

# 기대최대화 알고리즘을 활용한 도로노면 training 자료 자동추출에 관한 연구 - 감독분류를 통한 도로 네트워크의 자동추출을 위하여

## Automatic Extraction of Training Dataset Using Expectation Maximization Algorithm - for Automatic Supervised Classification of Road Networks

한유경<sup>1)</sup> · 최재완<sup>2)</sup> · 이재빈<sup>3)</sup> · 유기윤<sup>4)</sup> · 김용일<sup>5)</sup>

Han, You Kyung · Choi, Jae Wan · Lee, Jae Bin · Yu Ki Yun · Kim Yong Il

### Abstract

In the paper, we propose the methodology to extract training dataset automatically for supervised classification of road networks. For the preprocessing, we co-register the airborne photos, LIDAR data and large-scale digital maps and then, create orthophotos and intensity images. By overlaying the large-scale digital maps onto generated images, we can extract the initial training dataset for the supervised classification of road networks. However, the initial training information is distorted because there are errors propagated from registration process and, also, there are generally various objects in the road networks such as asphalt, road marks, vegetation, cars and so on. As such, to generate the training information only for the road surface, we apply the Expectation Maximization technique and finally, extract the training dataset of the road surface. For the accuracy test, we compare the training dataset with manually extracted ones. Through the statistical tests, we can identify that the developed method is valid.

Keywords : Road Extraction, Supervised Classification, Automatic Road Surface Training Data Extraction, Expectation Maximization Algorithm, Orthophoto, LIDAR Data, Digital Map

### 초 록

본 논문은 감독분류 기법을 활용한 도로 네트워크 추출의 기본 과정인 트레이닝 자료의 추출과정을 자동화함으로써 감독분류를 활용한 도로 네트워크 추출 과정의 자동화에 기여할 수 있는 방법론의 개발을 목적으로 한다. 이를 위해 본 연구에서는 상호 기하보정된 항공사진과 LIDAR 자료로부터 정사영상과 LIDAR 반사강도 영상을 제작하고, 기 구축된 수치지도를 활용하여 초기 트레이닝 자료를 자동으로 추출하였다. 하지만 위의 과정을 통하여 추출된 초기 트레이닝 자료는 기하보정과정에서 수반되는 기하학적 오차 및 다양한 개체들로 구성된 도로의 특성에 영향을 받아 다양한 분광특성을 포함하게 된다. 따라서 본 연구에서는 추출된 초기 트레이닝 자료에서 도로 추출의 기본이 되는 도로노면의 분광특성을 통계학적 기법인 기대최대화 알고리즘에 기초하여 효과적으로 결정하기 위한 방법론을 제안하였다. 또한 개발된 방법론의 평가를 위하여 동일지역에 대해 수동으로 취득한 트레이닝 자료와 본 연구에서 자동으로 추출한 자료를 비교·평가하여 정확도를 분석하였다. 실험결과에 대한 통계검증결과 본 논문에서 제안한 도로노면 트레이닝 자료 자동추출기법의 효용성을 증명하였다.

핵심어 : 도로추출, 감독분류, 자동 도로노면 트레이닝 자료 추출, 기대최대화 알고리즘, 정사영상, LIDAR 자료, 대축척 수치지도

- 1) 학생회원 · 서울대학교 건설환경시스템 공학부(E-mail:han602@snu.ac.kr)
- 2) 학생회원 · 서울대학교 건설환경시스템 공학부(E-mail:choijw11@snu.ac.kr)
- 3) 교신저자 · 정희원 · 서울대학교 건설환경시스템 공학부 박사후 과정(E-mail:dama77@snu.ac.kr)
- 4) 정희원 · 서울대학교 건설환경시스템 공학부 부교수(E-mail:kiyun@snu.ac.kr)
- 5) 정희원 · 서울대학교 건설환경시스템 공학부 교수(E-mail:yik@snu.ac.kr)

## 1. 서 론

도심지 내의 도로 네트워크 정보는 도시환경 분석, 자동차 자동항법, 도시 계획 및 설계 등에서 사용되는 기본 지리정보자료이다. 따라서 도로 네트워크의 추출 및 갱신을 통한 지속적인 관리는 공간정보 분야에서 필수적이라고 할 수 있다. 일반적인 도로의 추출 및 갱신은 항공사진, 위성영상 혹은 실측을 통하여 반자동 혹은 수동으로 이루어진다. 이러한 과정은 많은 시간 및 인력자원을 필요로 하며 때때로 개인의 부주의로 인하여 데이터베이스 내의 도로 네트워크에 기하학적 오차를 발생시킬 가능성이 있다. 또한, 현대의 도심지는 빠른 속도로 변화하기 때문에 도로정보 갱신 주기가 도심지의 변화 속도에 부합하지 못하는 문제가 발생하기도 한다. 이러한 문제를 해결하기 위해서 다양한 알고리즘을 이용한 도로 네트워크 갱신 알고리즘이 제안되고 있으며, 최근에는 LIDAR(Light Detection and Ranging) 시스템, 4S Van 등의 도입을 통해 최신 센서들로부터 취득된 자료들의 활용이 가능해짐에 따라 도로 갱신 기술에 대한 연구는 더욱 활발하게 이루어지고 있다(황태현 등, 2005; Hu, 2003).

도로 네트워크의 자동갱신 및 구축기법은 크게 다양한 영상처리 알고리즘에 기반을 둔 자동 개체인식, 추출 알고리즘을 이용한 방법과 영상의 분광 및 공간정보에 기초한 분류기법을 활용한 방법론으로 구성된다(Kim 등, 2006; Laptv 등, 2004; Gerke 등, 2004). 일반적으로 대부분의 개체 인식, 추출 알고리즘의 경우 사용자의 개입에 의하여 정확도가 결정되며, 이로 인한 시간·비용적인 측면에서 단점을 가지고 있다. 이러한 단점들로 인하여 영상의 분광정보에 기초한 분류기법을 활용하여 사용자의 개입을 최소화 하고 도로를 자동으로 추출하기 위한 연구들이 많이 수행되고 있다. Song 등(2004)은 영상 내에서 수동으로 취득한 도로와 비도로의 training 자료를 SVM(Support Vector Machine) 분류자에 적용하여 도로 지역을 추출하고 이를 벡터화 시켜 도로를 갱신하는 방법을 제안하였으며, Zhang 등(2006)은 무감독 분류자인 K-means 클러스터링 기법을 기반으로 하여 영상내의 도로 화소를 추출하는 방법을 제안하기도 하였다. Zhu 등(2005), Long 등(2005), Hu(2003)는 영상처리 알고리즘의 일종인 수리형태학적(morphology) 연산을 기반으로 하여 항공사진 및 LIDAR 자료 내의 도로 화소를 분류 및 추출하는 방법론을 제안하기도 하였다.

이러한 분류 알고리즘을 활용한 도로 추출·갱신 방법론들의 경우, 해당 지역 내의 도로 네트워크에 대한 분광 및 공간적 특징에 관한 사전정보가 있어야 하며 이러한 정보의 추출에 사용자의 개입이 수반된다는 문제점이 지적되어 왔다. 특히 도로 갱신에 있어 효과적인 방법으로 알려져 있는 감독분류(Supervised classification) 기반의 기법의 경우, training 자료의 추출과정을 사용자의 개입에 의존해 왔으며 이는 감독분류를 활용한 도로 네트워크 추출 방법론을 자동화하는데 있어서 개선해야 될 문제점으로 지적되어 왔다. 따라서 본 논문에서는 감독분류 기법을 활용한 도로 네트워크 추출의 기본 과정인 training 자료의 추출과정을 자동화함으로써 감독분류를 활용한 도로 네트워크 추출 과정의 자동화에 기여할 수 있는 알고리즘의 개발을 목적으로 한다.

이를 위해 본 연구에서는 상호 기하보정된 항공사진과 LIDAR 자료로부터 정사영상과 항공LIDAR 반사강도 영상을 제작하고, 기 구축된 수치지도를 활용하여 초기 training 자료를 자동으로 추출하였다. 하지만 위의 과정을 통하여 추출된 초기 training 자료는 기하보정과정에서 수반되는 기하학적 오차와 다양한 개체들로 구성된 도로의 특성에 영향을 받아 다양한 분광특성을 포함하게 된다. 따라서 본 연구에서는 추출된 초기 training 자료에서 도로노면의 분광특성을 통계학적 기법에 기초하여 효과적으로 결정하기 위해 기대최대화 기법(Expectation Maximization Algorithm)을 활용한 자동 training 자료 추출 방법론을 제안하였다. 또한 개발된 방법론의 평가를 위하여 동일지역에 대해 수동으로 취득한 training 자료와 본 연구에서 자동으로 추출한 자료를 비교 평가하여 정확도를 분석하였으며, 결과에 대한 통계검증을 통해 본 연구에서 제안한 자동 training 자료 생성 방법론이 유용함을 증명하였다.

## 2. 연구내용

### 2.1 자료의 전처리

본 연구에서는 첨단 지형정보 획득 기법인 항공사진 측량, LIDAR 자료로부터 구축된 항공사진과 LIDAR 자료, 1:5,000 ver. 1.0 대축척 수치지도를 활용하여 개발된 알고리즘의 적용 평가를 수행하였다. 이를 위해 우선 자료의 전처리 과정으로서 서로 다른 센서 및 측량 기술로 구축된 LIDAR 자료, 항공사진, 대축척 수치지도를 공통의 좌표계 상에 상호 기하보정 하였다. 특히 기하보정

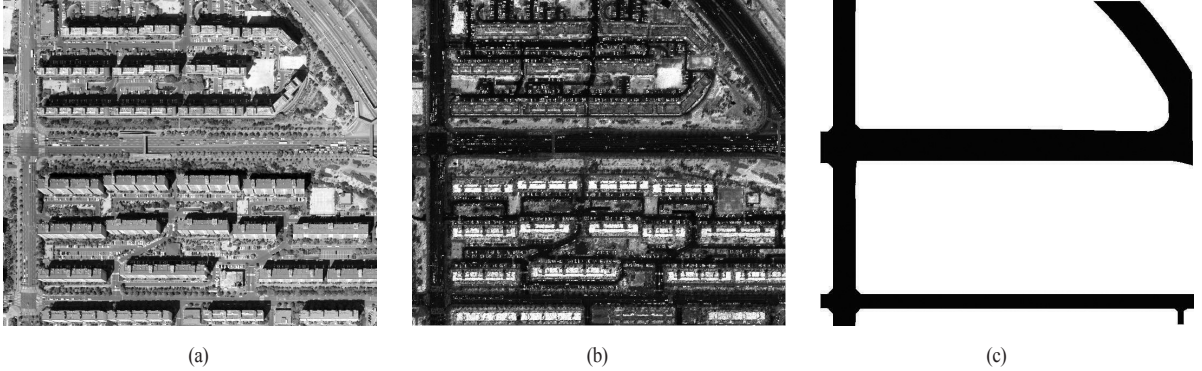


그림 1. (a) 정사영상 (b) LIDAR 반사강도 영상 (c) 대축척 수치지도 도로레이어

과정에서 이종의 측량 자료들 간의 상호 기하보정을 효과적으로 수행하기 위하여, 선형기하요소를 기하보정의 기본요소로 이용하는 알고리즘을 적용하여 이들 자료들에 대한 기하보정을 수행하였다(이재빈 등, 2006; 이재빈 등, 2007). 상호 기하보정 된 LIDAR 자료와 항공사진을 활용하여 본 알고리즘의 적용·평가에 사용될 정사영상을 제작하였으며, 또한 LIDAR의 반사강도 자료를 활용하여 정사영상과 같은 공간해상력을 가지는 대상지역에 대한 LIDAR 반사강도 영상을 제작하였다. 다음의 그림 1은 본 연구에서 사용될 정사영상, LIDAR 반사강도 영상, 대축척의 수치지도 도로레이어의 예를 보여준다.

## 2.2 Training 자료의 자동 추출

영상 내에 존재하는 도로에 해당하는 화소를 자동으로 추출할 수 있는 가장 좋은 방법은 상호 기하보정 된 대축척 수치지도의 도로 레이어를 이용하여 추출하는 방법이다. 하지만, 수치지도 작성 시 발생하는 도화 오차, 상호 기하보정 과정에서 발생하는 기하학적 오차, 도로 내에 존재하는 도로 이외의 수목, 그림자, 자동차와 같은 개체 등에 의하여 자동으로 추출된 초기 training 자료는 왜곡되게 된다. 따라서 본 연구에서는 기대최대화 알고리즘을 적용하여 수치지도 내의 도로 지역의 분광특성들의 최적분포를 추출하였다.

### 2.2.1 초기 training 자료 추출

대축척 수치지도 내에 존재하는 도로 레이어를 정사영상과 LIDAR 반사강도 영상에 적용하고 중첩(overlay) 연산자를 통하여 도로지역에 해당하는 자료를 추출할 수 있다. 그러나 수치지도 자체의 오차 및 이종 자료들

간의 상호 기하보정과정에서 발생하는 오차의 전파 등으로 인하여 수치지도 도로 레이어의 경계지역이 정사영상 및 LIDAR 반사강도 영상의 도로 경계와 정확하게 일치하지 않는 문제가 발생할 수 있다. 이러한 경계의 오차를 제거하기 위하여 수치지도로부터 획득한 초기 도로 지역의 폭을 축소하고자 하였으며 이를 위해 수리형태학적 연산처리를 적용하였다.

일반적으로 수리형태학 연산은 영상에서 경계, 골격 등과 같은 영역 형태를 표현하는데 있어 유용한 요소들을 추출하기 위한 도구로서, 영상 개체의 구조를 명확히 하기 위하여 사용된다(Gonzalez 등, 2002). 수리 형태학적 연산은 다양한 연산이 존재하며 본 연구에서는 침식 연산(erosion operator)을 사용하였다. 침식연산은 이진 영상에서 각 개체의 크기를 축소하거나 얇게 만드는 알고리즘으로, 팽창연산과 마찬가지로 구조요소 B에 의하여 결과는 제어된다. 구조요소 B에 대한 영상 A의 침식은  $A \ominus B$ 로 표현되며, 다음의 식 (1)과 같이 정의할 수 있다.

$$A \ominus B = \{z | (\hat{B})_z \cap A^c = \emptyset\} \quad (1)$$

B에 의한 A의 침식연산은 이동된 B가 모든 A의 일부와 겹치는 구조 요소의 원점의 위치들로 구성되는 집합이며, B가 (x,y)의 차원으로 이루어진 매트릭스로 구성된다고 하고 이를 이진 영상 내에 적용할 때, 침식연산은 다음과 같은 식 (2)로 표현할 수 있다.

$$(A \ominus B)(x,y) = \min \{A(x-x',y-y') * B(x',y') | (x',y') \in D_B\} \quad (2)$$

여기서,  $D_B$ 는 구조요소  $B$ 의 도메인이다. 이진영상 내에서 침식연산의 결과는 이진영상 A 내의 각 개체들이 구조요소 B의 모양에 인접하는 크기만큼 축소되는 결과를 보인다. 침식연산을 래스터화 된 도로지역에 적용하여 경계부근을 축소시키고 중첩연산을 이종의 자료들에 적용하여 초기 training 자료를 생성하였다. 그림 2는 반지름이 11화소인 원형(disk)의 구조요소를 이용한 수리형태학적 연산을 통해 폭이 축소된 도로의 형상을 보여주며 그림 3은 중첩연산을 통하여 정사영상 및 LIDAR 반사강도 영상으로부터 초기 training 자료가 추출되는 과정을 보여준다.

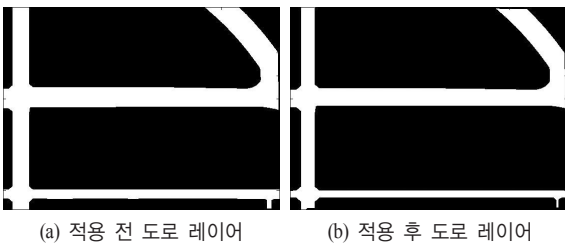


그림 2. 수리형태학적 연산의 적용

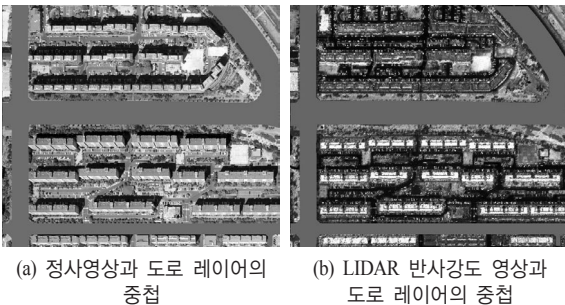


그림 3. 중첩 연산을 통한 초기 training 자료의 추출

2.2.2 기대최대화 알고리즘을 이용한 training 자료 추출

(1) 초기 training 자료의 히스토그램 분석

중첩연산을 통하여 추출된 초기 도로 training 지역은 아스팔트와 같은 도로노면을 구성하는 분광정보 외에도 도로중심선, 횡단보도 등의 노면 표지와 같은 개체 정보를 비롯하여 차량, 가로수 등의 도로 내에 포함되어 있지만 도로정보로 분류될 수 없는 개체들이 혼재되어 있다. 따라서 추출된 training 자료의 히스토그램은 그림 4와 같이 다양한 자료들의 분광정보가 혼재된 형태로 나타난다. 그림 4에서 가로축은 각 픽셀의 분광 값을 나타내며 세로축은 각 분광 값들의 빈도수/전체빈도수를 나타낸다.

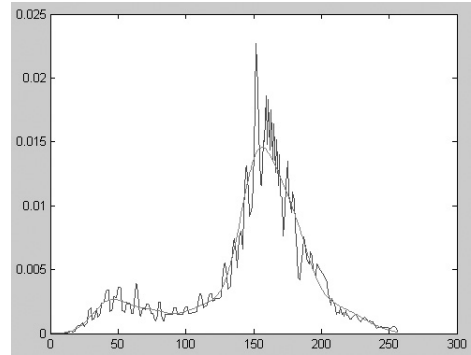


그림 4. 초기 도로 training 자료의 분포도

본 연구에서는 기대최대화 기법을 이용하여 도로 전체에 대한 training 자료 중 가장 지배적인 도로노면의 training 자료의 특징을 자동으로 추출하는 것을 목적으로 한다. 이를 위해 수치지도를 이용하여 추출된 도로 지역의 히스토그램을 도로노면, 도로중앙선 등 개의 클래스에 대한 가우시안 분포의 합으로 가정하고 이러한 가우시안 혼합 밀도모델에 기대최대화 기법을 적용하여 개의 클래스에 대한 파라미터를 추출한다.

(2) 기대최대화 알고리즘 (Expectation Maximization algorithm)

기대최대화 알고리즘은 관측되지 않은 변수의 확률모델의 파라미터를 최대우도를 갖는 값으로 반복연산을 통하여 찾아내는 통계적 기법으로, Expectation(E) step과 Maximization(M) step의 두 단계로 구성된다(Redner 등, 1984; 박노옥 등, 2003).  $K$ 개의 가우시안 분포를 가지는 개의 자료로 이루어진 확률밀도함수는 다음의 식 (3)과 같이 구성할 수 있다.

$$f(x; \theta) = \sum_{k=1}^K p_k g(x; m_k; \sigma_k) \tag{3}$$

여기서,  $(x; m_k; \sigma_k)$ 는 평균  $m_k$ , 표준편차  $\sigma_k$ 를 가지는 정규분포함수이다.

E step은 자료  $x$ 가 혼합 가우시안 확률모델을 가진다고 가정하고 관측 자료가 존재할 경우, 파라미터  $\theta$ 에 대한 최대우도를 계산할 수 있는데,  $i$ 번째 파라미터  $p_k^{(i)}$ ,  $m_k^{(i)}$ ,  $\sigma_k^{(i)}$ 에 대한 최대우도 계산식은 다음의 식 (4)와 같다.

$$p_k^{(i)}(k|n) = \frac{p_k^{(i)} g(x_n; m_k^{(i)}; \sigma_k^{(i)})}{\sum_{m=1}^K p_k^{(i)} g(x_n; m_k^{(i)}; \sigma_k^{(i)})} \quad (4)$$

M step에서는 E step에서 결정된 사후 확률을 기반으로 하여 새로운 가우시안 혼합 밀도 모델이 결정되며, 분포의 특성을 나타내는 새로운 파라미터( $p_k^{(i+1)}, m_k^{(i+1)}, \sigma_k^{(i+1)}$ ) 들은 다음의 식 (5), 식 (6), 식 (7)을 통해 구할 수 있다.

$$p_k^{(i+1)} = \frac{1}{N} \sum_{n=1}^N p^{(i)}(k|n) \quad (5)$$

$$m_k^{(i+1)} = \frac{\sum_{n=1}^N p^{(i)}(k|n) x_n}{\sum_{n=1}^N p^{(i)}(k|n)} \quad (6)$$

$$\sigma_k^{(i+1)} = \sqrt{\frac{1}{D} \frac{\sum_{n=1}^N p^{(i)}(k|n) \|x_n - m_k^{(i+1)}\|^2}{\sum_{n=1}^N p^{(i)}(k|n)}} \quad (7)$$

각각의 계산과정은 관측 자료를 이용하여 초기 파라미터를 결정한 후, 최대 우도의 변화가 없을 때까지 반복적으로 적용함으로써, 자료 내에서 최대우도 추정치를 가지는 개의 가우시안 모델을 생성해 낼 수 있다.

(3) 기대최대화 알고리즘을 활용한 도로표면 training 자료의 추출

초기 도로 training 자료 내에 존재하는 노면의 분포에 대한 최적의 파라미터를 추출하기 위해서는 기대최대화 알고리즘에서 사전에 결정되어야 하는 클래스의 개수  $K$ 를 정확하게 결정하여야 한다. 그러나 도로 training 자료 내의 클래스 개수는 지역에 따라서 변화할 수 있기 때문에, 의 변화에 따른 독립적인 가우시안 모델의 파라미터들을 분석하여 training 자료에 따른 최적의 를 결정하였다. 기대최대화 기법에 의해서 개의 가우시안 모델로 training 자료가 분리되었을 때, 도로의 특성에 의하여 가장 높은 비율( $\text{argmax}(p_k)$ )을 가지는 클래스는 도로노면이 되어야 한다. 또한, 도로노면의 클래스 분포의 평균  $m_k$ 는 전체 training 자료의 최빈값에 수렴되며 도로노면의 분광특성은 전역적으로 유사하기 때문에 표준편차  $\sigma_k$ 는 다른 클래스들에 비하여 작은 값을 가져야 할 것이다. 클래스 개수  $K$ 에 따른 도로노면으로 추정되는 특정

클래스에 대한 파라미터  $p_k, m_k, \sigma_k$ 를 평가하여, 자동으로 도로노면에 대한 최적의 training 자료 분포를 추출하였다. 다음의 그림 5는 초기 training 자료의 히스토그램으로부터 본 도로노면에 대한 최적의 training 자료 분포를 추출한 결과의 예를 보여준다.

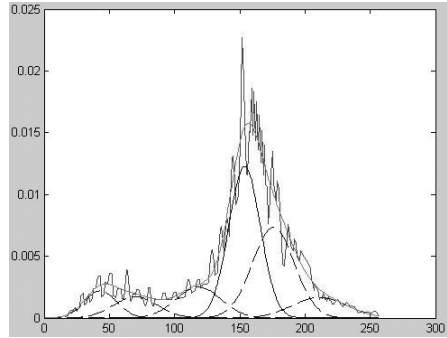


그림 5. 기대최대화 알고리즘 적용 결과의 예(K=6)

$K=6$ 일 때 형성된 전체 가우시안 모델 중, 실선으로 표현된 클래스의 가우시안 모델이 전체 training 자료에서 가장 높은 비율을 차지하며, 평균  $m_k$ 는 training 전체의 최빈값과 가장 유사함을 확인하여 이를 도로노면의 training 자료로 추출할 수 있다.

### 3. 실험 및 평가

#### 3.1 실험 대상지역

본 연구에서 제안한 방법론의 효용성을 평가하기 위하여 대전광역시 지역에 대해 수집된 항공사진과 LIDAR 자료 및 동일지역에 대한 1:5,000 수치지도를 사용하여 실험을 수행하였다. 수집된 항공사진의 공간해상력은 25cm 이고, LIDAR 자료의 경우 대상지역에 대해 2.5pt/m<sup>2</sup>의 점밀도로 수집되었다. 수집된 자료의 제한에 따르면 기대되는 수직, 수평 정확도는 각각 15cm, 30cm 이다. 자료의 전처리 과정을 통해 상호 기하보정을 수행하였으며 기하보정 된 자료들을 사용하여 초기 training 자료 추출을 위한 정사영상, LIDAR 반사강도 영상이 제작되었다. 실험을 위하여 2개의 대상지역(대상지역-I, 대상지역-II)이 선정되었으며, training 자료를 자동으로 추출하기 위하여 동일 지역의 수치지도에 존재하는 도로레이어를 정사영상과 LIDAR 반사강도 영상에 중첩하여 초기 training 자료를 추출하였다. 이 과정에서 제 2.2.1절에서 기술한 바와 같이 수치지도 자체에 존재

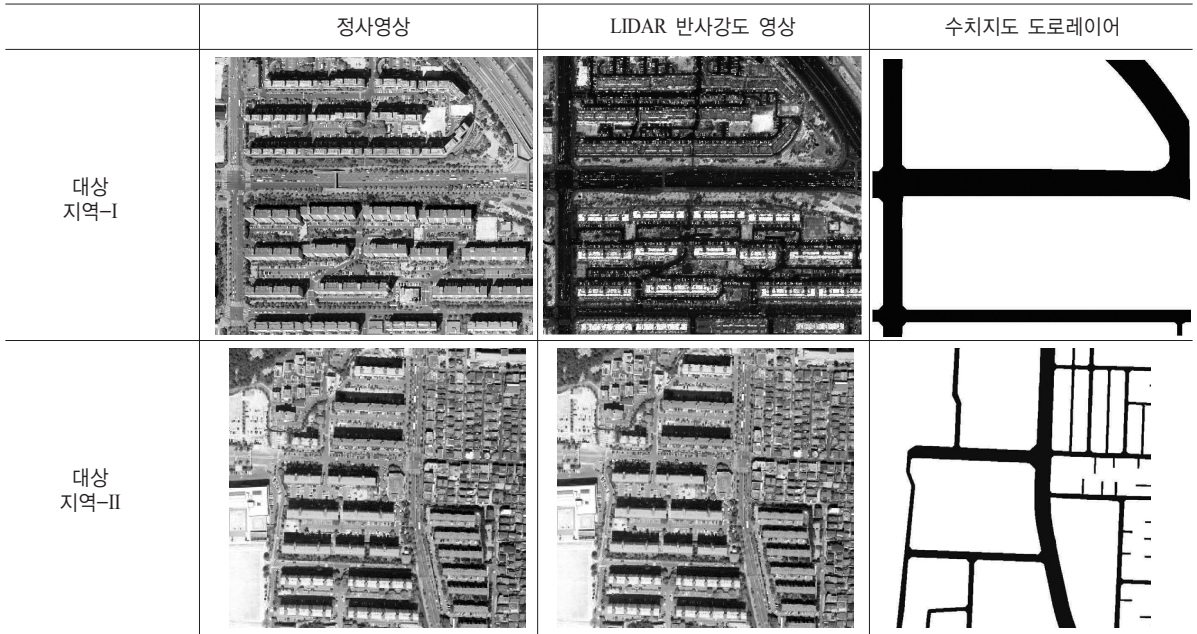


그림 6. 연구 대상지역

하는 오차와 자료들의 상호 기하보정 과정에서 전파되는 오차의 영향을 최소화하기 위해 도로레이어에 수리 형태학적 필터를 적용하고 이를 이용하여 초기 training 자료를 추출하였다. 다음의 그림 6은 대상지역-I과 대상 지역-II에 대해 각각 정사영상, 수치지도의 도로레이어, LIDAR 반사강도 영상을 보여준다.

### 3.2 기대최대화 알고리즘을 활용한 도로노면 training 자료 자동 추출

추출된 도로의 초기 training 자료를 기대최대화 알고리즘을 이용하여 K개의 정규분포를 갖는 성분으로 분리하였다. 일반적인 도로의 구성을 살펴보면 도로의 노면, 중심선, 도로 내에 존재하는 그림자 및 수목 등의 특성이 주를 이루고 그 외에 자동차나 횡단보도 등이 적은 분포를 이루고 있다. 따라서 초기 training 자료를 나누는 기준인 K의 최소값을 2로 설정하였다. 그 중에서 관심대상인 도로노면의 training 자료는 도로를 구성하는 가장 주된 성분이므로 K개의 정규분포를 갖는 성분 중에 가장 큰 비중을 차지하는 성분, 즉 가장 높은 비율인  $\text{argmax}(p_k)$ 의 값을 갖는 클래스를 도로노면의 training 자료로 판단하였다. 또한 K개의 성분이 전체 분포에 대해서 어느 정도 이상을 차지해야 의미가 있는 것이므로 각각의 클래스의 성분 비율(즉, 전체분포에 대해서 해당

클래스의 빈도수 비율을 의미)인  $p_k$ 의 값이 0.05가 넘지 못하는 값이 생기는 경우, 더 이상 클래스의 개수를 증가시켜도 의미를 갖지 못한다고 판단하였다.

이러한 절차를 따라 초기 training 자료로부터 K의 값에 따른 도로 노면의 training을 추출할 수 있다. 앞서 언급하였다시피, 도로의 training 자료에서 가장 주된 성분을 이루는 것이 도로 노면이므로, 도로의 전체 training 자료에서 나타나는 최빈값과 가장 유사한 평균을 가질 때의 K를 최적의 training으로 선별하였다. 하지만 서로 다른 K 값을 가지는 경우에도 비슷한 범위 내의 평균값을 가지는 분포가 존재할 수 있으므로 이러한 경우 좀 더 큰  $p_k$  값을 가지는 분포가 도로노면의 training 자료로 판단할 수 있도록 알고리즘을 구성하였다. 위와 같은 결정규칙에 의거하여 자동으로 추출된 대상지역-I, 대상

표 1. 대상지역-I에 대한 기대최대화 알고리즘 적용 결과

	K	p	평균	표준 편차	초기 training 자료의 최빈값
Band 1	3	0.54	142.7	23.4	138
Band 2	3	0.57	143.7	24.1	134
Band 3	3	0.52	147.7	19.2	139
반사강도	2	0.62	10.2	2.1	10

표 2. 대상지역-II에 대한 기대최대화 알고리즘 적용 결과

	K	p	평균	표준 편차	초기 training 자료의 최빈값
Band 1	4	0.55	157.2	15.4	153
Band 2	4	0.55	157.1	16.1	153
Band 3	4	0.57	159.3	16.5	155
반사강도	3	0.54	8.9	2.4	8

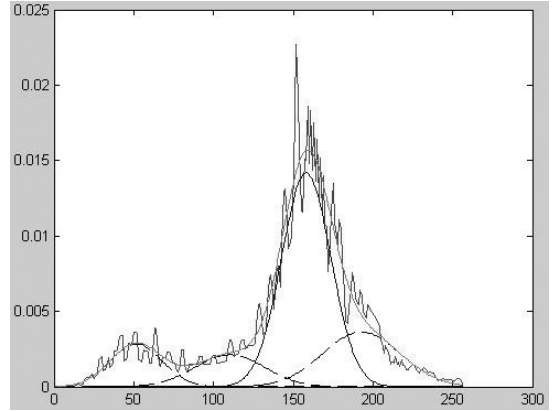
지역-II의 도로노면에 대한 training 자료는 다음의 표 1, 표 2와 같다.

대상지역-I의 경우 training 자료의 최빈값과 항공사진의 컬러 Band의 화소값의 평균에서 어느 정도 차이를 보이는 것을 알 수 있다. 이것은 대상 지역이 큰 도로 중심으로 구성이 되어있기 때문에 도로 내에 존재하는 차량의 빈도수가 많고 도로 주변의 가로수나 고층 건물에 의한 그림자 등에 의한 영향에 의한 것으로 보인다. LIDAR 반사강도에 대해 추출된 training 자료의 평균값은 초기 training 자료의 최빈값과 유사한 분포를 보여주고 있다. 상대적으로 대상지역-I에 비해 대상지역-II의 경우에는 항공 컬러영상의 training 자료의 최빈값과 자동 추출된 도로노면의 training 자료의 평균이 유사하게 나온 것을 알 수 있다. 다음의 그림 7은 대상지역-II 정사영상 Band 1과 LIDAR 반사강도 영상에 대해 기대최대화 알고리즘을 적용한 결과에 대한 히스토그램 분포를 보여준다.

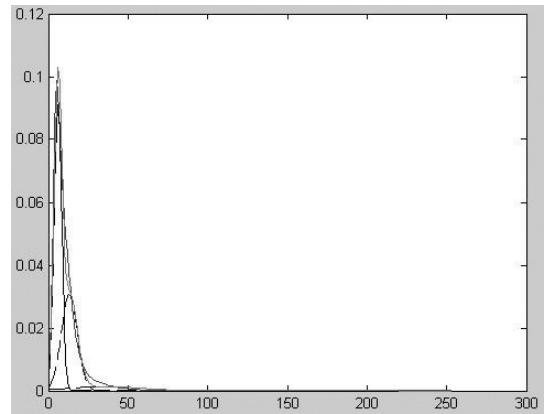
K=4일 때의 Band 1 정사영상과 K=3일 때의 반사강도 영상의 결과에서 형성된 전체 가우시안 모델 중, 실선의 클래스로 표현된 가우시안 모델이 전체 training 자료에서 가장 높은 비율을 차지하며, 평균 는 training 자료 전체의 최빈값과 가장 유사함을 확인하고, 이때의 값이 다른 경우보다 높음을 확인하여 이를 도로노면의 training 자료로 추출할 수 있었다.

표 3. 도로노면 training 자료의 수동 및 자동 추출 결과(대상지역-I)

	Band 1		Band 2		Band 3		반사강도	
	평균	표준편차	평균	표준편차	평균	표준편차	평균	표준편차
Training 자동 추출	142.7	23.4	143.7	24.1	147.7	19.2	10.2	2.1
Training 수동 추출	143.6	21.6	143.3	22.4	148.9	19.9	9.8	1.9



(a) 정사영상의 Band 1(K=4)



(b) LIDAR 반사강도 영상(K=3)

그림 7. 대상지역-II 기대최대화 알고리즘 적용 결과 예

### 3.3 정확도 평가를 위한 training 자료 수동추출 및 분석

정확도 평가를 위하여 대상지역-I, 대상지역-II 각각에 대하여 도로노면에 대한 training 자료를 수동으로 추

표 3. 도로노면 training 자료의 수동 및 자동 추출 결과(대상지역-I)

	Band 1		Band 2		Band 3		반사강도	
	평균	표준편차	평균	표준편차	평균	표준편차	평균	표준편차
Training 자동 추출	157.2	15.4	157.1	16.1	159.3	16.5	8.9	2.4
Training 수동 추출	153.0	16.7	152.6	17.1	153.7	17.9	8.7	2.1

출하였다. 대상지역-I, 대상지역-II의 도로노면에 대해 자동 및 수동을 통해 추출된 training 자료의 Band 별 평균 및 표준편차는 표 3, 표 4와 같다.

제안한 기법을 통해 추출된 도로노면의 training 자료의 평균과 표준편차가 도로노면을 대표할 수 있는 값인지 알아보기 위해서 평균과 표준편차 각각에 대하여 검정을 수행하였다. 우선, 자동으로 추출한 도로노면의 training 자료를 모집단으로 하고, 같은 클래스에 대해서 수동으로 추출한 데이터의 평균이 모집단의 평균과 일치하는지를 검정을 통해 알아보고자 하였다. 정규모집단의 모평균이  $\mu_0$ , 표본집단의 분산이  $S^2$ 이고 크기 n인 확률표본으로부터의 표본평균이  $\bar{X}$ 일 때의 검정통계량은 식 (8)과 같다.

$$Z = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \quad (8)$$

식 (8)을 토대로, 자동 추출한 평균과 수동 추출한 평균이 같을 때를 귀무가설로 설정했을 때의 양측검정 결과, 95%의 신뢰도에서 대상지역-I이 귀무가설을 채택하였고, 98%의 신뢰도에서 대상지역-II이 귀무가설을 채택함을 증명할 수 있었다.

마찬가지로 분산에 대한 검정을 수행하기 위하여 카이제곱검정법을 수행하였다. 크기 n인 확률표본으로부터의 표본분산이  $S^2$ , 모집단의 분산이 일  $\sigma_0^2$  때의 검정통계량은 식 (9)와 같다.

$$\chi^2 = (n-1)S^2/\sigma_0^2 \quad (9)$$

귀무가설이 모분산과 표본분산이 같을 때의 기각역 (식 (10))

$$\chi^2 \geq \chi^2(n-1, \alpha/2) \text{ or } \chi^2 \leq \chi^2(n-1, 1-\alpha/2) \quad (10)$$

을 통하여 추출된 분산에 대한 검정을 수행할 수 있다. 검정 결과, 대상지역-I, 대상지역-II 모두 95%의 신뢰구간에서 귀무가설을 기각할 수 없었고, 따라서 이를 통해 두 training 자료의 분산 또한 허용범위 내에서 다르지 않다는 것을 증명하였다. 즉, 본 연구를 통하여 자동으로 추출한 도로노면의 training 자료의 평균과 분산이 수동으로 추출한 도로노면 training 자료의 평균 및 분산과 통계학적으로 유의미한 수준에서 다르지 않다는 결론을 도출할 수 있었다.

#### 4. 결 론

본 연구를 통하여 도출된 결론 및 향후 계획은 다음과 같다.

첫째, 본 연구에서는 기대최대화 기법을 통하여 초기 training 자료로부터 도로노면에 해당하는 training 자료를 자동으로 분리·추출하는 방법론을 구축하였고, 수동으로 추출된 도로노면 training 자료와의 비교·평가를 통계학적 기법을 활용하여 수행하였으며, 이를 통해 제안된 방법론의 효용성을 증명하였다.

둘째, 본 연구를 통해 감독분류 기법을 활용한 도로 네트워크 자동추출의 기본 과정인 training 자료의 추출 과정을 자동화함으로써 감독분류를 활용한 도로 네트워크 추출 과정의 자동화에 기여할 수 있는 방법론의 구축이 가능함을 증명하였다.

셋째, 본 논문은 도로를 구성하는 주요소인 도로노면 training 자료의 자동추출 방법론의 구축에 연구의 초점이 제한되어 있다. 추가 연구를 통하여 도로표지, 자동차, 수목 등 도로를 구성하는 다양한 요소들에 대한 자동 추출·분류 기법의 추가 개발이 필요하다고 판단된다.



마지막으로, 본 논문에서는 항공사진, LIDAR 자료에 개발된 방법론을 적용하고 평가를 수행하였으나 추후 4S Van, 현장 도로에 설치된 CCD(Charge Coupled Device) 카메라 등의 측량장비로부터 획득되는 자료에의 확대 적용을 통해 개발된 방법론의 효용성을 증대시키기 위한 연구가 필요하리라 판단된다.

### 감사의 글

본 연구는 국토해양부 첨단도시기술개발사업-지능형 국토정보기술혁신사업과제의 연구비지원(07국토정보 C04)에 의해 수행되었습니다.

### 참고문헌

박노옥, 지광훈, 이광재, 권병두 (2003), 다중시기 원격 탐사 화상의 변화탐지를 위한 임계치 자동 추정, 대한원격탐사학회지, 대한원격탐사학회, 제 19권, 제 6호, pp. 465-478.

이재빈, 배환성, 유기운 (2007), 3차원 수치영상 지형지도의 제작에 관한 연구: 항공레이저측량자료, 항공사진, 대축척 수치지도를 활용하여, 국토연구, 국토연구원, 제 52권, pp. 136-150.

이재빈, 유기운 (2006), 선형기하보정 요소를 이용한 항공레이저측량 자료, 항공사진, 대축척 수치지도의 기하보정에 관한 연구, 한국지형공간정보학회지, 한국지형공간정보학회, 제 14권, 제 4호, pp. 37-44.

황태현, 주인학, 최경호 (2005), 4S-VAN을 이용한 공간정보 구축과 갱신을 위한 변화탐지 시스템의 개발, GIS/RS 공동 춘계학술대회 논문집, 한국 GIS 학회, pp. 47-52.

Gerke, M. Butenuth, M. Heipke, C. and Willrich, F. (2004), Graph-supported verification of road databases, *ISPRS Journal of Photogrammetry & Remote Sensing*, Vol. 58, pp. 152-165.

Hu, Y. (2003), *Automated extraction of digital terrain models, roads and buildings using airborne lidar data*, PhD dissertation, Department of Geomatics Engineering, University of Calgary.

Kim, J. Han, D. Yu, K. Kim, Y. and Rhee, S. (2006) Efficient extraction of road information for car navigation applications using road pavement markings obtained form aerial images, *Canadian Journal of Civil Engineering*, Vol. 33, No. 10, pp. 1320-1331.

Laptey, I. Mayer, H. Lindeberg, T. Eckstein, W. Steger, C. and Baumgartner, A. (2004), Automatic extraction of roads from aerial images based on scale space and snakes, *Machine Vision and Applications*, Vol. 12, pp. 23-31.

Long, H. and Zhao, Z. (2005), Urban road extraction from high-resolution optical satellite images, *International Journal of Remote Sensing*, Vol. 26, No. 22, pp. 4907-4921.

Redner, R. A. and Walker, H. F. (1984), Mixture densities, maximum likelihood and the EM algorithm, *SIAM Review*, Vol. 26, No. 2, pp.195-239.

Gonzalez, R. C. and Woods, R. E. (2002), *Digital image processing*, Prentice Hall, pp. 523-527.

Song, M. and Civco, D. (2004), *Road extraction using SVM and image segmentation*, PE&RS, Vol. 70, No. 12, pp. 1365-1371.

Zhang, Q. and Couloigner, I. (2006), Automated road network extraction from high resolution multi-spectral imagery, *In ASPRS Proceedings*.

Zhu, C., Shi, W. Pesaresi, M. Liu, L. Chen, X. and King, B. (2005), The recognition of road network from high-resolution satellite remotely sensed data using image morphological characteristics, *International Journal of Remote Sensing*, Vol. 26, No. 24, pp. 5493-5508.

Ziems, M. Gerke, M. and Heipke, C. (2007), Automatic road extraction from remote sensing imagery incorporating prior information and colour segmentation, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 141-147.