An algorithm for integrating heterogeneous urban traffic data

Andy H.F. Chow*, Ioannis Tsapakis, Ed Manley
Garavig Tanaksaranond, Ben Heydecker, Tao Cheng
Civil, Environmental and Geomatic Engineering,
University College London, London WC1E 6BT, United Kingdom

Andy Emmonds
Network Performance, Traffic Directorate,
Transport for London, London SE1 8NJ, United Kingdom
*Email: andy.chow@ucl.ac.uk

April 4, 2012

1 Introduction

A detailed and reliable spatio-temporal picture of traffic is essential for understanding and managing urban congestion. This paper presents a data fusion algorithm for analysing urban road networks. Compared with motorways, less research has been done on urban area due to the limited availability of relevant data and complexity of the problem. Recently, the increasing availability of urban traffic data provides new research opportunities. The objective of the present study is to produce a refined and accurate picture of traffic through processing and integrating heterogeneous data collected from different sources in urban area. The fusion algorithm can work with data collected in different spatio-temporal granularity, with different level of accuracy, and from different kinds of sensors. The methodology is illustrated through an application in London, UK.

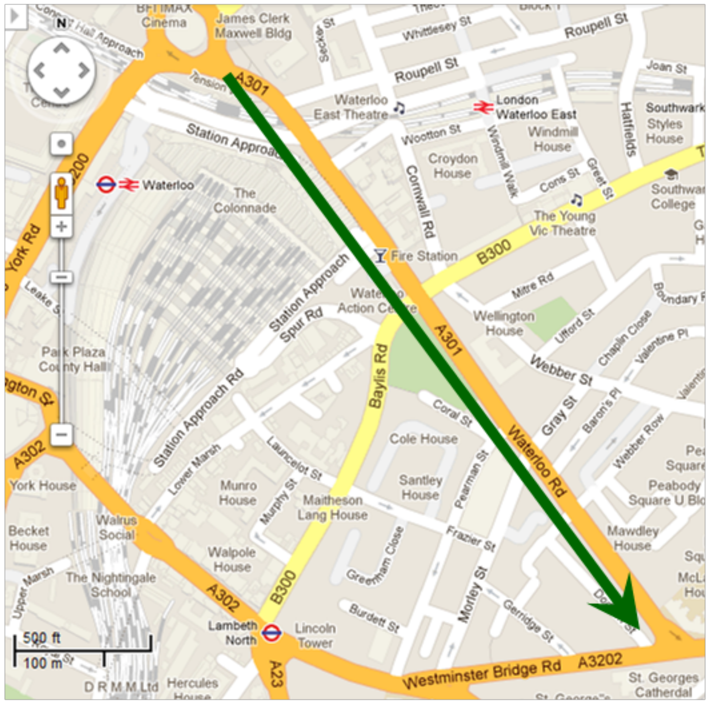

Figure 1: Waterloo Road, London (UK)

2 Urban traffic data

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

Figure 1 shows a 1200-meter stretch of Waterloo Road (A301) in Central London, and Figure 2 shows the associated speeds (i.e. reciprocals of journey times) measured along the road in April 2010. In the figure, the position and width of the bar at each 5-min interval reflects the average and dispersion of the journey times in the month at that particular interval.

Figure 3 shows the speed contour over time and space produced with ANPR journey times. As shown in the figure, a major weakness with the ANPR journey time estimates is

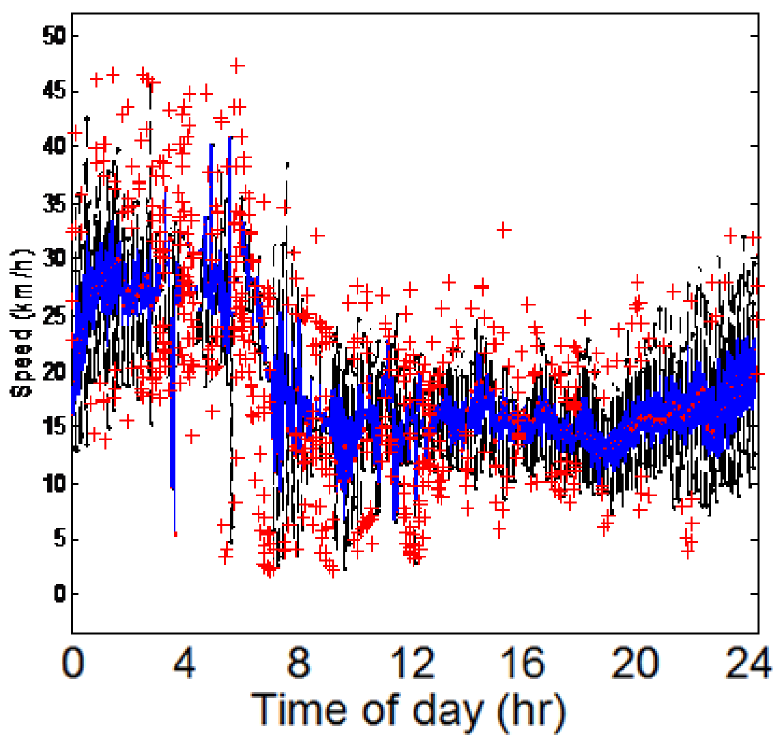

Figure 2: Variations of traffic speeds in April 2010 along Waterloo Road, London (UK)

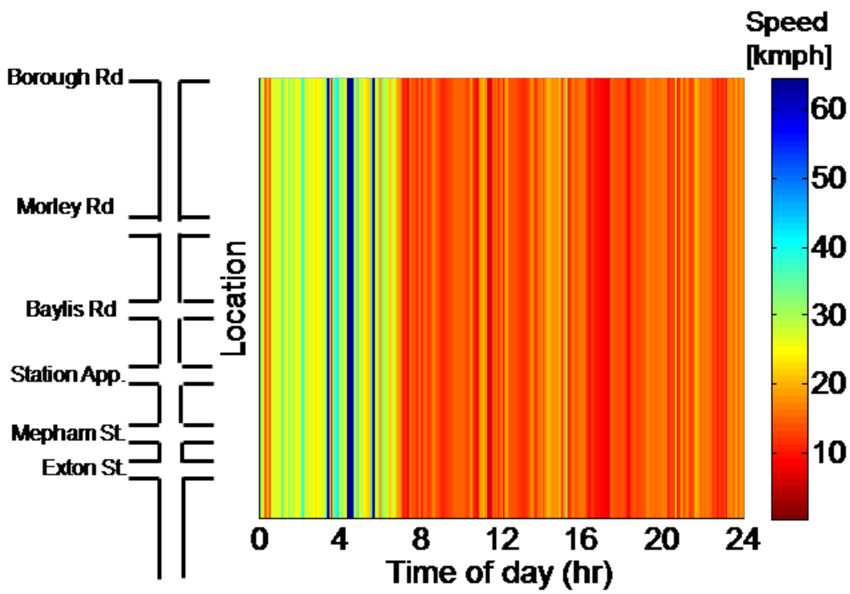

Figure 3: ANPR-Speed contour along Waterloo Road, London (UK)

that they do not capture much spatial feature of traffic. It is because the distance between a pair of ANPA camera sites is typically in the range of kilometres, which mean it could miss a lot of spatial variations in urban context.

To extract further spatial feature of traffic, UK Transport for London (TfL) uses floating car information provided by `Trafficmaster`[1]. `Trafficmaster` derives journey times from vehicles equipped with their GPS (Global Positioning System) devices. The `Trafficmaster` GPS devices report the locations of the vehicles on a regular basis (∼8-10 seconds). `Trafficmaster` further processes and stores the derived journey times into 15-min averages.

Figure 4 shows the corresponding speed contour with `Trafficmaster` data, which can be seen to reveal much more spatial feature than ANPR. Nevertheless, there are only very limited samples of `Trafficmaster` vehicles on the road (about 1,500 such vehicles in London Area). With such small sample size, TrafficMaster can only produce quarterly averages for meaningful statistical estimates. This implies that `Trafficmaster` can only reveal very limited temporal characteristics of traffic.

[1] http://www.trafficmaster.co.uk/

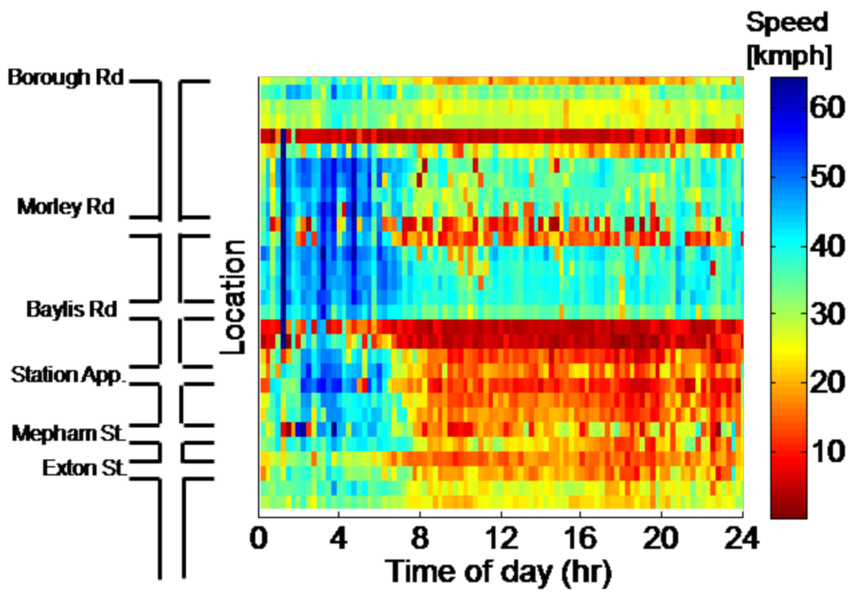

Figure 4: Trafficmaster-Speed contour along Waterloo Road, London (UK)

3 Data fusion

We present an algorithm to process and combine the ANPR and `Trafficmaster` data to produce a traffic pattern with fine spatial and temporal granularity. It should be emphasized that the algorithmic framework presented herein is general, which can incorporate other data kinds such as loop detectors and floating car trajectories apart from ANPR and `Trafficmaster` journey time data. The data fusion algorithm consists of two components: data 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 traffic information on a common space-time domain. We adopt the adaptive smoothing method (ASM) proposed by Treiber and Helbing [1] for this purpose.

Consider a set of traffic measurements u_i from a particular data source taken as location x_i and time t_i, where i = 1,2,, n and n is the total number of measurements taken by the source. Let (x, t) be the (space-time) coordinate on a new space-time domain defined by the analyst, the associated traffic state $u(x, t)$ on this new space-time domain can be derived from the measurements as

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

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 function, which depends on the spatio and temporal difference between the data point (x_i, t_i) and the point of interest (x,t). $M(x,t)$ is number of data points that we consider when calculating $u(x,t)$. The choice of $M(x,t)$ is a trade-off between computational speed and accuracy. Considering causality, we adopt a dynamic programming technique for solving $u(x,t)$ forward in time to minimize $M(x,t)$ and hence the computational effort.

3.2 Data integration

After reconstructing the traffic data onto a new and common space-time domain for all data sources, we can then integrate the data from these sources as:

$$\tilde{u}(x,t) = \frac{1}{E(x,t)} \sum_{k=1}^{K(x,t)} \epsilon_k(x,t) u_k(x,t), \tag{2}$$

where $u_k(x,t)$ is the smoothed and reconstructed data field from data source k; $\tilde{u}(x,t)$ is the eventual integrated data field; $E(x,t) = \sum_{k=1}^{K(x,t)} \epsilon_k(x,t)$ is a normalizing factor, in which $\epsilon_k(x,t)$ reflects the accuracy or reliability of data source k at (x,t); $K(x,t)$ is the number of data sources contributing to the estimate $\tilde{u}$ at (x,t). The data accuracy (or error) ϵ_k depends on various factors such as the sample size, variance of observations, and quality of the source data (e.g. level of imputation and patching).

4 Results and further validation

Using the data fusion algorithm presented above, Figure 5 shows the corresponding fused traffic pattern of Waterloo Road. It shows that the fusion algorithm is able to retrieve much spatio-temporal feature of traffic. It is also able to smooth out the corners (see Figure 4) observed in the `Trafficmaster` dataset through the underlying Treiber and Helbing's ASM.

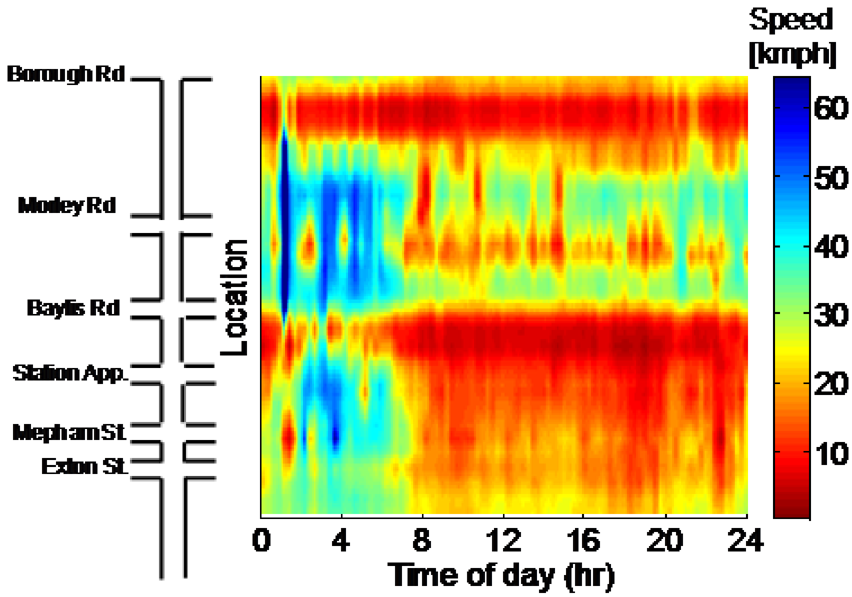

Figure 5: Fused traffic pattern along Waterloo Road, London (UK) with ANPR and Trafficmaster data

Sensitivity on the formulations (e.g. use of different kernels) and the associated parameters will be explored. The traffic pattern produced by the fusion algorithm will also be validated against external datasets such as urban loop detector (under SCOOT adaptive signal control system) and floating car data. In particular, there are two set of GPS floating car data available for validation purpose: taxi trajectories provided by `Addison Lee` and bus trajectories produced by `iBus` system. Details of the sensitivity analysis and validation will be reported in the full paper.

References

[1] Treiber, M., and D. Helbing, "Reconstructing the spatio-temporal traffic dynamics from stationary detector data", *Cooper@tive Tr@nsport@tion Dyn@mics* 1, 3.13.24, 2002.