

항공 라이다 데이터로부터 타일 단위의 샘플링을 이용한 효율적인 지형 분류 기법¹⁾

An Efficient Topographical Classification Scheme Using Tile-based Sampling from Airborne LiDAR Data

이성규* · 성철웅 · 박창후 · 이호준 · 김유성

Sung Gyu Lee* · Chulwoong Sung · Chang Hoo Park · Ho Jun Lee ·
Yoo-Sung Kim

인하대학교 정보통신공학과

darksha518@naver.com* · sung333@yahoo.co.kr · longing333@naver.com ·
lhjun2001@hanmail.net · yskim@inha.ac.kr

요 약

본 연구에서는 항공 라이다 데이터로부터 지표의 유형을 분류하는 과정에서 필요한 처리시간을 감소시켜 보다 효율적인 지형 분류가 가능하도록 타일 단위의 샘플링을 이용한 지형 분류 기법을 제안하였다. 또한 실험을 통하여 샘플링 정도에 따라 정확도와 시간효율을 비교분석하여 최적의 샘플링 정도를 제시하였다. 본 연구에서 제안한 타일 단위의 샘플링을 이용한 지형 분류기법을 이용해 자연재해나 도시 구조 변화 등의 지형 정보의 변화를 탐지하기 위한 모니터링 목적으로 활용할 경우에 지형분류의 정확도 손실을 최소화하면서 빠른 처리가 가능하기 때문에 실시간 서비스가 가능해 질 수 있다.

1. 서론

최근 산림의 분포, 자연재해, 해안선의 변화, 도심지 구조의 변경 등을 모니터링 하거나 관광자원으로서 활용 가능한 3차원 도시 모델링을 위해 대량의 지형정보를 고속으로 정확하게 취득할 수 있는 항공 라이다 (LiDAR : Light Detection And Ranging) 시스템이 각광받고 있다.[1] 항공 라이다 시스템은 레이저 스캐너를 이용하여 대량의 3차원의 지형정보를 고속으로 취득할 수 있는 장비로서 레이저의 반사정보를 이용하여 스캐닝 목적지의 정확한 좌표와 레이저광의 반사정도를 기록하여 정확한 지형 정보를 얻을 수 있도록 해준다. 이러한 항공 라이다 시스템을 이용하여 고속으로 대량의 데이터를 취득하는 것이 가능해졌으며 이에 따라 얻어

지는 대량의 점 데이터로부터 의미있는 정보를 추출하기 위해 처리해야하는 소용연산량도 점차 증가하는 추세이다.[1]

항공 라이다 데이터로부터 지형의 유형을 파악하기 위한 분류 방법에는 일반적으로 점 기반 방법[2]이 널리 활용되고 있는데 이는 점들 사이의 인접성, 기울기 등의 속성을 조합하여 분류기준을 수립하고 지표면과 비 지표면을 구분하는 방법이다. 기존에 사용된 점 기반 방법은 한 점을 중심으로 제한된 크기의 윈도우를 설정하고 그 내부에 존재하는 점들로부터 이방성, 평면성, 선형성, 곡률, 경사도 등의 특징을 추출하여 분류기준으로서 활용한다.[2] 하지만 이러한 방법으로 모든 점들을 분류 대상으로서 고려하는 것은 각각의 점들과 설정된 일정 영역내의 주변

1) 본 연구는 국토해양부 첨단도시기술개발사업 지능형 국토정보기술혁신 사업과제의 연구비 지원(07국토정보 C03)에 의해 수행되었습니다.

점들 사이의 중복적인 연산 소요를 발생시켜 연산상의 비효율성을 야기한다는 단점을 가지고 있다.[3]

이를 개선하기 위해 [3]은 항공 라이다의 점과 점들간의 인접성 관계를 그래프(graph) 구조로 모델링하여 저장하고 이를 패치형태로 그룹화하여 지형을 분류하는 방법을 제안하였다. 이러한 방법은 항공 라이다 데이터의 인접점들의 특성을 고려하여 면으로 발전시킴으로써 인접성에 기반하여 오류에 대한 견고성을 향상시킬 수 있으며, 기존 점 단위 방법보다 연산량과 시간적 측면에서 비용 절감 효과가 있다는 장점이 있다.[3] 하지만 이 방법 역시 그래프를 구성하는 과정에서 점 단위의 인접성 판단이 이루어져야 하며, 점의 개수만큼 힙 구조를 생성하여 점 상호간의 거리를 저장하여야 한다는 점에서 분류영역의 크기에 비례하는 연산량과 저장 공간의 비효율성을 야기한다. 이러한 문제점을 해결하기 위해 최근의 연구에서는 지형을 임의 크기의 셀로 분할하여 셀 단위 특징을 사용하여 인식의 정확성 및 처리의 효율성을 개선하려고 하였다.[4,5]

본 연구에서는 점 기반 분류 방법의 정확도를 크게 손상시키지 않으면서 넓은 지역의 데이터를 빠르게 분류 할 수 있도록 타일 단위의 샘플링을 이용한 분류기법을 제안하였다. 타일 단위의 분류 방법은 분류 대상이 되는 지역을 일정한 크기의 타일로 나누고 각 타일 단위로 지형을 잘 구분할 수 있는 특징들을 규정하여 이 특징값을 기준으로 지형의 유형을 분류하기 때문에 각각의 점들에 대한 중복 연산을 피할 수 있어서 처리 시간을 줄일 수 있다.[5] 또한, 타일 내부의 전체 점 데이터를 이용하지 않고 적은 수의 점 데이터만을 이용하여 분류하기 위하여 일정한 비율의 점 데이터들을 랜덤 샘플링하는 기법을 적용하여 처리 시간을 줄일 수 있도록 하였다. 본 연구에서는 제안된 타일 단위의 샘플링을 이용한 분류기법의 시간

적 효율성을 측정하기 위하여 전체 점 데이터를 사용했을 경우와 75%, 50%, 25%, 20%, 10%의 점들만을 샘플링한 경우를 비교 실험하였다. 이와 같은 실험을 통하여 샘플링의 비율이 낮아질수록 전체 점 데이터를 사용하여 분류했을 때와 비교하여 속도는 증가하고 정확도는 낮아짐을 확인할 수 있었다. 따라서 본 연구에서는 지형 분류에 소요되는 시간과 정확도를 이용하여 최적의 샘플링 비율을 제안하였다. 본 논문에서 제안한 타일 단위의 샘플링을 이용한 지형 분류기법을 이용해 자연재해나 도시 구조 변화 등의 모니터링 시스템을 위한 지형 분류를 수행할 경우 정확도의 손실을 최소화하면서 빠른 속도를 유지할 수 있을 것으로 예측된다.

본 논문은 다음과 같이 구성되었다. 2장에서는 기존의 점기반 지형 분류 연구들에 대한 소개와 문제점을 정리하였다. 3장에서는 기존 점기반 지형 분류 연구들의 문제점들을 보완한 타일 단위의 샘플링을 이용한 지형 분류기법을 소개한다. 4장에서는 제안된 타일 단위의 샘플링을 이용한 지형 분류기법의 샘플링 비율에 따른 수행 시간과 정확도에 대한 실험의 결과를 설명하며 5장에서는 본 연구의 결론 및 향후 연구를 소개한다.

2. 관련연구

항공 라이다 데이터로부터 얻을 수 있는 주요 정보는 포인트의 좌표(X축 좌표, Y축 좌표, Z축 좌표)정보와 레이저의 반사회수 및 반사정도등이 포함되어 있다. 이러한 항공 라이다 데이터를 이용하여 지형을 분류하는 방법으로는 개개의 점 단위로 다른 점들과의 인접성을 판단하거나 점들을 그룹화(grouping)하여 처리하는 점 기반 방법과 임의 크기의 셀 혹은 타일 단위로 분할하고 이를 분류 단위로 사용하는 셀, 타일 기반 방법으로 구분할 수 있다.

먼저 점 단위 지형 분류 기법으로서

[2,6]은 각 점을 중심으로 인접성이 수립되는 원통형의 영역을 설정하고 해당 평면내의 공분산 행렬(covariance matrix)의 고유 값(eigen value)을 이용하여 이방성, 평면성, 구형성, 선형성 등의 특징을 정의하고 이를 이용하여 점 단위 분류를 시도하였다. 하지만 이러한 점 기반 분류 방법의 경우 각 점 사이의 인접성을 정의하기 위한 과정에서 점의 개수 혹은 그 제곱에 비례하는 계산량을 필요로 하기 때문에 연산량의 측면에서 비효율성을 초래한다.[3]

이를 개선하기 위한 방법으로 점을 그룹화(grouping)하고 분할(segmentation)하여 군집(cluster)을 생성하고 생성된 군집의 특징정보를 이용하여 지표점과 비지표점을 구분하는 방법이 연구되었다. 특히, [3]은 2차원 영상의 픽셀단위 그룹화 방법에 착안하여 3차원 점 집합을 그래프(graph)형태의 구조로 저장하고 인접성이 수립되는 점들 사이의 거리를 힙 구조로 저장하여 처리속도 면에서 효율성을 향상시키는 방법을 제시하였다. 이후 평면계수를 이용하여 근사오차가 작은 초기패치들을 생성하고 반복확장(iterative growing)이라는 과정을 통해 점차적으로 표면 패치로 성장시켜 최종적으로 95%이상의 지역으로부터 패치들이 성공적으로 분할되었다고 분석하였다.[2] 하지만 이 방법은 초기 점 상호간의 인접성을 파악하여 그래프 모델을 생성하는 과정과 각 점들 사이의 거리를 힙 구조로 정리하는 과정에서 점의 개수에 비례하는 연산량과 저장 공간을 필요로 할 뿐만 아니라 패치와 점 혹은 패치와 패치 사이의 인접성을 파악하고 성장시키는 과정에서 점 단위로 인접성을 판단하여 추가적인 연산 소요를 발생시킨다는 문제점을 가지고 있다.

셀 기반의 지형분류 방법은 연산량이 많고 느린 단점을 가진 점 단위 분류기법에서 연산의 속도를 향상시키기 위한 방법의 일환으로 셀(혹은 타일)로 정의되는

일정 사각형 모양의 영역으로 전체 데이터를 분할하여 처리되는 기법들이 연구되었다. A.S. Antonarakis 등은 분류를 위해 항공 라이다 데이터를 5m 혹은 10m 단위의 셀로 분할하고 지표면을 9가지 영역으로 정의하였다. 이후 분할된 셀로부터 사행도(skewness), 첨예도(kurtosis), 수목의 높이(vegetation height) 등의 여섯가지 특징을 추출하고 이를 경험적 모델을 통해 분류하여 약 94% 정도의 분류 정확도를 얻는 것으로 분석하였다.[4] 또한 본 연구의 선행연구에서는 항공 라이다 데이터를 적정 크기의 타일 단위로 분할하여 지형 분류 모델을 설계하는 효과를 분석하기 위해 전라남도 장성지역의 항공 라이다 데이터를 각각 5m×5m, 10m×10m, 20m×20m로 분할하여 지형 분류 예측 정확도를 비교하여 10m×10m의 경우 약 97%의 가장 높은 지형 분류 정확도를 갖는 것으로 분석한 바 있다.

이러한 타일 기반 방법은 점 기반 방법에 비해 효율적인 지형 분류가 가능하지만 특징정보로서 타일 내 점 상호간의 인접성을 이용하는 경우 연산량의 증가를 불러올 수 있다. 따라서 본 논문에서는 정확도를 크게 손상시키지 않는 범위에서 연산의 효율성을 증대시키기 위해 타일 내부의 데이터를 샘플링하여 활용하는 지형 분류 기법을 제안하였다.

3. 타일 단위 샘플링을 이용한 지형 분류 기법

본 연구에서 활용한 타일 기반 분류 방법의 전체 구조는 그림 1에 표시한 바와 같이 항공 라이다 데이터를 타일 단위로 분할하고 특징을 추출하여 분류모델을 생성하고 분류를 시도하는 과정으로 구분할 수 있다. 먼저 3.1절에서는 학습용의 항공 라이다 데이터를 타일 단위로 분할하고 타일 단위로 지형 분류를 위한 특징값을 추출한 다음 데이터 마이닝 도구인 WEKA를 이용하여 타일 단위의 지형 분

류기를 생성하는 “학습과정”을 설명한다.[7] 3.2절에서는 실제 지형분류를 할 항공 라이다 데이터를 이용하여 타일 단위 분류기를 이용하여 지형 분류를 실시하는 과정을 설명한다.

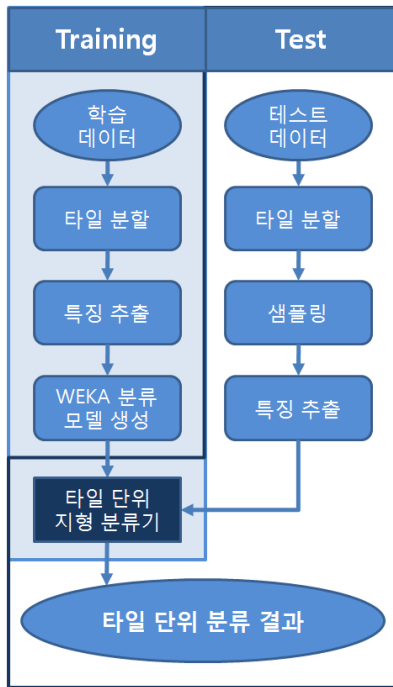


그림 1. 타일단위 지형 분류 방법 개요도

3.1 학습 과정

본 연구에서는 선행연구[5]에서 진행한 바와 같이 타일을 분할하고 분할된 타일로부터 특징정보를 추출하여 분류모델을 생성하는 과정으로 학습이 이루어진다.

타일의 크기를 선정함에 있어서 선행연구[5]를 고려하면 항공 라이다 데이터를 적정 크기의 타일 단위로 분할하여 지형 분류 모델을 설계하는 효과를 분석하기 위해 5m×5m, 10m×10m, 20m×20m의 세 종류로 타일의 크기를 분할하고 정확도를 분석한 결과 10m×10m 타일로 분할하여 지형 분류 모델을 구축하는 경우가 가장 높은 예측 정확도를 갖는 것으로 분

석한 바 있다. 따라서 본 연구에서도 항공 라이다 데이터를 분할함에 있어서 10m×10m의 크기의 타일을 활용하였다.

타일의 크기가 결정되면 해당 타일로부터 지역적 특성정보들을 추출해내는 과정을 거치게 되는데 이를 위해 선행연구[5]를 통해 검증된 32개의 특징정보뿐만 아니라 샘플링 과정에서의 정확도 감소를 완화할 수 있도록 추가로 표1에 제시한 6가지 특징정보를 포함하여 총 38가지 특징정보를 추출하였다.

표 1. 추가 선정 특징정보

특징	설명
Distance_MaxMin_Z	타일 내에서 고도의 최소값과 최대값을 갖는 점들 사이의 거리
Distance_MaxMin_I	타일 내에서 반사값의 최소값과 최대값을 갖는 점들 사이의 거리
return_avrage	반사 횟수의 평균
first_return_z_else_return_z	반사 횟수가 1인 점들의 고도평균에서 리턴넘버가 1이 아닌 점들의 고도평균을 뺀 값
center_of_massX	고도를 무게로 삼아 측정한 타일 무게중심의 X좌표
center_of_massY	고도를 무게로 삼아 측정한 타일 무게중심의 Y좌표

선행연구[5]에서는 WEKA를 이용하여 데이터마이닝의 대표적인 분류기법인 결정트리(decision tree), 규칙 집합(rule set), 베이저안 네트워크(bayesian network), 신경망(neural network)등의 기법을 분석하였다. 그 중 베이저안 네트워크는 조건부 확률에 기반한 통계적 기법을 활용하여 그 정확도가 비교적 높을 뿐만 아니라 타 분류 기법에 비해 모델의 구축 시간 및 분류과정에서 속도가 빠르다는 장점을 가지고 있다. 본 연구에서는 지형 분류기 생성 단계의 효율성 증진을 위해 분류 모델을 베이저안 네트워크로 채택하여 사용하여 타일 단위의 지형 분류기를 생성하

었다.

3.2 테스트 과정

테스트 과정은 학습과정과 대체로 동일한 과정으로 진행된다. 먼저 3.1에서 제시한 바와 같은 이유로 10m×10m 크기의 타일로 분할하여 타일단위의 점 데이터를 확보하고 확보된 점 데이터를 이용하여 샘플링 과정을 진행한다. 하지만 샘플링 과정에서 의미있는 데이터를 찾아내기 위해 타일 내 연산을 수행한다면 추가적인 연산 소요가 발생하며, 이 경우 오히려 샘플링 과정으로 인한 연산상의 비효율성을 초래할 수 있다. 따라서 본 연구에서는 무작위 샘플링(random sampling)기법을 활용하여 추가적인 연산소요를 배제하고 샘플링에 따른 연산량과 정확도의 관계만을 파악하였다. 일반적으로 무작위 샘플링을 거친 타일단위의 데이터는 원본 데이터로부터 샘플링 비율(sampling rate)에 비례하는 정보의 손실이 발생하며 이에 따른 정확도의 하락이 발생하게 된다. 샘플링 비율을 낮게 유지하면서 타일의 정확한 지형 분류가 가능하도록 하기 위한 샘플링 방법은 향후 제시될 예정이다.

샘플링과정을 거친 데이터는 학습 과정에서 기 구축된 지형분류기를 통해 분류 정확도와 소요 시간을 평가하게 되는데 이는 4절 성능평가 부분에서 자세히 다루었다.

4. 성능 평가

본 절에서는 2008년 11월에 ALTM Gemini 167을 이용하여 전라남도 장성을 점밀도 5-6pt/m2으로 촬영한 데이터를 학습 데이터로 이용하여 샘플링의 비율에 따른 지형분류의 수행시간과 정확성 예측을 분석하였다. 본 연구에서는 지형을 수목, 도로, 초지, 수역, 건물의 순수타일 5종류와 숲과 초지, 숲과 수역, 숲과 도로, 건물과 초지, 건물과 도로, 건물과 숲, 도로와 초지, 초지와 수역으로 구성되어

있는 복합타일 8 종류로 총 13종류의 지역을 10m×10m크기의 타일로 생성하여 학습데이터로 이용하였다.

성능 평가를 위한 샘플링의 비율은 10m×10m크기의 타일에 포함된 전체 점 데이터를 100%로 정하고 점 데이터를 75%, 50%, 25%, 20%, 15%, 10%로 샘플링 하였다.

샘플링의 비율에 따른 시간 절약 성능을 평가하기 위하여 10m×10m 크기의 타일에 포함되어 있는 점들을 각각의 비율만큼 랜덤 샘플링 기법을 이용하여 분류 모델 적용에 필요한 38개의 특징들을 추출하기 위해 소요되는 시간의 평균을 측정하여 표 2와 같은 결과를 얻었다.

표 2. 샘플링 비율에 따른 소요 시간

샘플링 비율(%)	100	75	50	25	20	15	10
시간(ms)	31.7	19.3	9.3	3.1	2.2	1.6	0.9

표 2에 따르면 각각의 샘플링 비율에 따라 10m×10m타일 하나당 소요되는 시간으로 샘플링 비율이 낮아질수록 특징들을 계산하는데 소요되는 시간이 줄어드는 것으로 분석되었다. 이는, 샘플링의 비율이 낮아질수록 특징들을 계산하는데 사용되는 점 데이터의 숫자가 줄어들기 때문으로 분석되었다.

샘플링의 비율에 따른 정확도를 판단하기 위하여 WEKA에서 제공하고 있는 대표적인 분류 기법 중 하나인 베이지안 네트워크를 이용하였다. 10m×10m크기의 타일 내부의 전체 점 데이터를 이용하여 분류모델을 생성하고 각각의 샘플링 비율에 따라서 랜덤하게 선택된 점 데이터들을 이용하여 예측 분류 정확률을 측정하여 표 3와 같은 결과를 얻었다.

표 3. 샘플링 비율에 따른 예측 정확률

샘플링 비율(%)	100	75	50	25	20	15	10
예측 정확률(%)	90.0	88.0	85.4	80.4	78.5	75.1	69.1

표 3에 따르면 각각의 샘플링 비율이 낮아질수록 예측 정확률이 낮아지는 것으로 분석되었다. 이는, 샘플링 비율에 따라 상대적으로 적은 수의 점 데이터를 사용하기 때문에 타일의 특징이 정확하게 계산되지 못하여 오분류된 것으로 분석되었다.

본 연구에서는 타일 단위로 지형을 분류할 때 최적의 샘플링 비율을 제안하기 위하여 샘플링 비율에 따른 소요시간과 예측 에러율을 이용하여 그림 2의 그래프를 얻었다. 그림2에서 분류 에러율은 베이지안 네트워크를 이용하여 본 실험을 위해 테스트 데이터로서 확보된 13,176개의 타일을 분류했을 때 오분류된 에러 비율을 의미하며 특징 추출시간은 지형 분류를 위해 필요한 38개의 특징 값을 추출하기 위해 소요되는 평균 시간을 의미한다.

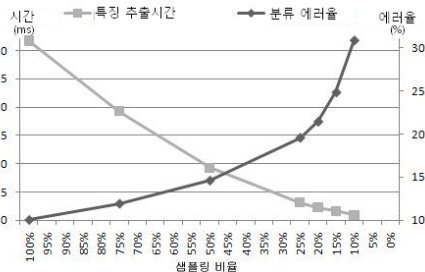


그림 2. 샘플링 비율에 따른 소요시간과 에러율

그림 2에 따르면 타일 단위로 지형을 분류하기 위해서는 40~50% 정도로 점 데이터를 샘플링을 할 때 최적의 효율을 갖는 것으로 분석되었으며, (점밀도 : 2~3 point/m²) 이때 각각의 타일은 85% 정도의

예측 정확률로 타일 당 약 9ms 정도의 시간을 소요하여 분류할 수 있는 것으로 분석되었다.

5. 결론 및 향후연구

본 연구에서는 항공 라이다 데이터로부터 지형을 분류하는 과정에서 발생하는 연산량의 줄이고 빠르게 분류를 수행할 수 있도록 샘플링을 이용한 타일 단위 지형 분류 기법을 제안하였다. 이를 이용하여 기존 항공 라이다 데이터 분류에 소요되는 연산량을 감소시켜 분류 속도 증진에 기여할 수 있을 것으로 기대된다.

향후연구로는 연산량에 과도한 부하를 주지 않는 범위에서 새로운 샘플링 기법을 적용하여 샘플링에 따른 정확도를 보정하고, 점 기반 접근방법과 같이 윤곽선을 추출해 보다 실제에 가까운 지형정보를 얻어내기 위해 복합타일에 대한 경계선 추출 알고리즘을 개발할 예정이다.

참고문헌

- [1] 서용철, 최운수, 허민, “항공 레이저 측량 기초의 응용”, 대한측량협회
- [2] Gross, H. and Thoennessen, U., “Extraction of lines from laser point clouds”, ISPRS, Vol. 36 Part 3A, pp. 87-91, 2006.
- [3] 이임평, “항공 라이다 데이터의 분할 : 점에서 패치로”, 한국측량학회, 제 24권 제 1호, pp. 111~121, 2006.
- [4] A.S. Antonarakis, K.S. Richards, J. Brasington, “Object-based land cover classification using airborne LiDAR”, Remote Sensing of Environment, Vol.112, pp.2988-2998, 2008.
- [5] 이성규, 이호준, 성철웅, 박창후, 조우석, 김유성, “항공 라이다 데이터로부터 데이터마이닝 도구 WEKA를 이용한 지형 분류기 제작 지원 시스템”, 한국측량학회, 제 28권 제 1호, pp133-142, 2010.

- [6] Nesrine Chehata, Li Guo, Clement
mallet, "Airborne lidar feature
selection for urban classification using
random forests", IAPRS, vol. XXXVIII,
pp.207-212, Sep. 2009.
- [7] Witten, I. H. and Frank, E., "Data
Mining : Practical Machine Learning
Tools and Technique"(2E), Elsevier,
2005.
- [8] 이임평, "라이다 데이터로부터
지표점 추출을 위한 피쳐 기반
방법", 대한 원격탐사학회, 제 22권,
제 4호, pp.265-274, 2006.
- [9] 최연웅, 조기성, "Entropy 이론을
이용한 라이다 자료 분류기법 연구",
대한 토목학회, 제 27권 2D호, pp.
225-232, 2007.
- [10] P. Charaniya, Roberto Manduchi, and
Suresh K. Lodha, "Supervised
Parametric Classification of Aerial
LiDAR Data", IEEE Computer
Society, Conference Paper, 2004.
- [11] M. Morgan, K. Tempfli, "Automatic
building extraction from airborne
laser scanning data", In
International Archives of
Photogrammetry and Remote
Sensing, Vol. XXXIII, B3, Amsterdam,
2000.
- [12] Song, J.-H., Han, S.-H, Yu, K., Kim,
Y.-I., "Assessing the possibility of
land-cover classification using lidar
intensity data", International
Archives of Photogrammetry,
Remote Sensing and Spatial
Information Sciences 34 (Part 3B),
259-262, 2002