converting journey times into point measurements

3.2 Data Integration

After smoothing and reconstructing the traffic data onto a common space-time grid, we can then integrate the data with the following formulation:

1
2
3 3194
5 320

$$\tilde{u}(x,t) = \sum_{k=1}^{K(x,t)} \beta_k u_k(x,t),$$

6
7 321

(5)

8 322

9 323

10 324

11 325

12 326

13 327

14 328

15 329

16 330

17 331

18 332

19 333

20 334

21 335

22 336

23 337

24 338

25 339

26 340

27 341

28 342

29 343

30 344

31 345

32 346

33 347

34 348

35 349

36 350

37 351

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

where $u_k(x,t)$ is the smoothed and reconstructed data field from data source k ; $\tilde{u}(x,t)$ is eventual the integrated data field; the weighting factor $\beta_k \in [0,1]$ associated with each source data can be related to various factors such as the accuracy, reliability, number and variance of measurements of data source k . This combination method is known as voting technique in data fusion literature (Olkin, 1992; Choi and Chung, 2002), which is essentially a weighted linear combination of information from different sources.

4. Simulation experiment

Before proceeding to the real world application, it is necessary to conduct analysis on the sensitivity and accuracy of the fusion algorithm. Due to the lack of ground truth data, we conduct the analysis on a micro-simulation test-bed. This study chooses VISSIM simulation package after considering the plausibility of the VISSIM model for replicating complex traffic dynamics.

The Waterloo Road section (Figure 1) is coded into VISSIM which is used to generate a synthetic scenario. The VISSIM simulation time step is set to be one second and the simulation period is one hour. According to field observations, the demand rates are set to be 900 vph and 100 vph respectively for the mainline (Waterloo Road) and the cross-streets. At the two signal-controlled vehicle intersections, 15% of the mainline traffic will be turning into the cross-streets, while 60% of cross-street traffic turning into the mainline. The signal timings are also set based upon observations from the field. A total of 1,516 vehicles are generated during the simulation period, from which we derive the corresponding speed field (Figure 6) with a space-time resolution at 50-m (Δx) and 1-min (Δt) respectively. At each space-time coordinate (x, t) in the figure, the associated speed $v(x,t)$ is calculated as the *average speed* of all vehicles detected within that (x,t) . We regard this speed field $v(x,t)$ as the 'ground truth' for comparison.

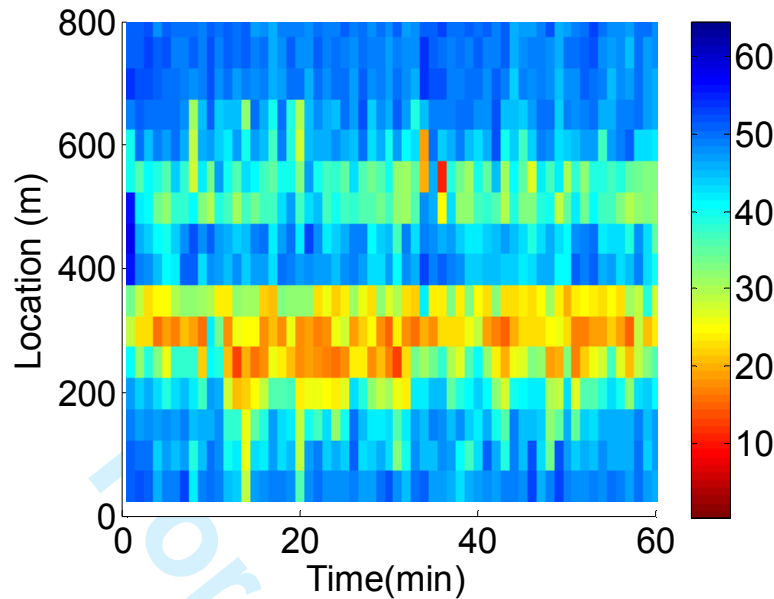


Figure 6 Speed field of VISSIM simulated data
(spatial resolution: 50-m; temporal resolution: 1-min)

To simulate the trafficmaster data, we randomly select 100 out of the 1,516 vehicles (i.e. a penetration rate of 6.6%) and assume that they are equipped with GPS devices that can report the associated positions every second.

To simulate the AVI system, we place two virtual ‘cameras’ at the entrance and the exit of the road stretch. It is found that 283 out of 1,516 vehicles would travel all way through the stretch and hence their associate journey times will be regarded as the ‘ANPR’ journey time herein. The journey times of these 283 vehicles are processed into 1-min averages.

The ‘raw’ trafficmaster and ANPR data are first processed by the ASM specified by formulae (1) and (3) in Section 3.1 and projected onto a common user-defined space-time grid with resolutions at 50-m (Δx) and 1-min (Δt). Following expression (5), the smoothed data are then integrated by using the following linear combination:

$$\tilde{u}(x, t) = \beta[u_{ANPR}(x, t)] + (1 - \beta)[u_{GPS}(x, t)], \quad (6)$$

for all x and t , where the weighting factor β lies between 0 and 1. We adopt a ‘data-data consistency’ concept (Ou, 2011) to estimate the value of this β . We first regard the overall journey time through the arterial given by ANPR is reliable. The parameter β is determined such that the difference between the corresponding journal times given by the fused data and those given by the ANPR is minimised.

With the ‘ground truth’ given by the VISSIM simulation, we test this concept and analyse the corresponding the overall RMSNE (Root-Mean-Square-Percentage-Error) with respect to the ‘ground truth’, where RMSNE is calculated as

383
384

$$RSMPE = \sqrt{\frac{1}{N} \sum_x \sum_t \left[\frac{\hat{v}(x,t) - v(x,t)}{v(x,t)} \right]^2},$$

385

386

(7)

387

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

5. Real world application – Waterloo Road, London

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

We now present the application of the data fusion algorithm with the real world data. Given the ANPR and trafficmaster data, we find β to be 0.27 using the method described previously, and it is found that this β value is not significantly different from the determined by using simulated data. The corresponding overall RSMPE is found to be 9.6% which is slightly higher than the one we obtained from the VISSIM simulation test-bed. We suggest this is due to the complicated nature (e.g. vehicles may stop unexpectedly) in real world scenario. More sophisticated systems, such as model based fault identification and state prediction, will be developed to improve the estimation errors.

Figure 7 shows the resulting fused speed field along Waterloo Road. It shows that the fusion algorithm is able to retrieve much hidden spatio-temporal feature of traffic with either ANPR or trafficmaster data alone. Moreover, the result also shows that the fusion algorithm is able to smooth out the ‘corners’ (compared with Figure 4) observed in the Trafficmaster dataset through the underlying kernel function.

We note that there are several areas that require further analysis. In particular, it will be interesting and important to explore whether the fusion algorithm can restore the queue formation and dissipation process – which are important characteristics - in urban networks. Nevertheless, this will require traffic data with finer granularity - say with spatial granularity down to metres and temporal resolution down to seconds. Moreover, it is desirable to validate the performance of the fusion algorithm with real world data rather than simulated ones. We are currently exploring new datasets for further investigation into this. Meanwhile, we are conducting some research into loop detector data collected under the SCOOT urban traffic control system (see Heydecker et al., 2012) which collect flow and concentration information at up to a frequency of 4Hz (i.e. every 0.25 sec). Moreover, we are exploring the possibility

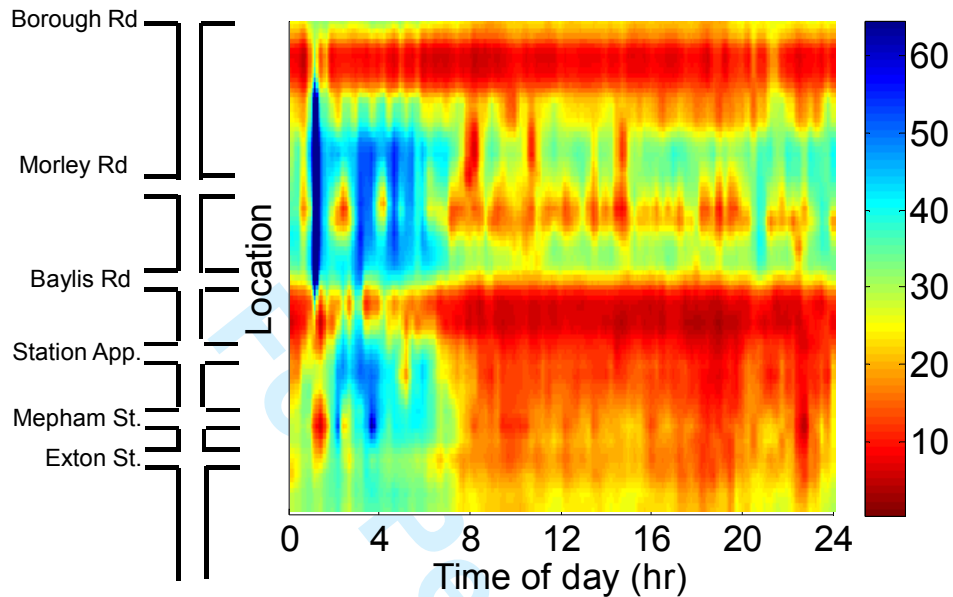
58

59

60

429 of obtaining some other higher quality GPS vehicle data including taxi trajectories from
 430 Addison Lee and bus trajectories from London ibus system, which both contain more samples
 431 than the trafficmaster dataset.

432



433

434

435 Figure 7 Speed field along Waterloo Road after fusing the ANPR and Trafficmaster data

436

437

438 6. Concluding remarks

439

440 This paper presents an application of kernel based interpolation method to urban traffic data
 441 fusion. The application is illustration through a case study on Central London road network.
 442 Due to the lack of ground truth data, the performance of the fusion algorithm is analysed by
 443 using a virtual test-bed generated by VISSIM micro-simulation. It is interesting to note that
 444 the anisotropic filter formulation proposed by Treiber and Helbing (2002) does not show
 445 significant improvements over the original isotropic one. This is different from the freeway
 446 case where Treiber and Helbing's formulation shows significant improvement. An
 447 explanation for this is that the spatial and temporal granularity of traffic data required in
 448 urban networks is finer than that in motorway (say, minutes in freeway case versus seconds in
 449 urban cases), hence taking the difference in wave speeds into consideration does not have a
 450 noteworthy effect.

451

452 The kernel-based algorithm presented herein does not require assumption and calibration of
 453 any traffic model. It is easy to implement and parallelize. For example, the fusion computing
 454 task can be parallelized such that each computer can be made be responsible for one specific
 455 data source over a specific space-time domain. Such properties make the fusion algorithm
 456 suitable for practical large-scale applications. The research contributes to the application of
 457 Big-Data analytics to infrastructure management.

458

459

460

461 **References**

- 462
463 Balijepalli, N, Ngoduy, D. Watling, D. (2013). Dynamic Network Loading Model using the
464 Two-regime Transmission Model. *Transportmetrica*, In press.
- 465 Camps-Valls, G., Gomez-Chova, L, Munoz-Mari, J, Rojo-Alvarez, J, Martinez-Ramon, M
466 (2008). Kernel-based framework for multi-temporal and multi-source remote sensing
467 data classification and change detection. *IEEE Transactions on Geoscience and*
468 *Remote Sensing*. 46(6), 1822 – 1835.
- 469 Choi, K and Chung, Y (2002) A data fusion algorithm for estimating link travel time. *Journal*
470 *of Intelligent Transportation Systems*, 7(3), 235 – 260.
- 471 Chow, AHF, Santacreu, A., Tsapakis, I., Tanaksaranond, G., Cheng, T. (2014) Empirical
472 analysis of urban congestion. *Journal of Advanced Transportation*, 48(8), 1000-1016.
- 473 Herrera, JC and Bayen, A (2010) Incorporation of Lagrangian measurements in freeway
474 traffic state estimation. *Transportation Research Part B*, 44(4), 160-481.
- 475 Heydecker, BG, Tsapakis, I., Chow, AHF (2012) Exploring Macroscopic Fundamental
476 Diagrams for the Transportation Network in Central London. *Proceedings of the 1st*
477 *European Symposium on Quantitative Methods in Transportation Systems*, September
478 4-7. Lausanne, Switzerland.
- 479 Kalman, RE (1960) A new approach to linear filtering and prediction problems. *Journal of*
480 *Basic Engineering*, 82, 35 – 48.
- 481 Kwon, J, Varaiya, P (2005) The congestion pie: delay from collisions, potential ramp
482 metering gain, and excess demand. *Proceedings of the 84th Annual Meeting*
483 *Transportation Research Board*, 9-13 January, Washington, D.C.
- 484 Lanckriet, GR, Deng, M, Cristianini, N, Jordan, M, Noble, W (2004) Kernel-based data
485 fusion and its application to protein function prediction in yeast. *Proceedings of*
486 *Pacific Symposium on Biocomputing*.
- 487 Mihaylova, L, Boel, R, Hegyi, A (2007) Freeway traffic estimation with particle filtering
488 framework. *Automatica*, 43(2). 290 – 300.
- 489 Ngoduy, D. (2011). Low rank unscented Kalman filter for freeway traffic estimation
490 problems. *Transportation Research Record*, 2260, 113-122.
- 491 Olkin, I (1992) Meta-analysis: Methods for combining independent studies, *Statistical*
492 *Science*, 7(2), 226 – 236.
- 493 Ou, Q (2011) Fusing Heterogeneous Traffic Data: Parsimonious Approaches using Data-Data
494 Consistency. *PhD thesis, Delft University of Technology, The Netherlands*.
- 495 Papageorgiou, M, Blosseville, J, Haj-Salem, H (1990) Modelling and real-time control of
496 traffic flow on the southern part of Boulevard Peripherique in Paris. *Transportation*
497 *Research Part A*, 24, 345 – 359.
- 498 Robinson, S, Polak, J (2006) Overtaking rule method for the cleaning of matched license
499 plate data. *ASCE Journal of Transportation Engineering*, 132(8), 609 – 617.
- 500 Tanaka, H, Uejima, S, Asai, K (1982) Linear regression analysis with fuzzy model. *IEEE*
501 *Transaction on Systems Management and Cybernetics*, 12(6), 903 – 907.
- 502 Treiber, M, Helbing, D (2002a) Reconstructing the spatio-temporal traffic dynamics from
503 stationary detector data. *Cooper@tive Tr@nsport@tion Dyn@mics*, 1, 3.1–3.24.
- 504 Treiber, M, Helbing, D (2002b) An adaptive smoothing method for traffic state identification
505 from incomplete information. *arXiv:cond-mat/0210050*.
- 506 Treiber, M, Kesting, A, Wilson, RE (2009) Reconstructing the traffic state from fusion of
507 heterogeneous data. *Computer-aided Civil and Infrastructure Engineering*, 26(6),
508 408–419.

- 1
2
3 509 Tsapakis, I., Turner, J, Heydecker, B, Emmonds, A, Cheng, T, Bolbol, A (2012) Effects of
4 510 Tube Strikes on Journey Times in Transport Network of London. *Transportation*
5 511 *Research Record*, 2274, 84 – 92.
6 512 Van Lint, JWC (2010) Empirical evaluation of new robust travel time estimation algorithms.
7 513 *Transportation Research Record*, **2160**, 50-59.
8 514 Van Lint, JWC, Hoogendoorn, SP (2009) A robust and efficient method for fusing
9 515 heterogeneous data from traffic sensors on freeways. *Computer-aided Civil and*
10 516 *Infrastructure Engineering*, **25**(8), 596-612.
11 517 Wang, Y and Papageorgiou, M (2005) Real-time freeway traffic state estimation based on
12 518 extended kalman filter: a general approach. *Transportation Research Part B*, 39(2),
13 519 141-167.
14 520
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review