

■ 論 文 ■

# 국도의 동질구간 선정과 이상치 제거 방법에 관한 연구

Development of Homogeneous Road Section Determination and Outlier Filter Algorithm

**도 명 식**

(한밭대학교 도시공학과 조교수)

**배 현 숙**

(한밭대학교 도시공학과 석사과정)

**김 성 현**

(한국건설기술연구원 선임연구원)

**김 종 식**

(한국건설기술연구원 연구원)

## 목 차

- I. 서론
- II. 동질구간 선정과 이상치 제거 방법
  - 1. 동질구간의 정의
  - 2. 번호판 대조법에 의한 구간소요시간 수집
  - 3. 기존 이상치 제거 방법의 평가
- III. 교통량과 소요시간에 의한 동질구간 선정
  - 1. 국도 3호선의 교통특성
  - 2. 동질구간 결정 프로세서
- IV. 결론
- 참고문헌

Key Words : 동질구간, 번호판 대조법, 이상치 제거, 통행시간, 국도3호선

## 요 약

국도의 단속교통류의 특성을 고려한 교통량 동질구간이란 도로의 공급측면에서 주요 신호교차로의 분기, g/C 비율, 신호교차로간 거리 등으로 분할할 수 있으며, 교통수요측면에서는 교통량 변화 패턴이 서로 유사하면서 정량적인 교통특성(예를 들어, 교통량의 크기나 속도의 범위 등)이 거의 같은 인근 검지기들로 이루어진 구간을 말한다.

본 연구에서는 국도 3호선의 곧지암 IC에서 장지 IC간의 10.72km를 대상으로 수집한 구간통행시간 자료와 검지기에서 수집한 지점 교통량과 속도 자료를 이용하여 기존의 이상치 제거방법에서의 문제점을 제시하고 참조자료를 활용한 새로운 유효 데이터 확보방안을 제시한다. 나아가, 향후 통행시간 추정모형 제작 및 검지기 자료의 이상치 및 결측 자료의 보정을 위한 동질구간의 선정방법을 KHCM 방식에 의한 공급자 측면과 교통류의 특성을 고려한 수요자 측면을 동시에 고려하여 제시하였다. 이러한 구간자료와 지점자료를 공유한 국도 구간의 교통류 특성에 관한 연구는 향후 소요시간 추정과 결측 및 데이터 보정에 필요한 기초 자료로 활용될 것으로 기대된다.

## I. 서론

늘어나는 차량 대수에 비해 턱없이 부족한 도로의 건설로 인해 교통혼잡과 소음, 공해 등 환경적인 문제가 날로 심각해지고 있다. 특히, 수도권과 연결되는 국도의 경우 이동성 확보와 지역간 통행특성이라는 본래의 기능을 상실해가고 있는 실정이다. 이천에서 성남시로 연결되는 국도 3호선의 경우 주중에는 서울로 출·퇴근하는 통행이 대부분이며, 주말에는 관광 및 레저 등의 통행이 특히 많은 구간이다.

특히, 성남시와 이천간 국도 50km를 시범구간으로 선정하여 1997년부터 건설교통부와 한국건설기술연구원에서 교통상황의 감지와 운전자에게 유효한 교통정보의 제공을 목적으로 국도교통관리시스템(NHTMS, National Highway Traffic Management System)을 구축 운영 중에 있다(서울지방국토관리청, 2001). 그러나, 구축 초기의 국도 인근의 교통특성과는 달리, 날로 도로변을 따라 도심화되어가고 있어 설치되어 있는 검지기의 자료를 교차로(혹은 정보제공을 위한 블록구간) 사이의 교통 특성을 대표한다고 간주하기 어렵게 되었다. 즉, 고속국도와는 달리 단속류의 특성을 가지는 국도라는 점에서 신호의 영향, 대기행렬의 영향 및 구간 내부 교통량(택시 등) 등 다양한 외부 변수의 영향을 받게 된다.

동질구간에 관한 연구는 주로 시설물의 유지관리나 보수를 위한 기능성의 등급을 매기기 위한 연구가 대부분이다(한국건설기술연구원, 2000). 그러나, 소요시간 예측이나 트랜드 추정을 위한 DB관리 등에 필요한 데이터 특성을 그 대상으로 하는 연구는 전무한 형편이며(Okutani and Sephanedes, 1984; FHWA, 2001), 시설물의 기능적 분류 등을 위한 통계학적 연구도 이루어져 왔다(김주현, 2002; Spath, 1980). 도명식 등(2004)은 국도 3호선을 대상으로 교차로의 특성, 인근 토지이용특성 및 검지기 자료를 종합하여 상·하행을 구분하여 선정한 바 있다. 그러나 검지기의 설치 위치, 경사도, 대기행렬에 크게 영향을 받는 지점 속도자료를 이용할 수 없어 교통량 자료만을 기준으로 선정하였다.

한편, 이상치 제거 및 보정에 관한 연구는 주로 통계적인 분석과 시계열 자료 등 연속자료의 결측이나 유효값의 범위선정 등에 많은 연구가 축적되어 있다(Washington et. al., 2003; 이지연 외, 2003).

따라서 본 연구에서는 향후 통행시간 추정모형 제작 및 검지기 자료의 이상치 및 결측 자료의 보정을 위한 동질구간의 선정방법의 제시와 번호판 대조법에 의한 구간통행시간 자료의 이상치 제거방안을 마련하는 것을 목적으로 한다.

## II. 동질구간 선정과 이상치 제거 방법

### 1. 동질구간의 정의

교통류를 대상으로 하는 동질구간(homogeneous road section)은 유사한 교통 특성을 가지는 구간이라고 정의할 수 있다. 즉, 통행속도에 영향을 미치는 변수들이 일정한 구간을 구분하는 데 적용되는데 도로 주변 개발 현황(도시 지역과 지방 지역), 교통량, 신호교차로, 도로 조건 등이 구분기준에 포함된다(대한교통학회, 2004).

따라서 국도의 단속교통류의 특성을 고려한 교통량 동질구간이란 도로의 공급측면에서 주요 신호교차로의 분기, g/C 비율(0.6 기준), 신호교차로간 거리(3.0km) 등으로 분할할 수 있으며, 교통수요측면에서는 하루 평균적인 교통량 변화 패턴이 서로 유사하면서 정량적인 교통특성(예를 들어, 교통량의 크기나 속도의 범위 등)이 거의 같은 인근 검지기들로 이루어진 구간을 말한다.

일반적으로 동질구간의 결정 목적은 아래와 같이 들 수 있다. 먼저, 대상구간 내 최적의 차량검지 시스템 설치 지점 결정을 위한 기초 자료로 활용할 수 있으며, 둘째는 비 관측 지점의 교통량, 속도 추정 및 결측 자료 보정자료로 활용할 수 있는 장점이 있기 때문이다.

본 연구에서의 동질구간 결정은 위에서 언급한 바와 같이 누락 및 특이자료의 보정을 위한 기초자료로의 활용과 향후 실시간이면서 정확한 교통정보(통행시간 정보 포함)의 제공을 위한 자료의 확보를 위해 단속교통류의 특성을 가지는 국도 3호선을 대상으로 효율적인 동질구간 결정을 위한 방안을 제시하는 것이다.

### 2. 번호판 대조법에 의한 구간소요시간 수집

수도권 남부 국도교통관리시스템의 대상구간 가운데 현재 AVI(Automatic Vehicle Identification) 장비가 설치되어 있는 곳은 1호선, 43호선에 설치 운영중

〈표 1〉 조사 대상구간의 특성 및 조사 장비 개요

구간	1	2	3	4	5	6	7	8	9	10	11	12
거리(m)	500	460	217	423	550	502	1728	510	400	1490	1540	2400
조사장비	NC-97	D-13	NC-97	NC-97	D-12	NC-97	NC-97	D-10	NC-97	D-9	D-8	D-6
비고	관지암IC										경안IC	장지IC

이며, 향후 46호선에도 설치할 예정이다. 그러나 본 연구의 대상구간인 국도 3호선은 97년에 구축되었기 때문에 AVI 검지기가 설치되어 있지 않다.

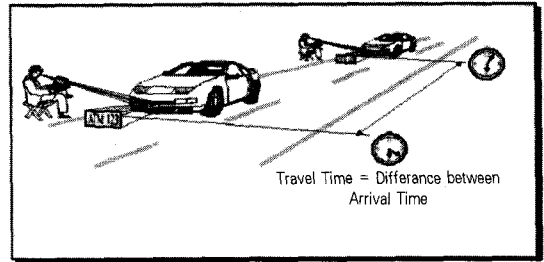
AVI가 없는 경우 구간의 소요시간을 측정하기 위해 번호판 대조법(number plate matching method)이나 Video를 이용한 촬영법이 많이 이용된다. 이 가운데 Video를 이용하여 번호판을 촬영한 후 실내에서 판독하는 방법은 차로 수가 많은 경우 연속진행이나 차량의 중첩현상 등으로 샘플수의 확보에 큰 어려움이 있으며, 번호판 대조법의 경우, 조사비가 많이 소요되며 조사사 사고의 위험 등 장단점을 가지고 있다(장진환, 2004; 安井, 1995).

따라서 본 연구에서는 국도 3호선 관지암 IC에서 장지 IC사이의 연장 10.72km를 대상으로 기 구축된 검지기의 자료와 연계하여 교통량과 소요시간 그리고 소요시간 자료를 이용한 구간속도를 획득하기 위해 번호판 대조법을 사용하여 데이터를 수집하였다.

조사 대상구간의 특성 및 조사 내용(일시, 시간대)에 대한 내용을 정리하면 〈표 1〉 및 〈표 2〉와 같다.

평일인 화·수·목요일을 대상으로 한 이유는 기존의 연구결과 국도 3호선에서 수도권으로 유출·입되는 교통류의 월요일 특성과 주말 및 공휴일 특성이 평일과 유의한 차이가 존재함이 확인되었기 때문이다(이지연 외, 2003; 도명식 외, 2004).

차량번호판 조사방법은 〈그림 1〉에서와 같이 대상구간의 각 교차로를 중심으로 2~3인의 조사원(차로당 최소 1인)을 배치하여 차량번호를 녹음기를 통해 차로



〈그림 1〉 차량 번호판 대조법 개념도

별로 초 단위의 데이터를 녹음한다. 수집한 자료를 실내에서 코딩하고 코딩한 결과값을 엑셀을 이용하여 매칭하여 통행시간을 산출하였다.

한편, 조사 대상인 12구간(10.72km) 가운데 영상 검지기가 설치되어 있지 않는 구간은 NC-97 교통조사 장비를 도로노면에 설치하여 교통자료를 수집하였다. NC-97로부터 수집한 자료는 실내에서 번호판 대조법과 마찬가지로 5분 단위로 재구성하여 각 구간별 데이터 셋을 만들었다.

### 3. 기존 이상치 제거 방법의 평가

신뢰할 수 있는 교통정보를 제공하기 위해서는 검지기로부터 관측되는 데이터의 정도가 매우 중요하다. 이는 모든 교통정보가 관측 데이터로부터 가공되거나 알고리즘을 이용하여 처리된 후 제공되기 때문이다. 최근 기존의 검지기로부터의 데이터 수집 및 Probe data 등 다양한 방법으로 자료 수집에 대한 연구가 시도되고 있다(김영찬 외, 2001; 장진기 외, 2002).

한편, 이러한 자료의 수집방법뿐만 아니라 이상치(outlier)의 제거 및 보정모직에 대한 연구도 병행하여 이루어지고 있다(장진환, 2004). 특히, 최윤혁(2003)은 택시에 GPS 수신기를 장착하여 수집한 프로브 자료의 이상치 제거방법의 가능성을 제시하였고, 이지연 외(2003)는 국도 3호선을 대상으로 검지기로부터 수집한 교통량 자료의 보정방안을 제시하였다. 그러나,

〈표 2〉 조사 일시 및 시간대

횟수	조사 일	조사 시간대
1	7월 20일(화)	17:30~19:00(오후 침두)
2	7월 21일(수)	07:30~09:00(오전 침두)
3	7월 21일(수)	12:00~14:00(오후 비침두)
4	7월 21일(수)	17:00~19:00(오후 침두)
5	7월 22일(목)	07:30~09:00(오전 침두)
6	7월 22일(목)	12:00~14:00(오후 비침두)
7	7월 22일(목)	17:00~19:00(오후 침두)

기존의 이상치 제거방법으로는 거의 대부분의 연구에서 통계적인 방법을 이용하고 있다. 그러나 Box-Plot법, Shapiro-Wilk 검정법 등 통계적인 방법은 편향되거나 중심에서 많이 벗어난 자료를 이상치로 취급하여 버리는 과정에서 분포의 정규성을 임의로 가정하지만, 구간 소요시간 등의 교통 데이터는 정규분포 대신 우편향된 분포를 가지는 등, 기존의 통계적인 방법을 그대로 적용하기는 많은 문제가 있다(Washington et. al., 2003).

1) 통계적인 방법을 이용한 이상치 제거(방법 I)

대표적인 통계적 방법을 이용하여 이상치를 제거하는 순서는 다음과 같다. 먼저, 개별차량의 구간통행시간을 측정(산정)한 후, 상한값과 하한값을 초과한 값을 제거하는 데, 이때 강진기 외(2002)가 사용한 상한 값은 「구간의 설계속도의 2배를 초과하는 구간통행시간」이며, 하한 값은 「해당구간을 10km/h 통행할 때 통행시간이며 단, 이러한 하한값을 보이는 구간 통행시간값들의 개수가 전체 구간 통행시간 값들의 50%이상을 초과할 경우는 포함」하는 방법으로 설정하였다. 그리고 신뢰구간 68%를 초과한 값을 이상치로 간주하여 제거하였다. 본 연구에서는 위의 통계적인 방법을 방법 I이라 부르기로 한다.

즉, 방법 I은 수집된 개별 차량들 ( $n$ 대)의 구간통행시간 ( $T_i$ )으로부터, 표준편차 ( $\sigma$ )의 1배 범위를 만족하지 못하는 개별 차량의 구간여행시간값들을 이상치로 간주하여 제거한 후 나머지 유효 차량들 ( $n_s$ 대)의 구간 여행시간값 ( $T_{is}$ )을 산술평균하여 구간 대표여행 시간 ( $T_s$ )으로 적용하는 것으로 다음과 같이 나타낼 수 있다.

$$T_s = \frac{\sum T_{is}}{n_s} \tag{1}$$

where,  $T_{is} = |T_i - \bar{T}| \leq \sigma$

$$\bar{T} = \frac{\sum T_i}{n}$$

2) 시계열 기법을 이용한 이상치 제거(방법 II)

바로 이전 수집주기 ( $k-1$ )에 측정한 지점 A와 B 간의 평균통행시간을 이용하여 현재 수집주기 ( $k$ )의 유효 통행시간  $(tt_{AB})_k$ 을 산정하는 방법은 다음과 같다.

$$(tt_{AB})_k = \frac{\sum_{i=1}^{Stt_{ABk}} (t_{Bi} - t_{Ai})}{Stt_{ABk}} \tag{2}$$

where,

$$Stt_{ABk} = \{t_{Bi} - t_{Ai} \mid t_k - t_{k-1} < t_{Bi} \leq t_k \text{ and}$$

$$(tt_{ABmin})_k \leq t_{Bi} - t_{Ai} \leq (tt_{ABmax})_k\},$$

$$(tt_{ABmin})_k = tts_{AB_{k-1}}(1 - l_{th}),$$

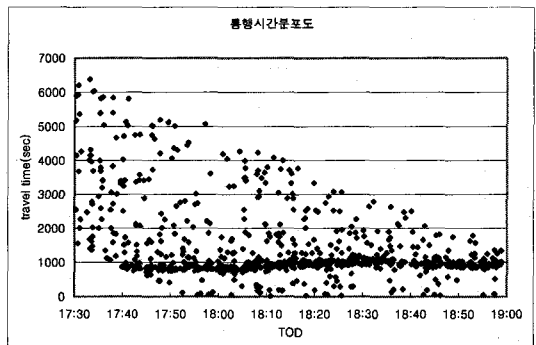
$$(tt_{ABmax})_k = tts_{AB_{k-1}}(1 + l_{th}),$$

$$(tts_{AB})_k = \alpha \cdot (tt_{AB})_{k-1} + (1 - \alpha)(tts_{AB})_{k-1}$$

여기서,  $Stt_{ABk}$ 는  $k$ 번째 수집주기에서 A지점과 B지점에서 관측된 유효 통행시간 수,  $t_{Ai}$ 와  $t_{Bi}$ 는 각각 A지점과 B지점에서 관측된  $i$ 차량의 관측시간,  $(tts_{AB})_k$ 는  $k$ 번째 수집주기에서 A지점과 B지점의 기대(평활) 평균통행시간을 나타내며,  $(tt_{AB})_k$ 는  $k$ 번째 수집주기에서 A지점과 B지점에서 관측된 차량의 평균통행시간을 나타낸다.  $t_k$ 는  $k$ 번째 수집주기의 종료시간,  $(tt_{ABmin})_k$ 와  $(tt_{ABmax})_k$ 은  $k$ 번째 수집주기에서 A지점과 B지점 최소 통행시간과 최대통행시간을 나타낸다. 한편,  $\alpha = (1 - (1 - \beta)^{Stt_{AB}})$ 는 지수평활계수이며,  $\beta$ 는 민감도 계수로써 0.2를 적용하였으며,  $l_{th}$ 는 링크 초기 통행시간 파라미터로 0.3을 적용하였다(Dion, 2003; 장진환, 2004).

여기서, 기존 이상치 제거방법을 비교 평가하기 위해 2004년 7월 20일 오후첨두(17:30~19:00) 시간대에 번호판 대조법으로 수집한 전체구간(10.72km)의 통행시간 분포를 이용하기로 한다. 먼저, 시간대별 수집된 소요시간 데이터의 분포도는 <그림 2>와 같다.

한편, 그림 2의 오른쪽 상부부분의 데이터가 없는



<그림 2> 매칭된 차량의 통행시간 분포도

것은 번호판 대조법의 특징으로 조사가 거의 동시에 종료되기 때문에 업무나 휴식 등으로 인해 통과시간이 낮은 차량의 자료는 포함되지 않기 때문이다.

〈그림 2〉에서 알 수 있는 바와 같이, 국도의 단속류 교통흐름의 특성상, 통행도중 휴식 및 업무시간을 포함한 통행시간의 연장 등 연속류 교통류와는 크게 다른 분포형태를 보이고 있다. 따라서 기존의 이상치 제거방법에서의 정규분포 가정 등은 적용하기가 어렵다는 것을 알 수 있다. 또한 번호를 녹음기로 입력하였다가 재생하면서 입력하는 과정에서 오류, 같은 번호판을 가진 차량의 존재 등 코딩과정에서의 오류도 이상치로 간주하여 제외하여야 할 것이다.

〈그림 3〉은 기존의 통계적인 이상치 제거방법(I)을 통해 제거되고 남은 유효데이터를 나타낸 것이다. 먼저, 전체구간을 상한값을 241.22sec, 하한값을 3856.12sec 기준으로 하여 상·하한 초과값을 제거한 후, 신뢰구간 68% 범위를 벗어나는 초과값을 제거하고 유효값을 획득하였다.

한편, 〈그림 4〉에는 시계열 기법을 이용하여 이상치를 제거한 유효 통행시간 데이터의 분포도를 나타내고

있다. 그러나, 앞 절에서도 언급한 바와 같이, 방법(II)의 경우 이전 주기에서 수집된 유효 통행시간의 데이터 수와 값이 다음 주기의 데이터에 영향을 미치게 된다. 즉, 최초 수집한 데이터가 편향된 크기의 값을 가지는 경우에는 참값이 제거되고 이상치가 남게 될 가능성이 많다. 이를 확인하기 위해 최초 5분의 샘플 데이터가 참값의 분포보다 큰 폭의 값을 취하는 경우(II-1, ◆의 형태로 표시)와 참값의 분포와 비슷한 폭과 값의 샘플 데이터를 취하는 경우(II-2, ▲의 형태로 표시) 그리고 참값 보다 큰 값의 샘플을 취하는 경우(II-3, —의 형태로 표시)의 3가지로 구분하여 이상치를 제거하였다.

본 연구에서는 이상치 제거의 기준이 되는 참값으로써는 정적인 국도통행시간 추정모델(KICT 모델)에서 얻은 추정통행시간 자료를 사용하였다.

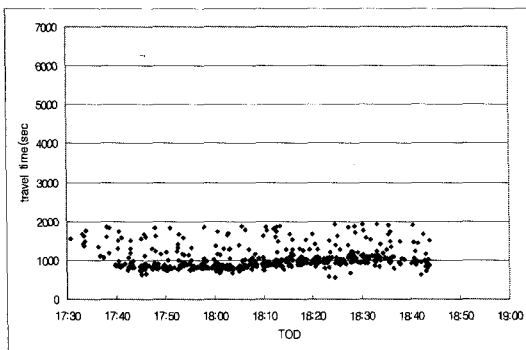
시계열 기법을 이용한 방법II의 3가지의 경우의 이상치 제거결과를 정리하면 방법(II-1) 즉, 샘플데이터의 폭이 큰 경우 초기에 매우 큰 폭으로 진동하여 이상치가 잘 제거되지 않음을 알 수 있다. 또한 이러한 초기의 영향으로 유효통행시간이 과대 추정됨을 볼 수 있다. 한편, 방법(II-2)의 경우에는 이상치를 제거하고 남은 유효데이터의 분포가 참값과 매우 유사하게 나타남을 확인하였다. 그러나 방법(II-3)의 경우와 같이 초기 샘플데이터가 상대적으로 큰 값으로 이루어진 경우 유효통행시간이 큰 값이 남게 되는 결과를 보였다.

따라서, 통계적인 방법으로 이상치를 제거하는 방법은 샘플의 정규성 가정의 문제점과 상대적으로 큰 폭의 유효데이터가 남게되어 전체적으로 평균값이 크게 나타나는 단점이 있으며, 이에 반해 시계열 기법을 이용한 이상치 제거는 초기에 이상치 제거의 기준이 되는 샘플값의 크기나 폭에 따라 매우 상이한 결과 값을 가질 가능성이 있음을 확인하였다.

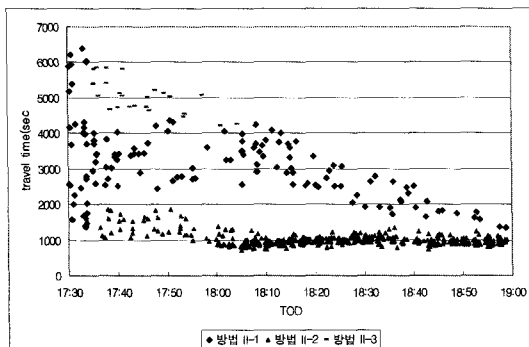
또한, 앞에서 언급한 바와 같이 번호판 대조법으로 구한 데이터의 특성상 조사 종료시간에 가까워질수록 데이터의 분산이 줄어들지만(참값에 가까워 지기 때문), 만약 조사 시작때와 같이 데이터가 분산되어 있는 경우라면 기존의 방법들은 훨씬 정도가 나쁜 결과를 얻게 될 것이다.

#### 4. 이상치 제거 방법의 개발 및 평가

현재 5분 주기의 자료를 기준으로 이상치를 제거하는 방법은 실시간 정보제공을 위해서 참조자료(reference



〈그림 3〉 통계적인 방법으로 이상치 제거(방법I)



〈그림 4〉 시계열 기법을 이용한 이상치 제거(방법II)

data)의 개수가 적거나 편향된 경우 및 주기를 짧게 하는 경우에는 초기에 획득한 참조자료에 지나치게 의존하여 이상치 제거방법에 큰 문제가 있음을 확인하였다.

따라서, 본 연구에서는 시계열 기법을 이용한 방법(II)가 최초 주기에 영향을 많이 받는 단점을 극복하기 위해 국도를 대상으로 한 통행시간 추정모델(이하, KICT 모델이라 함)에서 구한 데이터를 기준으로 이상치를 제거하는 방법을 제안하고자 한다. 또한 이러한 참조자료를 이용하는 방법은 향후 국도를 대상으로 하는 ITS사업이 활발해 짐에 따라 교통량, 속도 및 점유율 등 실시간으로 데이터를 얻을 수 있는 환경이 되면 실시간으로 이상치를 제거 및 보정할 수 있을 것으로 기대된다.

먼저, 이상치 제거를 위한 초기 단계로 식(3)은 국도구간의 통행시간 추정을 위한 KICT 모델의 기본 식이다.

$$TT_i^i = 3600 \times \frac{L_i}{v_i} + D_i^i \quad (3)$$

여기서,  $TT_i^i$  는  $i$  시간대  $l$  링크(조사구간)의 통행시간(초),  $L_i$  는 링크  $l$  의 길이(km),  $v_i$  는 링크  $l$  의 순행속도,  $D_i^i$  는 한 시간대별 차량 한 대당 제어 지체(초/대)를 나타낸다.

즉, 식(3)에서 알 수 있는 바와 같이 KICT 모델의 경우, 각 구간별 순행속도 기반 통행시간과 교차로에서의 지체를 고려한 통행시간 추정모형이다.

다음 단계로써 유효데이터의 범위는 최대, 최소값을 정하는 것으로, 식(4)에서와 같이 평활화된 평균통행시간을 이용하여 구하며, 식(4)를 만족하는 데이터를 대상으로 한다.

$$count_k = \arg \{ (tt_{ABmin})_k \leq t_{Bi} - t_{Ai} \leq (tt_{ABmax})_k \} \quad (4)$$

여기서, 최대값은  $(tt_{ABmin})_k = tts_{AB_{k-1}}(1 - l_{th})$ 이며, 최소값은  $(tt_{ABmax})_k = tts_{AB_{k-1}}(1 + l_{th})$ 으로  $l_{th}$ 은 0.3이며,  $(tts_{AB})_k = \alpha \cdot (tts_{AB})_{k-1} + (1 - \alpha) \cdot (tts_{AB})_{k-1}$ 이다. 한편,  $\alpha = 1 - (1 - \beta)^{count_k}$ 는 지수평활 계수이며,  $\beta$ 는 민감도계수로 제안한 방법을 기존방법과 효율을 비

교하기 위해 파라미터의 값을 기존의 값을 그대로 사용하기로 한다.

따라서 각 구간별로 A지점과 B지점을 통과하는 차량의 평균통행시간은 식(5)와 같이 구할 수 있다.

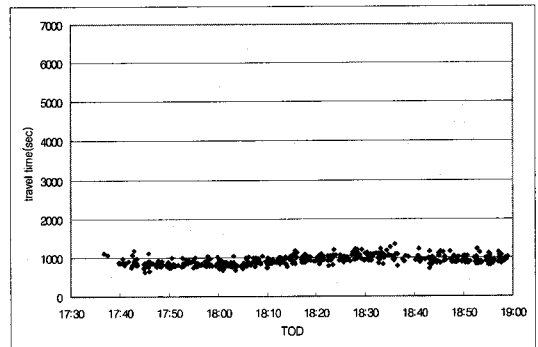
$$(tt_{AB})_k = \frac{\sum_{i=1}^{count_k} (t_{Bi} - t_{Ai})}{count_k} \quad (5)$$

<표 3>에는 각각의 이상치 제거방법을 통해 구한 유효 통행시간 데이터의 특성을 조사된 Probe-car의 평균값(901sec)과 비교한 결과이다. 통계적인 방법을 이용한 경우(방법I)과 시계열 기법가운데 참조값과 비슷한 샘플을 기준으로 이상치를 제거한 방법(II-2)의 경우, 평균은 조금 크게 나타났으나 사용에 큰 무리는 없을 것으로 판단된다. 그러나, 이전 주기의 값을 참조자료로 활용하는 시계열적 기법의 경우 방법(II-1)과 방법(II-3)의 경우와 같이 초기참조 자료가 편향되어 있는 경우에는 이상치를 제거하고 남은 유효값이 참조값과 매우 큰 차이를 보이는 것을 알 수 있다.

따라서 <그림 5>에서 알 수 있는 바와 같이, 시계열적 기법을 이용하는 경우에는 단순히 데이터의 분포 형태에서 상·하한값의 범위를 정해 이상치를 제거하는 것 보다는 본 연구에서 제안한 방법과 같이 참조 가능

<표 3> 각 이상치 제거방법의 결과 비교

구분	평균(sec)	표준편차	비고
probe-car	901	60.34	기준
방법(I)	1054	154.45	우수
방법(II-1)	2793	743.13	과다
방법(II-2)	1102	212.57	보통
방법(II-3)	3360	1401.28	과다
제안	926	70.82	우수



<그림 5> KICT 모델이용 이상치 제거(제안 모델)

한 샘플자료(예를 들어, Probe 차량, 통행시간 추정모델 등)를 최대한 활용하여 이상치를 제거하는 것이 가장 바람직하다고 판단된다.

### III. 교통량과 소요시간에 의한 동질구간 선정

#### 1. 국도 3호선의 교통특성

국도 3호선의 교통특성은 앞서서도 언급한 바와 같이, 평일의 수도권으로의 통근 및 통학을 위한 통행과 주말의 레크리에이션을 위한 통행이 주류를 이루고 있다. 또한 국도 3호선에는 수도권을 우회하는 고속국도와의 연결 I.C.가 많이 있기 때문에 각 구간별로 교차로 및 I.C.를 기준으로 회전비율이 매우 높은 편이다(서울지방국도관리청, 2001; 도명식 외, 2004).

수집된 5분 교통량 데이터들을 이용하여 인근 검지기를 사이에 상관성 여부를 알아보기로 한다. 상관관계의 정도를 재는 지표가 상관계수이며, 절대값이 1에 가까울수록 상관관계가 강하다고 판단하며, 상관계수의 값이 양수일 때, 양의 상관관계를 의미하고 음수일 때는 음의 상관관계가 있다는 것을 의미한다(SPSS 아카데미, 1999).

대상구간의 교통량과 구간속도 자료를 이용하여 상관분석을 하였다. 교통량의 경우 검지기 자료는 6월 22일~7월 22일의 화·수·목요일 자료(15일간)의 첨두별 2시간 5분평균교통량을 사용하였고, NC-97의 경우 7월 20일~7월 22일(3일)의 첨두별 2시간 5분평균교통량을 사용하였다.

구간속도의 경우 7월 20일~7월 22일(3일간) 조사된 번호판 대조법으로 구한 첨두별 1시간 자료의 5분평균 구간통행시간을 구간거리로 나누어 구하였다.

동질구간을 결정하는 기준은 구해진 상관계수를 기준으로 상관계수가 0.8 이상인 인근의 검지기를 유사검지기로 간주하는 것을 원칙으로 하였다. 따라서 상관계수가 높은 인근검지기들은 구간별로의 유사한 교통특성을 가지는 것으로 해석할 수 있으며, 데이터의 결측이나 보정에 유용하게 사용할 수 있다.

교통량자료를 사용하여 상관분석을 한 결과는 <표 4>와 같다. 여기서 오전첨두 시간대의 경우, 모든 구간에서 상관성이 높은 것으로 나타났다. 이는 오전 출근 시간대에 상행의 경우, 서울방향의 출근통행이 전 구간에 걸쳐 분포되어 있기 때문으로 해석할 수 있다. 이와

<표 4> 교통량 자료를 사용한 상관분석 결과

구간	유사 구간(상관계수)		
	오전첨두 (07:00~09:00)	비첨두 (12:00~14:00)	오후첨두 (17:00~19:00)
1구간	2구간(0.934)		
2구간	1구간(0.934) 3구간(0.913)	3구간(0.830)	
3구간	2구간(0.913) 4구간(0.978)	2구간(0.830) 4구간(0.928)	4구간(0.934)
4구간	3구간(0.978) 5구간(0.936)	3구간(0.928) 5구간(0.806)	3구간(0.934)
5구간	4구간(0.936) 6구간(0.933)	4구간(0.806) 6구간(0.833)	
6구간	5구간(0.933) 7구간(0.980)	5구간(0.833) 7구간(0.925)	7구간(0.959)
7구간	6구간(0.980) 8구간(0.937)	6구간(0.925) 8구간(0.867)	6구간(0.959)
8구간	7구간(0.937) 9구간(0.903)	7구간(0.867)	
9구간	8구간(0.903) 10구간(0.888)		
10구간	9구간(0.888) 11구간(0.966)	11구간(0.928)	11구간(0.802)
11구간	10구간(0.966) 12구간(0.839)	10구간(0.928) 12구간(0.908)	10구간(0.802)
12구간	11구간(0.839)	11구간(0.908)	

는 대조적으로 퇴근시간대의 경우는 오전첨두만큼 인접 구간과의 상관성이 높지는 않았다.

구간속도자료를 사용하여 상관분석을 한 결과 인접구간간에 상관성이 거의 없는 것으로 나타났다. 기존 연구의 결과에서 언급한 바와 같이, 지점속도 데이터의 경우에는 검지기가 설치된 위치에서의 도로의 경사도, 대기행렬에 의한 영향 및 연도의 토지이용특성에 많은 영향을 받아 신뢰성이 크게 떨어졌다(도명식 등, 2004).

그러나 상관계수를 고려한 인근 검지기(혹은 인근 구간)간의 상관성은 교통량이나 속도의 증가 및 감소의 연속성을 검증하기 위한 것이며 변수들의 절대값의 크기와는 상관없기 때문에 결측 데이터의 보정이나 누락에 대비한 자료의 축적 및 활용면에서 필요한 연구이다. 따라서 본 연구에서는 각 구간별로 구한 교통량과 구간속도를 동시에 고려할 수 있는 군집분석을 이용한 동질구간의 선정방안을 제시하고자 한다.

#### 2. 동질구간 결정 프로세스

##### 1) 공급측면(KHCM 방식)에 의한 동질구간 선정

도로용량편람(2004)에 의하면 다차로도로를 구간별

〈표 5〉 공급측면(KHCM)을 고려한 동질구간 선정

구간	시점 → 종점			접속도로	구간거리 (km)	신호주기 (초)	유효녹색 시간(초)	녹색시간비 (g/C)	신호등밀도 (개/km)	동질 구간
1	곤지암 IC	→	삼리2리		0.50	150	125	0.83	1.0	KHCM-1
2	삼리2리	→	나이키앞	지방도337교차	0.46	150	88	0.59		
3	나이키앞	→	벽산아파트		0.24	150	107	0.71		
4	벽산아파트	→	고려산업개발앞		0.40	150	125	0.83	1.1	KHCM-2
5	고려산업개발앞	→	쌍동1리		0.55	150	113	0.75		
6	쌍동1리	→	롯데아파트앞		0.40	150	122	0.81		
7	롯데아파트앞	→	초월파출소앞	지방도337교차	1.83	140~160	124~144	0.78	0.7	KHCM-3
8	초월파출소앞	→	소방파출소앞		0.51	140~160	140~159	0.88		
9	소방파출소앞	→	남촌풀장앞		0.40	140~160	124~144	0.78		
10	남촌풀장앞	→	쌍령2리		1.49	231	175	0.76		
11	쌍령2리	→	경안 IC	국도43교차	1.54	-	-	-		
12	경안 IC	→	장지 IC		2.40	-	-	-	0.0	KHCM-4

로 분할하는 과정을 간단히 요약하면 1) 도로주변 개발현황에 따라 도시 및 지방지역으로 구분하고, 2) 지방지역은 주요 신호교차로를 중심으로 재분할, 3) 직진 주 방향 g/C 비율이 0.6미만의 신호교차로, 또는 각 신호교차로간의 거리가 3.0km이상이면 다시 세분하게 된다. 4) 최종적으로 평균 g/C 비율, 신호밀도를 토대로 도로유형을 결정하는 과정이다.

국도 3호선의 대상구간을 대상으로 KHCM 방식으로 구간을 분할하면 〈표 5〉와 같이 정리할 수 있다. 즉, 주요 신호교차로(결절점)와 g/C 비율, 신호밀도를 고려하여 4구간으로 분류할 수 있음을 알 수 있다.

2) 수요측면을 고려한 동질구간 선정(군집분석)

군집분석(cluster analysis)은 N개의 개체들을 대상으로 p개의 변수를 측정하였을 때 관측한 p개의 변수들을 이용하여 N개 개체들 사이의 유사성 또는 비유사성의 정도를 측정하여 개체들을 가까운 순서대로 군집화하는 통계적 분석방법이다(한글 SPSS, 2002).

본 연구에서는 12개의 국도구간에 대해서 각 구간별 5분 단위로 집계한 교통량과 속도의 자료를 이용하여 각 구간들간의 거리 또는 비유사성을 구한 뒤 가까운 순서대로 군집화하였다.

군집분석을 위한 과정은 다음과 같다. 먼저, 1) 12개의 구간에 대해 2개의 변수(지점교통량, 구간속도)를 측정한다. 2) 모든 구간들간의 거리 또는 비유사성을 계산하여 모든 구간들 사이의 비유사성을 나타내는 거리행렬을 구한다. 이와 같은 군집간의 거리를 구하는 방법으로는 최단거리법, 최장거리법 및 평균연결법이 일반적으로 이용되고 있다. 본 연구에서는 분석방법간의

장단점을 고려하기 위해 3가지 방법으로 구한 결과를 통해 최적의 수요측면에서의 구간을 분류하기로 한다(도명식 외 2003). 3) 거리행렬에 근거하여 군집을 형성한다. 4) 각 군집의 성격이나 상호관계를 파악한다.

군집의 수를 선정함에 있어서는 기존의 연구 및 문헌(예를 들어, 도로용량편람)과의 일관성 유지 및 상호 비교를 위해 2개에서 4개의 군집으로 분류하였다.

〈표 6〉에서 알 수 있는 바와 같이 수요측면을 대상으로 군집분석한 결과, 군집의 수, 방법에 상관없이 10구간과 11구간(남촌 풀장~경안 IC)은 다른 구간들과 상이한 교통류의 특성을 가지고 있음을 확인하였다. 이 구간은 국도의 특성을 보이는 타 구간과는 달리 경안 IC의 영향으로 단속류와 연속류의 접속기능을 가지는 구간으로 회전교통량의 비율이 특히 높은 구간의 특성을 보이는 구간으로 해석할 수 있다.

한편, 1구간과 2구간은 최단연결법과 평균연결법에서 타 구간과 상이한 특성을 가지는 것으로 분류되는데, 곤지암IC와 지방도 337호선을 연결하는 구간으로 이 구간도 국도의 단속류 특성과 연속류에서의 유입교통량의 특성을 가지는 구간으로 회전교통량이 많은 구간임을 알 수 있다.

〈표 6〉에서 알 수 있는 바와 같이, 군집의 수에 따라 여러 개의 기능으로 분류할 수 있으며, 유사한 기능을 가진 도로구간이 가지는 성격을 파악할 수 있다.

따라서, 3가지 군집분석 방법과 교통류의 특성을 함께 고려한 결과 수요측면에서는 4개의 동질구간(1~2, 3~9, 10~11, 12) 구분할 수 있음을 알 수 있다.

결론적으로 공급적인(하드웨어적인) 측면과 수요적인 측면(교통류 특성)을 동시에 고려한 최종적인 동질



〈표 6〉 군집분석 결과(수요측면 기준)

구간	최장연결법			최단연결법			평균연결법(집단 간)			동질구간
	4군집	3군집	2군집	4군집	3군집	2군집	4군집	3군집	2군집	
1	1	1	1	1	1	1	1	1	1	수요-1
2	2	1	1	1	1	1	2	1	1	
3	2	1	1	2	2	1	3	2	1	수요-2
4	2	1	1	3	2	1	3	2	1	
5	2	1	1	3	2	1	3	2	1	
6	2	1	1	3	2	1	3	2	1	
7	2	1	1	3	2	1	3	2	1	
8	2	1	1	2	2	1	3	2	1	
9	2	1	1	2	2	1	3	2	1	
10	3	2	2	4	3	2	4	3	2	수요-3
11	4	3	2	4	3	2	4	3	2	
12	2	1	1	3	2	1	3	2	1	수요-4

〈표 7〉 수요·공급 측면을 고려한 동질구간 선정

구간	1	2	3	4	5	6	7	8	9	10	11	12
구간거리(m)	500	460	217	423	550	502	1728	510	400	1490	1540	2400
공급측면 동질구간(m)	동질 1 (960)		동질 2 (3420)				동질 3 (3940)				동질 4 (2400)	
수요측면 동질구간(m)	동질 1 (960)		동질 2 (4330)				동질 3 (3030)				동질 4 (2400)	
최종 동질구간(m)	동질 1 (960)		동질 2 (3420)				동질 3 (910)		동질 4 (3030)		동질 5 (2400)	

구간은 〈표 5〉와 〈표 6〉을 종합하여 다음과 같이 5개의 구간으로 분류할 수 있을 것이다.

제1동질구간은 1~2구간, 0.96km로 곤지암IC에서 나이키앞까지이며, 제2동질구간은 3~7구간, 3.42km로 나이키앞에서 초월파출소앞까지, 제3동질구간은 8~9구간, 0.91km로 초월파출소앞에서 남촌폴장앞 교차로까지, 제4동질구간은 10~11구간, 3.03km로 남촌폴장앞에서 경안IC까지, 제5동질구간은 12구간, 2.4km로 경안IC에서 장지IC까지 구간이다. (〈표 7〉 참조)

위의 결과로부터 향후 국도구간을 대상으로 한 검지기 체계의 배치와 연관시켜 언급하기로 한다. 동질구간으로 선정되는 구간에서는 최소한 1개의 검지기가 필요하지만, 결측 데이터의 발생 및 검지기의 고장 등으로 인한 리스크를 고려 즉, 운전자에게 유효한 정보의 제공과 효율적인 DB의 관리 등 가공되어 제공되는 교통정보에 대한 품질보증을 위해서는 인근 검지기와의 간격 등을 고려하여 적정간격으로 설치하여야 할 것이다.

즉, 동질구간으로 구분된 각 국도구간들은 교통류의 특성 및 주위의 연도 특성 등 기능적인 측면이 서로 상이하다. 따라서, 이러한 검지기 배치 체계 및 대수의 결정에 있어 사전검토 자료로 본 연구에서 제안하는 동

질구간 선정과정과 그 결과가 기초자료로 활용될 수 있을 것으로 기대한다.

#### IV. 결론

본 연구에서는 국도 3호선의 곤지암 I.C.~장지 I.C.간의 10.72km를 대상으로 수집한 검지기 자료와 번호판 대조법을 이용한 구간통행시간과 구간통행속도 자료를 이용하여 기존의 이상치 제거방법에서의 문제점을 제시하고 참조자료를 활용한 새로운 유효 데이터 확보방안을 제시하였다. 또한 이 과정을 통해 얻은 유효 데이터를 이용하여 평일을 대상으로 한 시간대별 동질구간을 선정하였다.

본 연구를 통해, 지역간 교통류의 이동성확보라는 국도 고유의 기능을 상실해 가는 국도구간에서의 교통류의 특성이 시간대별(오전첨두, 오후비첨두, 오후첨두)로 많은 차이가 있음을 확인하였다. 따라서, 운전자에게 효율적인 교통정보(소요시간 추정, 우회 경로안내 등)의 제공을 위해서는 대상구간이 가지는 시간대별, 요일별, 계절별 특성을 파악하여 대안을 마련하는 것이 바람직할 것이다. 또한, 이러한 교통류의 특성이나 패

턴을 대상으로 하는 연구는 향후 소요시간 추정과 결측 및 데이터 보정에 필요한 기초 자료로 활용될 것으로 기대된다.

향후, 연구 과제로는 검지기 및 번호판 대조법(혹은 AVI)으로 구한 교통관련 자료의 신뢰성 확보 방안 즉, 점유율 데이터의 활용을 통한 이상치 제거방법과 동질 구간 선정방안에 대한 연구가 필요할 것으로 판단된다. 나아가, 요일별, 계절별 교통 데이터를 대상으로 한 이상치 제거 및 동질구간 선정 방법에 대한 연구도 병행되어야 할 것이다.

### 참고문헌

1. 강진기·손영태·윤여환·변상철(2002), 비매설식 자동차량인식장치를 이용한 구간교통정보 산출방법 연구, 한국 ITS학회 논문집, pp.22~32.
2. 김영찬·김태용(2001), 검지자료합성을 통한 도시 간선도로 실시간 통행시간 추정모형, 대한교통학회지 제19권 제6호, 대한교통학회, pp.171~182.
3. 김주현·도명식·정재은(2002), 국도 기능분류를 위한 그룹핑 방법론에 관한 연구, 대한교통학회지 제20권 제5호, 대한교통학회, pp.131~144.
4. 대한교통학회(2004), 도로용량편람, pp.133~163.
5. 도명식·김성현·문학룡·김명수(2004), 국도구간의 동질성 결정을 위한 방법에 관한 연구, 대한토목학회논문집, 제24권 제4D, pp.523~533.
6. 서울지방국토관리청(2001), 수도권 국도교통관리 시스템 운영·관리 업무대행사업최종보고서.
7. 이지연·도명식·김성현·류승기(2003), 교통량 데이터의 실시간 보정로직 - 국도 3호선을 중심으로, 응용통계연구 제16권 제2호, pp.203~215.
8. 장진환(2004), 자동차량인식장치 자료 필터링 알고리즘 개발, 서울시립대학 석사학위논문.
9. 최윤혁(2003), 택시 GPS Probe 자료의 실시간 이상치 제거 알고리즘 개발, 아주대학교 석사학위논문.
10. 한국건설기술연구원(2001), 통행시간 추정과 통계자료보정을 위한 결측데이터보정모형제작 최종보고서.
11. 한국건설기술연구원(2000), 도로포장 관리시스템 최종보고서.
12. SPSS 아카데미(1999), 통계자료분석을 위한 한글 SPSS, 고려정보산업.
13. 安井一彦·池之慶一郎(1995), 街路における旅行時間予測, 交通工學, Vol.30, No.6, pp.9~19.
14. Dion, F. and Rakha, H.(2003), Estimation spacial travel time using automatic vehicle identification data, TRB Annual Meeting.
15. FHWA(2001), "Traffic Monitoring Guide", U.S. DOT.
16. Okutani, I., Sephanedes, Y.J.(1984), Dynamic prediction of traffic volume through Kalman filtering theory, Transportation Research 18B, pp.1~11.
17. Spath, H.(1980), "Cluster analysis algorithms for data reduction and classification of objects", John Willy & Sons.
18. Washington, S.P., Karlaftis, M.S. and Mannering, F.L.(2003), Statistical and econometric methods for transportation data analysis, Chapman & Hall/CRC.

✉ 주 작 성 자 : 도명식

✉ 논문투고일 : 2004. 10. 13

논문심사일 : 2004. 10. 29 (1차)

심사판정일 : 2004. 10. 29

✉ 반론접수기한 : 2005. 4. 30