

LISA를 이용한 LIDAR 데이터로부터 건물 추출에 관한 연구 A Study on Building Extraction from LiDAR Data Using LISA

변영기¹⁾ · 이정호²⁾ · 손정훈³⁾ · 유기윤⁴⁾

Byun, Young Gi · Lee, Jeong Ho · Son, Jeong Hoon · Yu, Ki Yun

Abstract

This paper aims at developing an efficient method that extracts building using local spatial association of raw LiDAR data without setting up empirical variables such as a minimum building area, and applying the method to survey data to evaluate the efficiency of that. To do this, LISA(Local Indicators of Spatial Association) statistics are used which reflect local variations that can be appeared in the research area. It can be also a preprocess that detects spatial outliers through the significance test of LISA statistics and interpolate using kernel estimation. Boundaries of buildings as well as buildings can be extracted based on quadrant of Moran Scatterplot. Experimental results show that the proposed method is promising in extracting buildings from LiDAR data automatically.

Keywords : LISA, Local Moran, Moran ScatterPlot, LiDAR

초 록

본 연구에서는 항공레이저 스캐닝 데이터를 이용하여 기존의 최소건물면적등과 같은 경험적 변수의 설정 없이, 원시 라이다 자료의 국지적 공간상관(local spatial association) 특성을 이용하여 건물을 추출하는 효율적인 방법을 개발하고, 이를 실측데이터에 적용하여 개발된 방법의 성능을 평가하는데 목적이 있다. 이를 위하여 연구지역 내에서 발생할 수 있는 공간상관의 국지적 변이(local variations)를 고려하는 공간통계분석기법인 LISA(Local Indicators of Spatial Association) 통계치를 이용하였다. 전처리 과정으로 LISA 통계치의 유의성 검정 과정을 통하여 공간이상치를 검출하였고, 검출된 이상치를 kernel estimation통해 주변의 특성을 반영한 값으로 보간하였다. Moran Scatterplot의 사분면을 기준으로 건물은 물론 동시에 건물의 외곽선 정보까지 추출할 수 있다. 실험결과 본 연구에서 제안한 방법은 건물을 자동으로 추출할 수 있는 가능성을 제시하였다.

핵심어 : LISA, 공간상관, Local Moran, Moran Scatterplot, LiDAR

1. 서 론

최근 항공레이저스캐너로부터 취득한 표고정보로부터 다양한 3차원 공간정보 추출을 위해 지속적인 연구와 개발이 이루어지고 있다.

최근 몇 년간 라이다 데이터를 이용하여 건물을 추출하기 위해 데이터 형태 및 인접관계 구조, 데이터 처리 방식에 따른 다양한 연구가 진행되어 왔다.

김형태(2001)와 Rottensteiner et al.(2002)그리고 이정호 등(2005)은 래스터 기반의 이미지데이터에 많이 사용되는 opening 연산에 기반한 다양한 모폴로지 필터의 적용을 통해 건물추출을 수행하였다.

좌윤석(2003)은 원시 포인트의 특성을 그대로 유지하면서 정규격자의 좌표 검색 기능을 제공하는 가상격자(pseudo grid)를 이용한 건물추출을 수행하였다.

한수희 등(2005)은 원시 항공레이저스캐너 포인트 자

- 1) 정회원 · 서울대학교 공과대학 지구환경시스템 공학부 박사과정(E-mail:kko071@snu.ac.kr)
- 2) 정회원 · 서울대학교 공과대학 지구환경시스템 공학부 박사과정(E-mail:ilome79@snu.ac.kr)
- 3) 정회원 · 서울대학교 공과대학 지구환경시스템 공학부 석사과정(E-mail:zig19@snu.ac.kr)
- 4) 연결저자 · 정회원 · 서울대학교 공과대학 지구환경시스템 공학부 조교수(E-mail:kiyun@snu.ac.kr)

료로부터 인식 가능한 스캔라인 정보를 이용한 건물포인트 분류를 수행하였다.

그러나 기존의 대부분의 연구들의 경우 LiDAR 데이터로부터 지면과 비지면을 분리하기 위해서 일반적으로 지면의 경사, 최대건물의 크기, 건물의 최소높이 등의 경험적 변수를 사용하였다. 따라서 지형의 형태 및 사용된 변수가 분석결과에 많은 영향을 미치게 된다.

이에 본 연구에서는 위와 같은 변수의 설정 없이, 원시 라이다 자료의 국지적 공간상관(local spatial association) 특성만을 이용하여 건물을 추출하는 효율적인 방법을 개발하고, 이를 실측데이터에 적용하여 개발된 방법의 성능을 평가하는데 목적이 있다. 개발된 알고리즘의 주요내용은 공간가중치행렬의 구성, 공간 이상치 검출, 건물 추출 및 경계 검출로 구성된다. 본 연구에서 제안한 라이다 데이터의 분할 방식의 대략적인 구성은 그림 1과 같다.

2. 공간 이상치(Spatial Outlier) 검출

원시 라이다 데이터는 일반적으로 다양한 원인으로 인해 공간 이상치를 포함한다. 공간 이상치는 데이터의 관측값들 중에서 지역적으로 불안전(local instability)하거나 특이한 관측값을 말한다. 공간 이상치는 크게 실제값보다 낮게 나오는 low-outlier 와 실제 지형지물보다 높게 측정되는 high-outlier로 나눌 수가 있다. 이러한 공간 이

상치는 라이다 데이터를 이용한 건물 추출등과 같은 다양한 용도에 활용하기 위해서는 반드시 제거되어야 한다.

라이다 데이터와 같은 위치정보를 포함하고 있는 공간 데이터의 이상점 검출을 위하여 전통적인 통계분석 기법을 적용하는 것은 통계적 모형의 유효성에 대한 그릇된 판별력을 갖게 함으로써 오인된 통계적 추론을 유도할 수 있다(Anselin, 1988).

이에 본 연구에서는 공간실체(spatial entities)들의 상호 의존성 및 이질성과 같은 공간효과(spatial effect)를 고려하는 공간통계분석기법인 LISA(Local Indicators of Spatial Association)을 이용하여 공간이상치를 검출하였다.

일반적으로 공간통계기법을 채용하여 개별데이터의 국부적인 특성을 조사하기 위해서는 개별데이터의 상호작용을 개념화하는 공간가중치행렬이 구성 되어야 한다.

2.1 공간가중치행렬 구성

공간적 상관(spatial association)분석에 있어서 가장 중요한 요소중의 하나인 공간 가중치 행렬(spatial weight matrix)은 연구자가 공간데이터의 상호작용을 정량화 하는 방식에 따라 매우 다양하게 구성될 수 있다.

가장 단순하게는 래스터 자료에 대해 Rock, Bishop, Queen과 같이 인접한 화소의 방향에 따라 공간적 이웃을 정의하고 인접성에 기반한 이진 가중치를 부여하는 방식에서부터, 불규칙 포인트 자료나 복잡한 형태의 벡터 자료에 대해 거리, 위상정보 등을 이용하여 공간적 인접성을 정의하고, 면적, 길이 등을 이용하여 가중치 함수를 정의하는 방식까지 매우 다양하다.

대개의 라이다 자료의 처리과정은 크게 원시 항공레이저의 불규칙한 포인트 자료 형태를 사용하거나 취득된 포인트 데이터를 래스터의 정규격자 형태로 변환하여 처리한다.

본 연구에서는 공간가중치 행렬의 구성 및 영상처리 프로그램의 장점을 취하기 위해서 원시 라이다 데이터를 Nearest Neighbor 기법을 통해 변환한 격자형 데이터를 사용하였다.

변환된 격자형 데이터의 공간가중치 행렬의 구성을 위해 본 연구에서는 Queen 구조 및 1차 이웃(first-order spatial lag)을 이용하고 인접성에 기반한 이진 공간 가중치 행렬을 구성하였다. 즉 공간단위인 화소 i 와 j 가 서로 인접하면 1의 가중치를 부여하고, 인접하지 않으면 0의 가중치를 부여한다.

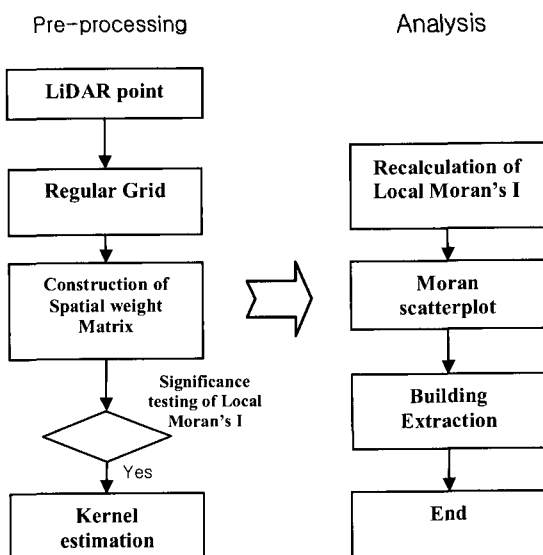


그림 1. 자료처리 흐름도

대다수의 공간통계 분석기법들은 계산의 편의성을 위해 행별 표준화된 공간 가중치 행렬(row-standardized spatial matrix)을 사용한다. 행별표준화된 공간가중치행렬은 가중치행렬을 구성하는 원소를 각행의 합으로 나누어 공간단위들이 갖는 상대적인 영향력을 표준화한 행렬을 말한다.

2.2 국지적 공간상관지표(LISA)

LISA 는 국지적 공간상관을 파악하는 시각화방법의 일종이다. 연구지역이 상대적으로 넓은 경우, 대상지역 안에서의 공간상관이 위치에 따라서 상이할 수 있다. 따라서 연구지역 내에서 발생할 수 있는 공간상관의 국지적 변이(local variations)를 고려한 시각적 지표가 LISA이다 (Anselin, 1995; Shekhar et al, 2003). 이 LISA를 이용하면 한 변수의 공간상관이 특정지점에서 매우 높게 나타나거나 매우 낮게 나타나는 공간 이상치를 찾을 수 있다.

국지적 규모에서 공간상관정도를 측정하기 위해서는, 각각의 공간단위에서 공간상관값이 계산되어야 하는데 여러 LISA 중 가장 손쉽게 활용될 수 있는 것이 Local Moran's I 이다.

2.3 Local Moran's I

대다수 통계학에서 사용되고 있는 통계치들은 표본공간에서의 관측값들 사이의 독립성을 가정하는 반면, Moran's I는 관측값들 사이의 상호의존성(spatial dependence) 또는 공간 자기상관(spatial autocorrelation)의 정도를 측정하는 통계치를 말한다.

Moran's I 측정은 크게 관측데이터의 전체적 경향을 보는 Global Moran's I와 개별위치에 따른 국지적 경향을 측정하는 Local Moran's I로 구분할 수 있다.

Local Moran's I은 Global Moran's I의 분자와 분모에 각각 n^2 을 나누고, 행별표준화된 공간 가중치행렬을 사용하면 식 (1)과 같이 Z-statistics 형태로 표현할 수 있다.

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} = \frac{1}{n} Z^T \tilde{Z} \quad (1)$$

여기서, n 은 전체데이터의 개수, \bar{x} 는 전체데이터의 평균, x 는 공간변수, w_{ij} 는 행별표준화된 공간가중치행렬의 원소를 말한다.

위 식을 개별위치에 해당하는 Moran's I 값으로 분해하

면 식 (2)와 같은 Local Moran's I 값을 구할 수 있다.

$$I_i = z_i \times \sum_j w_{ij} z_j \quad (2)$$

만일 특정한 지점의 값이 인접한 공간단위들과 유사한 값을 갖는 경우, Local Moran's I 계수는 정적공간상관을 갖는 반면, 서로 상이한 값들을 갖게 되면 부적공간상관을 갖게 된다.

2.4 Local Moran's I의 유의성 검정

앞서 측정된 Local Moran's I 값들 중에서 부적공간상관을 보이는 지점들을 공간이상치 후보점으로 생각할 수 있다. 본 연구에서는 측도의 신뢰성을 위해 유의성검정과정을 통해 최종적인 공간 이상치를 검출하였다. Local Moran's I의 유의성검정 방식은 모집단의 분포를 정규분포 또는 랜덤분포로 가정하느냐에 따라 모수의 추정치가 달라진다. 본 연구에서는 “국지적 규모에서의 공간 상관관계가 존재하지 않는다”는 귀무가설과 랜덤분포의 가정 아래 유의성 검정을 실시하였다.

랜덤분포 접근법(Randomization approach)은 주어진 변수값은 일정하게 고정되어 있고, 가능한 모든 공간적 배열을 생성하여 super-population을 구성하고, 이를 토대로 유의성검정을 실시하는 방법이다.

randomization 가정 및 귀무가설 가정하에서의 모수(parameter) 추정치는 다음과 같다(Anselin, 1995).

$$\begin{aligned} E[I_i] &= -\frac{1}{n-1} w_i \\ \text{Var}[I_i] &= \frac{w_i^{(2)}(n-b_2)}{n-1} + \frac{(w_i^2 - w_i^{(2)})(2b_2 - n)}{(n-1)(n-2)} \\ &\quad - [E(I_i)]^2 \\ b_2 &= m_4 / m_2^2, \quad m_4 = \sum_i z_i^4 / n \end{aligned} \quad (3)$$

여기서, n 은 전체데이터의 개수, b_2 은 첨도(kurtosis), $w_i = \sum_j w_{ij}$, $w_i^{(2)} = \sum_j w_{ij}^2$ 를 나타낸다. 이를 이용하여 유의수준 1%에서 통계적으로 유의한 값을 공간 이상치로 추출하였다.

2.5 Kernel estimation

본 연구에서는 4차의 커널(quartic kernel)을 이용하여

이상치 지점의 값을 보간하였다. 커널 추정은 주변값들을 이용하여 임의 지점에 대한 속성값을 추정한다는 점에서 일반적인 필터방식과 비슷하나, 주변값을 고려함에 있어서 거리에 따른 가중치를 고려할 수 있다는 장점이 있다. 사용된 quartic kernel function은 다음과 같다.

$$k_{\tau}(s) = \begin{cases} \frac{3}{\pi} \left(1 - \frac{d_{ij}^2}{\tau^2} \right)^2 & d_{ij} \leq \tau \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

여기서, d_{ij} 는 관측점 i 와 j 의 거리, τ 는 커널 밴드폭을 말한다. 본 논문에서는 $\tau=3$ 을 사용하였다.

3. 공간 패턴 분류

각 개별위치의 Local Moran' I 값을 시각화한 Moran Scatterplot에서의 위치를 표시한 일종의 정성적 주제도인 Moran scatterplot map을 이용하여 공간적 군집을 분석할 수 있다. 본 연구에서는 이 Moran scatterplot map을 이용하여 건물 및 건물의 경계를 추출하였다.

3.1 Moran ScatterPlot

Moran ScatterPlot은 식 (2)에서와 같이 Z-통계량으로 정규화된 개별위치들의 높이값을 X축에 정규화된($W \cdot Z$) 주변위치의 높이값을 Y축으로 하여 데이터를 2차원평면 상에 표현한 그래프이다. Moran ScatterPlot은 4개의 사분면으로 나누어 서로 다른 유형의 공간패턴을 표시할 수 있고 이를 이용하여 공간군집을 결정할 수 있다.

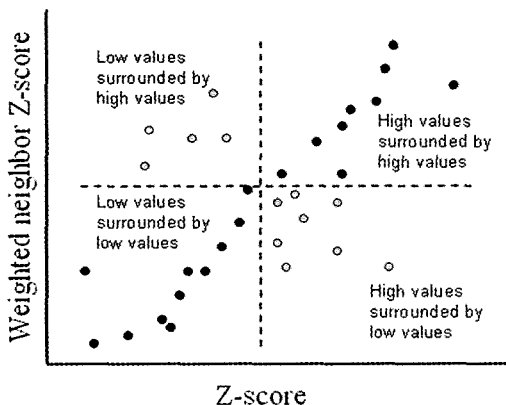


그림 2. Moran scatterplot

일반적으로 라이다 데이터에서 건물과 같은 비지면점의 경우 개별위치의 높이값뿐 만 아니라 주변위치의 높이 값이 높은 특성을 갖고 지면점은 그 반대의 특성을 갖는다. 또한 건물경계부분에서는 주변보다 높거나 낮은 특성을 갖는다. 그림 2에서 볼 수 있듯이, 일반적으로 Moran ScatterPlot 상에서 2사분면에 위치한 값들과 4사분면에 위치한 값들은 공간적으로 서로 상이한 값을 갖는 부적공간상관을 갖는다 할 수 있다. 하지만 1사분면과 3사분면에 위치한 화소들의 공간적 변동특성은 정반대의 경향을 갖는다. 즉, 1사분면에 위치한 값들의 경우 높이가 높은 값들로 둘러싸여있는 특성을 보이고, 3사분면에 위치한 값들은 그 반대의 특성을 갖는다. 따라서 본 연구에서는 Moran ScatterPlot상에서의 1사분면에 위치한 점들을 건물 점으로, 3사분면에 위치한 점들을 지면점으로, 2, 4 사분면에 위치한 점들을 건물경계로 표시한 Moran scatterplot map을 이용하여 건물점을 분리하고 동시에 건물경계를 분류하였다.

4. 실험 적용 및 평가

제안된 방법을 그림 3과 같이 Visual C++을 이용하여 객체지향프로그램으로 구현된 완전 자동화된 시스템으로 구축하였다.

4.1 실험 데이터

본 연구에서 사용한 자료는 대전시 일원으로 2005년 관측된 자료이며, 제안된 알고리즘을 적용할 대상지역은 고층아파트가 주를 이루는 고층아파트 지역과 저층주택

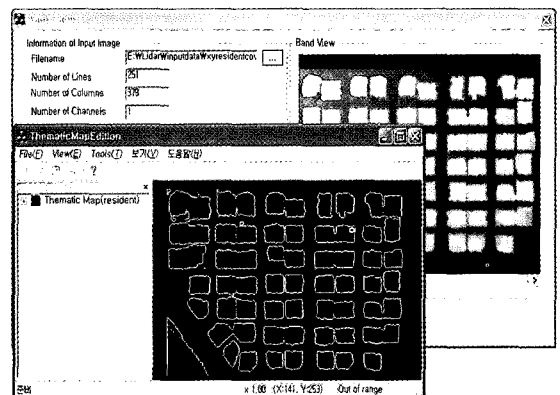


그림 3. 프로그램 구현

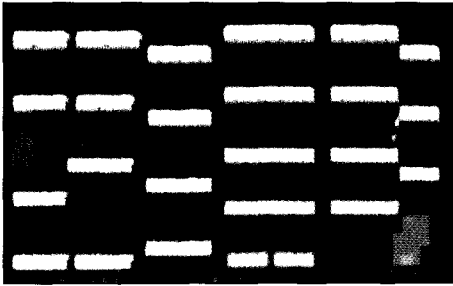


그림 4. 고층아파트 지역의 DSM

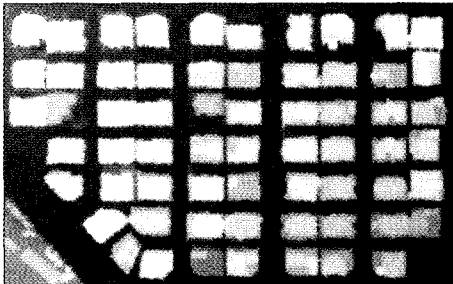


그림 5. 저층주거지역의 DSM

이 밀집한 주거지역을 선정하였다. 다음 그림 4와 5는 두 대상지역의 DSM영상이다.

4.2 공간이상치검출

대상지역의 시각검사 결과 명확한 공간이상점을 찾아 내기가 힘들었다. 따라서 본 연구에서는 제안한 알고리즘의 공간이상점검출의 성능을 알아보기 위해 라이다 원시 자료에 40개의 가상 이상점을 생성하였다.

가상 이상점은 Visual C++의 Random 함수를 사용하여 대상지역의 공간적 범위 내에서 난수를 발생시켜 생성하였다.

x와 y 좌표는 대상지역의 범위 내에서 임의(random)로 발생 시키고 이 점들의 z 좌표는 일정한 범위 내에서 임의로 20개의 high-outlier와 20개의 low-outlier를 생성하였다.

실험결과에는 표 1과 같으며 대상지 1과 대상지2의 경우 각각 총 42개, 41개의 점이 이상점으로 검출되었다. 검출된 점들은 약간의 Commission오차가 존재하였지만 대다수가 이상점으로 부여한 데이터였다. 그림 6의 경우 대상지1의 데이터에 이상점을 부여한 영상이고, 그림 7은 검출된 이상점들을 kernel estimation과정을 통해 보간한 영상이다.

표 1. 이상점 검출 분석 결과

	Site 1	Site 2
탐지된 이상점	42	41
탐지된 이상점 중 실제 이상점	40	40
User accuracy(%)	95.2	97.5
Producer accuracy(%)	100	100

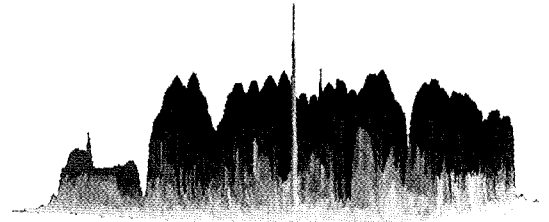


그림 6. 가상 이상점 생성 영상

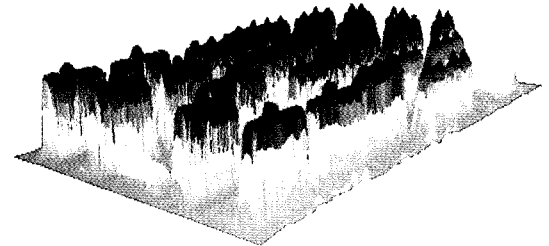


그림 7. 이상점 보간 영상

4.3 분할(segmentation)

본 연구에서 적용한 알고리즘의 경우 Moran scatterplot의 사분면 특성만을 이용하므로, 건물을 추출하기 위한 기존의 다양한 연구에서와 같이 필터의 크기나 경험적 임계값의 설정이 필요치 않다.

그림 8의 항공사진에서 볼 수 있듯이 대상지1의 경우 높이 30m 내외의 고층아파트와 부분적으로 단층건물들이 존재하는 지역이다. 처리결과 그림 10에서 보듯이 고층아파트의 경우 건물과 지면과의 높이차가 명확하기 때문에 다른 지역에 비해 쉽게 추출됨을 알 수 있다. 하지만 부분적으로 존재하는 저층건물의 경우 온전히 분리되지 않음을 확인 할 수 있다.

그림 9에서 알 수 있듯이 대상지 2는 높이 10m 이하의 연립주택들이 밀집한 지역으로, 주택 사이에는 좁은 통로가 존재한다. 처리결과 그림 11에서 보듯이 건물이 잘 추출됨을 볼 수 있으나 건물과 건물의 경계영역의 분리가

명확하게 이루어지지 않았음을 알 수 있다. 이는 래스터 생성을 위한 보간 과정에서 개체의 외곽선이 모호해지는 정보의 왜곡현상으로 기인한 것이다.



그림 8. 대상지 1의 항공사진

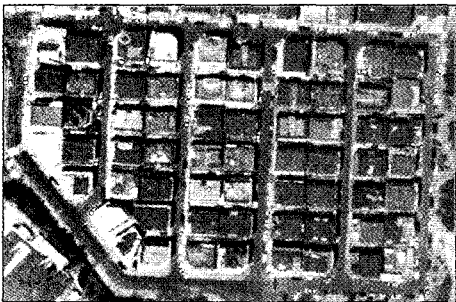


그림 9. 대상지 2의 항공사진

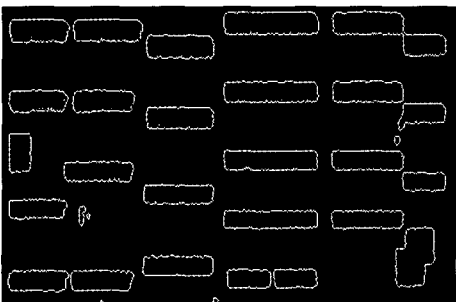


그림 10. 대상지 1의 분할 영상

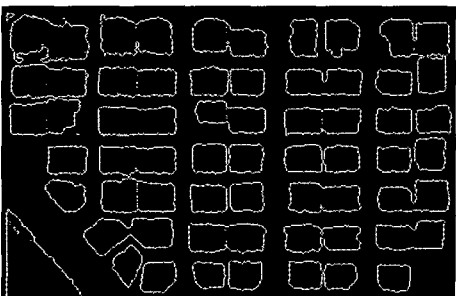


그림 11. 대상지 2의 분할 영상

본 연구에서 제안한 알고리즘을 통해 추출된 건물의 비율을 평가하기 위해서 식 (5)를 사용하였다(Avrabami et al; 2004).

$$\text{Building Extraction Rate} = \frac{BSE + k \times BPE}{BSE + BPE + BNE} \times 100 \quad (5)$$

여기서 BSE(buildings successfully extracted)는 성공적으로 추출된 건물의 수를 의미하고, BPM(buildings partially extracted)와 BNE(buildings not extracted)는 각각 부분적으로 추출된 건물의 수와 추출되지 않은 건물의 수를 의미하고 $k = 0.5$ 을 사용하였다. 대상지 2에 대한 평가에서 두 개의 건물이 붙어서 추출된 것들 중, DSM 데이터 자체에서 건물이 붙어있는 경우는 성공적으로 추출된 것으로 간주하였다. 이는 LiDAR 원시 데이터의 점밀도가 향상되면 해결될 수 있는 문제이기 때문에 본 논문에서 제안한 알고리즘의 성능을 평가하는데 있어서 문제가 없다고 판단된다.

표 2. 건물 추출 결과 정확도 분석

Test Area	Site 1	Site 2
BCE	24	57
BPE	1	6
BNE	2	0
Extraction Rate(%)	90.74	95.23

5. 결 론

본 연구에서는 원시 항공라이다 데이터의 국지적 공간상관특성을 이용하여 건물을 추출하였고 실험을 통하여 다음과 같은 결론을 얻을 수 있었다.

첫째, 라이다 데이터의 국지적 공간상관특성을 이용하여 경험적 변수 및 임계값의 설정 없이도 건물을 추출할 수 있었다.

둘째, 국지적 공간상관지표의 유의성검정을 통하여 라이다 데이터에서의 처리과정에 많은 영향을 미치는 공간이상치를 검출할 수 있었다.

셋째, 본 연구에서 적용한 공간군집과정을 통해 건물 속성정보 추출과 같은 객체인식의 후처리 과정에서 유용하게 활용될 수 있을 것이다.

공간상관분석은 공간구조의 정의 및 가중치행렬에 많은 영향을 받는다. 따라서 향후 다양한 지형의 특성에 따

라 이러한 인자가 분석에 얼마나 많은 영향을 미치는지 고려하는 것도 유용한 연구가 될 것으로 생각된다.

감사의 글

본 연구에 대한 공학연구소의 지원에 감사 드리며, 본 연구의 수행에 필요한 LiDAR 데이터와 항공사진을 제공해주신 (주)한진정보통신에 감사 드립니다.

참고문헌

김형태 (2000), GIS 건물레이어 자동 구축을 위한 LiDAR 데이터와 항공사진의 융합, 박사학위논문, 서울대학교.
이정호, 한수희, 유기운, 김용일, 이병길 (2005), LiDAR 데이터를 이용한 DEM 자동생성기법, 한국지형공간정보학회지, 제 13권, 제4호, pp. 27-32.
좌윤석(2003), 항공레이저 스캐닝 데이터를 이용한 건물 자동추

출에 관한 연구, 석사학위논문, 인하대학교.
한수희, 이정호, 유기운, 김용일, 이병길 (2005), 항공레이저측량 자료의 스캐라인 특성을 활용한 건물 포인트 분리에 관한 연구, 한국지형공간정보학회지, 제13권, 제4호, pp. 33-38.
Anselin, L. (1988), *Spatial Econometrics: Methods and Models*. Boston: Kluwer.
Anselin, L. (1995), Local Indicators of Spatial Association-LISA, *Geographical Analysis* 27, pp. 93-115.
Avrabami, Y, Raizman, Y. and Doytsher, Y. (2004), Semiautomatic 3D Mapping of Buildings From Medium Scale (1:40,000) Aerial Photographs. *International Society for Photogrammetry and Remote Sensing Commission III Istanbul* pp. 96-105.
Rottensteiner F. and Briese C. (2002), A New Method for Building Extraction in Urban Areas from High-Resolution LiDAR Data, *Proceeding of the ISPRS Commission III Symposium*.
Shekhar, S., Lu, C. T., and Zhang, P. (2003), A unified approach to detecting spatial outliers, *GeoInformatica*, Vol. 7, No. 2, pp. 139-166.

(접수일 2006. 9. 11, 심사일 2006. 9. 26, 심사완료일 2006. 10. 23)

