ZHENYU MEI, Ph.D.
E-mail: meizhenyu@zju.edu.cn
Zhejiang University
Department of Civil Engineering
Zijingang Campus, Hangzhou, 310058, China
DIANHAI WANG, Ph.D.
E-mail: wangdianhai@zju.edu.cn
Zhejiang University
School of Traffic and Transportation Engineering
Zijingang Campus, Hangzhou, 310058, China
JUN CHEN, Ph.D.
E-mail: chenjun@seu.edu.cn
WEI WANG, Ph.D.
E-mail: wangwei@seu.edu.cn
Southeast University
School of Traffic and Transportation Engineering
Sipailou Campus, Nanjing, 210096, China

INVESTIGATION OF BICYCLE TRAVEL TIME ESTIMATION USING BLUETOOTH SENSORS FOR LOW SAMPLING RATES

ABSTRACT

Filtering the data for bicycle travel time using Bluetooth sensors is crucial to the estimation of link travel times on a corridor. The current paper describes an adaptive filtering algorithm for estimating bicycle travel times using Bluetooth data, with consideration of low sampling rates. The data for bicycle travel time using Bluetooth sensors has two characteristics. First, the bicycle flow contains stable and unstable conditions. Second, the collected data have low sampling rates (less than 1%). To avoid erroneous inference, filters are introduced to "purify" multiple time series. The valid data are identified within a dynamically varying validity window with the use of a robust data-filtering procedure. The size of the validity window varies based on the number of preceding sampling intervals without a Bluetooth record. Applications of the proposed algorithm to the dataset from Genshan East Road and Moganshan Road in Hangzhou demonstrate its ability to track typical variations in bicycle travel time efficiently, while suppressing high frequency noise signals.

KEY WORDS

Bluetooth sensor; bicycle travel time; low sampling rates; filtering algorithm;

1. INTRODUCTION

Bicycles are the most popular travel means in China, which contribute to 30% to 70% of daily trips. Accurate travel time information acquisition is essential to the effective planning and management of bicycle travel conditions. Traditionally, video camera data have been used as the primary source for measuring the quality of bicycle travel time, which have been expensive and invasive to privacy. In recent years, a novel technology that uses bypassing Bluetooth devices to capture vehicle travel times has been proven to be effective [1-3]. With the increased penetration of Bluetooth on the roadways in China, nearly 2% to 3% of bicycles in Hangzhou had active Bluetooth devices in 2010. These sensors became a viable alternative to other methods of bicycle travel time data collection [4-6].

Bluetooth is a short-range communications protocol developed by the Special Interests Group for inter-device communications. This protocol creates some challenges for its additional uses, which could be optimized for travel time collection based on Bluetooth MAC (Media Access Control) address matching [7].

Bluetooth sensor systems may provide time stamps at which bicycles pass successive monitoring sensors; however, the data gathered by these systems require some form of filtering to remove outlier readings. Frequently, the outlier data include bicycles making a stop or taking a detour between two locations. These bicycles experience atypical travel times; therefore, these readings should be removed from the dataset of valid readings to avoid producing erroneous travel time estimates. Filtering algorithms are often based on removing the readings with unreasonably low speeds [8-11].

However, this approach has several disadvantages. The algorithm does not handle both stable (constant mean) and unstable (varying mean) traffic conditions, especially for bicycles flow. Moreover, the algorithm cannot be applied to low levels of market penetration (less than 1%). In response to the above issues, the current research has been dedicated to the following aspects:

1. Analysis of bicycle travel time error using Bluetooth sensors and problem formulation;
2. Description of the proposed filtering algorithm for travel time estimation using Bluetooth sensors;
3. Development of a filtering algorithm that considers low sampling rates;
4. Verification and implementation of the proposed method to the data from Genshan East Road and Moganshan Road in Hangzhou.
   The advantages of the proposed method are:
1. The proposed algorithms are simple and do not need a lot of computer resources for their realization;
2. The proposed algorithms can handle both stable (constant mean) and unstable (varying mean) traffic conditions, and correct abrupt changes in the means of bicycle travel times and spurious spikes;
3. Using the obtained coefficients described in the section of considering low sampling rates, the proposed algorithms can be applied to low levels of market penetration (less than 1%);
4. The proposed algorithms can be used not only for offline, but also for online data filtering.

2. PROBLEM FORMULATION

Obtaining travel time measurements using Bluetooth devices involves matching an observed MAC address between two locations. The difference in time between the two readings is the travel time [12]. Bluetooth readers are capable of detecting MACs within a specific range; hence, the travel times obtained can be considered zone-to-zone. This is illustrated in *Figure 1* where the dashed lines represent the Bluetooth detection zone. The average travel times obtained from the Bluetooth sensors can be expressed as:

$$M_{BT}(j,L) = \frac{\sum_{j=1}^{n_L}(t_{j,A}(L) - t_{j,B}(L))}{n_L} \qquad (1)$$

where $M_{BT}(j,L)$ is the Bluetooth-based average travel time between nodes A and B during period L, respectively; n is the number of readings; t is the time stamp when a bicycle or Bluetooth device is detected. The Bluetooth-equipped bicycle could be detected any time within the detection zone. The timestamp, t, for the Bluetooth-based method is defined as the time of the first discovery of the Bluetooth device no matter how many times it is detected afterwards.
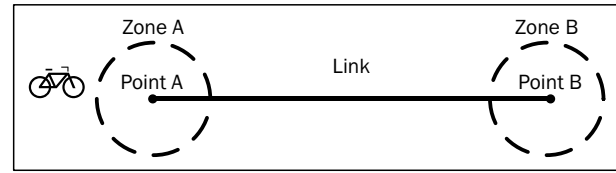

Figure 1 - Segment composition

The authors realized that the travel time reported by the Bluetooth device will be subject to the following sources of error: spatial error, temporal error and sampling error. Spatial error is caused by the variation of detection zone radius. Temporal error is caused by multiple times detection. Sampling error is caused by other mode of transportation being detached [13].

Bluetooth MAC address-based travel times are expected to contain outliers. These outliers are usually caused by drivers who detour or stop at certain places between detector locations rather than being delayed by the bicycle stream. The existing noise removal procedure can be relatively trivial because only bicycles comprise the relatively large travel time data set. However, the Bluetooth device still contains several errors. A median offline filtering algorithm is designed to extract ground truth from the pool of Bluetooth readings. A detailed comparison between the proposed filter and an alternative method by Dion and Rakha [14] can be found in Mei et al. [15].

3. PROPOSED FILTERING ALGORITHM

The proposed filtering algorithm is the Moving Median Filter (MMF). A window of size L is "moving" from the beginning to the end of the time series with unit step k and the median of L values falling within the window are continuously calculated. As shown in Eq. (2), the technique estimates the median of L successive travel time values. A similar process is used to estimate the standard deviation in Eq. (3).

$$M_{iL} = Median(V_{i-L/2}, \ldots, V_{i+L/2}) \qquad (2)$$

$$\sigma_k = \sqrt{\alpha(\sigma_{k-1}^2) + (1-\alpha)\cdot(\sigma_{Sk-1}^2)} \qquad (3)$$

where, $M_{i,L}$ is the Median of L successive travel time values, where time i is the middle point of window; V_i is the travel time of bicycles at moment i; σ_k is the Standard Deviation (SD) of time series in the sampling interval k; σ_{Sk} is the smoothed SD of time series that is estimated in the sampling interval k with the use of the filter; α is the smoothing factor (varying between 0 and 1, based on the number of valid readings in the current sampling interval).

The flowchart for MMF is presented in *Figure 2*, and a detailed description of the algorithm is introduced below.

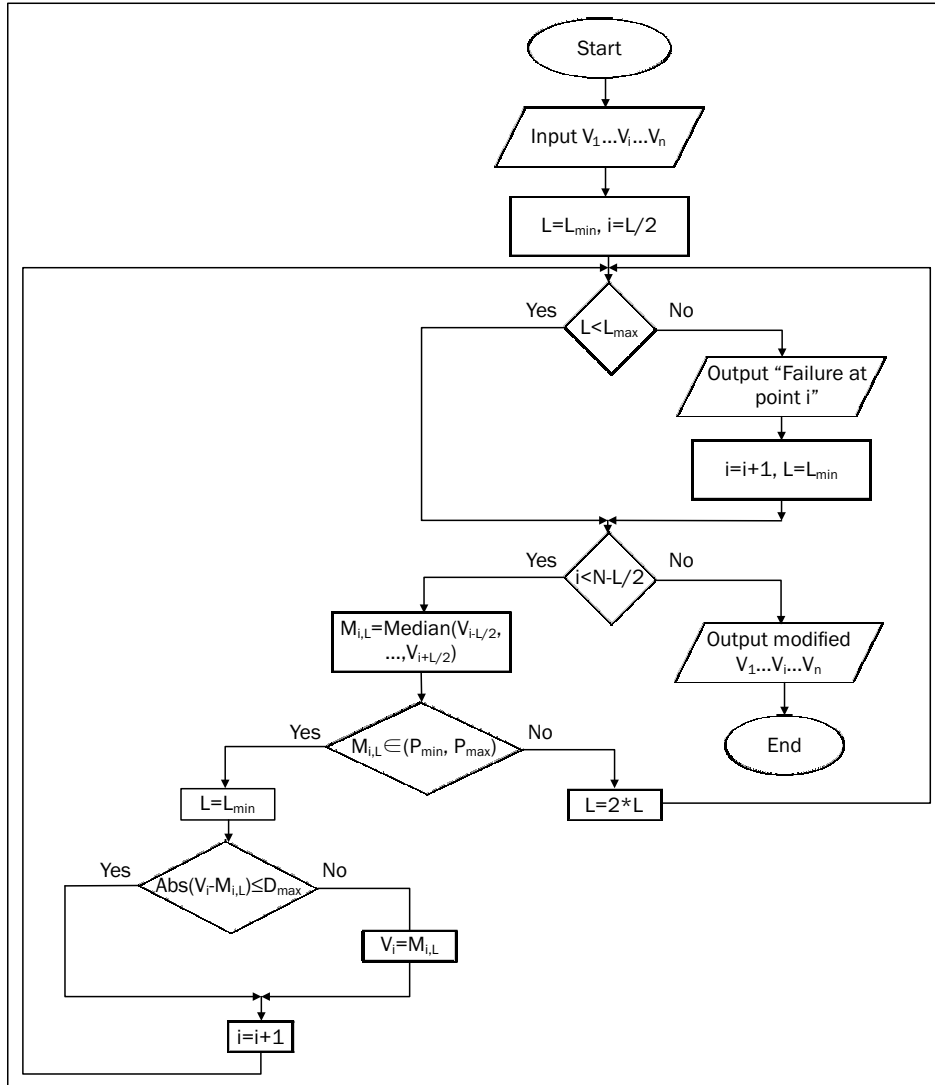Step 1: Select time series from database with N elements.

Figure 2 - Flowchart of the proposed MMF

Step 2: Start algorithm operation, assign $L = L_{min}$ and $i = L/2$. (L is the width of the moving window; L_{min} is the minimal width of the window).

Step 3: If $L < L_{max}$, continue; otherwise, report program failure at point *i*, increment *i* by 1, assign $L = L_{min}$. (L_{max} is the maximal width of the window),

Step 4: If $i < (N - L/2)$, calculate median value $M_{i,L}$ for interval $[i - L/2, i + L/2]$; otherwise, write filtered time series into database, stop.

Step 5: If $M_{i,L} \in [P_{min}, P_{max}]$, assign $L = L_{min}$, continue; otherwise, enlarge L by factor of 2 and go to step 3.(Maximal and minimal limits of $M_{i,L} - P_{max}, P_{min}$;)

Step 6: If absolute value of $(v_i - M_{i,L}) > D_{max}$, assign $v_i = M_{i,L}$. (Limit of maximal deviation of current count rate from median value – D_{max}.)

Step 7: Increment *i* by 1, go to step 3.

This filter is optimal for the time series containing abrupt changes in the mean, followed by recovery.

The flowchart of the proposed MMF is presented in *Figure 2*.

Figures 3 (a) and (b) illustrate the application of filtering algorithm to survey data with Bluetooth sensor on Genshan East Road, and Figures 3 (c) and (d) illustrate the application on Moganshan Road. Genshan East Road and Moganshan Road are both four-lane arterials with four inside vehicle lanes and two bicycle outside lanes. The average bicycle volumes of two roads exceed 1,000 bic/h which provides good conditions for detecting bicycle data. Figures 3 (a), (b), (c) and (d) illustrate how the search window enhances the performance of the proposed algorithm. The size of the validity window is defined by the number of standard deviations which is user-definable in the algorithm. However, the basic validity ranges encompassing two or three standard deviations are envisioned to be utilized. The use of two-standard-deviations wide search window would mean that all data points
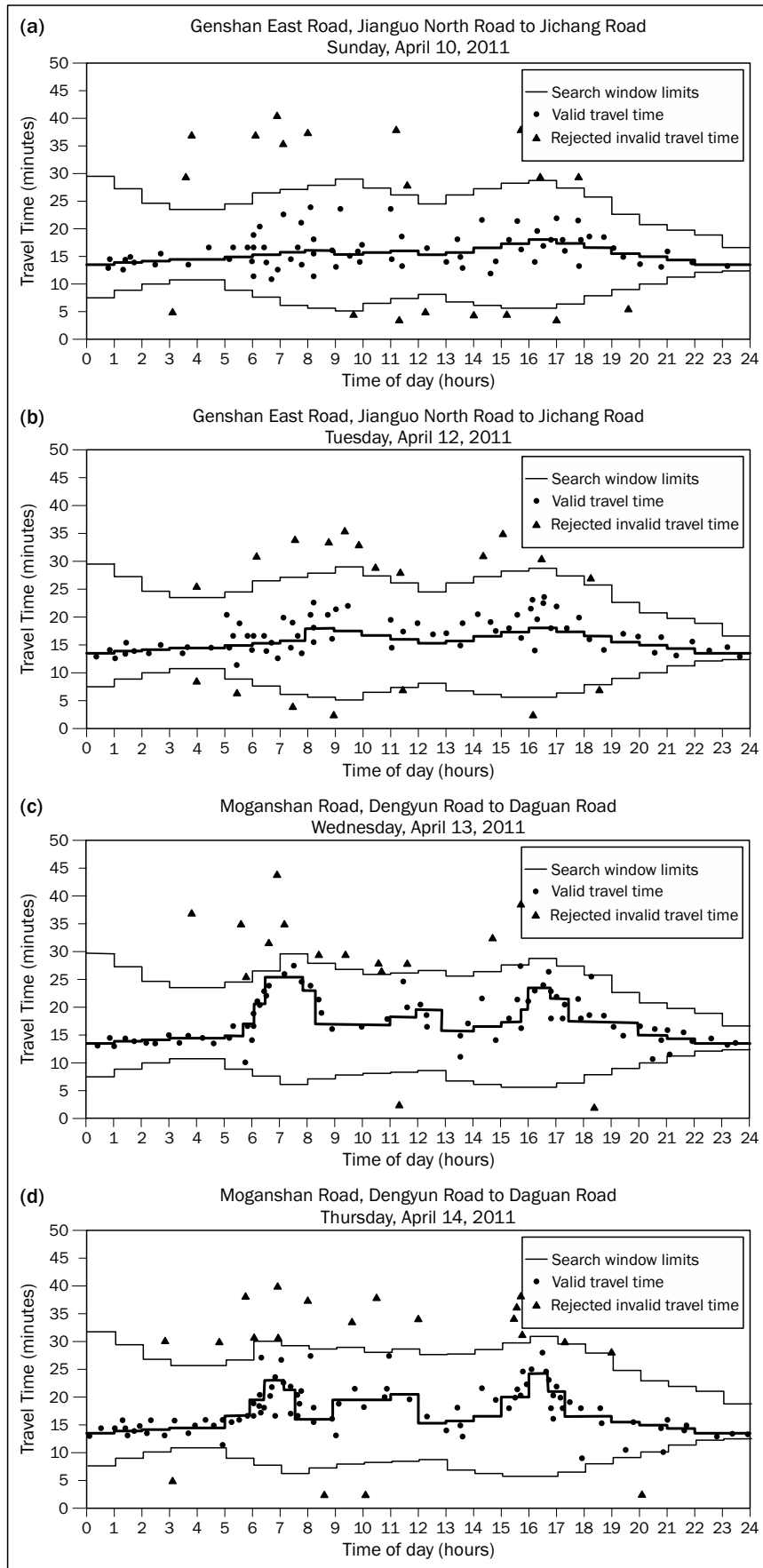
Figure 3 - Application of MMF filtering algorithm to survey data with Bluetooth sensor

within a 95% lognormal confidence interval would be considered valid, and that all other points falling outside this range would be rejected from consideration when estimating average link travel times. The use of a validity window three-standard-deviations wide validity window would mean that all data points within a 99% confidence interval would be considered valid. Through using the four sets of data of two arterials, the figures illustrate the variation in the interval average travel time, as well as the lower and upper limits of the validity window, after the consideration of trends outside the basic validity range.

4. CONSIDERATION OF LOW SAMPLING RATES

To further improve the filtering algorithm for the consideration of low sampling rates, additional enhancements are made. Given the variable nature of traffic flow, estimating the travel times during an interval while using data collected in the previous intervals does not ensure that the resulting estimates are truly representative of the real average trip time of the interval [5, 11, 16]. For instance, if bicycle flow is slowly increasing during a given portion of the day, the average travel time measured in consecutive intervals could be expected to increase gradually. Secondly, the assumption that the median travel time and standard deviation of the validity window would remain constant during intervals with no readings, as defined in Eqs. (2) and (3), could potentially result in the algorithm using outdated average travel times to determine the validity window limits.

As can be observed in *Figure 4*, the application of the filtering criterion defined by Eqs. (2) and (3) does not produce good results. The filtering criterion is unable to follow the underlying travel time increase between 6:30 and 8:30 a.m. This failure is caused by the combination of the abrupt increase in roadway travel times and no observations. Because travel times change rapidly at the onset of the period, a gap of a few minutes is sufficient in this case to raise the average link travel time beyond the validity bounds. The validity window would remain constant during intervals with no readings. Since all subsequent observations lie outside the validity window, they are considered invalid and rejected.

In this case, a period with no recorded travel time does not mean that there are no bicycles passing through the pair of Bluetooth sensors, but simply that no bicycles equipped with Bluetooth are travelling. To increase the responsiveness of the algorithm to the changes in traffic conditions and to situations with low sampling rates Eqs. (4) and (5) are introduced to modify the search window limits. Eqs. (4) and (5) compute for the number of standard deviations n_{σ_k} used in Eqs. (2) and (3) within each sampling interval k, based on the number of intervals with zero readings.

$$D_{\max} = n_{\sigma_k} \cdot \sigma_k \tag{4}$$

$$n_{\sigma_k} = \lambda + \lambda \left[1 - (1 - \beta_\sigma)^{n_{0_k}} \right] \tag{5}$$

where n_{0_k} is a user-defined parameter; λ represents the minimum number of standard deviations to consider, and a sensitivity parameter β_σ. Eq. (5) provides a model that dynamically adjusts the size of the validity window based on the number of preceding sampling intervals without Bluetooth record. For any interval in which at least one reading was made in the preceding interval, the equation defines a validity window that corresponds to the minimum size specified by the user, λ. If no readings were made in the previous interval, the size of the validity window is increased to $\lambda + \lambda(\beta_\sigma)$. If no readings were made in the two previous intervals, the validity window is further increased to $\lambda + \lambda \left[1 - (1 - \beta_\sigma)^2 \right]$. The size of the validity window will continue to increase with each preceding interval without readings, until the maximum size of 2λ is reached.
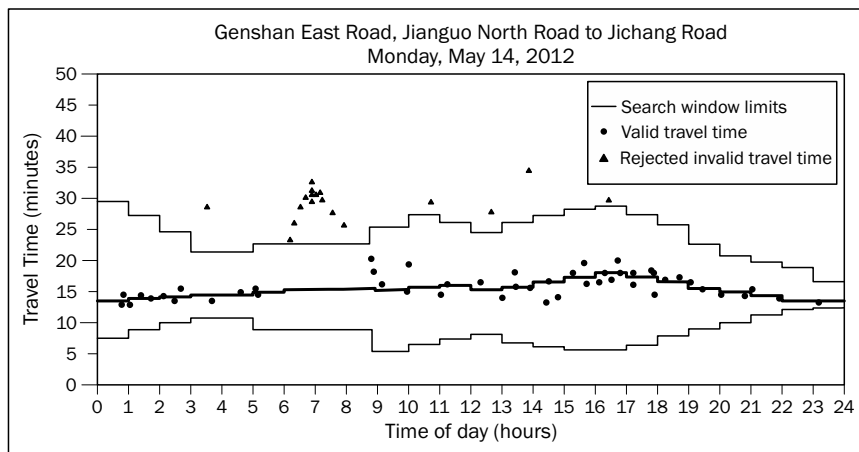

Figure 4 - Application of MMF filtering algorithm to survey data with Bluetooth sensor for low sampling rates
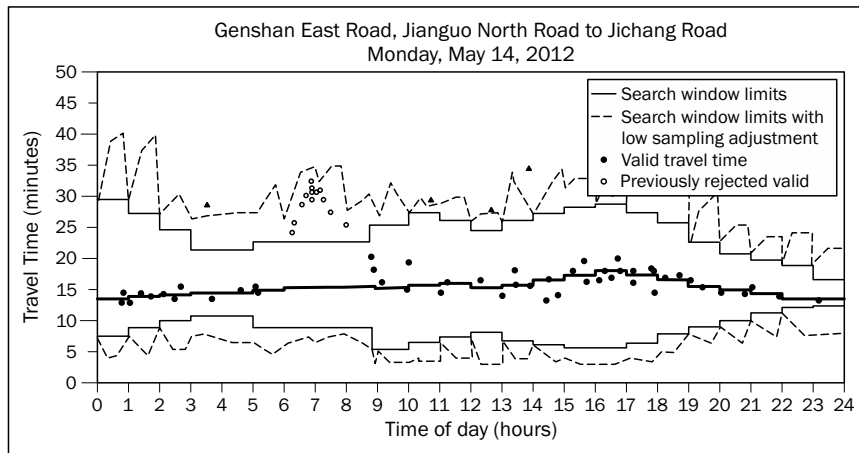
Figure 5 - Effect of low sampling search limits on the operation
of the proposed filtering algorithm

Figure 5 illustrates the effect of Eqs. (4) and (5) on the performance of the proposed algorithm. The figure compares various algorithms, considering an arterial road segment with a relatively low travel time sampling rate by applying a version of the filtering algorithm with only Eqs. (2) and (3), with a version that also includes Eqs. (4) and (5). There is a noticeable difference in the size of the validity window used by both versions of the filtering algorithm. Identical validity window limits are used by both versions of the filtering algorithm for all sampling intervals, for which travel time readings are made in the preceding interval. Increasing differences are observed for intervals preceded by an increasing number of intervals without readings.

In *Figure 5*, the effect of the modified filtering algorithm is particularly apparent in the intervals between 5:00 and 8:30. Within this period, thirteen travel times were observed, which were concentrated between 24 min. and 35 min. After the detection of the first bicycle passage, the minimum and maximum search window limits for the subsequent sampling interval, 5:00 to 6:00, are set at 8.66 min. and 24.44 min., respectively, by the MMF algorithm. As time passes without any additional readings, the filtering algorithm based only on Eqs. (2) and (3) maintains a fixed validity window, whereas the algorithm with Eqs. (4) and (5) gradually alters the size of the validity window to 4.02 min. and 35.62 min. by the time the 5:00 to 6:00 interval is reached. This thirteen travel times concentrated between 24 min. and 35 min. were considered as valid by the modified algorithm with Eqs. (4) and (5). A visual analysis of the time series of the recorded travel times appears to indicate that the thirteen travel time readings should be considered according to the Video cameras observations on spot.

5. ALGORITHM TESTING AND VALIDATION

To evaluate the ability of the proposed filtering algorithm to follow the fluctuations in the observed travel times correctly, the algorithms defined by Eqs. (2) and (3), and (4) and (5) were applied to two arterial roads in Hangzhou, Genshan East Road and Moganshan Road, as shown in *Figure 6*. The series consists of readings made on two arterial roads. The first is Genshan East Road, between Jianguo North Road and Jichang Road, over a period of two consecutive days in April 2011 and one day in May 2012. The second is Moganshan Road, between Daguan Road and Dengyun Road, over a period of two consecutive days in April 2011. The multiple days' readings of two arterial road segments are utilized to illustrate the performance of the proposed algorithm. The link of Genshan East Road is 2.28 km long and has an average bicycle volume of 1,125 bic/h, and the link of Moganshan Road is 2.34 km long and has an average bicycle volume of 1,260
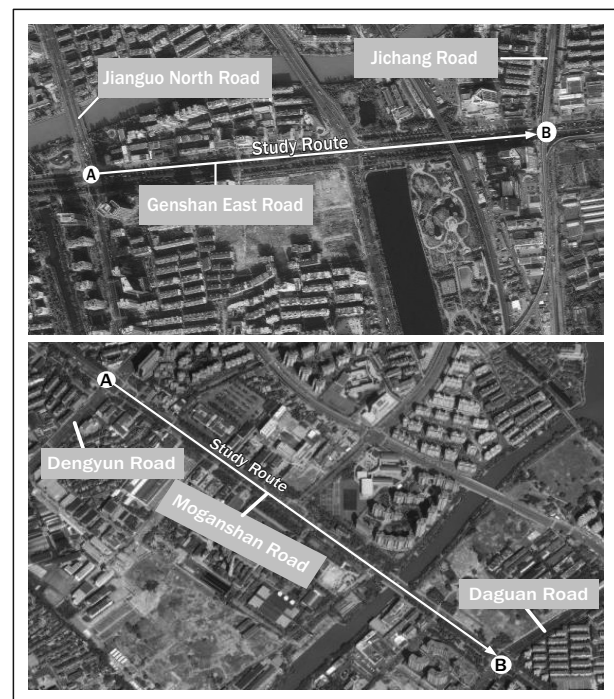

Figure 6 - Data collection sketch map

bic/h. Bicyclist behaviour was also captured through video cameras. Video cameras acquisition is one of the most reliable travel time measurement methods available and travel time data collected by video camera can be used for evaluating the performance of the Bluetooth-based approach.

Figures 7 (a), (b), (c), (d) and (e) illustrate the application results of the proposed filtering algorithm to the travel time series described above. The results shown in each figure are those associated with the calibrated parameters producing the best results. The average confidence interval is about 99.6%. The proposed filtering algorithm is able to track all major changes in the observed travel times correctly. There may be arguments about the validity of some readings; however, these are relatively few and do not appear to affect the underlying trend significantly. In particular, the figures show that the proposed filtering algorithm could effectively respond to abrupt increases and decreases in observed travel times. The ability to track these changes is particularly apparent in the data for April 13 in Figures 7 (a) and (b), which illustrate a significant change in travel times possibly caused by the onset of congestion due to an incident or illegal parking on the road.

To successfully demonstrate the ability to respond to the underlying trend, *Figure 7* shows, more specifi-
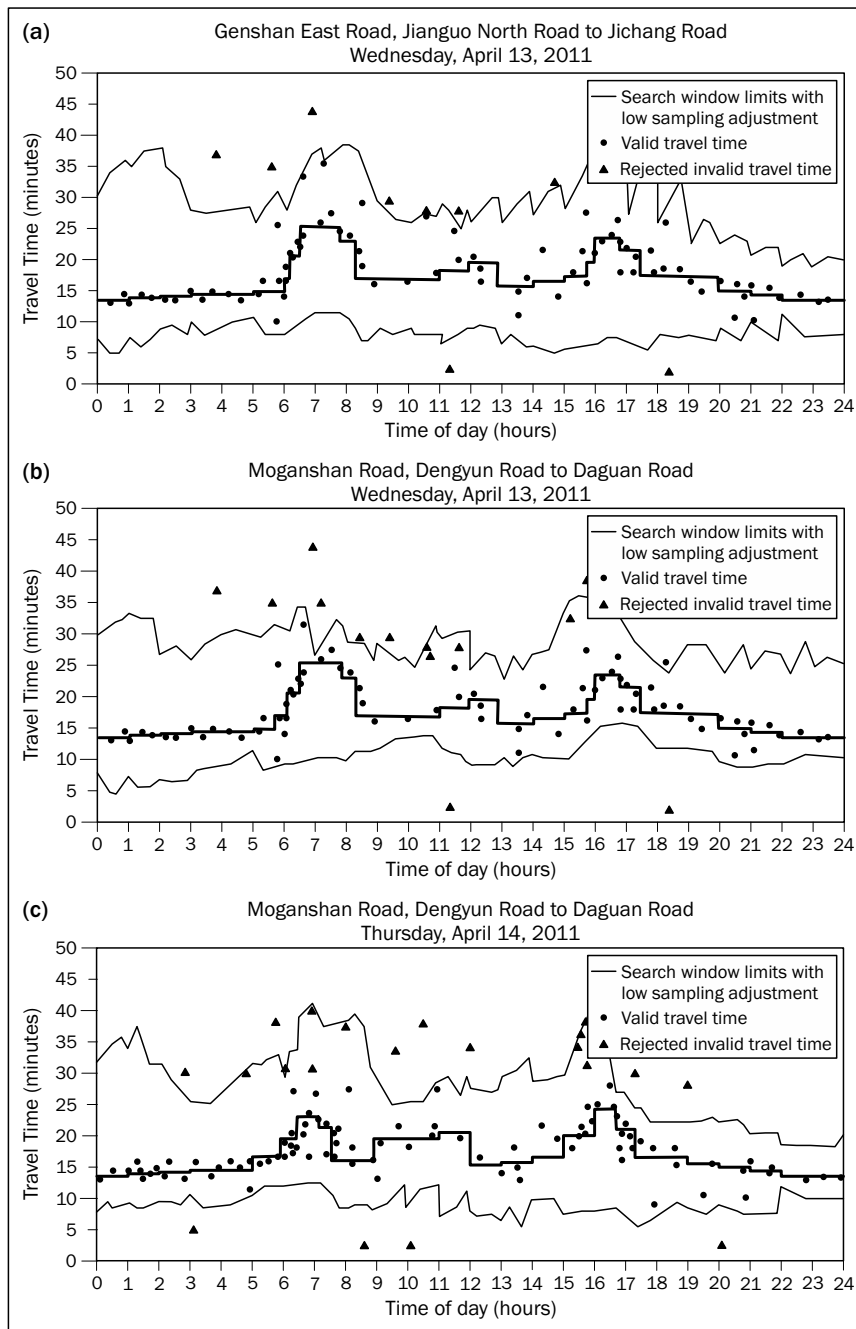

Figure 7 - Sample application to corridors for low sampling rates ($\lambda = 4$, $\beta_\sigma = 0.4$)
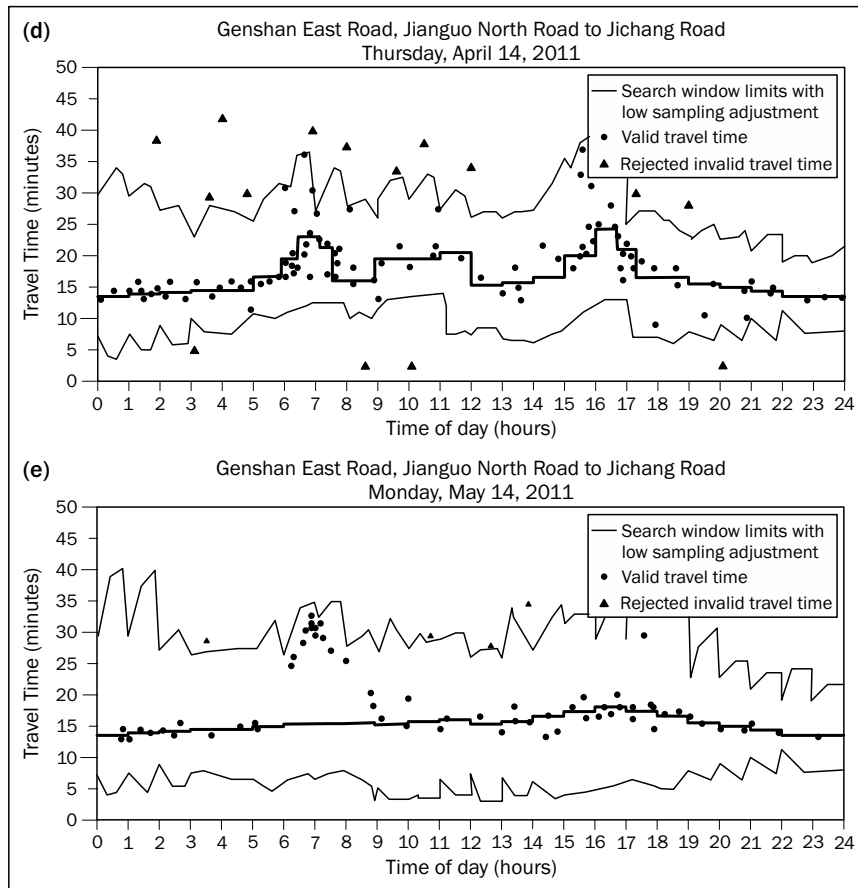
Figure 7 - Sample application to corridors for low sampling rates ($\lambda = 4$, $\beta_\sigma = 0.4$)
(continued)

cally, the ability of the proposed algorithm to follow general travel time fluctuations on links with very low sampling rates. As indicated earlier, this is an issue with the existing algorithms. As shown in the Figure, the total number of Bluetooth readings made throughout an entire day on the test arterial segments varies between 74 and 96, which are considered small numbers. As a result of these very low sampling rates, the diagrams in *Figure 7* of Genshan East Road in May 14, 2012 typically feature large intervals with no readings. In the diagram, there are typically no Bluetooth readings between 5:00 and 6:00. Aside from these two time points, periods of more than 30 min. without any reading is not uncommon. The presence of these large gaps greatly increases the task of the proposed filtering algorithm to identify underlying low frequency trends from the high frequency white noise, as well as to determine whether travel times observed after a large gap truly represent the existing bicycle flow conditions or should be considered invalid.

6. CONCLUSION

Filtering the data for bicycle travel time using Bluetooth sensors is crucial to the estimation of link travel times on a corridor. The current paper describes an adaptive filtering algorithm to estimate link travel times using Bluetooth data. The proposed algorithm overcomes a number of shortcomings in the existing algorithms by handling both stable and unstable bicycle flow conditions and functioning at different levels of market penetration (including less than 1% of the traffic volume). The proposed filtering algorithm determines the range of valid travel time readings within each sampling interval by considering the number of readings in the current and previous sampling intervals, as well as the number of consecutive readings outside the validity range. Applications of the proposed algorithm to the dataset of link travel times collected by Bluetooth sensors on Genshan East Road and Moganshan Road in Hangzhou demonstrate its ability to track the underlying roadway travel times correctly, while suppressing high frequency noise.

Despite the successful application of the proposed algorithm to the Bluetooth dataset, further tests are required to validate the findings of the current study. The use of historical data as an additional validation criterion or a means to provide travel time information during long intervals without travel time readings from Bluetooth readers should also be investigated.

ACKNOWLEDGMENT

梅振宇，博士，副教授，单位：中国浙江大学土木工程系
王殿海，博士，教授，单位：中国浙江大学土木工程系
陈峻，博士，教授，单位：中国东南大学交通学院
王炜，博士，教授，单位：中国东南大学交通学院

摘要

对蓝牙传感器采集的自行车旅行时间数据进行过滤对于主干路的连线间旅行时间预测非常重要。该论文提出一种可以适用于处理自行车低样本率的蓝牙数据来预测旅行时间的过滤算法。采集来的自行车蓝牙数据有两大特征，一是自行车车流数据包括稳定情况和不稳定情况。二是采集的数据为低抽样率（小于1%）。为了排除错误数据的干扰，首先提出一个过滤算法来处理多组数据，该算法通过设计一个鲁棒数据过滤器来控制动态的有效窗口大小，由此确定有效的数据。该动态窗口的大小变化是基于前面没有蓝牙数据的样本间隔数量。最后，将方法应用到杭州市的丘山东路和莫干山路，结果显示其能高效地应对自行车流的波动和数据的冗杂性。

关键词

蓝牙传感器，自行车旅行时间，低样本率，过滤算法

REFERENCES

[1] **Ahmed H**, **EL-Darieby M**, **Abdulhai B**, & **Morgan Y**. *Bluetooth and Wi-Fi-Based Mesh Network Platform for Traffic Monitoring*. Transportation Research Board 87th Annual Meeting. Transportation Research Board, Washington, D.C., 2008.

[2] **Haghani A**, **Hamedi M**, **Sadabadi, KF**, **Yound S**, & **Tarnoff PJ**. *Freeway Travel Time Ground Truth Data Collection Using Bluetooth Sensors*. Transportation Research Board 89th Annual Meeting. Transportation Research Board, Washington, D.C., 2010.

[3] **Sharifi E**, **Hamedi M**, & **Haghani A**. *Vehicle Detection Rate for Bluetooth Travel Time Sensors: A Case Study in Maryland and Delaware*. Transportation Research Board 89th Annual Meeting. Transportation Research Board, Washington, D.C., 2010.

[4] *Traffic Congestion and Reliability*, FHWA (Federal Highway Administration), U.S. Department of Transportation, Sept. 2005.

[5] *Urban Traffic Report*, China Transportation Institute. Sept. 2010.

[6] **Shan X**. *A Research on Urban Bicycle Transportation Rational Ridership and Road Resource Allocation* (Doctoral dissertation). Southeast University, China, 2007.

[7] *What Is Traffic Message Channel?* http://www.tmcforum.com. Accessed July 2009.

[8] **Brennan TM**, **Day CM**, **Wasson JS**, **Sturdevant JR**, & **Bullock DM**. *Assessing Signal Timing Plans for Winter Conditions*. ITE Learned Journal of Transportation,Washington,DC,1(1), 2011, p. 59-76.

[9] **Brennan TM**, **Ernst JM**, **Day CM**, **Bullock DM**, **Krogmeier JV**, & **Martchouk M**. *Influence of Vertical Sensor Placement on Data Collection Efficiency From Bluetooth MAC Address Collection Devices*. Journal of Transportation Engineering, 136(12), 2011, p. 1104–1109.

[10] **Bullock D**, **Haseman R**, **Wasson J**, & **Spitler R**. *Anonymous Bluetooth Probes for Measuring Airport Security Screening Passage Time: The Indianapolis Pilot Deployment*. Transportation Research Board 89th Annual Meeting. CD-ROM. Transportation Research Board, Washington D.C, 2010.

[11] **Wasson JS**, **Sturdevant JR**, & **Bullock DM**. *Real-Time Travel Time Estimates Using Media Access Control Address Matching*. Institute of Transportation Engineers Journal, 78(6), 2008, p.20–23.

[12] **Malinovskiy Y**, **Wu Y**, **Wang Y**, & **Lee U**. *Field Experiments on Bluetooth-based Travel Time Data Collection*. Transportation Research Board 87th Annual Meeting. CD-ROM. Transportation Research Board, Washington, D.C., 2010.

[13] **Mei Z**, **Wang D**, **Chen J**. *Investigation with Bluetooth Sensors of Bicycle Travel Time Estimation on a Short Corridor*. International Journal of Distributed Sensor Networks, 2012, 2012.

[14] **Dion F**, & **Rakha H**. *Estimating Dynamic Roadway Travel Times Using Automatic Vehicle Identification Data for Low Sampling Rates*. Transportation Research Part B, 40(9), 2006, p.745–766.

[15] **Mei Z**, & **Tian B**. *Real-Time Travel Time Estimation: Filtering Raw Data in an Automatic Vehicle Identification Setting*. In Proceedings of the First International Conference on Transportation Engineering, Chengdu, China, 2007, p.34-39.

[16] **Kothuri SM**, **Tufte KA**, **Fayed E**, & **Bertini RL**. *Toward Understanding and Reducing Errors in Real-Time Estimation of Travel Times*. Transportation Research Record: Journal of the Transportation Research Board, 2049, 2008, p. 21-28.