

통행시간 추정을 위한 Voting Rule과 중위절대편차법 기반의 복합 필터링 모형

Combined Filtering Model Using Voting Rule and Median Absolute Deviation for Travel Time Estimation

정 영 제* 박 현 석** 김 병 화*** 김 영 찬****
(Youngje Jeong) (Hyun Suk Park) (Byung Hwa Kim) (Youngchan Kim)

요 약

본 연구에서는 교통정보시스템에서 통행시간의 이상치 자료를 제거하기 위한 복합 필터링 모형을 제시하였으며, 이는 중위절대편차법과 Voting Rule을 기반으로 하는 이중화된 필터링 모형에 해당한다. 본 모형은 중위절대편차법을 이용해 표본을 정규분포화 시키기 위한 1차 필터링을 수행하며, 이후 Voting Rule을 이용해 중위절대편차법의 적용 이후에도 남아 있는 이상치 자료를 제거하는 방식에 해당한다. 이때 Voting Rule은 표본의 통행시간과 평균통행시간의 차이가 임계치를 초과하는 경우 해당 표본을 이상치로 판정하며, 다수결의 원칙을 이용하여 이상치 자료의 비율에 따라 이상치에 대한 제거 여부를 결정한다. 일반국도 3호선의 경기도 광주시 구간을 대상으로 한 사례분석을 통해 복합 필터링 모형이 이상치 표본만을 선택적으로 제거하여 통행시간 추정의 정확도 개선에 기여할 수 있음을 확인하였다.

핵심어 : 통행시간, 이상치 제거, 필터링, 중위절대편차, Voting Rule

Abstract

This study suggested combined filtering model to eliminate outlier travel time data in transportation information system, and it was based on Median Absolute Deviation and Voting Rule. This model applied Median Absolute Deviation (MAD) method to follow normal distribution as first filtering process. After that, Voting rule is applied to eliminate remaining outlier travel time data after Median Absolute Deviation. In Voting Rule, travel time samples are judged as outliers according to travel-time difference between sample data and mean data. Elimination or not of outliers are determined using a majority rule. In case study of national highway No. 3, combined filtering model selectively eliminated outliers only and could improve accuracy of estimated travel time.

Keywords : Travel Time, Outlier Elimination, Filtering Model, Median Absolute Deviation, Voting Rule

* 주저자 : 서울시립대학교 교통공학과 연구교수
** 공저자 : 한국건설기술연구원 연구원
*** 공저자 : 한국건설기술연구원 연구위원
**** 공저자 및 교신저자 : 서울시립대학교 교통공학과 교수
† 논문접수일 : 2013년 10월 21일
‡ 논문접수일 : 2013년 11월 11일
‡ 논문접수일 : 2013년 11월 22일

I. 서론

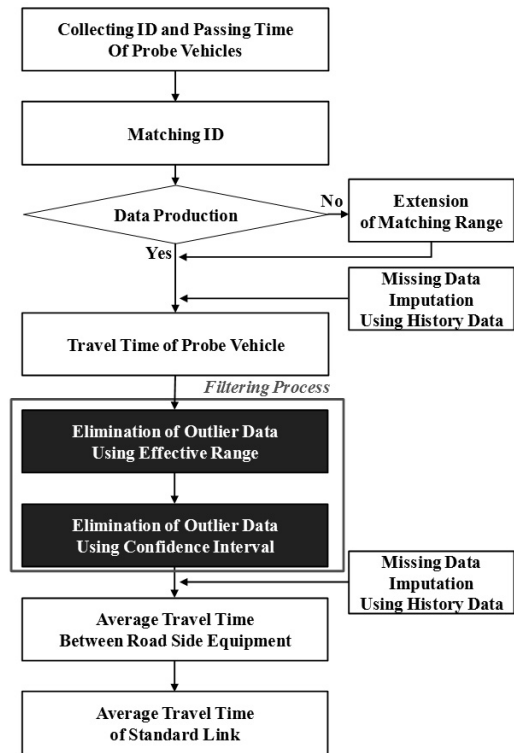
국내 일반국도의 지능형교통체계 구축은 1997년 국도 3호선 성남~이천 구간에서 시작되었으며, 2012년 현재 수도권 전역의 17개 노선, 859km에서 ITS를 이용한 교통운영관리가 수행되고 있다[1, 2]. 국도 ITS 운영 구간 중 서울지방국토관리청 도로교통정보센터를 기준으로 DSRC를 주검지체계로 이용하는 경우 하이패스 단말기 부착차량을 수집원으로 하며, Probe의 ID와 통과시간 정보를 수집하여 개별차량의 구간통행시간을 생성하게 된다[3].

DSRC 기반의 통행시간 가공체계에서는 하이패스 단말기 부착차량에 해당하는 Probe의 통행시간 정보를 이용함에 따라 현장의 소통상황이 반영된 유효한 표본을 수집하는 과정이 중요하다. 그러나 일반국도에서는 배달차량, 휴게소 이용차량 등으로 인해 통행시간 패턴과는 상이한 이상치 자료들이 수집될 수 있으며, 이를 제거하기 위한 필터링 과정이 수반된다. 표본의 통행시간을 이용하는 통행시간 가공방식에서는 필터링 성능이 통행시간 추정결과의 정확도에 영향을 미치게 되며, 수도권 일반국도와 같이 토지이용이 활발한 지역에서는 필터링 이후에도 여전히 제거되지 않은 이상치 자료로 인해 통행시간의 정확도가 저하될 수 있다. 이러한 현상은 필터링 방식에 따라 정도의 차이는 있으나, 이상치 자료만을 완벽히 제거하는 데는 한계가 있다.

본 연구에서는 표본의 정보로부터 구간통행시간을 추정하는 과정에서 사용되는 필터링 방식을 개선하기 위한 방법으로 복합 필터링 모형을 제시하였으며, 본 모형은 중위절대편차법과 다수 자료를 판단하는 Voting Rule을 결합한 방식에 해당한다. 중위절대편차법을 이용한 필터링 적용 이후에도 남아 있는 이상치 자료들을 Voting Rule로 제거하게 되며, 수도권 일반국도와 같이 이상치 자료의 비율이 높은 지역에서 이상치 자료만을 선택적으로 제거하기 위한 목적에 해당한다. 또한 본 연구에서는 국도 ITS를 운영 중에 있는 일반국도 3호선의 곧지암~장지구간에 대해 DSRC 원시자료들을 이용하여 복합 필터링 모형에 대한 효과분석을 실시하였다.

II. 수도권 국도ITS의 통행시간 산정과정

본 연구에서 개별차량의 검지기 통과정보에서부터 구간 통행시간을 산정하는 과정은 수도권 국도 ITS의 DSRC 기반 링크 통행시간 산정 프로세스를 기준으로 하였다. 수도권 국도 ITS에서 DSRC 기반의 통행시간 산정과정은 <그림 1>과 같이 개별차량 구간통행시간 생성, 결측보정, 필터링, 링크 통행시간의 산정 순서로 진행된다. DSRC의 검지영역을 통과하는 개별차량의 ID와 통과시간정보를 수집하며, 상하류부 검지기 간의 ID 매칭을 통해 개별차량의 통행시간을 생성한다. 이후, 생성된 통행시간 자료의 개수에 따라 누락보정 여부를 결정하며, ID 매칭에도 생성되는 개별차량 통행시간 정보가 없는 경우 검지기 노변장치 범위를 확대하여 ID 매칭의 과정을 반복 수행한다[2,3].



<그림 1> 수도권 국도 ITS의 통행시간 산정 과정

<Fig. 1> Travel time estimation process of national highway ITS

수도권 국도 ITS의 경우 전체 차량 중 Probe의 비율이 20% 내외로 일시적인 수집을 저하에 따른 누락 데이터의 보정과정을 수행한다. 개별차량 통행시간 정보들을 이용하여 유효범위에 따른 논리적 필터링과 표본의 신뢰구간을 이용한 통계적 필터링을 차례로 수행하며, 이후 결측보정을 수행하여 실시간 통행시간 생성을 위한 개별차량의 유효 표본정보의 생성을 완료한다. 여기에서 논리적 필터링은 최저속도 및 제한속도를 이용해 이를 벗어나는 통행시간 표본정보를 제거하는 작업을 의미하며, 통계적 필터링은 해당 표본 통행시간 정보의 평균과 분산을 이용해 95% 신뢰구간을 만족시키기 위한 표본만을 선택하는 과정에 해당한다. 이후 1분 주기로 이력 5분 간의 개별차량 통행시간 및 통행속도를 식 (1) 및 (2)와 같이 평균하여 DSRC 노변장치 간의 평균 구간통행속도 및 구간통행시간 소통정보를 생성한다[2].

$$\bar{u}_s = n / \sum_{i=1}^n \frac{1}{u_i} \quad (1)$$

$$\bar{t}_s = \frac{1}{n} \sum_{i=1}^n t_i \quad (2)$$

여기에서,

\bar{u}_s = 1분주기 5분 구간통행속도

u_i = 개별차량 i 의 통행속도

\bar{t}_s = 1분주기 5분 구간통행시간

t_i = 개별차량 i 의 통행시간

n = 이력 5분간 수집된 통행시간의 개수

수도권 국도ITS와 같이 DSRC 검지체계 기반의 통행시간 추정과정 중 필터링 프로세스는 통행시간 추정결과의 신뢰성에 영향을 미치는 가장 중요한 요소에 해당한다. 결측보정을 위해 사용되는 자료 또한 필터링이 수행된 이력정보에 해당하며, 20% 내외의 표본 통행시간을 이용함에 따라 필터링 프로세스의 성능에 따라 통행시간 추정의 정확도가 결정되는 구조에 해당한다.

III. 필터링 방법의 고찰

표본 통행시간을 수집하는 과정에서는 정상 교통류와는 전혀 다른 패턴을 나타내는 이상치 자료가 포함될 수 있다. 통행시간의 이상치 자료들이 발생하는 원인은 통신영역 문제와 같은 검지체계 오류, 개별차량 통행패턴의 오류, 정보수집 이후 가공과정의 오류 등 다양하다. 일례로 일반국도에서는 배달차량, 택시 등이 특정 구간을 반복주행하면서 동일 ID가 짧은 시간 동안에 반복 수집되면서 ID 매칭 상의 문제를 발생시킨다. 또한 국도를 따라 조성된 상업지역 이용 차량들로 인해 주정차 시간이 포함된 왜곡된 통행시간 정보가 수집되기도 한다. 이러한 이상치 자료가 차지하는 비율은 미미하나, 타 교통류와의 통행시간 차이가 극심해 통행시간 추정의 신뢰성을 저하시키기는 원인이 된다. 본 연구에서는 통행시간 추정시 사용되는 필터링방법을 정상 자료의 범위를 지정하는 방식, 시계열 자료의 예측정보를 이용하는 방식, 통계적 검정을 이용하는 방식으로 구분하여 특성을 확인하였다.

첫째, 정상 자료의 범위를 지정하여 이상치 자료를 판단하는 방식에는 백분율법(Percentile Method), 중위절대편차법(Median Absolute Deviation), 사분위편차법(Quartile Deviation Method), 신뢰구간법 등이 있다. 백분율법은 자료의 상하위 일정범위를 이상치로 지정하여 제거하는 방식으로 절사평균법 또는 절사대체평균법과 동일한 의미를 가진다. 일반적으로 상위 10% 이상, 하위 90% 이하의 자료를 이상치로 판정하며, 이상치 자료가 일정한 패턴을 나타내는 경우 간단한 방식으로 효과적인 필터링이 가능하나, 정상 자료만으로 구성되는 표본 집단의 경우 정상 자료 까지 함께 제거하는 문제가 있다[4, 5]. 중위절대편차법은 표본 자료의 중위값(median)을 이용하여 특정 범위를 초과하는 자료들을 제거하기 위한 방법에 해당하며, 표본을 정규분포로 만들기 위한 필터링 기법에 해당한다. 중위절대편차법에서는 식 (3)과 같이 중위값과의 절대평균오차를 의미하는 MAD를 산정한다. 이후 식 (4)와 같이 개별 표본들의 z_i^{MAD} 를 산정하여 식 (5)와 같이 z_i^{MAD} 의 절

대값이 제거변수 z_{out} 을 초과하는 경우 이상치로 판정하여 표본 x_i 를 제거한다.

$$MAD = b \times \frac{1}{N} \times \sum_{i=1}^N |x_i - x_{med}| \quad (3)$$

$$z_i^{MAD} = \frac{x_i - x_{med}}{MAD} \quad (4)$$

$$x_i = \begin{cases} \text{유지}; & (|z_i^{MAD}| \leq z_{out}) \\ \text{제거}; & (|z_i^{MAD}| > z_{out}) \end{cases} \quad (5)$$

여기에서,

x_i = 변수 x 의 관측값

x_{med} = 변수 x 의 중위값

b = 조정 계수

N = 표본의 관측 개수

z_{out} = 제거변수

중위절대편차법은 표본의 분포에 관계없이 적용이 가능하며, 비대칭성 분포를 가지는 표본에 대해 중위값을 이용하여 이상치를 제거함으로써 통행시간 정보의 필터링에 논리적으로 타당하다. 그러나 강건한(Robust) 특성을 나타내는 필터링 기법으로 이상치를 덜 제거하는 경향이 있다[6]. 사분위편차법은 중위절대편차법과 유사하나 사분위수의 상하위값과 t-통계량을 이용하며, 95% 신뢰수준을 만족시키기 위한 변이계수가 적용된다[6]. 신뢰구간법은 자료가 정규분포를 따른다는 가정에서 평균에서 표준편차의 배수에 해당하는 상하위 범위를 초과하는 표본을 이상치로 판정하는 방식이며, 95% 신뢰수준을 만족시키기 위해 표준편차의 ± 1 배를 초과하는 상하위 표본을 제거한다[7]. 신뢰구간법의 변형된 방식으로 수도권 국토 ITS의 DSRC 기반 통행시간 가공시스템에서는 평균 대비 표준편차의 비율(표준편차/평균)이 5% 미만인 경우 상위 3%, 하위 2%, 10% 미만은 상위 5%, 하위 5%, 15% 미만은 상위 8%, 하위 7%의 표본들을 이상치로 판정하여 제거하며, 평균 대비 표준편차의 비율이 15% 이상에 해당하는 경우 평균에서 표준편차 크기를 초과하는(평균

\pm 표준편차) 모든 표본을 이상치로 판정하여 제거한다[2, 3]. 본 방식은 신뢰구간법과 Percentile Method를 결합한 방식으로 통행시간의 통계적 신뢰성을 확보하기 위해 간략한 규칙에 따라 효과적인 이상치 제어가 가능하다. 그러나 DSRC를 이용해 수집한 개별차량 통행시간을 정규분포로 가정함에 따라 특정 영역으로 치우친 분포에서는 치우친 영역의 표본이 더 많이 제거되는 오류가 있으며, Percentile Method에서와 동일한 경향으로 정상 표본까지 제거하는 문제가 있다.

둘째, 시계열자료의 예측정보를 이용한 필터링 방식에는 이동평균법을 이용하는 ARIMA (Autoregressive Moving Average)와 Wavelet Transformation을 이용하는 기법들이 제시된바 있다. ARIMA 모형을 이용한 방식에서는 일반적으로 최소제곱법, 최우추정법 등으로 모수를 추정하며, 모수로부터 특정 신뢰수준을 만족시키기 위한 상하위 제거 포인트를 결정하는 방식이 적용된다[8,9]. Choi(2004)는 ARIMA 모형을 기반으로 시계열 GPS 자료의 이상치 제거를 위해 시계열 모형의 신뢰구간 추정부문을 이상치 제거모형과 결합하여 통행시간 이상치의 제거 알고리즘을 제시한 바 있다.

셋째, 이상치 제거를 위해 통계적 검정을 이용하는 방법이 있다. Seo(2002)는 Modified Z-Test를 이용한 이상치 제거방법을 제시하였으며, 전통적인 Z-Test에서 z-score의 민감도가 큰 문제를 해결하기 위해 중위값을 이용함으로써 보다 강건한(robust) 알고리즘을 제시한 바 있다. 이외 표본의 평균과 표준편차로부터 최대절대편차를 산정하여 이상치의 임계값을 추정하게 되는 Grubbs' Test, Chauvenet's Criterion 등이 있다[12,13].

개별차량 통행시간 정보의 필터링을 위해 적용되고 있는 개별적 방법들은 표본의 분포, 이상치의 발생패턴, 이상치의 발생비율, 전체 표본의 개수, 이상치의 크기 등에 따라 영향을 받게 된다. 이에 본 연구에서는 일반국도의 통행시간 표본과 같이 다양한 분포를 가지며, 이상치 자료가 특정한 패턴이 없이 무작위로 발생하는 경우에 적합한 필터링 방법으로 중위절대편차법과 Voting Rule 필터링 모형을 결합

한 복합 필터링 방법을 제시하였다.

IV. Voting Rule을 이용한 복합필터링

일반국도의 통행시간 산정과정에서 이용되는 필터링 방법의 개발을 위해 통행시간 자료의 특성을 반영하여 다음과 같은 필터링기준을 제시하였다.

첫째, 일반국도에서의 개별차량 통행시간 표본은 정체상황, 신호운영 패턴 등의 요인으로 다양한 분포를 가진다. 일반적으로 통행시간은 비포화 상황에서는 정규분포를, 과포화에서는 로그를 취한 이후 정규분포를 나타내는 것으로 알려져 있다[14]. 그러나 일반국도의 경우 신호운영 상태에 따라 연동차량군과 비연동차량군으로 구분되어 좌 또는 우로 치우친 형태의 다양한 분포를 나타낼 수 있으며, 이때 통행시간 표본의 분포와 관계없이 적용이 가능한 범용의 필터링 알고리즘이 요구된다.

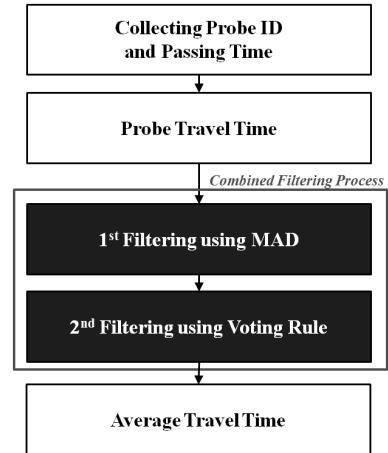
둘째, 통행시간 이상치 자료는 특정한 패턴을 나타내지 않는다. 수도권 국도의 통행시간 이상치 자료는 상업시설 이용차량, 배달차량 등 특이 형태의 통행패턴이 주된 원인이며, 이에 이상치의 패턴과 관계없이 적용이 가능한 필터링 기법이 요구된다.

셋째, 정상 표본의 제거를 최소화하기 위한 필터링 기법이 요구된다. 표본의 통행시간을 이용하여 링크 통행시간을 산정하는데 있어 표본수는 통행시간의 신뢰성을 결정하는 중요한 요소이며, 일반적으로 표준편차와 평균의 크기에 따라 신뢰수준에 따른 최소 표본수가 결정된다[15, 16]. 수도권 국도 ITS에서는 침두시 기준으로 20% 내외 표본비율이 유지되고 있으나 [2], 정상 표본 수를 최대한 확보하기 위한 필터링 기법이 요구된다.

표본의 분포와 이상치 만의 효과적인 제거를 위해 본 연구에서는 중위절대편차법과 Voting Rule을 이용한 이중화된 필터링 방법을 제시하였다. 중위절대편차법은 표본의 분포에 관계없이 적용이 가능하며, 이상치 만을 효과적으로 제거할 수 있다. 그러나 강건한 필터링 기법으로 이상치를 덜 제거하려는 경향이 강하게 나타난다. 이때 이상치가 집중되어 중위절대편차법 이후에도 남아 있는 이상치 자료에

대해 다수결의 원칙을 응용한 Voting Rule을 적용하여 효과적인 필터링이 가능하다.

본 연구에서 제시하는 복합필터링 모형은 <그림 2>와 같이 중위절대편차법으로 1차적인 통계적 필터링을 수행하며, 이후 패턴이 서로 다른 표본을 선택적으로 제거하기 위해 Voting Rule을 이용한 2차 필터링이 수행되는 방식이다.



<그림 2> 본 연구에서의 필터링 과정
<Fig. 2> Filtering process in this study

Voting Rule 또는 Voting System이란 선거 또는 선호도를 결정하는 과정에서 투표자의 최종 의견을 결정하는 방법을 의미한다. Voting Rule은 득표 수의 산정 방식과 결론의 도출 방식 등에 따라 다양한 기법으로 구분되며, 최다 득표자를 선출하는 다수결 원칙(Majority Rule), 득표수에 비례해 당선자를 결정하는 비례대표제(Proportional Representation), 순위를 지정하는 상대다수투표제(Plurality Voting)와 같은 다양한 방법들이 있다. Voting Rule은 고대 그리스의 민주정치에서부터 시작하여 1700년대 프랑스 혁명을 계기로 이론적 토대가 형성되기 시작하였으며, 사회과학 및 경제학에서 하나의 학문분야로 발전해 왔다[17,18].

본 연구에서 제시하는 Voting Rule 기반의 필터링 프로세스는 패턴이 현저히 다른 통행시간 표본을 탐색하기 위한 방법에 해당한다. 본 모형에서는 먼

저 해당 표본의 이상치 여부를 판단하기 위해 이진 수 형태의 보팅값 vt_i 를 식 (6)과 같이 평균통행속도를 이용해 산정한다.

중위절대편차기법을 적용하여 1차 필터링이 완료된 개별차량 통행시간 ft_i 와 평균통행시간 t_{avg} 간 차이의 절대값이 자유속도 통행시간 t_{min} 과 정체상태 통행시간 t_{max} 의 차이를 초과하는 경우 1을 부여하여 이상치 표본으로 판정하며, 반대의 경우 0을 부여하여 정상 표본으로 판단하게 된다. 이때 Voting Rule 기반의 필터링은 정체상태의 차량과 자유속도 상태의 차량이 동시에 관측될 수 없다는 조건이 적용됨에 따라 t_{min} 는 해당 링크의 자유속도 통행시간을 적용하며, t_{max} 는 해당 링크의 정체시 최저속도를 이용하여 상수의 형태로 적용된다.

$$vt_i = \begin{cases} 0; (|ft_i - t_{avg}| \leq t_{max} - t_{min}) \\ 1; (|ft_i - t_{avg}| > t_{max} - t_{min}) \end{cases} \quad (6)$$

여기에서,

- vt_i = i 번째 차량 통행시간의 보팅값
- ft_i = 1차 필터링 이후 i 번째 차량 통행시간
- t_{avg} = 표본의 평균통행시간
- t_{min} = 자유속도 통행시간
- t_{max} = 정체속도 통행시간

이후, 이력 5분간의 통행시간 중 수식 (7)과 같이 이상치 표본의 비율을 산정하여 이상치의 비율 기준을 의미하는 VR 과의 비교를 통해 개별차량 통행

시간 ft_i 의 제거여부를 최종적으로 결정한다.

$$ft_i = \begin{cases} \text{유지}; (\frac{1}{n} \sum_{i=1}^n vt_i \geq VR) \\ \text{제거}; (\frac{1}{n} \sum_{i=1}^n vt_i < VR \text{ and } vt_i = 1) \end{cases} \quad (7)$$

여기에서,

VR = 이상치가 다수 자료임을 판정하는 비율

이상치 표본 비율이 보팅 비율 VR 을 초과하는 경우 이상치 표본은 다수 자료에 해당하며, 모든 통행시간 표본은 정상치로 판정한다. 이와 반대로 이상치 표본의 비율이 VR 을 초과하지 않는 경우 다수자료의 원칙에 따라 보팅값 vt_i 가 1로 부여된 개별차량 통행시간 ft_i 는 이상치로 판정하여 제거한다. 이때 보팅비율을 산정과정에서 이력 5분간의 표본을 이용함에 따라 이전 주기에 Voting Rule을 통해 제거된 이상치 표본을 포함시켜 적용하며, 이를 통해 급격한 통행시간의 변화를 수용할 수 있다.

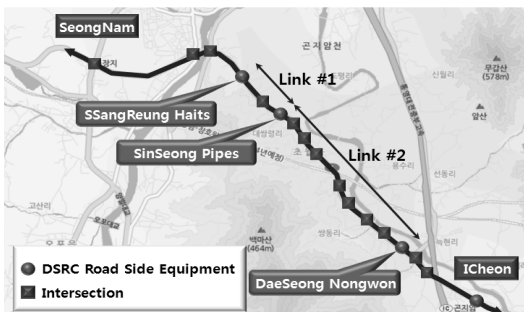
Voting Rule을 이용한 2차 필터링 과정을 <그림 3>과 같이 예시로 표현하였다. 1분 주기 5분 통행시간의 산정 과정 중 첫 1분 주기에는 $ft_{i=8}$ 이 1,200초로 보팅값 vt_i 가 1에 해당하며, 이때 보팅비율은 9%로 VR 30% 미만에 해당하여 해당 표본은 제거된다. 다음 주기에는 이전 주기에 제거된 자료를 다시 이용한다. 본 과정을 반복하여 이상치가 다수 자료가 되는 경우 이상치를 정상표본으로 판정하여 유지한다.

												$t_{min} = t_{60kph} = 90 \text{ sec}$	$t_{max} = t_{10kph} = 540 \text{ sec}$	$VR = 30\%$									
												1 min		2 min		3 min		4 min		5 min			
ft_i	150	116	241	240	350	120	268	1200	321	261	280												
i	1	2	3	4	5	6	7	8	9	10	11												
$ ft_i - t_{avg} $	172	206	81	82	28	202	54	878	1	61	42												
vt_i	0	0	0	0	0	0	0	1	0	0	0												
												2 min		3 min		4 min		5 min		6 min			
ft_i	241		240	350	120	268	1200	321	261	280	361	305	245										
vt_i	0		0	0	0	0	1	0	0	0	0	0	0										
																						$\frac{1}{n} \sum vote_i = 1/11 = 9\% < VR$	
																						$ft_8 \rightarrow \text{Elimination}$	
																						$\frac{1}{n} \sum vote_i < VR$	
																						$ft_6 \rightarrow \text{Elimination}$	

<그림 3> Voting Rule을 이용한 필터링 과정 예시
<Fig. 3> Example of filtering process based on Voting Rule

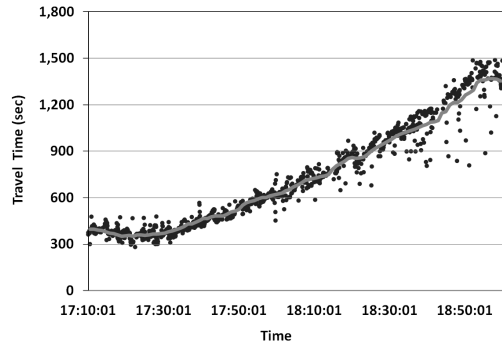
V. 복합 필터링모형의 성능평가

Voting Rule 기반 복합필터링 모형의 평가를 위해 수도권 국도 ITS를 운영 중에 있는 일반국도 3호선의 경기도 광주시 구간을 대상으로 필터링 모형별 성능평가를 시행하였다. 분석 구간은 <그림 4>와 같이 일반국도 3호선의 이천과 성남 구간으로 곤지암 IC와 장지IC 사이의 3개 DSRC 노변장치를 대상으로 하며, 대성농원앞에서 부터 쌍령하이츠 앞까지의 성남 방향 2개 링크를 대상으로 하였다.



<그림 4> 성능평가 대상구간
<Fig. 4> Study site for performance evaluation

복합 필터링 모형의 비교평가를 위해 Percentile Method, 신뢰구간법, 중위절대편차법, Voting Rule 기반의 복합필터링의 분석을 시행하였으며, 수도권 국도 ITS의 2012년 9월 5일 오후 침두 2시간 자료를 이용하였다. 교통정보시스템에서 노변장치 간 ID 매칭이 완료된 Probe 통행시간 정보를 이용하여 필터링 전후를 비교하였으며, 실측 통행시간 자료와의 비교를 통해 필터링에 따른 통행시간 가공의 정확도 개선 정도를 분석하였다. 통행시간 실측 자료는 링크 2의 상류부 검지기 지점과 (대성농원) 링크 1의 하류부 검지기 지점에서 (쌍령하이츠) 비디오 촬영을 통해 번호판 정보를 수집하였으며, 번호판 매칭을 통해 개별차량 통행시간을 실측하였다. 번호판 촬영을 통한 통행시간 수집과정에서도 이상치가 존재함에 따라 통행시간의 크기를 기준으로 상하위 각각 30%의 표본을 제거하여 <그림 5>와 같이 개별 차량 통행시간의 표본 정보를 수집하였다.

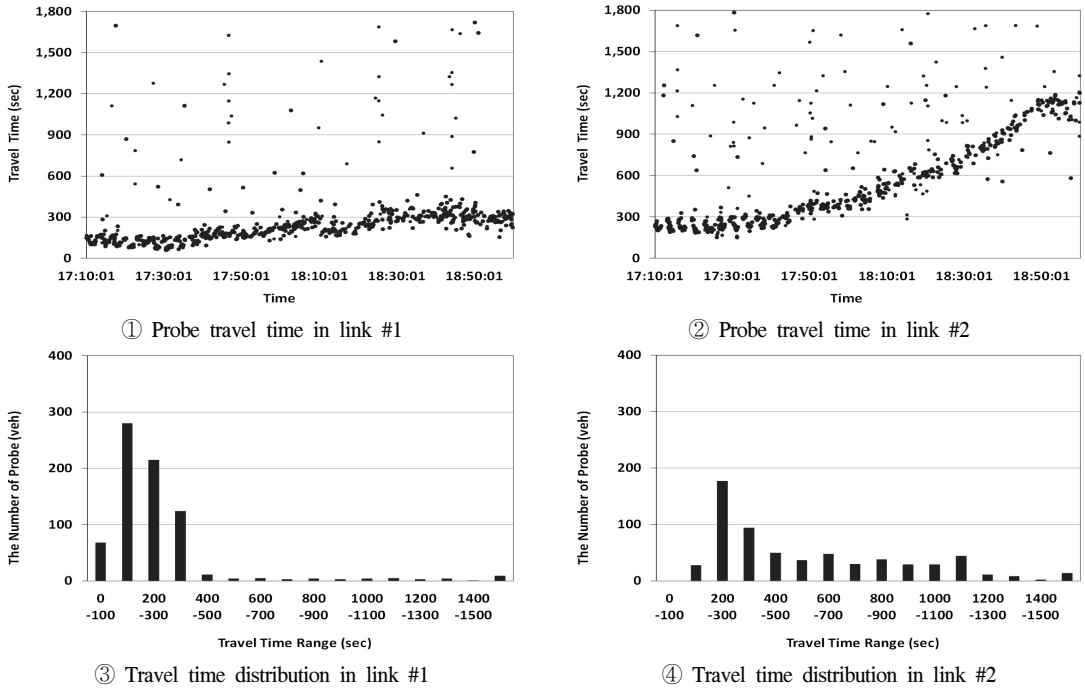


<그림 5> 통행시간의 실측조사 결과
<Fig. 5> Travel time survey result

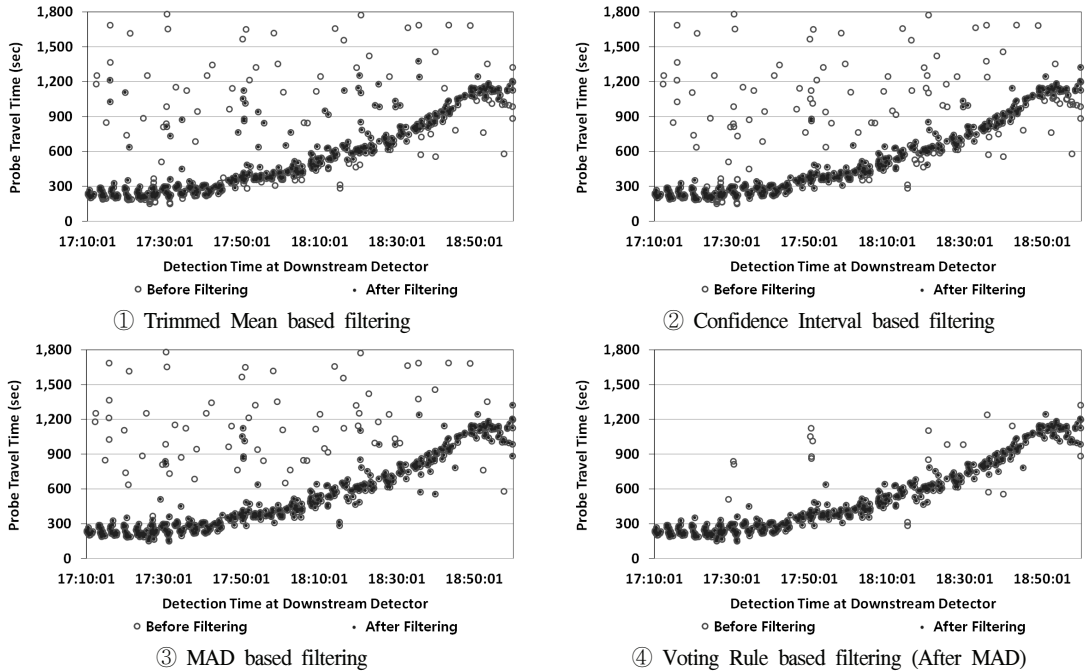
차량 번호판 조사를 수행한 결과 5분당 평균 300대의 통과교통량 중 70대의 구간통행시간 정보가 수집되었으며, 평균 20% 이상의 표본 비율을 나타내었다. 실측한 개별차량 통행시간을 수도권 국도 ITS의 DSRC 기반 통행시간 가공에서와 동일한 방법으로 식 (1)과 같이 1분 주기 5분 통행시간의 산술평균을 통해 실측 구간통행시간으로 산정하였으며, 이를 구간통행시간의 참값으로 이용하였다.

필터링 전 Probe 통행시간은 <그림 6>과 같이 링크 1 및 2로 구분하여 수도권 국도 ITS의 통행시간 가공 시스템으로부터 수집하였다. Probe의 통행시간은 링크 1, 2 모두에서 정체와 비정체시가 구분되며, 통행시간의 패턴과는 상이한 이상치 자료들이 관측되었다. 이때 링크 2에서 상대적으로 높은 통행시간 이상치의 관측빈도는 교차로 및 토지이용의 밀도 증가가 원인으로 판단된다.

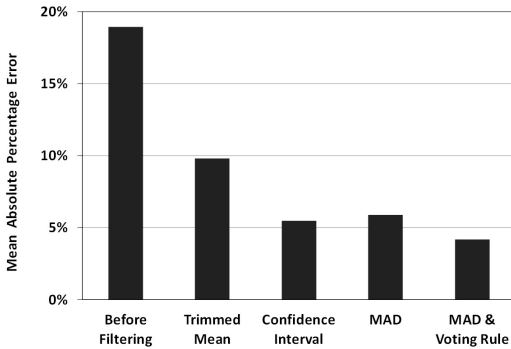
DSRC 검지체계로부터 수집된 Probe 통행시간을 이용하여 필터링 모형별 적용결과를 <그림 7>과 같이 제시하였다. 절사평균법(Trimmed Mean)에서는 상하위 10% 만을 이상치로 제거함에 따라 이상치가 집중되는 시간대에서 정상 패턴의 자료들과 200초 이상 차이를 나타내는 통행시간이 9개 이상 관측되었다. 또한 통행시간의 편차가 작은 시간대에서는 정상 패턴의 자료들이 제거되는 현상을 나타내었다. 신뢰구간법(C Confidence Interval)에서는 표준편차와 평균의 비율에 따라 제거범위를 결정함에 따라 비교



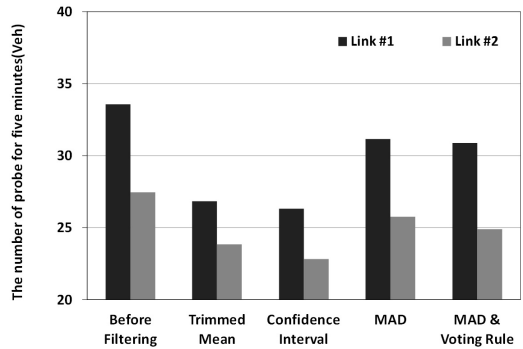
〈그림 6〉 수도권 국도 ITS의 Probe 통행시간
 〈Fig. 6〉 Probe travel time of metropolitan area national highway ITS



〈그림 7〉 링크 2의 필터링 수행결과
 〈Fig. 7〉 Filtering results in link #2



① Mean Absolute Percentage Error(MAPE)



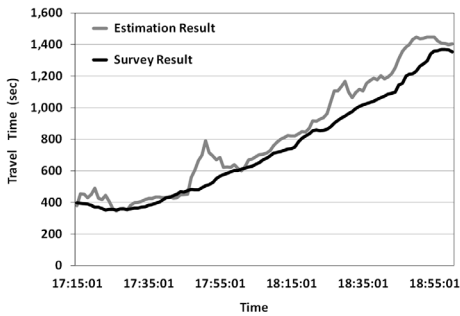
② The number of probe for five minutes

〈그림 8〉 필터링 방법의 평가결과
 〈Fig. 8〉 Evaluation results of filtering methods

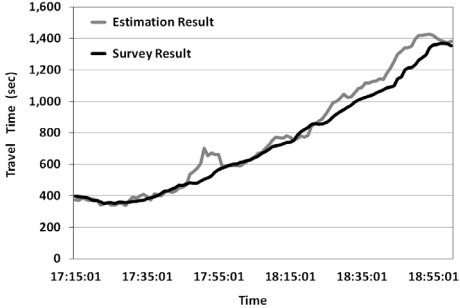
적 우수한 이상치 제거능력을 나타내었다.

그러나 절사평균법과 동일한 경향으로 이상치는 주로 상위집단에서 발생되나 상하위 표본을 모두 제거함에 따라 정상 패턴의 표본들이 제거되는 경향을 나타내었다. 이는 Probe 통행시간의 분포가 <그림 6>에서 같이 정규분포가 아닌 비대칭적 형태

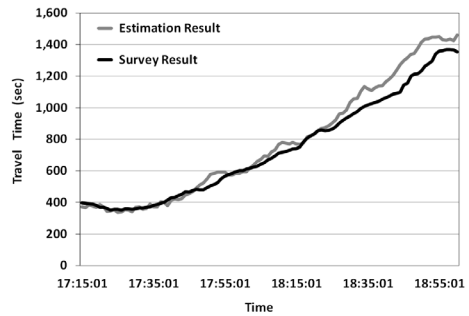
를 나타내기 때문이다. 신뢰구간법의 경우 이상치 제거능력은 우수한 결과를 나타내었으나, 통행시간 표본이 정규분포임을 가정하는 방법으로 표본분포의 치우침에 따라 특정 영역의 제거율이 높아지는 논리적 모순을 가진다. 중위절대편차법(MAD)을 이용한 필터링에서는 이상치 표본만을 효과적으로 제



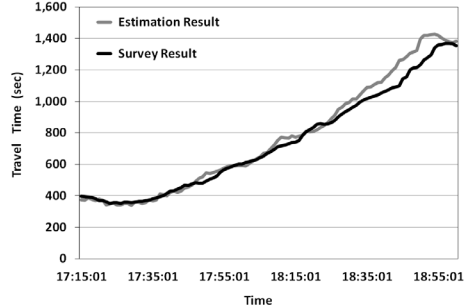
① Trimmed Mean based travel time



③ MAD based travel time



② Confidence Interval based travel time



④ Voting Rule based travel time (After MAD)

〈그림 9〉 통행시간의 추정 결과
 〈Fig. 9〉 Estimation results of travel time

거하는 결과를 나타내었으나, 이상치가 집중되어 중위값이 증가하는 특정 시간대에서 중위절대편차법의 강건한 특성으로 인해 이상치가 효과적으로 제거되지 않는 경향을 나타내었다.

마지막으로 중위절대편차법과 Voting Rule의 복합 필터링 모형을 적용한 경우 중위절대편차법에서 제거하지 못한 이상치 자료만이 모두 제거된 결과를 나타내었다. 필터링 이후 1분 주기 5분 통행시간은 <그림 9>에서와 같이 절사평균법과 중위절대편차법의 경우 필터링 이후에도 남아있는 이상치 자료들로 인해 특정 시간대에서 실측 통행시간과의 격차를 나타내었으며, 신뢰구간법과 Voting Rule을 이용한 복합 필터링 모형의 경우 실측통행시간과 유사한 패턴의 통행시간 산정 결과를 나타내었다.

실측 통행시간과 필터링 방법별 가공 통행시간과의 평균절대백분율오차(MAPE)는 그림 8과 같이 필터링 전 18.9%에서 절사평균법이 9.8%, 신뢰구간법이 5.5%, 중위절대편차법이 5.9%, 복합 필터링 모형이 4.2%를 나타내어 복합필터링 모형이 가장 우수한 오차수준을 나타내었다. 또한 이력 5분간의 표본 개수는 링크 1을 기준으로 필터링 전 33.6개에서 절사평균법이 26.8개, 신뢰구간법이 26.3개, 중위절대편차법이 31.2개, 복합 필터링 모형이 31.0개를 나타내어 복합 필터링을 이용해 이상치 만의 효과적가능함을 확인하였다.

필터링 모형에 따른 통행시간의 산정결과 중위절대편차법과 Voting Rule을 이용한 복합필터링 모형은 표본의 분포와 관계없이 이상치 만을 선택적으로 제거함에 따라 정상 패턴을 나타내는 유효 표본을 유지함과 동시에 통행시간의 오차를 개선할 수 있는 유용한 방법이 될 수 있음을 확인하였다.

VI. 결론

본 연구에서는 일반국도에서 Probe 통행시간을 이용해 구간통행시간을 생성하는 과정에서 이상치 표본을 제거하기 위한 복합 필터링 모형을 제시하였으며, 이는 중위절대편차법과 Voting Rule의 이중화된 필터링 방법에 해당한다. 본 필터링 모형은 표

본수를 유지함과 동시에 이상치 자료만을 선택적으로 제거하기 위한 목적을 가지고 있으며, 이때 Voting Rule은 다수결의 원칙이라는 직관적인 다수 자료의 판단 방법으로서 상이한 패턴을 나타내는 통행시간 이상치 만을 효과적으로 제거할 수 있다. 일반국도 3호선의 광주시 구간을 대상으로 한 성능평가에서는 절사평균법, 신뢰구간법, 중위절대편차법, 복합 필터링 모형의 비교분석을 통해 복합필터링 모형이 이상치 만을 선택적으로 제거하여 정상 패턴의 자료들을 유지하는데 가장 우수한 성능을 나타냄을 확인하였다.

이상치 수집의 문제를 해결하는 대안으로 토지이용이 활발한 구간에서 검지체계의 노변장치 간격을 조밀하게 구성하는 하드웨어적인 방법과 Probe ID 매칭을 위한 시간범위를 짧게 지정하는 소프트웨어적인 방법이 적용될 수 있으나, 이들 또한 모두 근본적인 해결책이 되지 못하여 검지체계 증가에 따른 비용의 문제와 과포화 상태에서 정보수집을 저하를 일으킬 수 있다. 따라서 본 연구에서 제시하는 이중화된 필터링 프로세스는 간략하지만 현장적용이 용이하며, 다수결의 원리라는 논리적 타당성 또한 포함하고 있어 통행시간 추정의 신뢰성 제고를 위한 유용한 방법으로 판단된다.

향후 연구로서 Voting Rule 기반의 필터링 성능개선을 위해 이상치 자료의 판정기준으로 이용되는 t_{avg} , t_{min} , t_{max} 의 결정방법에 대한 보완이 필요하며, 이와 함께 Probe 통행시간의 최종 제거여부를 결정하기 위해 적용되는 보팅 비율 VR 에 대한 민감도 분석 등의 논리적 산정방법에 대한 보완이 요구된다.

참고문헌

- [1] S. H. Lim, "Situation of Planning and Prospects of National Highway ITS", *Transportation Technology and Policy*, vol. 26, no. 5, pp.61-68, 2008.
- [2] University of Seoul, *Transportation Information Quality Assurance for Performance Improvement of Transportation Information Center*, pp.1-2, 2012.
- [3] Korea Institute of Construction Technology, 2012

- Work Manual for Operation and Management of Transportation Information Center*, pp.155, 2012.
- [4] V. Barnett and T. Lewis, *Outliers in Statistical Data*, John Wiley & Sons, 1985.
- [5] B. Iglewicz and D. Hoaglin, *How to detect and handle outliers*, ASQC Quality Press, 1993.
- [6] S. D. Clark, S. Grant-Muller, and H. Chen, "Cleaning of Matched License Plate Data", *Transportation Research Record 1804*, pp.1-7, 2002.
- [7] J. K. Kang, Y. T. Son, Y. H. Yoon, and S. C. Byun, "Regional Traffic Information Acquisition by Non-intrusive Automatic Vehicle Identification", *Journal of Korea ITS Society*, vol. 1, no. 1, pp.22-32, 2002.
- [8] J. M. Roth, *A Time Series Approach To Removing Outlying Data Points From Bluetooth Vehicle Speed Data*, Thesis of Master Degree, The University of Akron, 2010.
- [9] C. Bilen and S. Huzurbazar, "Wavelet-Based detection of outliers in time series", *Journal of Computational and Graphical Statistics*, vol 21, pp.311-327, 2012.
- [10] K. J. Choi and J. A. Jang, "A Time Series-based Algorithm for Eliminating Outliers of GPS Probe Data", *Journal of Korea Society of Transportation*, vol. 22, no. 6, pp.67-75, 2004
- [11] S. W. Seo, *A Review and Comparison of Methods for Detecting Outliers in Univariate Data Sets*, Thesis of Master Degree, Kyung Hee University, 2002.
- [12] F. Grubbs, "Procedures for Detecting Outlying Observations in Samples", *Technometrics*, vol. 11, no. 1, pp.1-21, 1969.
- [13] M. S. Ross, "Peirce's criterion for the elimination of suspect experimental data", *Journal of Engineering Technology*, vol. 20, no. 2, pp.1-12, 2003.
- [14] F. Dion and H. Rakha, "Estimating dynamic roadway travel times using automatic vehicle identification data for low sampling rates", *Transportation Research Part B*, vol. 40, no. 9, pp.745-766, 2006.
- [15] K. K. Srinivasan and P. P. Jovanis, "Determination of Number of Probe Vehicles Required for Reliable Travel Time Measurement in Urban Network", *Transportation Research Record 1537*, pp.15-22, 1996.
- [16] Y. I. Lee and J. H. Lee, "A Study on Link Travel Time Estimating Methodology for Traffic Information Service (Determination of an Adequate Sample Size)", *Journal of Korea Society of Transportation*, vol. 20, no. 3, pp.55-67, 2002.
- [17] M. Lines, "Approval Voting and Strategy Analysis: A Venetian Example", *Theory and Decision*, vol. 20, pp.155-172, 1986.
- [18] H. Nurmi, *Comparing Voting Systems*, D. Reidel, Dordrecht, Holland, 1987.

저자소개



정 영 제 (Jeong, Youngje)

2011년 3월 ~ 현 재 : 서울시립대학교 교통공학과 연구교수

2011년 2월 : 서울시립대학교 교통공학과 박사

e-mail : sleep108@hanmail.net

연락처 : 02-2210-2671



박 현 석 (Park, Hyun Suk)

1999년 1월 ~ 현재 : 한국건설기술연구원 도로교통정보사업단 연구원
2007년 2월 : 서울시립대학교 교통공학과 박사수료
1999년 2월 : 인천대학교 토목공학과 석사
e-mail : hysupark@kict.re.kr
연락처 : 02-753-0946



김 병 화 (Kim, Byung Hwa)

1984년 9월 ~ 현재 : 한국건설기술연구원 도로교통정보사업단 연구위원
2004년 2월 : 한양대학교 기계공학과 박사
1988년 8월 : 한양대학교 기계공학과 석사
e-mail : bhkim@kict.re.kr
연락처 : 031-910-0640



김 영 찬 (Kim, Youngchan)

1996년 ~ 현재 : 서울시립대학교 교통공학과 교수
1996년 : 명지대학교 교통공학과 교수
1993년 : 도로교통안전협회 연구소 연구위원
1991년 : 교통개발연구원 선임연구원
1990년 : Texas A&M University 토목공학과 박사
e-mail : yckimm@uos.ac.kr
연락처 : 02-2210-2768