

# 일반국도 상시조사 교통량 자료의 이상치 판정 알고리즘 개발

신재명\* · 이상협\*\* · 김현석\*\*\*

Shin, Jae Myong\*, Lee, Sang Hyup\*\*, Kim, Hyun Suk\*\*\*

## The Outlier-Filtering Algorithm for National Highway Continuous Traffic Counts Data

### ABSTRACT

In this study the quantitative outlier-filtering algorithm has been developed using the smoothing method based on the day-of-the-week traffic volume variation pattern and then, in order to test the effectiveness of the algorithm, it has been used to identify outliers from the traffic volume data collected at 14 continuous traffic counts sites on the national highways in the year 2010. The test results are satisfactory since the filtering rate is 98.2% for normal days and the mis-filtering rate is 8.0% for abnormal days. Therefore, the algorithm will be able to be used for roughly-but-quickly filtering outliers from the collected traffic volume data.

**Keywords :** smoothing method, quantitative, outlier, filtering, algorithm

### 초 록

본 연구에서는 요일별 교통량 변동 패턴 기반 평활화법을 활용하여 정량적 이상치 판정 알고리즘을 개발하였다. 또한 개발된 알고리즘을 활용하여 2010년 일반국도 상시조사 지점 중 14개 지점의 교통량 자료에 대한 이상치 필터링을 수행하여 알고리즘의 적합성 여부를 평가하였다. 그 결과 정상일 필터링율은 98.2%, 이상일 중 오판터링율은 8.0%로 평가되었다. 따라서 본 연구에서 개발된 알고리즘은 수집된 교통량 자료의 1차적인 이상치 필터링에 충분히 적용 가능할 것이다.

**검색어 :** 평활화법, 정량적, 이상치, 필터링, 알고리즘

## 1. 연구 배경 및 목적

도로교통량이란 도로의 특정 지점 또는 구간을 단위 시간 동안 통과하는 각종 차량의 수를 의미한다. 따라서 도로교통량조사(Traffic Counts, 이하 '교통량조사'로도 표현한다)는 도로를 이용하는 각종 차량의 단위 시간 당 통과 대수를 종류별, 방향별 및 시간대별로 관측하는 조사이다. 조사된 도로교통량(Traffic Volume, 이하 '교통량'으로도 표현한다) 자료는 도로 설계, 포장 설계, 도로 운영 및 유지관리 등에 폭넓게 이용되며 도로망 계획, 교통망 계획 및 교통관리 계획 수립을 위한 기초적 정보가 된다.

현재 우리나라에서는 국토해양부, 지방자치단체, 민간투자도로주식회사, 한국도로공사 등이 교통량조사를 시행하고 있지만 국토해양부의 교통량조사 이외에는 모두 산발적이며 소규모로 이루어지고 있다. 국토해양부에서 시행하고 있는 교통량조사는 크게 일반국도 수시조사(Short Duration Counts), 일반국도 상시조사(Continuous Counts), 그리고 고속국도와 국가지원지방도·지방도에 대하여 10월 3째 주 목요일에 이루어지는 수시조사로 나누어진다.

\* 명지대학교 교통공학과 (firstrain1@naver.com)

\*\* 정회원·교신저자·한국건설기술연구원 (Corresponding Author·Korea Institute of Construction Technology·infohi2@hanmail.net)

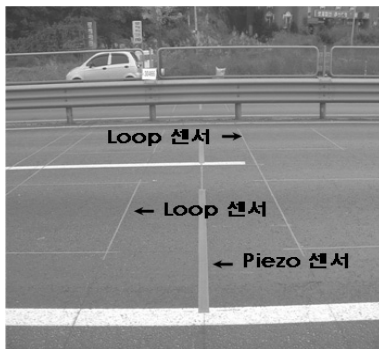
\*\*\* 정회원·한국건설기술연구원 (hskim3@kict.re.kr)

Received June 1 2012, Revised June 22 2012, Accepted December 13 2012

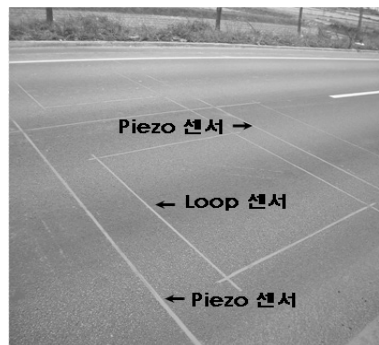
일반국도 수시조사(이하 ‘수시조사’로도 표현한다.)는 각 조사 지점에 이동식 교통량조사 장비(Portable Traffic Counter)를 설치하여 07:00시부터 익일 07:00시까지 24시간 동안 통과한 차량의 대수를 조사하는 것으로 동일 조사 지점에 대하여 1년에 1~5회 실시된다. 단, 차종조사(Vehicle Classification Counts)는 1회만 실시된다. 일반국도 상시조사(이하 ‘상시조사’로도 표현한다.)는 각 조사 지점에 고정식 교통량조사 장비(Permanent Traffic Counter)를 설치하여 1년 이상의 장기간에 걸쳐 그 지점을 통과한 차량의 차종별 대수 및 총 대수를 조사하는 것으로 해당 지점의 교통량 변동에 대하여 요일별, 월별, 계절별 특성 등을 포함하여 시계열적으로 파악하는 데 목적이 있다(Lee, Ha and Yoon, 2012).

현재 일반국도에 설치되어 운영되고 있는 상시조사 장비의 구성은 크게 제어기부, 센서부, 통신부, 전원부로 나눌 수 있으며 이중 장비의 유형을 구분 짓는 것은 제어기부와 센서부이다. 제어기는 프랑스 E사의 HESTIA 모델과 영국 T사의 HI-TRAC 모델을 사용한다. 2011년 11월 말 기준으로 가동 중인 제어기 대수는 HESTIA 모델 319대와 HI-TRAC 모델 166대로 총 485대이다.

센서는 Fig. 1에서 보듯이 Loop-Piezo-Loop와 Piezo-Loop-Piezo의 두 가지 배치형태를 사용하며 Piezo 센서는 미국 M사의 Roadtrax BL Piezoelectric Traffic Sensor Class II를 사용한다.



Loop-Piezo-Loop Configuration



Piezo-Loop-Piezo Configuration

Fig. 1. Types of the Sensor Configurations for the Permanent Traffic Counter(Lee, Ha and Yoon, 2012)

센서를 도로에 매설하기 위하여 Resin을 사용하는데 영국 G사의 PU-200을 사용한다(Lee, Ha and Yoon, 2012).

상시조사 장비의 제어기와 센서는 차량이 다니는 도로변과 도로 표면에 설치되어 있어서 온도, 습도, 먼지, 진동, 벼락, 압력 등에 항상 노출되어 있다. 따라서 노후화가 빨리 진행되며 고장이 자주 발생하게 된다. 특히 도로표면에 매설된 센서는 도로 덧씌우기, 중차량으로 인한 도로 파손, 중앙분리대 설치 공사, 갓길에 가스관 또는 통신관로 매설 공사 등으로 인하여 파손되는 경우가 빈번하게 일어나며 일반국도의 경우 고속국도에 비하여 사용되는 포장 재료의 질이 떨어져서 레진의 파손이 쉽게 일어나고 이는 곧 불량, 감도 저하 등 센서의 고장 또는 단선 등 파손으로 이어지게 된다. 이처럼 제어기나 센서가 고장 나거나 파손될 경우 교통량 자료를 수집할 수 없거나 수집된다 하더라도 자료의 정확도가 저하될 수 있다.

제어기 고장의 경우는 부품 교체나 제어기 교체 등으로 단기간 내에 정상 가동이 가능하여 자료의 손실 기간이 짧을 수 있으나 센서 불량, 감도 저하나 단선 등 센서 고장의 경우는 도로포장면 절개 등을 수반하는 센서 교체 공사가 이루어져야 하므로 단기간 내에 정상 가동이 불가능하여 자료의 손실 기간이 길게 된다.

또 하나의 고장 유형인 통신 고장은 통신용 모뎀의 회로기판 등 구성부품이 오랜 기간 사용되어 노후화되거나, 열악한 환경에 오랜 기간 노출되어 관련 회로들이 단선이 되거나, 전기적 쇼크, 벼락 등에 의하여 보드의 구성부품들이 완전히 타버리거나, 모뎀에 내장된 프로그램이 얽혀버려서 제어기에서 신호처리된 교통 자료가 센터의 서버에 제대로 전송되지 못하는 상태를 의미한다.

통신용 모뎀 고장의 경우는 제어기에 내장되어 있는 메모리에 많게는 3개 월 분량의 교통량 자료가 저장될 수 있으므로 이 기간 내에 제어기에 직접 노트북 컴퓨터 등을 연결하여 자료를 받게 되면 자료의 손실 기간은 제로(Zero)가 된다.

마지막으로 전원 고장은 전압이 안정적이지 않거나 전신주로부터 제어기에 연결된 전선이 통신관로나 가스관 매설공사로 인하여 끊어져서 전신주로부터 제어기에 전력 공급이 제대로 되지 않아 제어기, 통신용 모뎀 등이 오작동하거나 작동을 멈추버리는 상태를 의미한다.

전선 단선의 경우는 전기공사 전문업체에 의뢰하여 연결하면 되므로 단기간 내에 정상 가동이 가능하여 자료의 손실 기간이 짧을 수 있으나 누전이나 전압 불량 등의 경우는 접지 공사, 누전 원인 제거, 전압 안정화 등의 조치가 이루어져야 하며 이러한 조치는 단기간 내에 이루어지기 어려우므로 자료의 손실 기간이 길게 된다.

위에서 언급한 것처럼 일반국도에 설치되어 있는 상시조사 장비는 여러 가지 원인으로 인하여 오작동이 발생하거나 작동이 멈추어서 정상적이지 못한 자료가 수집되거나 자료가 누락된다. 한국건설기술연구원에서 운영·관리하고 있는 상시조사 장비는 지역적으로 전국의 일반국도망에 분포되어 있으며 설치 대수 또한 2011년

11월 기준 485대로 적지 않으므로 아무리 철저하고 신속하게 파손, 오작동 등 고장을 처리한다고 하더라도 장비별로 편차가 있기는 하지만 연간 며칠 이상의 기간은 자료가 누락되거나 정상적이지 않은 자료 즉 이상치(Outlier)가 수집된다.

한국건설기술연구원에서 일반국도 상시조사 자료의 이상치 점검을 위하여 사용하고 있는 방법은 크게 정량적 기준에 의한 점검 방법과 정성적 기준에 의한 방법의 2가지를 들 수 있는데 실제로는 이 2가지 방법을 병행하여 이상치 점검을 수행하고 있다.

정량적 기준에 의한 점검 방법은 특정 기간 동안의 평균일교통량(Average Daily Traffic, ADT) 또는 월평균일교통량(Monthly Average Daily Traffic, MADT)±3σ(표준편차)를 상한 값과 하한 값으로 설정하고 이 범위를 벗어나는 경우를 이상치로 판단하는 방법과 최대 기본 용량(연속류 : 2,200 대/시/차로)을 상한 값으로 설정하여 이 값을 초과하는 경우를 이상치로 판단하는 방법이다. 정성적 기준에 의한 점검 방법은 차종별 교통량과 시간대별 교통량의 차이가 큰 경우, 한 방향 또는 한 차로의 자료가 없는 경우, 상행과 하행의 교통량의 차이가 큰 경우 등을 이상치로 판단하는 방법이다(Korea Institute of Construction Technology, 2001).

위에 언급한 이상치 점검 방법은 이해하기 쉽고 적용하기 쉬운 장점이 있는 반면 정량적 기준은 엄격하지 않은 기준으로 이상치가 걸러질 가능성이 낮아질 수 있으며, 정성적 기준은 점검자의 주관적인 요소가 많이 개입되므로 일관성 및 객관성이 결여된 기준이라 할 수 있다. 게다가 신규로 설치되는 장비이거나 또는 과거 정상적으로 수집된 자료가 확보되어 있어야 적용이 가능하다는 단점이 있다.

이러한 단점들을 극복하기 위하여 본 연구에서는 일반국도 교통량의 시간적 변동 특성을 설명할 수 있는 평활화법 즉, 요일별 교통량 변동 패턴(Pattern) 기반 평활화법을 활용하여 연속적으로 수집되는 자료의 이상 유무를 판정할 수 있는 정량적 이상치 판정 알고리즘을 개발하였다.

또한 개발된 알고리즘을 활용하여 2010년 일반국도 상시조사 지점 중 14개 지점의 교통량 자료에 대한 이상치 필터링을 수행하여 필터링과 오픈터링을 산출함으로써 알고리즘의 적합성 여부를 평가하였다.

## 2. 선행 연구 고찰

지금까지 연속적으로 수집되는 교통량 자료의 이상치 판정에 관한 연구는 이루어진 적이 거의 없으며 대부분의 연구는 통행시간 자료의 이상치 판정에 관하여 이루어졌다. 이들 교통 관련 자료의 이상치 판정에 이용된 기법으로는 통계적 기법과 시계열 기법이 대표적이다.

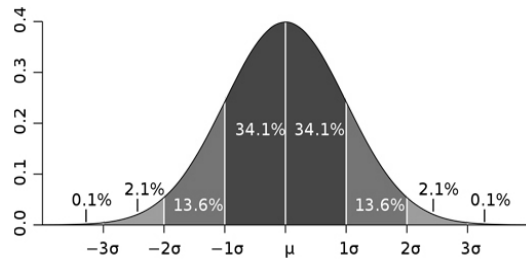


Fig. 2. An Example of the Mean±3σ Filtering

### 2.1 통계적 기법

통계적 기법 중 첫 번째 기법은 정규분포를 이용하는 방법으로 수집되는 교통량 자료는 정규분포를 이룬다는 가정 하에 평균을 중심으로 ±3σ의 범위를 유효범위로 설정한다. 따라서 평균(μ)을 중심으로 ±3σ의 범위는 전체의 99.8%를 의미하므로 0.02%만이 이상치로 결정된다(Korea Institute of Construction Technology, 2001).

두 번째 기법은 중위절대편차(MAD : Median Absolute Deviation)를 이용하는 방법으로 수집되는 자료의 분포 형태를 가정하지 않아도 되는 장점이 있으며 국내에서는 수집된 통행시간의 이상치 제거에 활용되었다. 이상치 판정 기준은  $|Z_i^{mad}| > Z_{cut}$  이면 이상치로 판정한다. 일반적으로  $Z_{cut}$ 은 3으로 선정한다(Seoul Metropolitan Government, 2003).

$$MAD = 1.4826 \times \text{median}|x_i - x_{med}| \quad (1)$$

$$Z_i^{MAD} = \frac{|x_i - x_{med}|}{MAD} \quad (2)$$

여기서, MAD : 중위절대편차

$Z_i^{MAD}$  : MAD에 의한 표준점수(Z score)

$x_i$  : i번째 데이터

$x_{med}$  : x데이터 계열의 중위값

1.4826 : MAD를 정규분포의 표준편차와 동일하게 만드는 조정계수

세 번째 기법은 신뢰구간 추출법으로 Kang et al.(2002)가 비매설식 자동차량번호판인식장치(AVNPRS: Automatic Vehicle Number Plate Recognition System, 보통 'AVI'로 불린다.)를 개발하면서 제시한 방법이다. 개별 차량의 구간통행시간이 극단적인 측정치들에 의하여 편중되지 않도록 상한 값(설계속도의 2배를 초과하여 통행하는 시간)과 하한 값(해당 구간을 10 km/시로 통행할 때의 통행시간)들을 상수 값으로 설정하여 이 범위를 벗어나는

구간통행시간을 이상치로 제거한 후 신뢰구간 95% 또는 68%의 범위를 초과하는 값들을 제거하는 방법이다.

$$\bar{T} = \frac{\sum T_i}{n} \quad (3)$$

$$\sigma = \sqrt{\frac{\sum T_i^2 - n\bar{T}^2}{n-1}} \quad (4)$$

$$T_{is} = |T_i - \bar{T}| \leq \sigma \quad (5)$$

여기서,  $\bar{T}$  : 구간교통정보 수집장치로부터 수집된 개별 차량들의 구간통행시간의 산술평균

$T_i$  : 구간교통정보 수집장치로부터 수집된 개별 차량의 구간통행시간

$\sigma$  : 구간교통정보 수집장치로부터 수집된 개별 차량들의 구간통행시간 표준편차

$T_{is}$  : 구간교통정보 수집장치로부터 수집된 개별 차량 중 유효한 차량의 통행시간

$n$  : 구간교통정보 수집장치로부터 자료가 수집된 차량 중 상·하한값을 제외한 차량 수

본 절에서 언급된 통계적 기법은 이해와 적용이 쉬울 수 있지만 정량적 기준은 기준 자체가 엄격하지 않아 이상치가 걸러질 가능성이 낮을 수 있으며, 정성적 기준은 점검자의 주관에 개입될 여지가 많아서 일관성에 문제가 있을 수 있다.

### 2.2 시계열 기법

시계열 기법은 연속적으로 수집된 자료의 시계열적 특성을 활용하는 방법으로 평활화 기법을 응용하는 것이다. 시계열 기법을 이용한 예로는 TransGuide, TranStar, Transmit 등의 시스템이 있는데 이 세 가지 시스템은 1990년 후반에 개발·설치되었으며 2002년까지는 미국에서 구간통행시간 자료 수집을 위하여 운영 중인 유일한 AVNPR 시스템이었다.

TransGuide는 미국의 San Antonio 고속도로 교통관리시스템에서 운영 중인 AVNPR 시스템으로 구간통행시간 추정을 위하여 수집주기 내에서 일정 범위를 초과하는 통행시간 값을 자동적으로 제거하는 이동 평균 알고리즘을 도입하였다. 유효 통행시간의 평균을 구하는 방법은 다음과 같다.

$$\begin{aligned} Stt_{ABi} = \{ & t_{Bi} - t_{Ai} | t - t_w < t_{Bi} < t \\ & \text{and } t'_{ABi}(1-l_{th}) \leq t_{Bi} - t_{Ai} \leq t'_{ABi}(1+l_{th}) \} \end{aligned} \quad (6)$$

$$tt_{ABi} = \frac{\sum_{i=1}^{Stt_{ABi}} (t_{Bi} - t_{Ai})}{|Stt_{ABi}|} \quad (7)$$

여기서,  $Stt_{ABi}$  : 시간 t에서 AB구간에서 관측된 유효차량 수

$t_{Ai}$  : A지점에서 관측된 차량 관측시간

$t_{Bi}$  : B지점에서 관측된 차량 관측시간

t : 수집주기

$t_w$  : 이동 평균 창(rolling average window)

$l_{th}$  : 링크 초기 통행시간 파라미터(0.2)

$tt_{ABi}$  : 시간t에서 AB구간에서 관측된 유효차량의 평균 통행시간

$t'_{ABi}$  : t-1시간에서 AB구간에서 관측된 유효차량의 평균통행시간

TranStar 알고리즘은 TransGuide 알고리즘과 유사하지만 이상치 제거의 기준이 되는 통행시간 값이 주기적으로 갱신이 된다는 점이 TransGuide 알고리즘과의 주된 차이점이다. 앞에 언급한 두 알고리즘이 관측치에서 이상치를 제거한 후 평균통행시간을 계산하는 반면 Transmit 알고리즘은 이상치를 제거하지 않고 포함하며 현재 수집주기 관측치의 평균을 10% 그리고 이전 평균통행시간을 90% 반영하여 현재 수집주기의 평균통행시간을 계산한다 (Dion and Rakha, 2004; Jang, Byun, Baik and Kim, 2005).

위에서 살펴 본 바와 같이 교통량 자료에 대하여 이상치 판정이 적용된 사례는 거의 없다. 그나마 교통량 자료에 대하여 이상치 판정이 적용된 사례로 예를 든 통계적 기법의 경우도 수집된 자료가 이상치로 판정될 확률이 매우 낮으며 또한, 수집된 자료가 정규분포를 가진다는 가정이 성립되어야만 적용할 수 있다. 시계열 기법은 현재 수집 주기의 정상치 여부를 판정함에 있어서 바로 직전 수집 주기의 평균통행시간만을 이용하여 최대최소값으로 유효 범위를 정한 후 유효 범위에 속하는 자료를 유효 자료로 판정하므로 초기 통행시간 파라미터가 중요한 요소로 작용한다.

### 3. 이상치 판정 알고리즘 개발

상시조사 장비에서 연속적으로 수집되는 교통량 자료에 대한 이상치 판정 방법은 교통량 자료의 시계열적인 변동 특성을 반영하여 개발하며 기본적인 고려사항은 1) 교통량 자료의 시계열적 특성 반영, 2) 연속적으로 수집되는 자료에 대하여 특정 주기로 유효성 검증 가능, 3) 신뢰성(변별력)이 높고 이해하기 쉬우며 시스템 구현이 간단한 알고리즘 등이다.

Table 1. Types of the Abnormal Data Produced by the Permanent Traffic Counter and Their Causes

Type of Abnormal Data		Type of Missing Value and Outlier	Direction		Cause
			Up	Down	
Missing Value	Hourly Volume	Missing Type 1)	(×)	(×)	• Breaking or malfunctioning of sensor
		Missing Type 2)	×	×	• Breaking or malfunctioning of sensor, breaking down of controller, breaking down of power
	Daily Volume	Missing Type 3)	(×)	(×)	• Depending upon Missing Type 1) (when hourly volumes are not collected for 24 consecutive hours)
		Missing Type 4)	×	×	• Depending upon Missing Type 2) (when hourly volumes are not collected for 24 consecutive hours)
Outlier	Hourly Volume	Outlier Type 1)	(△)	(△)	• Malfunctioning or sensitivity degradation of sensor
		Outlier Type 2)	△	△	• Malfunctioning or sensitivity degradation of sensor
	Daily Volume	Outlier Type 3)	(△)	(△)	• Depending upon Missing Type 1) (when hourly volumes are not collected for several hours) • Depending upon Outlier Type 1)
		Outlier Type 4)	△	△	• Depending upon Missing Type 2) (when hourly volumes are not collected for several hours) • Depending upon Outlier Type

Note : '×' indicates the missing of data. '△' indicates that the volume is out of the normal range. '( )' indicates that it applies only for one direction whether it is upstream or downstream.

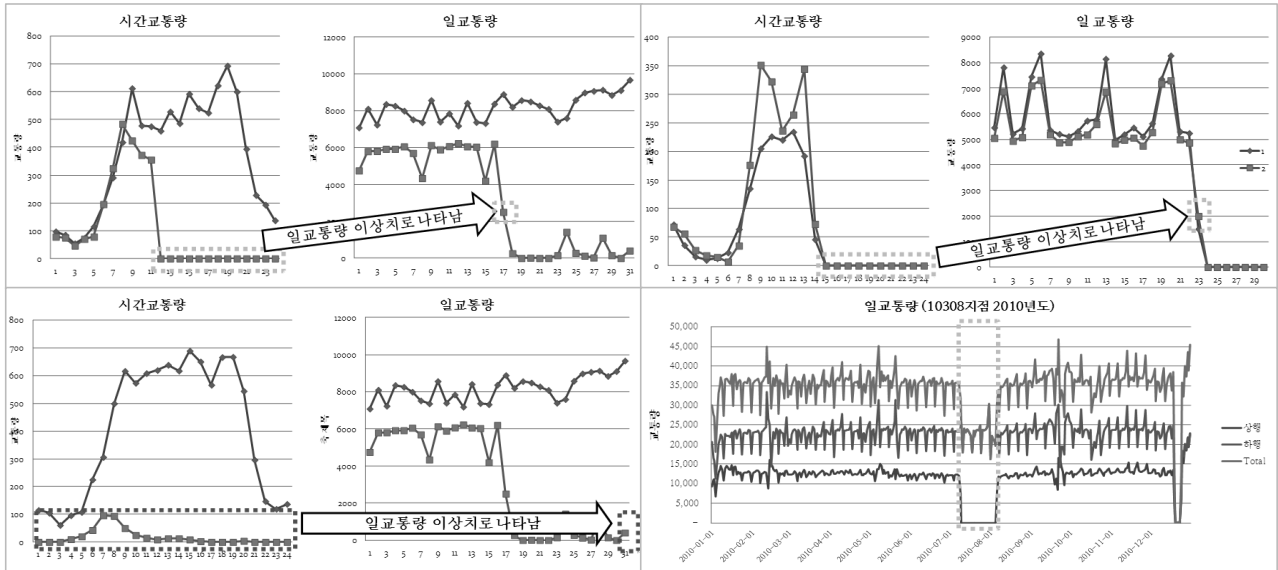


Fig. 3. Effects of Hourly Traffic's Abnormality on Daily Traffic

### 3.1 상시조사 장비의 이상치 유형과 발생원인

본 논문 제1장에서 언급하였듯이 한국건설기술연구원에서 사용하고 있는 상시조사 장비(Permanent Traffic Counter)인 AVC (Automatic Vehicle Classifier)는 교통량을 검지하는 루프 및 피에조 센서가 도로포장 면에 매설되는 장비로서 이상 발생의 원인은 리드선을 포함한 센서의 단선 및 불량, 감도 저하가 높은 빈도로 발생하고 있으며, 이 경우 통과하는 교통량을 누락하게

되어 실제 교통량보다 적은 교통량을 검지한 것으로 나타나게 된다. 또한 제어기 고장이나 전원 고장의 경우도 마찬가지로 결과로 나타나게 된다.

상시조사 장비의 최소 수집단위인 시간교통량이 결측되거나 수집된 자료에 이상치가 포함될 경우에는 일교통량에 영향을 주어 일교통량 또한 이상치로 나타나게 된다.

이러한 특성을 이용하면 일교통량의 점검을 통하여 이상치를

Table 2. Examples of the Directional Traffic Deviation for Continuous Traffic Counts Sites

Site ID	Day of Week/ Direction		Average (veh/day)	Standard Deviation (veh/day)	Coeff. of Variation(CV)	Site ID	Day of Week/ Direction		Average (veh/day)	Standard Deviation (veh/day)	Coeff. of Variation(CV)
10002	Sun	1	10,472	1,442.54	0.14	20351	Sun	1	1,878	487.78	0.26
		2	12,633	1,789.60	0.14			2	2,030	488.99	0.24
	Mon	1	13,147	1,638.20	0.12		Mon	1	1,534	421.00	0.27
		2	13,312	1,574.58	0.12			2	1,508	462.07	0.31
	Tue	1	13,055	1,224.02	0.09		Tue	1	1,446	391.51	0.27
		2	13,019	1,447.02	0.11			2	1,476	436.21	0.30
	Wed	1	13,014	1,284.28	0.10		Wed	1	1,483	429.29	0.29
		2	13,057	1,427.32	0.11			2	1,536	439.72	0.29
	Thur	1	12,890	1,314.58	0.10		Thur	1	1,445	396.14	0.27
		2	12,933	1,407.59	0.11			2	1,534	560.23	0.37
	Fri	1	14,236	1,275.76	0.09		Fri	1	1,629	370.29	0.23
		2	13,585	1,145.66	0.08			2	1,678	353.74	0.21
	Sat	1	14,531	1,878.28	0.13		Sat	1	1,964	488.34	0.25
		2	13,207	1,465.27	0.11			2	1,918	401.71	0.21
Average	1	13,049	1,884.59	<b>0.14</b>	Average	1	1,626	468.43	<b>0.29</b>		
	2	13,107	1,490.38	<b>0.11</b>		2	1,669	493.54	<b>0.30</b>		

Note : 1) Coefficient of Variation(CV) =  $\frac{\text{Standard Deviation of Daily Volume}}{\text{AADT or MADT}} \times 100$

- 2) '1' indicates upstream and '2' indicates downstream. For the highway route of even number, the direction from east to west is expressed as upstream. For the highway route of odd number, the direction from south to north is expressed as upstream.
- 3) Site 10002 : Located at Yangbok-Rhi, Bogae-Myun, Ansung-Shi, Kyunggi-Do of the National Highway Route 38
- 4) Site 20351 : Located at Dongjeom-Dong, Taebaek-Shi, Kangwon-Do of the National Highway Route 31

효율적으로 구분할 수 있으며, 또한 알고리즘을 통하여 이상치로 판정된 자료가 이상치인지 아닌지를 확인할 수 있다.

Table 2와 Fig. 4는 일반국도 상시조사 지점들의 일교통량을 평활화하여 유효범위를 선정할 경우 요일 구분 없이 전체 수집일을 유효범위로 할 경우와 요일별을 유효범위로 할 경우 중 어느 경우가 보다 효과적인지를 알아보기 위하여 일반국도 38호선 경기도 안성시 보개면 양북리 소재 상시조사 지점과 일반국도 31호선 강원도 태백시 동점동 소재 상시조사 지점의 2개 지점에 대하여 요일별 구분 없이 전체 수집일에 대한 교통량의 분산 또는 변동과 요일별 교통량의 분산 또는 변동을 분석한 결과를 정리한 것이다.

일반국도 38호선 경기도 안성시 보개면 양북리 소재 상시조사 지점의 경우 상행 방향은 일요일의 변동계수가 0.14로 가장 크며 금요일은 0.09로 가장 작다. 한편 요일 구분 없이 전체 수집일의 변동계수는 0.14로 일요일의 변동계수와 같으며 그 외의 요일들의 변동계수보다는 크다. 하행 방향은 일요일의 변동계수가 0.14로 가장 크며 금요일은 0.08로 가장 작다. 한편 요일 구분 없이 전체 수집일의 변동계수는 0.11로 일요일과 월요일을 제외한 요일들의 변동계수보다는 크거나 같다.

일반국도 31호선 강원도 태백시 동점동 소재 상시조사 지점의 경우 상행 방향은 수요일의 변동계수가 0.29로 가장 크며 금요일은 0.23으로 가장 작다. 한편 요일 구분 없이 전체 수집일의 변동계수는 0.29로 수요일의 변동계수와 같으며 그 외의 요일들의 변동계수보다는 크다. 하행 방향은 목요일의 변동계수가 0.37로 가장 크며 금요일과 토요일은 0.21로 가장 작다. 한편 요일 구분 없이 전체 수집일의 변동계수는 0.30으로 월요일과 목요일을 제외한 요일들의 변동계수보다는 크거나 같다.

2개의 상시조사 지점에 대한 교통량 변동계수 분석 결과 요일별 구분 없이 전체 수집일의 변동계수 즉 분산은 요일별로 구분한 변동계수 즉 분산보다 크게 나타남을 알 수 있다. 따라서 일교통량을 평활화하여 유효범위를 선정할 경우 요일별로 구분하여 유효범위를 선정하는 것이 효과적임을 알 수 있다.

### 3.2 이상치 판정 알고리즘 개발

도로의 교통량은 시간의 흐름과 함께 변화하며 어느 기간 동안 증가 추세를 보이던 교통량이 어느 시점부터 감소 추세로 바뀔 수도 있고, 또한 어느 기간 동안 일정한 수준의 평균을 유지하던

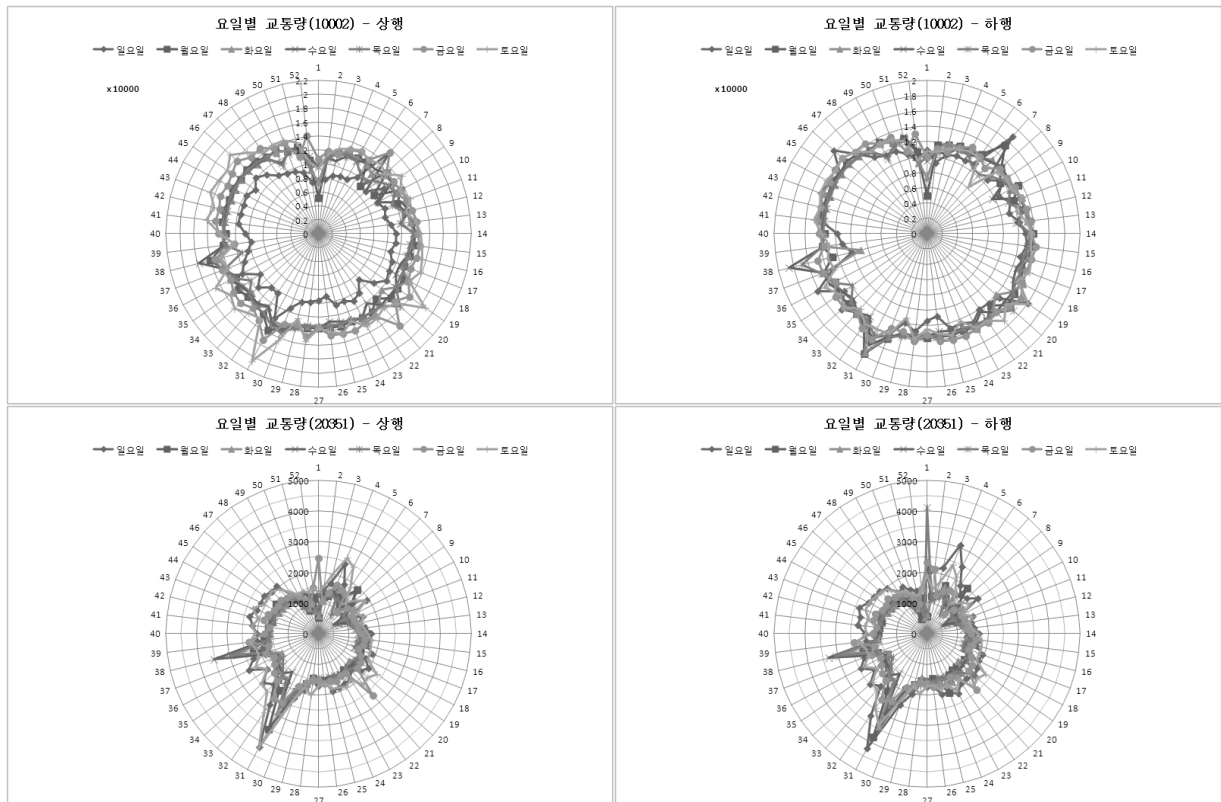


Fig. 4. Examples of the Day-of-the-Week Traffic Variation

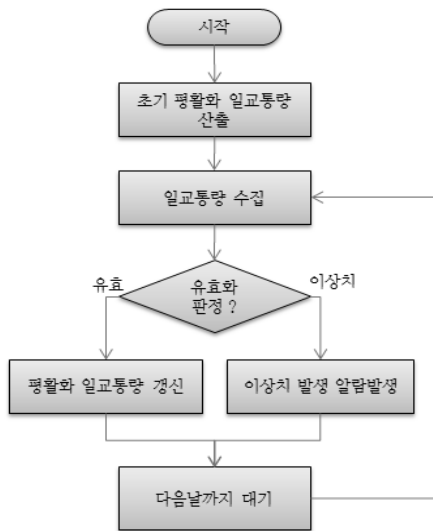


Fig. 5. Outlier-Filtering Algorithm

교통량이 어느 시점부터 얼마간 증가하여 새로운 평균을 유지할 수도 있다.

따라서, 이상치 판정 알고리즘 개발을 위하여 연속적으로 수집된 과거 자료와 최근 자료에 가중치를 부여하여 평활화한 일교통량을

기댓값으로 선정된 후 유효범위를 선정한다. 단계별 세부사항은 Fig. 5에 나타나 있다.

단계는 초기 평활화된 일교통량을 산출하는 단계로 현재 시점에서 수집된 일교통량과 이전 시점에서 유효화된 일교통량에 가중치를 곱하여 기댓값을 산출한다.  $k=0$  시점의 평활화 교통량은 과거 연도의 이력자료가 있는 경우에는 과거 연도 연평균일교통량(AADT)을 적용하고, 과거 연도의 이력자료가 없는 신규 지점의 경우에는 최초로 수집된 일교통량(ADT)을 적용한다.

평활화 계수( $\alpha$ )는 기본적으로 0.5를 적용하되 지점의 교통량 변동 특성에 따라 조정하여 사용한다. 즉 교통량의 변동이 클 경우는  $\alpha$ 를 크게 적용하고 교통량의 변동이 작을 경우는  $\alpha$ 를 작게 적용한다.

$$V_{ij}^k = \alpha \cdot v_{ij}^k + (1 - \alpha) \cdot V_{ij}^{k-1} \quad (8)$$

여기서,  $\alpha$  = 평활화 계수(0.5) 적용,  $0 \leq \alpha \leq 1.0$

$V_{ij}^{k-1}$  =  $i$ 지점  $j$ 요일의  $(k-1)$ 번째 평활화 일교통량

$v_{ij}^k$  =  $i$ 지점  $j$ 요일의  $k$ 번째 유효 일교통량

$i$  = 지점

$j$  = 요일

Table 3. Criteria for  $V_i^{capa}$

Type of Site	Calculation of $V_i^{capa}$
Newly installed site	<ul style="list-style-type: none"> <li>Calculating based on hourly volume</li> <li>- 3,200 veh/hr for two-lane national highway(1,600 veh/hr for one way)</li> <li>- 8,400 veh/hr for four-lane national highway</li> </ul>
Already-in-operation site	<ul style="list-style-type: none"> <li>Maximum value of daily volumes collected with no error(directional)</li> </ul>

Table 4. Continuous Traffic Counts Sites for the Outlier-Filtering

Site ID	Address	National Highway Route No.	No. of Lanes	AADT(veh/day)
10002	Yangbok-Rhi, Bogae-Myun, Ansong-Shi, Kyunggi-Do	38	4	26,169
10343	Doonjeon-Rhi, Pokok-Eup, Yongin-Shi, Kyunggi-Do	45	4	20,499
20351	Dongjeom-Dong, Taebaek-Shi, Kangwon-Do	31	4	3,281
20356	Yeonsang-Rhi, Joongdong-Myun, Yungwol-Khoon, Kangwon-Do	31	4	14,796
30457	Imhan-Rhi, Tanboo-Myun, Boeun-Khoon, Choongchungbuk-Do	25	2	2,750
30466	Daeso-Rhi, Iryu-Myun, Choongju-Shi, Choongchungbuk-Do	3	4	32,151
40001	Nasung-Rhi, Nam-Myun, Yeonki-Khoon, Choongchungnam-Do	1	4	24,441
40009	Eosan-Rhi, Palbong-Myun, Seosan-Shi, Choongchungnam-Do	32	4	22,701
50011	Osong-Rhi, Ongdong-Myun, Jungeup-Shi, Jeonrabuk-Do	1	4	12,865
60013	Yongjeon-Rhi, Miryuk-Myun, Bosung-Khoon, Jeonranam-Do	29	4	5,256
60604	Oechi-Rhi, Wolya-Myun, Hampyung-Khoon, Jeonranam-Do	22	4	16,892
70013	Sabu-Rhi, Sungsan-Myun, Koryung-Khoon, Kyungsangbuk-Do	26	4	11,785
70014	Buya-Rhi, Chungdo-Myun, Chungdo-Khoon, Kyungsangbuk-Do	20	2	4,035
80004	Sangmun-Rhi, Munsan-Myun, Jinju-Shi, Kyungsangnam-Do	2	4	16,518

2단계에서는 국도상 조사지점에 설치되어 있는 상시장비를 통하여 일교통량을 입력받는 단계로 1일 동안 수집된 교통량 자료가 점검대상이 된다.

3단계에서는 평활화된 일교통량을 이용하여 유효범위를 선정 한 후 수집된 일교통량과의 비교를 통하여 유효화를 판단한다.

$$\max\{0, V_{ij}^{k-1} \cdot (1-\delta)\} \leq v_{ij}^k \leq \min\{(V_{ij}^{k-1} \cdot 1+\delta), V_i^{capa}\} \quad (9)$$

여기서,  $v_{ij}^k = i$ 지점  $j$ 요일  $k$ 번째 일교통량

$V_{ij}^{k-1} = k-1$ 번째까지 평활화한 일교통량

$\delta =$  유효 계수(기본값: 0.2)

$V_i^{capa} = i$ 지점의 가능 용량 또는 기본 용량

유효범위의 상한값은 이전 일까지의 평활화한 일교통량에 유효 계수를 더하여 산출하며 지점의 가능 용량 또는 기본 용량보다 작아야 하며 하한값도 상한값과 유사하게 유효 계수를 곱하여 산출하되 0보다 커야 한다. 유효 계수는 기본값으로 0.2를 적용하되 지점의 교통량의 변동성에 따라 유동적으로 적용한다. 즉 일교통량

의 변동이 심한 경우에는 높은 값을 적용하며, 심하지 않은 경우에는 낮은 값을 적용한다.

4단계는 평활화 일교통량의 갱신을 수행하는 단계로 유효로 판정된 경우에만 1단계의 식을 적용하여 갱신된 평활화 일교통량을 산출한다.

## 4. 알고리즘의 평가

### 4.1 대상 지점 선정

일반국도 상시조사 지점 중 총 14개 지점을 선정하였으며, 선정된 지점은 Table 4와 같다. 대상 지점은 지역별로 1개 지점 이상을 선정하였으며, 차로수 별로는 4차로 지점 12개, 2차로 지점 2개로 구성하였다.

### 4.2 이상치 필터링 결과

선정된 14개 지점을 대상으로 2010년 수집된 교통량 자료에 대하여 365일간 방향별 요일별 패턴에 기반한 필터링을 수행하였다. 이상치로 필터링된 결과는 확인 가능한 장비의 유지관리 이력과 비교하여 필터링의 유효성을 평가하였다.



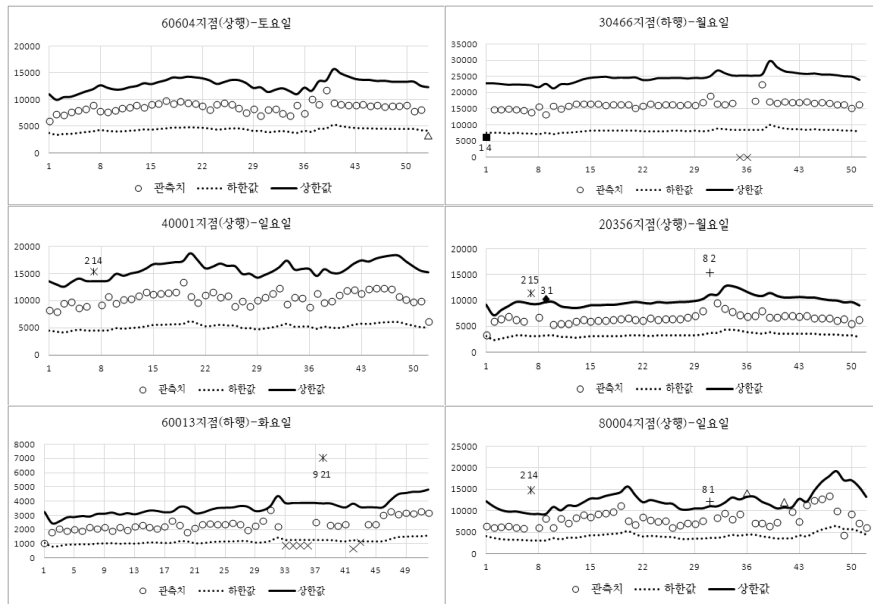


Fig. 6. Examples of the Outlier-Filtering for Continuous Traffic Counts Sites

Table 5. Summary of the Outlier-Filtering Results for Continuous Traffic Counts Sites

Site ID	Direction	Maintenance Record		Filtering Result		Filtering Rate for Normal Days(%)	
		No. of Days of Normal Operation	No. of Days of Abnormal Operation	No. of Normal Days	No. of Abnormal Days	Direction	Site
10002	Up	365	0	363	2	99.5	99.5
	Down	365	0	363	2	99.5	
10343	Up	365	0	363	2	99.5	<b>99.5</b>
	Down	365	0	363	2	99.5	
20351	Up	365	0	351	14	96.2	96.4
	Down	362	3	350	15	96.7	
20356	Up	365	0	355	10	97.3	97.5
	Down	365	0	357	8	97.8	
30457	Up	353	12	344	21	97.5	97.7
	Down	353	12	346	19	98.0	
30466	Up	355	10	353	12	99.4	99.3
	Down	356	9	353	12	99.2	
40009	Up	365	0	362	3	99.2	99.0
	Down	363	2	359	6	98.9	
40001	Up	362	3	360	5	99.4	99.4
	Down	362	3	360	5	99.4	
50011	Up	355	10	349	16	98.3	98.6
	Down	354	11	350	15	98.9	
60013	Up	330	35	319	46	96.7	<b>96.2</b>
	Down	327	38	313	52	95.7	
60604	Up	365	0	359	6	98.4	98.2
	Down	363	2	356	9	98.1	
70013	Up	365	0	358	7	98.1	98.0
	Down	345	20	338	27	98.0	
70014	Up	365	0	353	12	96.7	96.8
	Down	362	3	351	14	97.0	
80004	Up	359	6	350	15	97.5	97.9
	Down	365	0	359	6	98.4	
Total		10,041	179	9,857	363	<b>Average</b>	<b>98.2</b>

이상치 필터링을 통하여 이상치로 판정된 이상일과 장비의 유지 관리 이력 중 이상으로 확인된 일과 비교한 결과는 Table 5와 같으며, 전체적으로 이상일 수보다 많은 일이 이상일로 나타났다. 지점별 정상일에 대한 필터링율은 최소 96.2%에서 최고 99.5%이며, 전체 평균은 98.2%를 보이고 있다. 그러나 알고리즘을 통하여

이상치로 판정된 이상일에는 단순한 장비 이상 외의 다른 요인들이 있을 수 있으므로 원인분석을 수행하였다.

요일교통량 패턴의 변동에 큰 영향을 미치는 원인으로는 기상, 연휴, 명절기간, 하계휴가 기간 등이 있으며, 2010년 중 확인 가능한 기간을 정리하면 Table 6과 같다.

Table 6. Special Occasions in 2010

Type of Occasion	Period
Bad Weather(Heavy Snow)	2010. 1. 4
Holiday	2010. 1. 1 ~ 1. 3(New Year's Day, Friday ~ Sunday) 2010. 2.28 ~ 3. 1(Samiljeol, Sunday ~ Monday) 2010. 5. 5(Children's Day, Wednesday) 2010. 5.21 ~ 5.23(Buddha's Birthday, Friday ~ Sunday)
Festive Day(Seol, Chuseok)	2010. 2.12 ~ 2.16(Seol, Friday ~ Monday) 2010. 9.18 ~ 9.26(Chuseok, Saturday ~ Monday)
Vacation Season	2010. 7.23 ~ 8. 8(Summer Vacation, Friday ~ Sunday)

Table 7. The Analysis of Abnormal Days and the Outlier-Filtering Rate

Site ID	Direction	Abnormal Days Based upon Filtering	Bad Weather	Holiday	Festive Day	Vacation Season	Abnormal Operation	Unidentified Cause	Mis-Filtering Rate(%)
10002	Up	2	1	0	1	0	0	0	0.0
	Down	2	1	0	1	0	0	0	0.0
10343	Up	2	1	0	1	0	0	0	0.0
	Down	2	1	0	1	0	0	0	0.0
20351	Up	14	2	2	3	5	0	2	14.3
	Down	15	2	1	4	5	3	0	0.0
20356	Up	10	1	3	2	4	0	1	12.5
	Down	8	1	2	1	3	0	0	0.0
30457	Up	21	1	1	4	0	12	3	14.3
	Down	19	1	0	4	0	12	2	10.5
30466	Up	12	1	0	1	0	10	0	0.0
	Down	12	1	2	0	0	9	0	0.0
40009	Up	3	0	0	1	2	0	0	0.0
	Down	6	0	1	0	3	2	0	0.0
40001	Up	5	1	0	1	0	3	0	0.0
	Down	5	1	0	1	0	3	0	0.0
50011	Up	16	1	0	3	0	10	2	12.5
	Down	15	1	0	1	0	11	2	13.3
60013	Up	46	0	2	4	4	35	1	2.2
	Down	52	0	3	6	4	38	1	1.9
60604	Up	6	1	0	4	0	0	1	16.7
	Down	9	1	0	3	0	2	3	<b>33.3</b>
70013	Up	7	0	0	4	2	0	1	14.3
	Down	27	0	1	4	2	20	0	0.0
70014	Up	12	0	2	2	4	0	4	<b>33.3</b>
	Down	14	0	1	3	4	3	3	21.4
80004	Up	15	0	0	4	3	6	2	13.3
	Down	6	0	0	3	2	0	1	16.7
Total		363	20	21	67	47	179	29	<b>Average : 8.0</b>

필터링된 이상일을 특별상황 기간 및 장비 유지관리 이력에 따라 세분하면 Table 7과 같다. 이상치 필터링 알고리즘은 요일별 교통량 변동 패턴에 기반하므로 요일별 교통량 변동 패턴과 상이한 기상이상, 연휴기간, 명절기간, 하계휴가기간, 장비 이상은 당연히 이상일로 필터링 되는 것이 합리적이다. 따라서 미확인된 일수는 추가적인 확인이 필요하나 오퍼터링을 엄격히 산출하기 위하여 이상일 중 미확인된 일수를 기준으로 산정한다.

Table 7에 나타난 이상일의 원인 분포를 분석해 보면 Fig. 7과 같다. 즉, 전체 이상일 363일 중 가장 큰 비중을 차지하는 것은 장비 이상으로 179일, 49.3%를 차지하며 다음은 설, 추석 등의 명절 기간으로 67일, 18.5%, 그 다음은 하계휴가 기간으로 47일, 12.9%의 순서이며 가장 비중이 작은 것은 기상 이상으로 20일, 5.5%를 차지한다. 이상일의 원인 확인이 이루어지지 못한 경우는 29일로 8.0%를 차지한다.

위에서 언급한 것처럼 미확인된 일수를 기준으로 오퍼터링율을

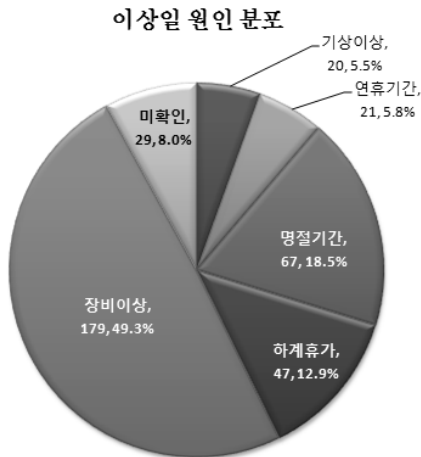


Fig. 7. Distribution of Causes for Abnormal Days

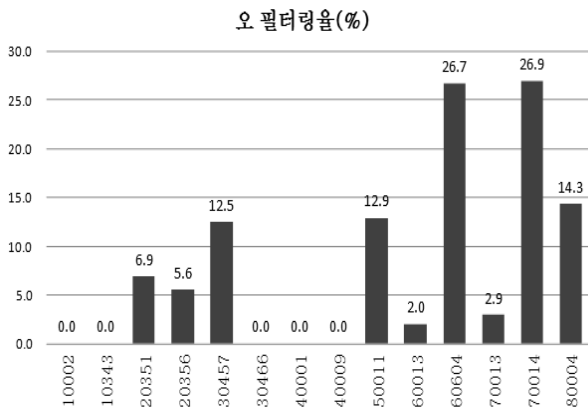


Fig. 8. Mis-Filtering Rates for the Continuous Traffic Counts Sites

산출하면 상행과 하행을 구분하였을 경우, 지점에 따라 0.0%~33.3%의 범위로 나타나며, 전체 평균은 8.0% 정도로 나타나고 있다. 상행과 하행을 구분하지 않고 양방향을 합쳐서 오퍼터링율을 산출한 결과는 Fig. 8에서 볼 수 있듯이 0.0%~26.9%의 범위로 나타나고 있다.

이상일 원인 중 장비 이상의 경우는 장비 유지관리 이력과 비교하여 볼 때 평가 대상 14개소 전체 지점에 대하여 전체 이상일을 다 걸러낸 것으로 확인되어서 필터링율이 100%라고 할 수 있다.

## 5. 결론 및 향후 연구 과제

### 5.1 결론

교통량 자료는 도로 설계, 포장 설계, 도로 운영 및 유지관리 등에 폭넓게 이용되며 도로망 계획, 교통망 계획 및 교통관리 계획 수립을 위한 기초적 정보가 된다. 이러한 중요한 자료를 수집하기 위하여 국토해양부는 한국건설기술연구원에 위탁하여 전국의 일반 국도에 대한 교통량 상시조사를 시행하고 있다.

이에 한국건설기술연구원은 일반국도 상시조사를 위하여 2011년 말 기준으로 485대의 고정식 교통량조사 장비인 AVC를 설치하여 1년 이상의 장기간에 걸쳐 그 지점을 통과한 차량의 차종별 대수 및 총 대수를 조사하고 있다.

하지만 AVC는 여러 가지 원인으로 인하여 오작동이 발생하거나 작동이 멈추게 되며 아무리 신속하게 고장을 처리한다고 하더라도 연간 며칠 이상의 기간은 자료가 누락되거나 정상적이지 않은 자료, 즉 이상치가 수집된다. AVC의 최소 수집단위인 시간교통량이 결측되거나 수집된 자료에 이상치가 포함될 경우는 일교통량에 영향을 주어 일교통량 또한 이상치로 나타나게 된다.

이러한 이상치를 효과적으로 걸러내기 위하여 기존에 한국건설기술연구원에서 사용해 온 이상치 점검 방법 즉 정량적 기준에 의한 점검 방법과 정성적 기준에 의한 점검 방법과는 달리 교통량의 시간적 변동 특성을 설명할 수 있는 평활화법을 활용하여 연속적으로 수집되는 자료의 이상 유무를 판정할 수 있는 정량적 이상치 판정 알고리즘을 개발하였다.

또한 개발된 알고리즘을 활용하여 2010년 일반국도 상시조사 지점 중 14개 지점의 교통량 자료에 대한 이상치 필터링을 수행하여 필터링율과 오퍼터링율을 산출함으로써 알고리즘의 적합성 여부를 평가하였다. 그 결과 정상일 필터링율은 98.2%, 이상일 중 오퍼터링율은 8.0%로 평가되었다.

따라서 본 연구에서 개발된 알고리즘은 수집된 교통량 자료의 1차적인 이상치 필터링에 충분히 적용 가능할 것이며, 이상치로 판단된 일에 대한 원인 분석을 통하여 최종 이상치로 판정하는

단계가 추가된다면 좀 더 신뢰성 있는 필터링이 가능할 것이다.

또한 본 연구에서 개발된 이상치 필터링 알고리즘을 활용하면 수집되는 교통량 자료의 이상 여부를 1일 주기로 점검이 가능하며, 이상치가 포함된 이상일의 원인 중 장비 이상의 경우는 평가대상 전체 지점에서 100% 필터링이 가능하므로 AVC 장비 유지관리 시스템과 접목 한다면 장비 유지관리의 효율성을 높일 수 있을 것이다.

## 5.2 향후 연구 과제

본 연구에서는 알고리즘의 필터링 범위를 결정하는 유효계수를 단일 값만 제시하였으나 실제 적용을 위해서는 보다 많은 지점에 대하여 적용 및 검증을 함으로써 알고리즘을 정교화할 필요가 있을 것이며, 경계치 처리의 정교한 조정이 필요할 것이다.

게다가 요일별 교통량 패턴과 구분되는 특징을 가진 기상 이상일, 연휴기간, 명절기간, 휴가기간에 대하여도 별도로 구분하여 적용 가능한 알고리즘에 대한 추가적인 연구가 필요할 것이다.

본 논문은 한국건설기술연구원의 2013년 주요사업 「효율적인 ITS 성능평가를 위한 Universal 기준장비 개발」의 지원으로 작성되었습니다.

## References

- Dion, F. and Rakha, H. (2004). "Estimating Spatial Travel Times Using Automatic Vehicle Identification Data", Presented at 83rd Annual Meeting of the Transportation Research Board, Washington DC.
- Jang, Jinhwan, Byun, Sangchul, Baik, Namchul and Kim, Sunghyun (2005). "A Data Processing Algorithm for AVI Equipment - Performance Tests on National Highway", *Journal of the Korean Society of Civil Engineers*, Vol. 25 No. 2D, pp. 233-240 (in Korean).
- Kang, Jin-Kee, Son Youngtae, Yoon, Yeo-Hwan and Byun, Sangchul (2002). "Regional Traffic Information Acquisition by Non-Intrusive Automatic Vehicle Identification", *Journal of the Korea Institute of Intelligent Transport Systems*, Vol. 1 No. 1, pp. 22-32 (in Korean).
- Korea Institute of Construction Technology (2001). *Developing the Missing Value Imputation Model for the Travel Time Estimation and Statistical Data Imputation* (in Korean).
- Lee, Sang Hyup, Ha, Jung Ah and Yoon, Taekwan (2012). "A Study on Improving the National Highway Traffic Counts System : with Focus on Short Duration Counts and Continuous Counts", *Journal of the Korean Society of Civil Engineers*, Vol. 32 No. 3D, pp. 205-212 (in Korean).
- Seoul Metropolitan Government (2003). *Report of the Basic Design for Installing the Urban Expressway Traffic Management System (Phase III)* (in Korean).