

# 고속도로 차량검지기 이력자료 활용을 위한 전처리과정 개선

## Improvement of A Preprocessing of Archived Traffic Data Collected by Expressway Vehicle Detection System

이 환 필\*  
(Hwan-Pil Lee)

남 궁 성\*\*  
(Seong Namkoong)

김 수 희\*\*\*  
(Soo-Hee Kim)

김 진\*\*\*\*  
(Jin Kim)

### 요 약

간 차량검지기로부터 수집되는 다양한 정보는 주로 실시간 자료로 이용되었으나 최근 교통데이터 이력자료의 활용방안에 대한 중요성이 증대되고 있다. 이러한 배경에서 본 연구는 차량검지기자료의 이력자료 활용을 위한 전처리 개선에 대한 연구를 수행하였다. 실제 교통현상과 가장 가까운 데이터 처리를 목적으로 세부처리로직을 개선하였다. 평가결과 기존 전처리 과정보다 개선 전처리 과정이 실제값에 가까운 결과를 나타내는 것으로 분석되었다.

핵심어 : 검지기자료, 전처리, 데이터 보정, 차량검지시스템, 이력자료

### Abstract

While the vehicle detector is collected from a variety of information was mainly used as a real-time data. Recently scheme of application for archived traffic data has become increasingly important. In this background, this research were conducted on the improvement of the preprocessing for archived traffic data application. The purpose of improving specific preprocessing was reflect transportation phenomena by traffic data. As evaluation result, improvement preprocessing was close to the actual value than exist preprocessing

**Key words** : Preprocessing, Vehicle Detection System, Historical Data, ADUS, Data Imputation

## I. 서 론

### 1. 연구의 배경 및 목적

교통관리를 수행하는 ITS(Intelligent Transportation

System)의 기반시설은 차량검지기이며 수집자료는 실시간 활용부문과 이력자료 활용부문에 구분할 수 있다.

수집자료의 실시간 활용부문은 주로 정체 및 돌

† 본 논문은 2012년 한국ITS학회 춘계학술대회 발표논문을 기초로 수정, 보완하였습니다.

\* 주저자 및 교신저자 : 한국도로공사 교통처 교통정보통합활용지원팀 Post doc.

\*\* 공저자 : 한국도로공사 도로교통연구원 교통연구실장

\*\*\* 공저자 : 한국도로공사 교통처 교통정보통합활용지원팀 선임연구원

\*\*\*\* 공저자 : DB정보통신 경기사업단 팀장

† 논문접수일 : 2012년 10월 24일

† 논문심사일 : 2013년 1월 10일

† 게재확정일 : 2013년 1월 31일

발상황관리, 이용자 정보제공 등과 같은 교통류관리분야로서 시스템의 관리 목적에 맞도록 가공처리하여 사용한다.

이력자료 활용부문은 관련연구, 정책결정 기초자료, 도로운영에 대한 평가 및 통계분석 등이 주요분야로서 아직 국내에서 활발히 적용되고 있지는 않다. 이력자료 활용 목적상 실시간성 활용 자료보다는 실제값에 가까워야 하며, 교통현상이 수집정보에 반영될 수 있는 가공처리 과정을 거쳐야 한다.

차량검지기 수집자료는 차로기반 자료로서 비정상자료나 누락자료를 보정하는 전처리 과정을 거친 후 시, 공간적으로 집계된다.<sup>1)</sup> 집계자료의 신뢰성은 전처리 과정에서 오류자료와 누락자료를 제대로 보정할 수 있는지에 따라 좌우된다.

기존 국내 고속도로 차량검지기 수집자료의 전처리 과정은 대부분 실시간 교통류 관리에 맞추어서 운영되어 왔으며 이력자료의 생성, 가공처리, 저장에 대한 관심이나 적용은 미흡한 실정이었다. 하지만, 최근의 교통자료 이력자료 활용방안에 관한 연구[1-3]와 공공정보 활용 활성화<sup>2)</sup> 등의 추진으로 차량검지기 이력자료 활용분야가 증대되고 있다. 이러한 배경에서 신뢰성, 일관성 확보를 위해 이력자료의 특성을 고려한 전처리 과정 수립이 요구되고 있다.

이에, 본 연구는 정확한 이력자료 생성을 위해 기존 실시간용 전처리 과정을 개선하여 교통이력자료용 전처리 과정을 제시하는 것을 목적으로 한다.

## 2. 연구의 내용 및 범위

본 연구는 고속도로에서 수집되는 차량검지기 자료의 전처리과정을 연구대상으로 한다. 연구 수행을 위해 차량검지기 자료의 전처리 과정에 관한 선행연구 분석으로 기존 전처리 과정의 문제점과 개선방향을 설정하였고 이를 토대로 이력자료 생성

을 위한 전처리 과정을 개선하였다.

평가를 위해 개선된 전처리과정과 기존 고속도로 교통관리시스템용으로 실시간 전처리된 자료를 다양한 지표로 비교하였다.

## II. 전처리 기법

### 1. 전처리 과정 개요

차량검지기는 최초 설치 시의 정확도기준을 만족시키더라도 설치 이후부터 기기결함, 유지관리 불량, 장비노후 및 손상 등 다양한 이유로 신뢰성 및 정확도가 저하된다. 따라서, 수집자료 활용을 위해서는 전처리 과정을 통해 비정상자료의 보정과정을 거쳐야만 하며 이러한 전처리 과정은 시스템 목적에 맞도록 구성된다.

차량검지기 원시자료(Raw Data)는 차로별 수집되며 검지기로부터 주기별<sup>3)</sup>로 수집되어 가공처리를 거치기 전 단계의 자료이다. 원시자료에는 검지기 작동상태 불량<sup>4)</sup>, 통신불량 등으로 불완전 자료가 포함될 수 있으며 검지값이 누락된 결측자료와 자료가 명확하지 않은 비정상 자료로 분류할 수 있다.

결측자료는 검지실패, 네트워크 장애, 통신오류 등에 의해 자료가 수집되지 않은 것으로 전처리 과정에서 자료를 채워넣게 된다.(Imputation)

비정상 자료는 오류(Error)와 이상치(Outlier)로 구분되며 검지기 감도 이상, 검지상태 불량, 현장제어기 처리 오류 등에 의해 발생하게 된다. 전처리 과정에서 필터링을 통해 비정상자료를 제외한 후 정상에 가까운 자료로 대치시키게 된다.

### 2. 전처리 기법

전처리 과정은 크게 3가지로 구분할 수 있으며 수집자료의 오류 판단과정과 판단된 오류자료와 결

1) 차량검지기 자료가공과정은 전처리과정과 집계과정으로 구분됨. 일반적으로 30초 차로자료 기반으로 30초 지점자료 생성과정이 전처리 과정에 해당하며 이후의 5분, 15분, 1시간 등의 자료를 생성하는 과정이 집계과정임

2) 공공DB 제공 및 이용에 관한 법률 발의, 2011. 9. 5

3) 국내 FTMS(Freeway Traffic Management System)에서 대부분이 30초를 주기로 사용

4) 국내에서 주로 사용되는 루프검지기의 작동상태 불량원인은 감도이상, 단선, 루프손상 등임

측자료의 보정처리 과정, 수집자료의 일시적 변동 을 상쇄시키기 위한 평활화 과정으로 구분된다.

1) 오류자료 판단기법

오류자료는 교통량, 속도, 점유율 등 수집항목간 에 상호 논리적이지만 않은 자료<sup>5)</sup>와 일반적인 수집범 위를 벗어난 자료로 구분할 수 있다. 이외에 쌍루프 검지기에서 루프헤드(head)간 자료의 차이가 큰 경 우<sup>6)</sup>, 동일한 자료가 지속적으로 수집되는 경우, 현 장제어기 자체적으로 오류자료로 판단하는 경우 등 도 포함된다. 이러한, 오류자료 비율은 자료 유효성 에 영향을 미치고 오류자료가 많을수록 유효성이 떨어지므로 활용가능한 자료가 적어진다.

오류자료 판단과정은 현장제어기에서 검지기 이 상유무를 판단하는 상태확인부분과 자료 범위를 검 토하는 자료 임계치 부분, 자료간 관계를 평가하는 자료관계 평가부분으로 구분된다. 오류자료는 식별

〈표 1〉 오류 판단 기법 세부내용  
〈Table 1〉 Description of Error Check Method

Category	Description
Local controller check	<ul style="list-style-type: none"> <li>• Check detector status by local Controller itself</li> <li>• Divided into normal and abnormal status</li> <li>• Report for Anormal status with tag, but data may be valid</li> </ul>
Univariate range check	<ul style="list-style-type: none"> <li>• Set threshold of volume, speed, occupancy rate, decide to error for out of threshold data</li> <li>• Check each variables</li> <li>• (ex) speed is over than 300km/h</li> </ul>
Data relation check	<ul style="list-style-type: none"> <li>• Process for logical error</li> <li>• Set logical relation between each variables, if some data aren't correspond to logical realtion, decide to error data</li> <li>• Additionally, speed-occupancy rate ration, vehicle length-speed, number of continuous same data, etc. are used to check</li> <li>• (ex) volume is 0 and speed is over than 0</li> </ul>

5) 예를 들어 검지기의 실측값인 교통량이 0일 때 산출값인 속도가 0이상의 값을 나타내는 경우

6) 국내 FTMS에서는 쌍루프 검지기를 대부분 사용

가능한 경우<sup>7)</sup>와 식별이 불가능하여 오류판단과정 을 거쳐야 하는 경우가 있다.<sup>8)</sup>

오류처리된 자료는 비정상자료로 전처리 과정에 서 결측자료와 함께 보정의 대상이 된다.

Korea Expressway Corporation Research Institute (2006)의 연구에서 검지기 오류를 다음과 같이 분류 하고 있으며[4], 검지기의 오류는 발생원인과 유형, 특성이 다양하다.

〈표 2〉 오류 발생 빈도와 지속정도  
〈Table 2〉 Error Occurrence Frequency and Duration

Type	Form	Description
Frequency	Collectively	• The error of data being sent continuously
	Intermittently	• The error of data being sent non continuously and randomly
Duration	Short term	• Collective error symptoms are being persisted in 15 minutes or less
	Long Term	• Collective error symptoms are being persisted more than 15 minutes
Note	If Short term symptoms are appeared collectively, treat as long term error symptoms	

장기적 오류상태는 검지기 및 통신상태 이상으 로 원인 및 오류발생에 대한 확인이 쉽고 지속적인 오류자료로 전처리 과정에서 예외처리하거나 이상 자료로 분류할 수 있다. 하지만 간헐적, 단기적인 오류는 일시적 이상이므로 원인 파악이 어려우며 보정처리를 위한 주 대상이 된다.

2) 보정처리 기법

오류자료와 결측자료는 불완전 자료로서 전처리 과정을 통해 보다 완전한 형태로 보정처리 된다. 주 로 사용하고 있는 보정방식은 공간보정, 시간보정, 이력자료(패턴자료)를 활용한 보정 방식 등이다.

7) [-999(교통량), -999(속도), -999(점유율)]와 같은 경우

8) [1(교통량), 110(속도), 0(점유율)]과 같은 경우

〈표 3〉 보정처리기법의 유형 및 설명  
 〈Table 3〉 Types and description of Correction Method

Category	Description
Spatial Method	<ul style="list-style-type: none"> <li>usually apply short term error status less than 5 min.</li> <li>Assume that target and adjacent detector's traffic characteristic are similar</li> <li>divide into same point data and adjacent point data correction method</li> </ul>
Temporal Method	<ul style="list-style-type: none"> <li>usually apply short term error status less than 5 min.</li> <li>using target detector's time series data</li> <li>Using moving average commonly that is convenient to apply</li> </ul>
Using historical data	<ul style="list-style-type: none"> <li>When spatial and temporal method are unavailable, use historical data method</li> <li>Suitable for long term error symptoms</li> <li>Use pattern data are accumulated by same characteristic data from same point and time</li> </ul>

3) 평활화

일반적으로 검지자료의 수집주기가 짧을수록 일시적 변동(Fluctuation)을 포함하고 자료 가공시 불안정을 야기할 수 있다. 평활화(Smoothing) 과정은 검지자료의 일시적 변동을 보정하기 위해 자료를 연마하는 과정으로 자료에 포함된 불안정을 제거하고 규칙적인 연속성을 확보하게 된다.

주로 실시간 교통류 관리의 교통상태 판정과정에서 일시적 변동에 의한 교통상태의 과도한 변화를 막기 위한 용도<sup>9)</sup>로 사용되고 있으며 국내에서는 지수평활법을 주로 적용하고 있다.

III. 선행연구 및 기존 전처리기법 분석

1. 전처리관련 선행연구 검토

Chao chen et al. (2003)은 단루프 검지기의 교통량, 점유율관계를 통해 오류자료<sup>10)</sup>를 검지하고 인

9) 평활화 과정 미적용시 FTMS시스템의 소통상태 판정에서 소통원활, 정체상태가 과다하게 변경될 수 있음

10) 비정상자료를 의미하며 오류자료와 이상치를 포괄한 의미임. 본 연구에서도 동일한 의미로 사용하며 필요시 구분하여 명시함

접 검지기 선형회귀분석으로 결측자료와 오류자료에 대한 보정알고리즘을 제시하였다<sup>5)</sup>.

Brian L. Smith et al. (2003)의 연구에서는 결측자료 보정을 위해 휴리스틱 기법과 통계기법에 대한 비교를 수행하였다. 동일 자료를 통한 분석 결과 통계기법이 보다 정확하다고 보고하고 있다<sup>6)</sup>.

Han D. C. et al. (2004)은 누락 및 결측자료 보정을 이력자료를 이용한 산술평균 보정, 자기회귀 보정방안에 대한 연구를 수행하였다. 분석결과 교통량은 유사지점 자료를 이용한 자기회귀 보정방안, 속도는 과거 이력자료를 이용한 산술평균 보정이 가장 유의한 것으로 보고하고 있다<sup>7)</sup>.

Kim J. Y. et al. (2006)의 연구에서는 고속도로 수집 교통자료의 결측보정처리 과정개선에 관한 연구를 수행하였다. 인접지점 참조방식과 이력자료를 활용한 방법을 적용하여 추정력을 비교한 결과 이력자료 기반 보정처리 결과가 상대적으로 원자료에 가깝게 도출되는 것으로 확인되었다<sup>8)</sup>.

Chung S. J. et al. (2007)의 연구에서는 교통자료에 대한 품질평가 및 자료처리 기법 구현을 위해 평가계산식을 도입한 품질평가와 오류판단, 결측보정의 자료처리 기법을 제안하였다. 품질평가 개선을 위해서 결측자료의 포함여부에 따라 품질평가 과정을 1, 2차로 구분하였다<sup>9)</sup>.

Texas Transportation Institute and Battelle(2007)의 연구에서 이력자료의 활용을 위한 데이터 품질관리 절차를 제안하였고 이력자료관리시스템(Archived Data Management System : ADMS)의 데이터 처리 절차를 검토하여 품질관리를 위한 타당성 지표를 제시하였다. 추가적으로 교통데이터의 품질관리에 관한 속성정보(Metadata)와 이력정보를 제공하는 것을 권장하고 있다<sup>10)</sup>.

Rafael J. et al. (2009)의 연구에서는 현재 사용하고 있는 검지기 자료의 결측기법에 대한 고찰과 분류를 실시하였다. 결측기법을 시계열, 평균값 적용 등의 독립적인 기법과 시공간분석, 회귀분석 등과 같은 연관기법으로 구분하였다<sup>11)</sup>.

Gang Chang and Tongmin Ge(2011)의 연구에서 결측자료보정을 위해 보정기법에 대한 비교를 실시하

였다. 결측을 불규칙 상황에서 발생하는 전체 결측, 일부결측으로 분류하고 이의는 별도로 결측이유가 분명한 결측상황으로 구분하였다. 분석결과 상황별로 베이시안, 최소자승법, Expectation Maximization 등의 기법이 RMSE를 최소화하는것으로 분석되었다[12].

Weihao Yin et al. (2012)은 비돌발상황에서의 오류데이터 보정을 위해 기존에 적용중인 시, 공간적인 보정방법을 비교하였다. 시간적 보정방법으로 이력데이터의 시간적 경향(trend)를 이용한 보정기법을 적용했고 공간적 보정방법으로 선형회귀(linear regression)과 커널회귀(kernel regression), 차로별 배분율 기법을 적용하여 분석하였다. 다양한 오류비율상에서의 평가결과 시간적 보정방법이 다양한 오류비율에서 공간적 보정방법보다 나은 것으로 분석되었다[13].

## 2. 기존 적용 전처리 기법 현황 및 문제점 분석

### 1) 국내외 FTMS 적용 전처리 기법 검토

국내의 FTMS용 전처리 과정의 구성은 수집데이터에 대한 오류판단 후 오류자료와 결측자료에 대한 보정을 실시하게 된다.

한국도로공사 관리구간, 민자고속도로, 고속국도 우회도로, 서울시고속도로 등 국내 고속도로 FTMS는 시스템별 특별한 차이없이 유사하게 전처리 과정이 구성되어있다. 전술한 것과 같이 자료임계치 평가와 자료관계 평가를 통하여 오류자료를 걸러낸 후 공간적, 시간적, 패턴자료 보정방식등의 간단한 연산처리를 통해 오류 및 결측자료를 채워 놓게 된다.

국외의 오류판단 과정은 국내와 유사하게 자료 임계치 평가 위주로 사용하고 있다. 단기간 보정은 시공간적 보정방법을 위주로 사용하고 장기간보정의 경우 국내는 보정을 실시하지 않지만 국외에서는 별도의 알고리즘을 이용하여 보정하고 있다.

### 2) 문제점 분석

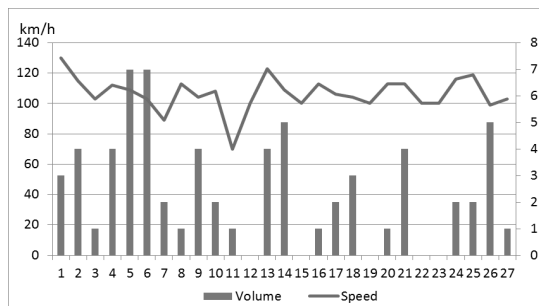
현재 국내에서는 이력자료 활용을 위한 별도의

전처리 과정이 존재하지 않으므로 실시간 교통류 관리를 위한 데이터 전처리 기법의 문제점을 다음과 같이 분류하였다.

첫째, 교통류 관리에 활용하기 위해 일부자료의 경우 강제보정을 실시하여 이에 따른 자료의 왜곡이 발생하고 있다.

고속도로 교통류 관리시스템은 검지기 자료를 기반으로 소통상황을 상황판에 표출하고 있으며, 주로 야간시간대에 통행량이 없어 상황판 표출자료가 없는 경우(Null) 제한속도로 보정하여 소통원활로 표시하고 있다. 이 경우 통과차량이 없는 원인을 추정하게 되며, 혼잡에 의한 경우와 실제 수요가 없는 경우로 구분하게 된다. 혼잡에 의한 경우는 설정된 최저속도, 수요가 없는 경우는 제한속도로 보정하고 있으며 이와 같은 방식으로 보정할 경우 교통량, 속도, 점유율이 논리적으로 비정상값인 [0, 100, 0]의 형태를 띄게 된다. 이러한 자료를 시, 공간적 집계할 경우 오차 누적원인으로 작용한다.

<그림 1>11)에서 교통량이 0일 12, 15, 22, 23주기는 차량이 통과하지 않아 교통량, 속도, 점유율 [0, 0, 0]인 자료를 전처리 과정에서 [0, 100, 0]로 보정한 자료이다.

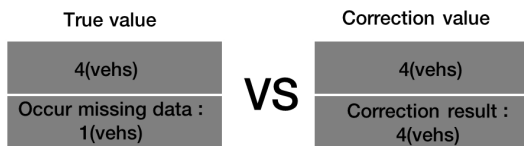


<그림 1> 속도자료 강제보정결과 예시  
<Fig. 1> Example of Compulsion Correction for Speed

둘째, 보정시 기하구조특성인 차로수를 반영하지 않고 획일적인 보정기법을 사용하고 있다.

11) 서울외곽순환고속도로 성남 TG 인근차량검지기(차량검지기ID: 1000VDS00800)의 2011년 11월 24일 02시 경 30초 차로 전처리 결과

대부분의 전처리 과정에서는 차로수에 대한 구분없이 동일한 보정방식을 적용하고 있다. 예를 들어, 동일지점 인접차로의 유효 데이터를 이용한 보정방식을 일괄적으로 적용할 시 참조차로가 1개(편도 2차로, 혹은 검지기 오류발생으로 유효차로가 1개일 경우)이면 과다, 과소보정의 가능성이 있다. 즉, 복수 개의 차로를 참조할 경우 대푯값을 사용할 수 있지만, 참조차로가 1개일 때에는 해당 차로값을 그대로 적용시키게 되므로 <그림 2>와 같이 과다 추정되는 현상이 발생한다.



<그림 2> 유효차로 보정에 따른 과다추정 예시  
<Fig. 2> Example of Over-Correction by Effective Lane

동일한 맥락에서 차로수가 바뀌거나 버스전용차로 등이 존재할 경우 구분하여 적용할 수 있도록 보정과정에 반영되어야 한다.

셋째, 전처리 과정을 거친 데이터가 일부 교통상황을 반영하지 못하는 문제점이 있다.

전처리 과정은 정체가 발생하더라도 차량이 이동가능하다는 전제로 설계되어 있다. 이러한 전제는 논리오류를 검지과정에서 “교통량이 없을 경우 점유율은 0”이라는 것에서 확인할 수 있다. 즉, 교통량, 속도, 점유율 [0, 0, 100]으로 수집될 경우 오류처리하여 보정하게 된다. 예를 들어, 대형사고 및 공사 등으로 인해 차로 폐쇄시 차량 정지상태가 발생하게 되어 검지기 헤드위에 차량이 정지할 경우 교통량, 속도, 점유율이 [0, 0, 100]으로 산출되게 되지만 이 값은 교통현상을 반영한 참값이다. 하지만, 대부분 전처리 과정에서는 오류값으로 보정처리되게 된다. 이러한 경우, 차량 정지상태를 전처리 후 산출자료로 표현하지 못하며, 자동돌발상황감지 (Automatic Incident Detection : AID)에서 돌발상황 검지를 어렵게 하는 요인으로 작용하게 된다.

## IV. 전처리 과정 개선

### 1. 개선방향 설정

전처리 과정은 시스템에서 요구하는 자료형태와 목적에 맞도록 설계되어하며 실시간 활용을 위한 전처리과정과 이력자료 활용을 위한 전처리 과정과 산출데이터는 차별화되어야 한다.

실시간 활용을 위한 전처리 과정은 교통류 관리에 활용할 수 있도록 다음 수집주기 전까지(주로 30초) 전처리 과정이 완료되어야 하며, 특정한 값 발생시<sup>12)</sup> 시스템 운영을 위해 미리 지정한 값으로 변환하거나 오류처리하기도 한다. 또한, 실시간 사용이 목적이므로 전처리 과정에서 발생한 오류데이터 유무와 보정이력을 제공할 필요가 없다.

이력자료 활용을 위한 전처리 과정은 실시간 전처리 과정보다 처리시간 확보가 용이하며, 실제 현상을 반영한 데이터의 산출을 최우선으로 하기 때문에 특정값의 강제보정이 필요하지 않다. 산출된 데이터는 통계, 정책, 연구자료로 활용되므로 이용자가 해당 데이터의 품질이나 보정내역을 확인할 수 있는 이력자료가 제공되는 것이 타당하다.

이러한 차이점을 통해 본 연구에서는 이력자료용 전처리 과정 개선 목표를 “실제 교통현상을 왜곡하지 않고 데이터 보정처리 내역을 확인할 수 있도록 하는 전처리 과정 수립”으로 하고 개선방향을 설정하였다.

#### 1) 실제 현상을 반영할 수 있는 전처리 과정 수립

교통류 관리 시스템에서는 운영 편의를 위해 평활화<sup>13)</sup>, 강제보정 등의 보정기법을 사용하고 있다. 반면, 이력자료 제공시스템에서는 전처리의 목표가 실제값을 반영한 자료 산출이므로 아래와 같은 자료 왜곡항목은 배제하였다.

12) <그림 1>과 같은 경우와 정체시 교통량, 속도, 점유율 [0, 0, 100]을 오류처리하는 경우

13) 교통류 관리시스템에서는 자료의 변동이 클 경우 소통상태의 판정결과가 매 주기마다 변하므로 평활화 과정을 대부분 운영중임

- 자료 변동추세를 완화하는 평활화 과정제거
- 과다, 과소 보정의 우려가 있는 보정방식 지양
- 교통류 관리를 위한 임의적 설정값 강제보정 과정 제거

## 2) 일반화된 판단과정 구축

일부 시스템에서 사용하는 속도-점유율비, 차량 길이-속도 등의 오류판단방식은 수집지점 및 수집 시간별 임계치 조정이 필요하고 획일적 적용시 문제를 발생시킨다. 개선된 오류판단과정에서는 임계치, 자료관계 설정 항목에 대해서 아래와 같이 일반적인 교통특성을 따를 수 있는 인자만을 사용하였다.

- 개별 변수(교통량, 속도, 점유율)의 임계치는 공학적 근거에 기반하여 모든 상황에 적용할 수 있도록 충분한 여유를 두어 설정
- 자료관계 설정시 시간적, 공간적 특성에 따라 변동되는 항목은 배제

## 3) 자료 속성정보 제공

Brian L. Smith는 “보정처리된 데이터에 대해서는 태그(Tag)를 삽입하여 데이터 이용시 이용자가 해당 데이터를 사용할 것인지를 결정할 수 있도록 해야 한다”<sup>[6]</sup>라고 제시하고 있다.

이력자료 활용시에도 이러한 자료 품질측면의 지표와 보정내역 등의 제공이 필요하며 개선 전처리 과정에서도 이를 반영하였다.

- 완전성, 유효성 등의 품질지표 산출
- 전처리 관련 정보를 태그화하여 제공

## 2. 전처리 과정 세부개선방안

### 1) 자료 형식

일부시스템은 실수형 수집자료를 연산의 용이성을 위해 정수형으로 변환하는 경우가 있다. 이런 경우 점유시간이 0미만일 때 0으로 변환시키므로 논리적 오류가 발생한다<sup>14)</sup>. 이를 방지하고자 수집되

는 자료의 형식을 실수형으로 정의하였다.

### 2) 오류판단과정

쌍루프검지기의 헤드별 수집자료는 1개 헤드만 정상작동하면 자료 수집이 가능하다. 그러나, 기존 전처리 방식에서는 1개 헤드의 오류가 발생할 시 해당 자료는 오류처리되고 있다.

검지기 발생오류의 대부분은 장기오류로 1개 헤드의 장기오류 발생시 상당량의 자료가 오류처리된다. 이를 개선하고자 오류판단 대상을 검지기 대륙 값에서 헤드별 자료로 판단방식으로 변경하였다.

오류판단 내용은 헤드별 점유율차, 교통량차에 따라 임계값을 검사하는 1차 오류판단 과정<sup>15)</sup>과 수집된 자료의 임계값을 평가하는 2차 오류판단 과정으로 구성하였다. 2차 오류 판단과정의 임계치 중 교통량은 도로용량편람상 이상적 용량에 여유값을 더한 값이며, 속도는 경험적 최대속도값, 점유율은 산출될 수 있는 최소값과 최대값이다

자료관계 판단과정에서는 논리적으로 발생이 불가능한 자료를 교통량, 속도, 점유율간의 논리적 관계를 통해 판단한다. 단, 차량 정지상태인 교통량, 점유율, 속도 [0, 100, 0]은 예외처리하여 잠재적 돌발상황 상태로 정의하였다. 통행량이 없어 교통량, 점유율, 속도가 모두 0인 경우를 제외하고 일정주기 이상 동일한 값이 반복되는 반복오류를 검출할 수 있도록 반복오류 판단과정을 추가하였다.

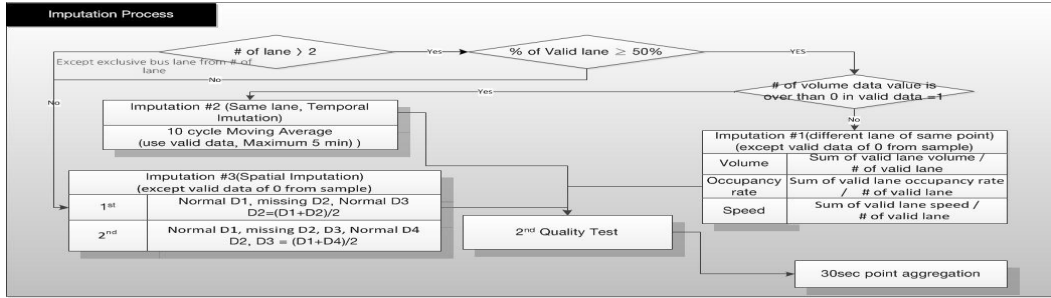
### 3) 차로보정과정

차로수별 획일화된 보정방식 개선을 위해 참조 차로가 1개인 경우를 구분하고자 편도 3차로 이상과 미만으로 구분하였다. 또한, 편도 3차로 이상 도로 중 정상차로 자료가 50%미만일 경우에도 편도 2차로 도로와 동일한 보정방식을 적용하였다.

편도 2차로 도로와 정상차로 자료가 50%미만일 경우 상, 하류부의 공간적 보정을 실시하고 보정대상 검지기의 범위에 따라 1, 2단계로 구성하였다.

14) 교통량, 점유율, 속도가 [1, 0.3, 120]과 같은 경우 정수형으로 변환하면 [1, 0, 120]으로 논리적 오류에 해당

15) 차로이탈주행, 차로변경차량 등의 비정상적인 자료를 헤드간 점유율, 교통량으로 검지



〈그림 3〉 차로보정과정  
 〈Fig. 3〉 Procedure for Lane Correction

편도 3차로 이상 도로는 “동일지점에서는 통행특성이 유사하다”라는 가정으로 동일지점 정상차로의 평균값을 이용하여 보정한다. 하지만, 야간 인접차로의 차량 미통행으로 인해 보정이 어려운 경우에는 해당 차로의 이동평균을 통해 보정을 실시한다.

4) 지점보정과정

지점보정은 차로보정에서 미보정된 자료때문에 집계가 불가능한 경우 적용한다. 상하류 인접지점을 이용한 공간적 보정과 이동평균에 의한 시간적 보정, 이력자료에 의한 패턴보정으로 구분되며 5분 이내 오류에 대해서만 실시한다.

5) 기타사항

개선 전처리 과정에는 오류자료 해당항목, 차량 정지상태, 보정처리여부등에 대한 정보를 태그화하여 제공하도록 구성하였다. 또한, 이러한 태그들을 이용하여 자료 완전성과 유효성 정보를 제공하여 자료 품질을 확인할 수 있도록 하였다.

V. 개선 전처리 과정의 평가

1. 평가개요

1) 평가환경

실제 수집된 차량검지기 자료를 개선 전처리 과정과 기존 FTMS 전처리 과정에 적용해서 집계하여 품질지표와 교통량을 비교 평가하였다. 품질지표는

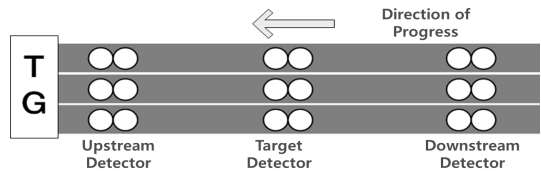
30초 차로원시자료와 비교했고 교통량은 톨게이트 교통량과 전처리 후 집계된 자료와 비교하였다.<sup>16)</sup>

오차를 제곱평균하는 지표인 평균오차 제곱근(RMSE:Root Mean Square Error, 이하 RMSE)과 평균 절대비율오차(MAPE:Mean Absolute Percentage Error, 이하 MAPE)을 통해 비교하였다.

2) 평가세부사항

기준값으로 사용한 톨게이트는 고속도로 본선에 위치하고 대상 검지기와 톨게이트 간(2km) 진출입 램프가 존재하지 않는다. 진출입램프가 없어서 교통량의 유출입이 없으며 1시간 단위 비교시 공간상 이격으로 인한 오차는 미비하다고 가정하였다.<sup>17)</sup>

〈그림 4〉와 같이 톨게이트 최인접 3개 검지기 중 중앙검지기를 대상검지기로 하였다. 이는, 공간보정을 위해 상, 하류 검지기가 필요하기 때문이다.



〈그림 4〉 평가지점 검지기 구성  
 〈Fig. 4〉 Evaluation Site Detector Layout

16) VDS자료의 경우 참값을 알 수가 없으므로 기준값을 톨게이트 교통량으로 설정

17) 5분 단위와 같이 짧은 집계간격의 비교시 공간상 이격으로 인해 오차가 발생하므로 1시간단위로 비교



상기의 조건에 따른 평가대상지점 선정결과는 <표 4>와 같다.

<표 4> 평가대상지점 선정결과  
<Table 4> Result for Evaluation Location

Line	Tollgate	Direction	Detector Location	# of lanes
Seohaean	Mokpo	Down	9.6km	2
Kyungbu	Pusan	Up	2.48km	3

## 2. 평가결과

분석자료는 2012년 2월 6일~10일, 5일간 자료이며 30초 차로원시자료를 전처리한 후 1시간 지점자료로 집계하였다.

### 1) 품질지표 분석결과

Off-Line 자료 처리방식인 개선과정에 비해 On-Line<sup>18)</sup> 처리방식인 기존과정이 결측건수가 많이 나타났다. 이는 실시간 처리를 위한 자료 수집시 통신 지연 및 오류 등으로 인한 손실이 주 원인으로 나타났다. 오류건수도 개선 전처리 과정이 많으며 이는 개선 과정이 오류판단기준과정을 강화시켰기 때문이다. 보정건수는 기존 과정은 보정내역 확인이 불가능한데 비해서 개선 과정은 전체 오류와 결측건수 중 약 95.2%에 해당하는 자료를 보정하였다.

<표 5> 30초 차로 오류건수 및 보정현황  
<Table 5> Result of 30s Data Evaluation

		2/6	2/7	2/8	2/9	2/10	Total
Total		14,400	14,400	14,400	14,400	14,400	72,000
Exist	Missing	253	224	267	289	244	1,277
	Error	1,595	1,527	1,583	1,227	1,272	7,204
	Correction	-					
Improved	Missing	52	44	90	113	42	341
	Error	1,641	1,577	1,615	1,242	1,319	7,394
	Correction	1,636	1,562	1,583	1,257	1,326	7,364

18) 전처리 과정에서는 실시간 처리를 on-line처리, 이력자료를 위한 비실시간 전처리를 off-line처리임

품질지표는 완전성<sup>19)</sup>과 유효성<sup>20)</sup>을 효과적으로 분석하였고 척도별 산출식은 식(1), (2)와 같다[14].

$$\text{완전성 (\%)} = \frac{X-m}{X} \times 100 \quad (1)$$

$$\text{유효성 (\%)} = \frac{((X-m)-e)}{X-m} \times 100 \quad (2)$$

여기서,  $m$  = 전체 자료 중 결측자료수  
 $e$  = 전체 자료 중 오류자료수  
 $X$  = 전체 자료 개수

품질지표 분석결과 결측자료수 관련 지표인 완전성은 개선된 과정이 높으며 결측자료와 오류자료수 관계인 유효성은 기존 전처리 과정이 약간 높은 것으로 나타났다. 이는 개선 전처리 과정의 강화된 오류판단기준 때문이라 판단된다.

<표 6> 30초 차로 품질검사 결과  
<Table 6> Result of 30s Data Quality Evaluation

		2/6	2/7	2/8	2/9	2/10	Avg.
Exist	completeness(%)	98.2	98.4	98.1	98.0	98.3	98.2
	validity(%)	88.7	89.2	88.8	91.3	91.0	89.8
Improved	completeness(%)	99.6	99.7	99.4	99.2	99.7	98.8
	validity(%)	88.6	89.0	88.7	91.3	90.8	89.7

차량 미통과시 속도를 강제보정<sup>21)</sup>한 횟수 분석 결과 전체 14,400건 중 1일 평균 약 3,400건이 강제보정되고 있으며 이로 인해 강제보정에 의한 속도 자료의 과다추정이 발생하는 것으로 판단된다.

19) 이론적 수집갯수 중 이용가능한 자료의 비율임

20) 수집된 이용가능한 자료 중 오류판단기준에 의거, 오류가 발생하지 않는 자료의 비율임

21) 기존 전처리 과정에만 해당하며 실제 차량이 없어서 미통과일 경우 속도를 100km/h, 110km/h(제한속도), 혼잡으로 미통과할 경우 속도를 60km/h로 보정

〈표 7〉 속도강제보정 분석결과  
 〈Table 7〉 Result of Speed Compulsion Correction Analysis

	2/6	2/7	2/8	2/9	2/10	Avg.
# of compulsion correction	3,196	3,393	3,873	3,522	2,921	3,381

2) 교통량 분석결과

전처리 과정을 거친 30초 차로자료를 1시간 지점 자료로 집계하여 톨게이트 1시간 통과교통량과 비교한 결과는 <표 8, 9> 및 <그림 5, 6>과 같다.

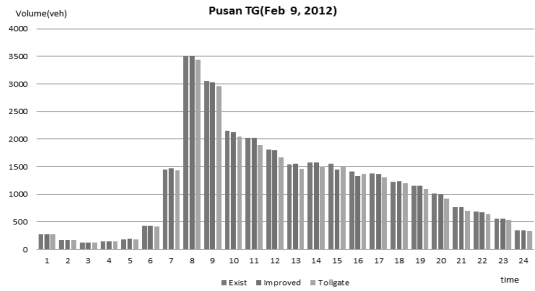
〈표 8〉 교통량 분석결과(부산톨게이트)  
 〈Table 8〉 Result of Traffic Volume Evaluation (Pusan TG)

Category	2/6	2/7	2/8	2/9	2/10	Avg.	
① (veh)	TCS	28,725	27,256	26,209	27,283	29,215	27,738
	Exist	29,839	28,324	27,125	28,476	30,443	28,841
	Impr.	29,816	28,255	26,876	28,251	30,344	28,708
② (veh)	TCS	1,197	1,136	1,092	1,137	1,217	1,156
	Exist	1,243	1,180	1,130	1,187	1,268	1,202
	Impr.	1,242	1,177	1,120	1,177	1,264	1,196
③ (veh)	Existst	59.7	56.6	50.7	63.5	62.1	58.5
	Impr.	57.3	53.2	46.4	58.6	58.1	54.7
④ (%)	Existst	4.25	3.98	3.66	4.20	4.31	4.08
	Impr.	3.96	3.72	3.35	4.24	4.17	3.89

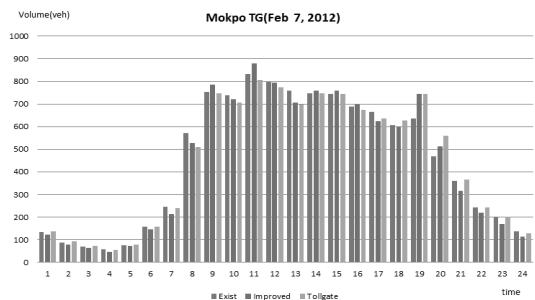
〈표 9〉 교통량 분석결과(목포톨게이트)  
 〈Table 9〉 Result of Traffic Volume Evaluation (Mokpo TG)

Category	2/6	2/7	2/8	2/9	2/10	Avg.	
① (veh)	TCS	11,993	10,734	8,947	9,392	12,356	10,684
	Exist	11,780	10,772	8,719	9,555	12,291	10,623
	Impr.	12,045	10,674	8,667	9,443	28,324	10,679
② (veh)	TCS	500	447	373	391	515	445
	Exist	491	449	363	398	512	443
	Impr.	502	445	361	393	524	445
③ (veh)	Exist	26.1	36.3	32.9	18.9	19.5	26.7
	Impr.	22.8	26.5	24.1	14.2	21.2	21.7
④ (%)	Exist	3.61	4.46	8.85	4.57	3.98	5.09
	Impr.	6.07	7.01	9.25	4.18	5.39	6.38

주 1: ①:1일 총 교통량, ②:1시간 평균교통량,  
 ③:RMSE, ④:MAPE  
 주 2: <표 8, 9> 동일



〈그림 5〉 부산톨게이트 교통량 비교결과(일부예시)  
 〈Fig. 5〉 Result of Traffic Volume Comparison for Pusan TG



〈그림 6〉 목포톨게이트 교통량 비교결과(일부예시)  
 〈Fig. 6〉 Result of Traffic Volume Comparison for Mokpo TG

1일 총교통량, 1시간 평균 교통량에 대한 분석결과 개선 전처리 과정이 기준값인 톨게이트 교통량과 좀 더 일치하는 것으로 분석되었다. 또한, 경부, 서해안 고속도로 모두 개선 전처리 과정이 기준모형보다 낮은 RMSE를 나타내고 있다. 이는 오차가 발생하더라도 개선 전처리 과정이 오차를 상대적으로 적게 발생시키는 것으로 해석할 수 있다.

MAPE 분석결과 경부고속도로에는 개선 전처리 과정이 우수하고 서해안고속도로는 일부 분석일자에 기준모형이 더 낮은 값을 나타내고 있다.

1일 총교통량, 1시간 평균교통량, RMSE, MAPE 등의 분석결과 전반적으로 개선 전처리 과정이 기존 전처리 과정에 비하여 참값에 가까운 값을 나타내는 것으로 분석되었다. 이는 개선 전처리 과정이 이력자료 활용을 위한 실제 교통현상을 보다 잘 반영하는 것으로 추론할 수 있다.

## VI. 결론 및 향후 연구과제

차량검지기 자료는 ITS 시스템의 근간이 되는 핵심자료로서 혼잡관리, 돌발상황관리, 교통상태 파악 등의 주요한 자료로 활용되고 있다. 국내외에서는 시스템의 목적에 부합하는 전처리 과정을 적용중이나 일부 시스템에서 과다, 과소 보정의 우려가 있는 보정기법을 사용하고 있다. 이런 경우 과다 과소 추정 및 자료 왜곡 등으로 실제에 근사한 교통자료를 생성하기는 어렵다.

이에 본 연구에서는 고속도로 차량검지기의 수집자료를 이력자료로 활용하기 위해 “실제 교통현상을 왜곡하지 않고 데이터 보정처리 내역을 확인할 수 있도록 하는 전처리 과정 수립”으로 목표로 설정하고 전처리 과정을 개선하였다.

기존 전처리과정과 개선 전처리 과정을 비교한 결과 개선 전처리 과정의 품질지표가 우수한 것으로 나타났다. 이는, 동일 자료를 사용할 경우 비실시간 Off-Line 방식을 적용하는 개선 전처리 과정이 보다 완전한 정보를 활용할 수 있음을 의미한다. 또한, 톨게이트 교통량을 기준자료로 분석한 결과 개선 전처리 과정으로 처리된 결과가 기준값에 가까운 것으로 확인되었다. 이는 실제현상을 그대로 반영하여야 하는 이력자료의 활용목적상 개선 전처리 과정이 더 부합하며 전처리과정에서 정확성을 확보하는 것으로 해석할 수 있다.

개선 전처리 과정을 시스템에 적용하기 위해서는 실제 다양한 교통현상이 반영된 충분한 기간 동안의 모니터링과 검증이 요구된다. 검지기 자료의 특성상 30초 단위로 수집되는 자료의 참값을 매주기마다 실측하기는 어려우므로 개선 전처리과정의 효과를 분석할 수 있도록 정기교통량 조사의 영상자료 분석결과와 연계하여 검증을 수행할 수 있는 방안도 고려해볼 필요가 있다.

향후, 해외에서 적용되고 있는 통계기반의 보정 기법인 시계열분석, Expectation Maximization 등과 같은 고급기법의 적용에 대해서도 시스템 처리시간, 가용시스템 리소스 고려 등 실제 시스템 적용시키기 위한 관점에서 실용성 있는 연구가 뒤따른다

면 전처리 관련 기법의 고도화가 가능할 것이다.

더불어, 본 연구에서 제시한 상황별 검지결과태그를 활용한 다양한 품질 지표 개발과 유지관리에 적용이 뒤따를 경우 자료 품질관리측면에서 공간 문제가 되었던 차량 검지기 수집자료의 신뢰성 향상을 위한 기반환경 조성이 가능할 것이다.

## 참고문헌

- [1] Han D. H., Lee Y. T., “Improvement of ITS Data management System”, *KST Transportation Technology and Policy*, vol. 3, no. 3, pp.179~189, Sep. 2006.
- [2] Kim H. S. et al., “Analysis of the Research Trend and Developmental Direction against the VDS Data”, *The Journal of The Korea Institute of Intelligent Transport Systems*, vol. 6, no. 1, pp. 13~26, 2007
- [3] Kim H. S. et al., “Investigation of Service Item for Archived VDS Data User Services: Focused on Expressway”, *The Journal of The Korea Institute of Intelligent Transport Systems*, vol. 6, no. 2, pp.1~14, 2007
- [4] Korea Expressway Corporation Research Institute, *Improvement of Traffic Data Collection, Analysis and Utilization from Vehicle Detection System on Freeway*, pp.30~38, 2006
- [5] Chao Chen et al., “Detecting Errors And Imputing Missing Data For Single Loop Surveillance Systems”, *Journal of the Transportation Research Board*, Vol. 1855, pp.160~167, 2003.
- [6] Brian L. Smith et al., “Exploring Imputation Techniques for Missing Data in Transportation Management Systems”, *Journal of Transportation Research Board*, Vol.1 836, pp.132~142, 2003
- [7] Han D. C et al., “A Study of imputation methods for missing data on ITS”, *The Korea Institute of Intelligent Transport Systems Autumn Conference*, pp.270~275, 2004

- [8] Kim J. Y. et al., "A Study on the Imputation for Missing Data in Dual-loop Vehicle Detector System", *Journal of Korean Society of Transportation*, Vol. 24, No. 7, pp.27~40, 2006
- [9] Chung S. J., "Implementation of Quality Evaluation, Error Filtering, Imputation for Traffic Missing Data", *Korean Institute of Information Scientists and Engineers Autumn Conference*, Vol. 34, No. 2, pp.185~190, 2007
- [10] Texas Transportation Institute(TTI), Battelle, *Quality Control Procedure for Archived Operations Traffic Data : Synthesis of Practice and Recommendation Final Report*, TTI, 2007
- [11] Rafael J. F. et al., "Toward Improved And Transparent Imputation Techniques For Online Traffic Data Streams And Archiving Applications", *88th Annual Meeting of the Transportation Research Board*, 2009
- [12] Gang Chang, Tongmin Ge, "Comparison of Missing Data Imputation Methods for Traffic flow", *International Conference on Transportation, Mechanical, and Electrical Engineering (TMEE)*, 2011
- [13] Weihao Yin et al., "Imputing Erroneous Data of Single-Station Loop Detectors for Nonincident Conditions : Comparison Between Temporal and Spatial Method", *Journal of Intelligent Transportation System*, vol. 16, no. 3, pp.159~176, 2012
- [14] Battelle, *Traffic Data Quality Measurement Final Report*, 2004

### 저자소개



**이 환 필 (Lee, Hwan-Pil)**

2003년 2월 : 아주대학교 교통공학 석사  
 2005년 3월 ~ 2012년 2월 : 아주대학교 교통연구센터(ATRI) 연구원  
 2011년 2월 : 아주대학교 교통공학 박사  
 2012년 2월 ~ 현재 : 한국도로공사 교통처 교통정보통합활용지원팀  
 e-mail : pogijaylee@gmail.com



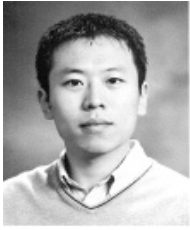
**남 궁 성 (Namkoong, Seong)**

1996년 : 한양대학교 도시공학과 박사(교통전공)  
 1996년 8월 ~ 2002년 8월 : 한국도로공사 도로연구소 교통연구 그룹장  
 2002년 9월 ~ 2004년 10월 : University of Virginia, Smart Travel Lab. Research Scientist  
 2007년 1월 ~ 현재 : 한국도로공사 교통정보통합활용지원센터(OASIS) 운영  
 2008년 1월 ~ 현재 : 한국도로공사 명절 교통예보관  
 2004년 11월 ~ 현재 : 한국도로공사 도로교통연구원 교통연구실장



**김 수 희 (Kim, Soo-Hee)**

1999년 3월 : 아주대학교 교통공학 석사  
 2002년 3월 ~ 2011년 5월 : 아주대학교 교통연구센터(ATRI) 연구원  
 2007년 8월 : 아주대학교 교통공학 박사  
 2011년 6월 ~ 현재 : 한국도로공사 교통처 교통정보통합활용지원팀



**김 진 (Kim, Jin)**

2002년 2월 : 건국대학교 국제무역학 학사

2005년 ~ 2007년 : 삼성전자 R3 시스템 구축

2007년 ~ 2009년 : 한국도로공사 원격유지관리시스템 구축

2009년 ~ 현재 : DB정보통신 경기사업단 팀장