韓國地形空間情報學會誌 第19卷 第4號 2011年 12月 pp. 139-144

연구논문

요인 및 군집분석을 이용한 지상 라이다 자료의 분류
Classification of Terrestrial LiDAR Data Using Factor and Cluster Analysis

최승필* · 조지현** · 김 열*** · 김준성****
Choi, Seung Pil · Cho, Ji Hyun · Kim, Yeol · Kim, Jun Seong

要 旨

본 연구는 지상라이다 자료에서 얻어진 색상정보(R, G, B)와 반사강도정보(I)를 동시에 이용하여 이를 통계학적 분류기법으로 서로의 연관성을 분석하여 라이다 자료에 대한 분류방법을 제시하였다.
이를 위하여 우선 변수 R,G,B 및 I를 사용하여 분산 을 극대화하는 요인을 추출하여 주요인과 각 변수들 간의 요인행렬을 산출하였다.
그러나 요인행렬은 기초자료를 축소시켜 보여주기는 하지만, 이로부터 어떤 변수들이 어떤 요인에 의해 높게 관계되는지 명확하게 알기 어렵기 때문에 직각회전방식 중에서 Varimax방법을 이용하여 회전된 요인행렬을 구하여 요인점수를 산출하였다. 그리고 비 계층적 군집화 방법인 K-평균법을 이용하여 요인분석으로 산출된 요인점수에 대하여 군집분석을 실시한 후, 지상라이다 자료의 분류 정확도를 평가하였다.

핵심용어 : 라이다 자료, 요인분석, 군집분석, 색상정보, 반사강도 정보

Abstract

This study proposed a classification method of LIDAR data by using simultaneously the color information (R, G, B) and reflection intensity information (I) obtained from terrestrial LIDAR and by analyzing the association between these data through the use of statistical classification methods.
To this end, first, the factors that maximize variance were calculated using the variables, R, G, B, and I, whereby the factor matrix between the principal factor and each variable was calculated.
However, although the factor matrix shows basic data by reducing them, it is difficult to know clearly which variables become highly associated by which factors; therefore, Varimax method from orthogonal rotation was used to obtain the factor matrix and then the factor scores were calculated. And, by using a non-hierarchical clustering method, K-mean method, a cluster analysis was performed on the factor scores obtained via K-mean method as factor analysis, and afterwards the classification accuracy of the terrestrial LiDAR data was evaluated.

Keywords : LiDAR Data, Factor Analysis, Cluster Analysis, Color Information, Reflection Intensity Information

1. 서 론

최근 라이다(Light Detection And Ranging)의 장비 성능과 측정 정확도의 향상, 자료처리 및 활용 등을 위한 꾸준한 기술개발로 인하여 현재 고밀도, 고정확도의 원시자료를 취득 할 수 있게 되면서 토목, 건축, 문화재, 지리정보시스템 등 다양한 분야에서 응용 및 활용이 되고 있다(김종석 등, 2008).

특히, 지상라이다는 측정 대상물의 표면에 조밀한 간격으로 무수히 많은 레이저 펄스를 주사하여 얻은 원시자료를 이용해 대상물의 형태를 직접 취득하는 첨단 레이저 측량장비이다. 또한 항공라이다에 비해 보다 높은

2011년 11월 10일 접수, 2011년 12월 9일 채택
* 정회원 · 관동대학교 공과대학 토목공학과 교수(spchoi@kwandong.ac.kr)
** (주)지트 GIS사업부 부장(jihyun7@gitt.com)
*** C3d 과장(kim-yeol@hanmail.net)
**** 교신저자 · 학생회원 · 관동대학교 토목공학과 석사과정(agogogo1@nate.com)

정밀도로 대상물의 원시자료를 획득할 수 있으며, 항공라이다에서 관측되기 어려운 사면지형, 구조물측량, 현환측량, 문화재측량, 해안선 변화측량, 체적측량 및 터널측량 등에 활용도가 높아지고 있다(조형식 등, 2008).

그리고 이 라이다의 원시자료는 3차원 위치정보(X, Y, Z)뿐만 아니라 내장된 카메라를 통하여 색상정보(R, G, B)와 레이저 펄스에 의해 대상물에 반사되어 돌아오는 반사강도정보(I)를 포함하고 있다. 그러나 지상라이다로부터 취득되는 원시자료는 고밀도, 고정확도로 신뢰성이 높지만, 특정 대상을 관측할 경우 주변의 식생 및 구조물과 같은 비지면 등 여러 장애물이 원시자료에 내포되어 있고 이를 그대로 활용하기에 어려움이 있으므로 지면과 비지면을 분류하여 이 비지면을 제거하기 위한 작업이 불가피한 실정이다(이정호 등, 2011, 國土地理院, 2001, 國土地理院, 2003, 日本測量調査技術協会, 2008).

현재 라이다 원시자료의 응용 및 활용을 위하여 지면과 비지면의 분류와 관련된 다양한 연구가 활발하게 진행되고 있는 가운데 대표적인 방법으로는 Morphologi, Maximum Local Slope, Elevation Threshold Expanding Window방법 등이 있다(김남운 등, 2009).

그러나 이러한 방법들은 지형적인 환경, 대상물의 형상 및 재질 등에 따라 분류정확도에 차이가 발생된다.

따라서, 본 연구는 지상라이다 자료의 색상정보(R, G, B)와 반사강도정보(I)를 동시에 이용하여 이를 통계학적인 기법으로 연관성을 분석하여 지상라이다 자료의 분류방법을 제시하고자 한다.

2. 연구방법

2.1 자료취득

본 연구에서는 라이다 자료에 대한 분류의 신뢰도를 평가하기 위하여 강원도 강릉시 강동면 통일공원 진입로 우측에 위치한 사면지형에 대하여 정면 30m떨어진 곳에 지상라이다 장비를 설치하여 1cm간격으로 스캔하였다.

그리고 스캔된 원시자료 중에서 식생, 지면 및 아스팔트 등 3개 영역에서 각 5,000포인트 등 총15,000포인트를 추출하여 이 포인트에 대한 3차원 정보(X, Y, Z), 색상정보(R, G, B) 및 반사강도정보(I)를 취득하였다. 그림 1은 지상라이다 장비로 관측하는 모습과 이때 얻어진 원시자료를 나타 낸 것이다.

표 1은 원시자료 3개 영역에서 추출된 3차원 정보(X, Y, Z), 반사강도 정보(I) 및 색상정보(R, G, B)를 나타낸 것이다.

2.2 연구절차

본 연구는 지상라이다 자료에서 얻어진 색상정보(R, G, B)와 반사강도정보(I)를 동시에 이용하여 이를 요인 및 군집분석 방법으로 서로의 연관성을 분석하여 라이다 자료에 대한 분류방법을 제시 하였다.

이를 위하여 우선 변수 R, G, B 및 I를 사용하여 분산을 극대화하는 요인을 추출하고, 이것의 고유값으로 Scree도표를 작성하여 요인의 수를 결정 하였고, 주요인과 각 변수들간의 요인행렬을 산출하였다.

그러나 요인행렬은 기초자료를 축소시켜 보여주기는 하지만, 이로부터 어떤 변수들이 어떤 요인에 의해 높게 관계되는지 명확하게 알기 어렵기 때문에 추출된 요인축을 회전하면 어떤 변수가 어떤 요인에서 높게 나타나는지 알 수 있게 된다. 이를 위하여 본 연구에서는 직각회전방식 중에서 Varimax방법을 이용하여 회전된 요인행렬을 구하여 요인 점수를 산출하였다(박광배, 2000, 고왕경, 2011). 그리고 비 계층적 군집화 방법인 K-평균법을 이용하여 요인분석으로 산출된 요인점수에 대하여 군집분석을 실시하였다(허명회, 2010, 고왕경, 2011). 즉, 요인분석에서 추출된 RGB요인과 반사강도

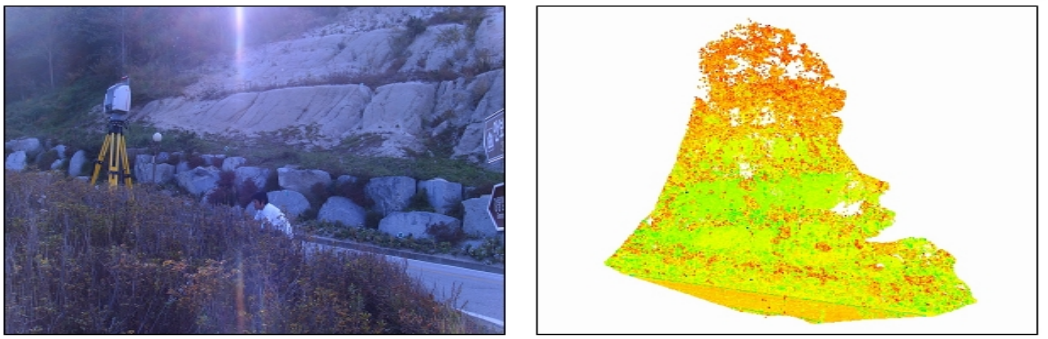

그림 1. 지상 LiDAR 관측전경 및 스캔된 원시자료

표 1. 영역별 포인트 자료

영역	포인트	3차원 정보			반사강도 정보	색상정보		
		X	Y	Z	I	R	G	B
식생	1	18.83543	6.80571	5.80519	0.31344	20	26	26
	2	18.81273	6.79750	5.80724	0.34128	20	26	26
	3	18.79866	6.79237	5.81688	0.25264	19	25	25
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	4998	20.07283	5.74501	5.87062	0.36484	17	18	23
	4999	20.01122	5.72740	5.86195	0.42771	18	19	24
	5000	19.85463	5.68260	5.82988	0.47208	20	21	26
지면	5001	18.85512	6.75078	4.57460	0.37009	36	39	30
	5002	18.86357	6.75377	4.58620	0.38553	37	40	31
	5003	18.87355	6.75734	4.59801	0.36527	39	42	33
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	9998	19.37215	5.54527	5.30858	0.38956	46	50	49
	9999	19.37825	5.54701	5.31981	0.34562	47	51	50
	10000	19.40077	5.55342	5.33571	0.37803	47	51	50
아스팔트	10001	8.42265	2.59093	-0.80153	0.24465	38	47	41
	10002	8.41808	2.59419	-0.80370	0.25203	39	48	42
	10003	8.41777	2.57549	-0.80690	0.27218	39	47	42
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	14998	7.82082	2.20482	-0.90611	0.25942	39	47	43
	14999	7.82076	1.67207	-0.98171	0.26046	42	50	44
	15000	7.82063	1.49529	-1.00993	0.25240	42	49	43

요인을 활용하여 군집분석으로 각 데이터의 유사성을 분석하여 성질이 유사한 데이터를 하나로 묶어내어서 자료를 분류한 후 정확도를 평가하였다.

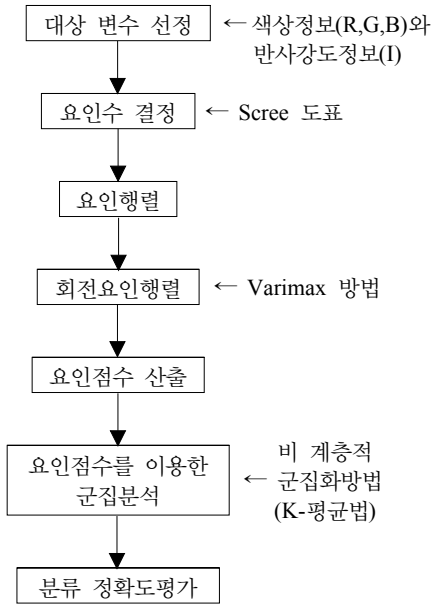

그림 2. 연구절차도

라이다 자료의 분류를 위한 전체적인 연구절차는 그림 2와 같다.

3. 통계분석기법 적용

3.1 요인분석

색상정보(R, G, B)와 반사강도정보(I)를 동시에 이용하여 요인분석을 실시하였다.

이를 위하여 4개 변수(R, G, B, 반사강도)로부터 추출된 각 요인의 초기 고유값, 추출된 요인 적재값 및 회전된 요인 적재값을 표 2에 나타내었다. 회전된 요인1(색상정보요인)에서 분산값 72.641%, 요인2(반사강도 요인)에서 분산값 26.129%로 총 누적 분산값은 98.770%로 나타났다. 고유값은 그 요인의 분산량을 나타내는 것으로 값이 클수록 중요한 요인이라 할 수 있다.

요인분석 시 최적의 요인수를 결정하기 위하여 그림 3과 같은 Scree도표를 작성하였다.

이 Scree도표를 이용하여 고유값이 줄어드는 정도를 기준으로 요인수를 결정 할 수 있다.

즉, 요인이 2에서 3으로 증가하면 고유값의 감소폭이 둔화되기 때문에 요인의 수를 2로 결정하였다.

표 3은 4개 변수에 대한 색상정보(R, G, B)와 반사강도정보(I)의 초기 요인행렬을 나타낸 것이다.

표 2. 요인 적재값

요인	초기 고유값			추출 적재값			회전 적재값		
	전체	분산 (%)	누적 (%)	전체	분산	누적 (%)	전체	분산 (%)	누적 (%)
1	3.056	76.394	76.394	3.056	76.394	76.394	2.906	72.641	72.641
2	0.895	22.376	98.770	0.895	22.376	98.770	1.045	26.129	98.770
3	0.040	1.012	99.783						
4	0.009	0.217	100.000						

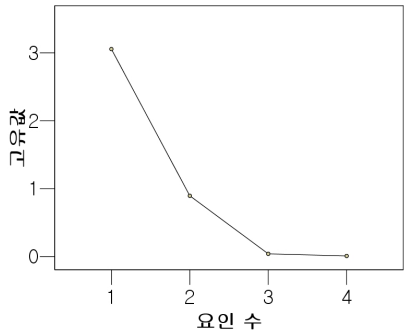

그림 3. Scree 도표

요인행렬은 색상정보(R, G, B)와 반사강도정보(I)를 축소시켜 보여 주기는 하지만, 이로부터 어떤 변수들이 어떤 요인에 의해 높게 관계되는지 명확하지 않기 때문에 Varimax방법으로 3회 반복 회전하여 얻어진 요인행

표 3. 요인행렬

요인 / 변수	색상정보 (R, G, B)	반사강도정보 (I)
I	-0.395	0.919
R	0.978	0.177
G	0.992	0.073
B	0.979	0.120
Cumulative Variance	76.394%	22.376%

표 4. 회전요인행렬

요인 / 변수	색상정보 (R, G, B)	반사강도정보 (I)
I	-0.139	0.990
R	0.990	-0.087
G	0.976	-0.192
B	0.976	-0.142
Cumulative Variance	72.641%	26.129%

표 5. 영역별 요인점수

영역	포인트	색상정보(R,G,B)요인	반사강도(I) 요인
지면	1	-1.06788	-0.40759
	2	-1.00203	0.05473
	3	-1.30011	-1.42948
	⋮	⋮	⋮
	4998	-1.33030	0.41070
	4999	-1.09316	1.46704
	5000	-0.81125	2.22895
식생	5001	-0.02654	0.67919
	5002	0.09846	0.94817
	5003	0.22753	0.63694
	⋮	⋮	⋮
	9998	1.25611	1.17069
	9999	1.24069	0.45369
	10000	1.31733	0.99177
아스팔트	10001	0.32405	-1.33865
	10002	0.43000	-1.20344
	10003	0.45392	-0.86852
	⋮	⋮	⋮
	14998	0.46021	-1.07554
	14999	0.65524	-1.03003
	15000	0.57604	-1.16802

렬을 표 4에 나타내었다.

표 4에서 보는 바와 같이 초기 요인행렬에 비하여 요인구조가 매우 뚜렷하게 나타나고 있음을 알 수 있다. 색상정보 요인에는 R, G, B값이, 반사강도정보 요인에는 반사강도값이 높게 적재되어 있으므로 색상정보 요인은 R, G, B값을, 반사강도정보 요인은 반사강도값을 잘 나타내고 있다.

4개 변수를 요인분석에 의하여 2개의 변수로 축소시킨 색상정보(R, G, B)요인, 반사강도(I)요인에 대하여 회전된 요인행렬과 변수를 이용하여 산출된 요인점수를 영역별로 표 5에 나타내었다.

색상정보(R, G, B)요인을 X축, 반사강도 요인(I)을 Y축으로 하여 15,000포인트에 대하여 요인점수 분포도를 그림 4에 나타내었다.

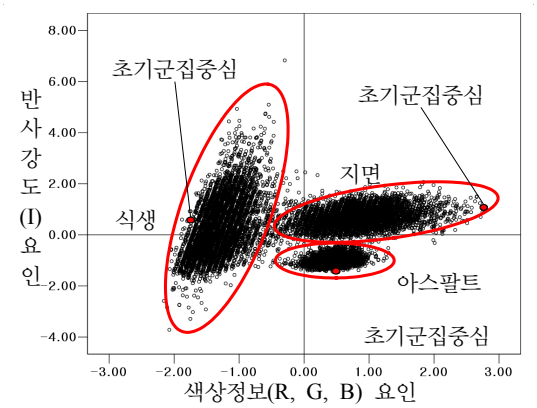

그림 4. 요인점수 분포도

3.2 군집분석

요인분석에서 추출된 색상정보(R, G, B)요인과 반사강도 요인(I)의 요인점수를 이용하여 군집분석을 실시하였다.

비 계층적 군집화 방법인 K-평균법에 의하여 요인점수를 3개의 군집으로 나누어 이 군집의 중심영역을 선정하여 라이다자료를 분류하였다.

이를 위하여 요인점수 분포도에서 식생의 196번 포인트, 지면의 9650번 포인트, 아스팔트의 10175번 포인트를 각각 1, 2, 3 군집의 초기 군집중심으로 선정하여 표 6에 나타내었다.

표 6. 초기군집중심

군집 / 요인	1(식생)	2(지면)	3(아스팔트)
색상정보(RGB) 요인	-1.69721	2.75717	0.53001
반사강도(I) 요인	0.65060	0.98218	-1.44529

표 7. 영역별로 분류된 군집

영역	포인트	소속 군집	분류 군집	영역	포인트	소속 군집	분류 군집	영역	포인트	소속 군집	분류 군집
식생	1	1	1	지면	5001	2	2	아스팔트	10001	3	3
	2	1	1		5002	2	2		10002	3	3
	3	1	1		5003	2	2		10003	3	3
	4	1	1		5004	2	2		10004	3	3
	5	1	1		5005	2	2		10005	3	3
	⋮	⋮	⋮		5006	2	2		10006	3	3
	38	1	2		5007	2	2		10007	3	3
	⋮	⋮	⋮		5008	2	2		10008	3	3
	147	1	3		⋮	⋮	⋮		10009	3	3
	⋮	⋮	⋮		5049	2	1		⋮	⋮	⋮
	4401	1	3		5050	2	1		12115	3	2
	⋮	⋮	⋮		⋮	⋮	⋮		⋮	⋮	⋮
	4428	1	2		5140	2	3		14992	3	3
	⋮	⋮	⋮		⋮	⋮	⋮		14993	3	3
	4994	1	1		5203	2	3		14994	3	3
	4995	1	1		⋮	⋮	⋮		14995	3	3
	4996	1	1		9324	2	3		14996	3	3
	4997	1	1		⋮	⋮	⋮		14997	3	3
	4998	1	1		9998	2	2		14998	3	3
	4999	1	1		9999	2	2		14999	3	3
	5000	1	1		10000	2	2		15000	3	3

표 8. 최종 군집중심

군집 / 요인	1(식생)	2(지면)	3(아스팔트)
색상정보(RGB) 요인	-1.28580	0.84439	0.43647
반사강도(I) 요인	0.26869	0.75492	-0.98618

각 영역들이 속한 군집에 대하여 해당군집의 최종군집중심에서 각 포인트 간의 거리를 이용하여 분류된 군집을 표 7에 나타내었다.

요인분석을 통하여 분류하기 어려웠던 부분들을 분집분석을 통하여 보다 명확하게 분류할 수 있었다.

그리고 군집화 된 영역들의 평균을 최종 군집중심으로 선정하여 표 8에 나타내었다.

3.3 분류 정확도

소속된 군집을 기준으로 요인분석과 군집분석에 의해서 분류된 식생, 지면 및 아스팔트에 대하여 분류정확도를 평가하여 표 9에 나타내었다.

식생 5,000포인트, 지면 5,000포인트 및 아스팔트 5,000포인트가 소속된 군집에 대하여 요인과 분집분석을 이용하여 식생 4,850포인트, 지면 4,813포인트, 아스팔트 4,999포인트로 분류되었다.

따라서 식생의 경우 97%, 지면의 경우 96.26%, 아스팔트의 경우 99.98%의 매우 높은 분류 정확도를 나타내었다.

특히, 아스팔트 영역이 식생이나 지면보다 분류 정확도가 높게 나타난 것은 선정한 곳이 매끈하고, 평편한 지역이기 때문이라고 생각된다.

4. 결 론

사면지형에서 지상라이다로 측정하여 얻어진 식생, 지면 및 아스팔트에 대한 원시자료를 통계학적 분석기법을 이용하여 분류한 자료를 평가하여 다음과 같은 결과를 얻을 수 있었다.

첫째, 지상라이다로 측정된 사면지형의 원시자료에 대하여 식생정보와 반사강도정보를 이용하여 97.75%의 높은 정확도로 식생, 지면 및 아스팔트 영역을 분류할 수 있었다.

둘째, 복잡한 지형요소가 많이 포함된 대상지일 경우 지형특성에 따라 서로 다른 결과가 나타 날 수 있을 것으로 생각되어 향후 보다 다양한 지형을 선정하여 측정된 지상라이다 자료에 대한 분류가 연구 되어야 할 것으로 생각된다.

참고문헌

1. 고왕경, 2011, *SPSS 단계별 입력에 의한 사회과학 데이터 분석*, 경문사, pp.261-348.
2. 김남운, 노이주, 정경훈, 김기두, 2009, “색상 정보와 호프변환을 이용한 3차원 점군데이터 구조물 추출 기법 연구”, *전자공학회 논문집*, 제46권, 제3호, pp.379-387.
3. 김종석, 손홍규, 김성삼, 조형식, 정성수, 2008, “지상 LiDAR 측정각도에 따른 오차분석”, *한국지형공간정보학회 공동추계학술대회* 2008.10, pp.324-328.
4. 박광배, 2000, *다변량분석*, 학지사, pp.87-124.
5. 이정호, 염준호, 김용일, 2011, “스캔 라인 기반의 인접 관계를 이용한 항공레이저측량 자료의 필터링”, *한국측량학회지*, 제29권, 제4호, pp.359-365.
6. 조형식, 손흥규, 김종석, 이석군, 2008, “지상 LiDAR를 이용한 터널의 Revers Engineering”, *대한토목학회논문집, 제28권, 제6D호*, pp.931-936.
7. 허명회, 2010, *사회과학을 위한 다변량자료 분석*, 자유아카데미, pp.112-128.
5. 國土地理院, 2003, “地上型スキヤソ式レーザ 測距儀による斜面地形計測 · 解析技術に関する研究作業報告書”, 日本測量調査技術協會, pp.55-74.
8. 國土地理院, 2001, “地上型スキヤソ式レーザ 測距儀による斜面地形計測 · 解析技術に関する研究作業報告書”, 日本測量調査技術協會, pp.105-106.
9. 財團法人 日本測量調査技術協会, 2008, “圖解 航空レーザ 計測”

표 9. 정확도 평가 (단위 : 포인트)

소속군집 / 분류된 군집	1(식생)	2(지면)	3(아스팔트)	계
1(식생)	4,850(97.00%)	116	0	4,966
2(지면)	103	4,813(96.26%)	1	4,917
3(아스팔트)	47	71	4,999(99.98%)	5,117
계	5,000	5,000	5,000	15,000(97.75%)