

Star Coordinates 기법을 이용한 원격탐사 데이터의 다차원 시각화 제안 및 클러스터링

Multi-Dimension Visualization Proposition and Clustering of Remote Sensing Data Using Star Coordinates Technique

김대성¹⁾ · 김용일²⁾ · 유기윤³⁾

Kim, Dae Sung · Kim, Yong Il · Yu, Ki Yun

¹⁾ 서울대학교 대학원 지구환경시스템공학부 박사과정(E-mail: mutul94@empal.com)

²⁾ 서울대학교 공과대학 지구환경시스템공학부 부교수(E-mail: yik@snu.ac.kr)

³⁾ 서울대학교 공과대학 지구환경시스템공학부 조교수(E-mail: kiyun@snu.ac.kr)

Abstract

단 밴드 영상과는 달리 다차원 데이터는 분광적인 특성을 이용한 자동화된 영상 분석을 수행하는 장점이 있는 반면, 3차원 이상의 데이터를 분광차원 상에 시각화 하는데 어려움이 따른다. 클러스터링 기법을 이용한 영상 정보 추출은 자동화된 영상 분석적인 측면에서 중요한 분야 중 하나로서, 분광차원에서 구 형태의 조밀한 클러스터를 분리하는데 효과적인 방법으로 알려져 있지만 부정형(不定形)의 클러스터를 추출하는 방법에는 한계를 가진다. 따라서 본 연구는 모든 차원의 데이터를 2차원 상에 시각화하여 화소간 인접성을 개략적으로 확인할 수 있는 Star Coordinates 기법을 제안한다. 데이터의 다차원 시각화를 통해, 부정형 클러스터를 제거하여 다음 단계의 영상 분석 시 발생할 수 있는 오류를 방지할 수 있고, 명확한 클러스터를 확인·지정하여 클러스터링 정확도를 높일 수 있을 것으로 기대된다. 부가적인 연구로서, Star Coordinates 기법을 적용하여 Plot된 영상 데이터를 K-Means 알고리즘을 이용한 무감독 분류를 수행하여 그 결과를 확인하였다.

1. 서론

다중분광 또는 하이퍼스펙트럴 원격탐사(Multispectral and Hyperspectral Remote Sensing)는 여러 개 또는 수백 개의 전자기파 파장대를 이용하여 대상체나 대상지역으로부터 반사(Reflected), 방사(Emitted) 및 산란(Scattered)된 에너지를 수집하는 것을 말한다(Jensen, 2000). 취득된 데이터의 밴드는 하나의 차원으로 간주되어 해석되며, 자동화된 영상 분석이 용이하게 된다. 하지만, 3차원 이상의 데이터를 시각적으로 분석하는 것은 한계가 있으며, 분광차원에서 군집화 기법을 이용한 분석에 있어 데이터의 특성을 시각적으로 파악하는 것은 다음

의 영상해석 단계를 위해 매우 유용한 기법이다. 또한 영상 분류를 위한 트레이닝 정보의 추출시 도움을 줄 수 있을 것이며, 실제 원격탐사 데이터 분석 프로그램에서 이런 다차원 시각화 기법을 지원하고 있다. 본 연구는 다중분광 데이터를 2차원 상에 투영하여 다차원 데이터를 시각화할 수 있는 Star Coordinates 기법을 제안하고 있다. Star Coordinates 기법을 적용하여 Plot된 영상 데이터를 K-Means 알고리즘을 이용한 무감독 분류를 수행함으로써 2차원 상에 투영된 데이터가 선형적인 특징을 나타냄을 보였다.

2. Star Coordinates를 이용한 다차원 시각화

2.1 다차원 시각화 기법들

대부분의 분류 작업에 관련한 시스템은 분석자의 개입을 최소화 하는 방향으로 개발되고는 있지만 분류정확도 증가와 같은 요구는 분석자의 개입 필요성을 점점 높이고 있다 (Ankerst et al, 2000). 다차원 시각화 기법은 매우 여러 분야에 적용되고 있으며, 사용자가 데이터에 대한 이해를 돕는 효과적인 방법 중 하나이다 (Soon et al, 2003). 다차원 시각화 기법에 대한 연구는 최근 많이 이루어지고 있으며, 그 대표적인 기법들로는 Parallel Coordinates (Inselberg, 1996), Worlds within Worlds (Feiner et al, 1990), Star Coordinates (Kandogan, 2001) 등을 들 수 있다.

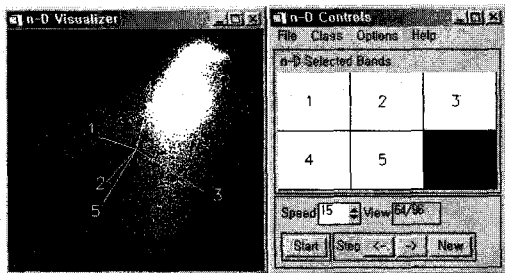


그림 386 N-Dimension Visualization 예 (ENVI 3.5)

ENVI나 TNTlite 프로그램 등은 현재 원격탐사 데이터를 다차원 시각화 기능을 지원하고 있으나, 사용 알고리즘에 대해서는 공개하고 있지 않은 실정이다. 그림 1은 ENVI 프로그램에서 지원하는 N-Dimension Visualization 기능을 보여준다.

2.2 Star Coordinates

Star Coordinates는 다차원의 한 점을 가상의 2차원 상에 투영하여 다차원의 군집과 경향성, Outlier 등을 보여주기를 위한 기법으로 Kandogan에 의해 처음으로 소개되었다 (Soon et al, 2003). Star Coordinates에서 각각의 차원은 데이터 값의 범위(8bit 영상

의 경우 0~255)를 반지름으로 하는 원 내에서 각각의 축으로 결정된다. 6개의 밴드 또는 차원을 가지는 데이터가 있다고 가정하고, 데이터 상에 존재하는 밴드별 한 화소의 밝기값을 $P = (x_1, x_2, x_3, x_4, x_5, x_6)$ 라고 한다면 그림 2와 같이 표현될 수 있다.

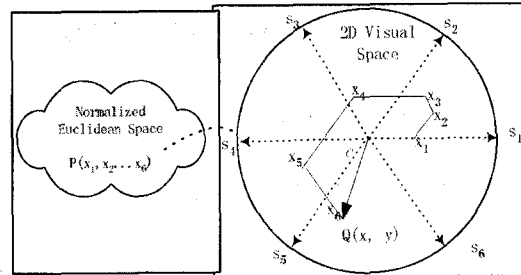


그림 2. Star Coordinates 개념도 (Keke et al, 2004)

임의의 점 P 를 2차원의 가상 면에 표현하기 위해 각 차원은 등간격으로 6 등분으로 축 ($s_1, s_2, s_3, s_4, s_5, s_6$)을 생성하게 된다. P 의 모든 값은 각 축에 평행하게 이동하여 최종 $Q(x, y)$ 값으로 계산되며, x 와 y 의 좌표계산은 식 1과 2를 통해 구할 수 있다. 여기서, k 는 밴드 수이다.

$$Q_x = x_0 + \sum_{i=1}^k \alpha \cdot x_i \cdot \cos(2\pi i/k) \quad (1)$$

$$Q_y = y_0 + \sum_{i=1}^k \alpha \cdot x_i \cdot \sin(2\pi i/k) \quad (2)$$

Star Coordinates의 가장 큰 특징은 선형적(Linear)이라는 것이다 (Keke et al, 2004). 군집화된 데이터들을 Star Coordinates 기법을 통해 2차원 상에 투영하여도 그 군집의 형태는 달라지지 않으며, 군집간 빈 공간에 대한 정보를 취득 가능하므로 군집간 인접성을 시각적으로 파악하는데 용이한 장점을 가진다. 또한 α 값을 설정함으로써 각 차원의 스케일을 조정할 수 있고, $\cos(2\pi i/k)$ 나 $\sin(2\pi i/k)$ 값을 등간격이 아닌 사용자가 지정한 회전각을 적용하여 데이터의 군집화 또는 트레이닝 정보

추출을 용이하게 할 수 있다. 하지만, 계산 과정을 통해 투영되는 화소 중 중복되는 데이터를 구분하는 것은 여전히 문제점으로 존재한다. 그림 3에서 좌측은 4차원 데이터를 Plot한 것이고, 중간은 A 차원을 회전하여 데이터 특성 파악의 이해를 높이는 것을 설명하며, 우측은 A, B, C 차원을 회전하여 구성된 데이터를 보여주고 있다.

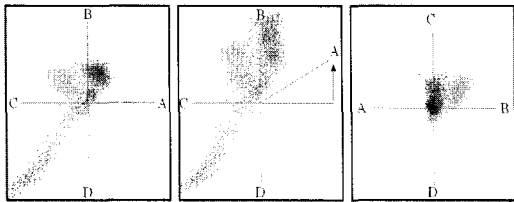


그림 3 Star Coordinates의 축 회전 (Ankerst et al, 2000)

3. 데이터 적용

3.1 개요

Star Coordinates 기법의 이해를 바탕으로 실제 데이터의 좌표값을 계산·투영 후 클러스터링 기법을 적용하여 간단한 분석을 수행하였다. 2차원 Star Coordinates 계산식과 3차원으로 확장하여 각 좌표값을 계산한 3차원 Star Coordinates 기법을 각각 적용하여 클러스터링 작업을 수행하였으며, 계산 과정을 통해 투영되는 중복되는 화소 문제점을 해결하기 위해 각 차원에서의 밝기값의 합을 이용하여 하나의 차원으로 추가하였다. 관계식은 다음과 같다.

$$Q_p = \sum_{i=1}^k x_i \quad (3)$$

3.2 3차원 Star Coordinates

2차원 상에 투영하는 Star Coordinates 기법과는 달리 3차원 Star Coordinates 기법은 구면좌표를 직각좌표로 변환하는 식을 기본으로 한다. 다음 식 4, 5, 6은 3차원 Star Coordinates 기법을 이용하여 3차원 상에 좌표를 투영하는 변환식이다. 식 3에

서 계산된 Q_p 값을 하나의 차원으로 추가하여 4개 밴드에 대해 클러스터링 작업을 수행하였다.

$$Q_x = x_0 + \sum_{i=1}^k \alpha x_i \sin\left(\frac{2\pi i}{k}\right) \cos\left(\frac{2\pi i}{k}\right) \quad (4)$$

$$Q_y = x_0 + \sum_{i=1}^k \alpha x_i \sin\left(\frac{2\pi i}{k}\right) \sin\left(\frac{2\pi i}{k}\right) \quad (5)$$

$$Q_z = y_0 + \sum_{i=1}^k \alpha x_i \cos\left(\frac{2\pi i}{k}\right) \quad (6)$$

3.3 클러스터링 기법 적용 및 결과

3.3.1 사용 데이터



그림 4 Study Area

알고리즘 적용을 위해 서울지역을 촬영한 ETM+ 영상을 사용하였다. ETM+ 센서는 30m의 공간해상도를 가지며, 약 185km × 185km 지역의 촬영이 가능하다. 전정색(Panchromatic) 영상과 열적외선(Thermal IR) 영역의 밴드를 제외한 6개의 밴드를 추출하였고, 한강 이남에서 관악산 부근까지 350Pixels × 200Pixels의 지역을 대상영역으로 선정하였으며, Flat Field 기법을 이용하여 대기효과를 보정하였다. 그림 4는 ETM+ 밴드를 조합하여 표현한 컬러 영상이다.

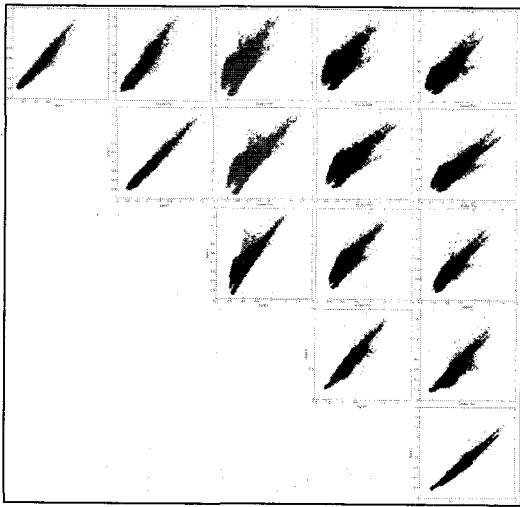


그림 5 밴드간 2차원 Scatter Plot

각 밴드 간 관계를 알아보기 위해 두 밴드를 이용하여 2차원 Scatter Plot을 수행하였다 (그림 5 참고). 하지만 이러한 작업은 $\sum_{i=1}^N (i-1)$ 개의 Scatter Plot 영상을 생성하여야 하며, 상당한 작업량을 필요로 한다 (여기서, N 은 밴드수).

그림 6과 7은 각각 Star Coordinates 기법을 적용하여 Scatter Plot 작업을 수행한 결과를 보여주며, 특히 그림 6은 6개의 밴드를 등간격으로 나눈 각각의 축을 Plot된 결과에 가상으로 그려본 것이다.

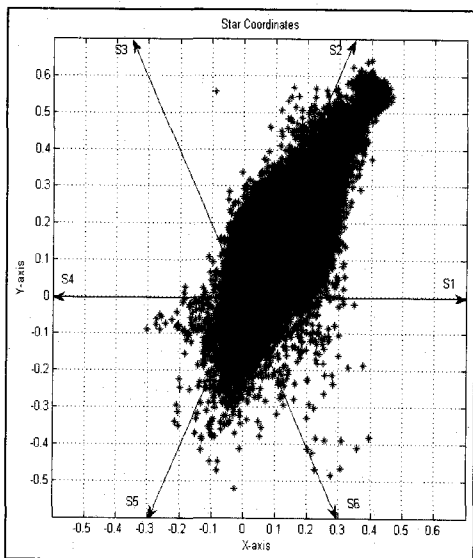


그림 6 대상지역의 2차원 Star Coordinates

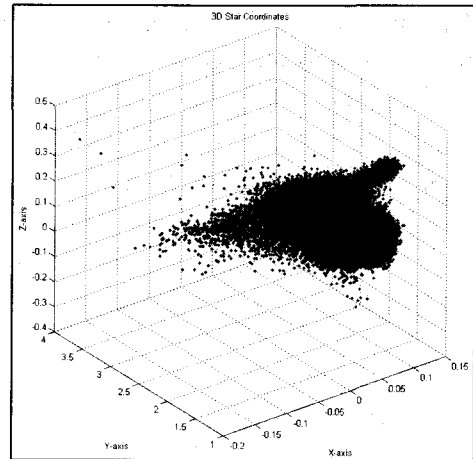


그림 7 대상지역의 3차원 Star Coordinates

3.3.2 K-Means 알고리즘 적용

K-Means 알고리즘은 ISODATA와 함께 많이 사용되는 클러스터링 기법 중 하나이다 (Duda and Hart, 1973). 본 연구에서는 3가지 경우(2차원 및 3차원 Star Coordinates로 변환된 영상과 변화처리를 하지 않은 영상)의 데이터에 대해 K-Means 알고리즘을 적용하여 결과를 분석 하였다.

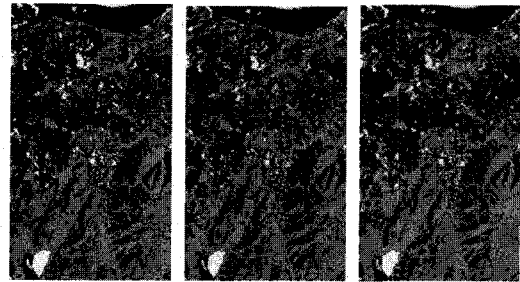


그림 8 K-Means 결과 영상 (4 클러스터)
좌: 2D Star Coordinates, 중: 3D Star Coordinates, 우: 변환 없음
(이하 동일 순서로 나열하였음)

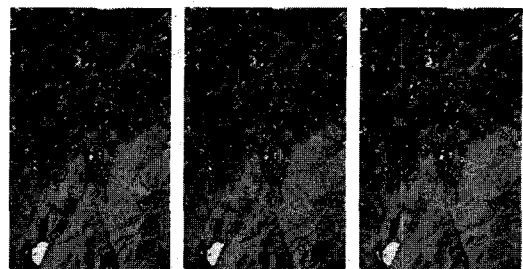


그림 9 K-Means 결과 영상 (6 클러스터)

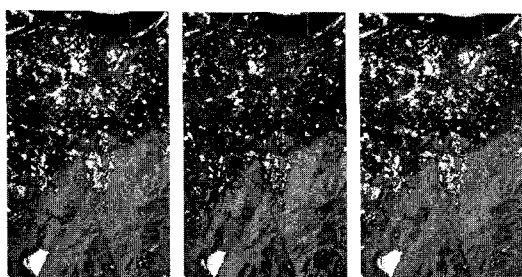


그림 10 K-Means 결과 영상 (8 클러스터)

4개, 6개, 8개 클러스터로 개수를 증가시키면서 알고리즘을 적용하였으며, 변환처리를 하지 않은 영상을 기준으로 2차원 및 3차원 Star Coordinates 변환을 수행하여 적용한 영상의 K-Means 결과와 차이를 비교하였다. 그림 8, 9, 10은 각각 4, 6, 8개의 클러스터를 설정하여 적용된 결과 영상으로서, 각 결과 영상에서 나타나는 클러스터를 기준으로 2진 영상을 생성하여 동일 클러스터를 지정하였고, 같은 색상을 지정하여 주제도를 작성하였다.

결과의 정량적인 평가를 위해, 변환처리를 하지 않은 영상을 기준으로 2차원 또는 3차원 Star Coordinates 변환을 수행한 K-Means 결과와의 차이값을 정확도 형태로 평가해 보았다. 결과에서 확인할 수 있듯, 2차원 Star Coordinates 적용시 실제 데이터와 높은 상관관계를 가지며, 전체 정확도에서도 높게 나타났다. 이것은 Star Coordinates의 특징인 선형성을 뒷받침 해주는 결과로 보인다.

표 126 정확도 평가

		Result (변환 없음)				
		1	2	3	4	합계
Res. (2D)	1	11,719	291	0	0	12,010
	2	513	22,080	1,017	0	23,610
	3	0	1,284	28,727	202	30,213
	4	0	0	134	4,033	4,167
	합계	12,232	23,655	29,878	4,235	70,000

Overall Accuracy = 95.084 %

표 1은 오차 행렬(Error Matrix)을 통해 4개의 클러스터에 대해서 변환처리를 하지 않은 영상과 2차원 Star Coordinates 변환

수행 결과의 정확도 평가 예이며, 표 2는 각각의 경우에 대해 계산한 전체정확도와 영상간 상관계수를 나타낸 것이다.

표 127 적용 Case 전체 정확도

적용 Case	설정 Cluster	전체 정확도 (%)	상관계수
2D	4Clusters	95.0840	0.9935
	6Clusters	83.8343	0.9916
	8Clusters	90.3971	0.9926
3D	4Clusters	91.3486	0.9869
	6Clusters	77.7800	0.9876
	8Clusters	57.8457	0.9821

4. 결론

본 연구는 다차원 시각화를 통해 분석하고자 하는 데이터의 이해도를 높여 군집화나 분류에 있어 정확도를 향상 시키고자 함을 목적으로 한다. 다차원 시각화를 위해 Star Coordinates 기법을 제한하고 있으며, 2차원 및 3차원 Star Coordinates 기법을 적용한 클러스터링 분석을 통해 2차원 Star Coordinates 투영이 선형적인 변환을 나타냄을 확인할 수 있었다. 원격탐사 데이터와 같이 다차원 밴드를 가진 경우 효과적인 분석 기법이 될 것으로 기대되며, 최근 하이퍼스펙트럴 데이터의 취득이 가능함에 따라 분석 시간과 밴드간 상관관계로 인한 잉여 정보 문제 해결의 보완 역할을 할 수 있을 것이다. 본 연구를 바탕으로, Star Coordinates 기법에 스케일과 회전 변환을 적용하여 부정형 클러스터의 확인이나 Signature 정보를 AOI(Area Of Interest) 과정을 통해 추출할 수 있는 시스템의 개발을 수행할 것이며, 하이퍼스펙트럴 데이터에 대해서도 Star Coordinate 기법의 유효성 확인에 관한 연구가 필요할 것으로 판단된다.

Reference

1. M. Ankerst, M. Ester and H.-P. Kriegel, 2000, Towards an Effective Cooperation of the User and the Computer for Classification, Proc. 6th Intl. Conf. on Knowledge Discovery and Data Mining (KDD'00)
2. E. Kandogan, 2001, Visualizing Multi-Dimensional Clusters, Trends, and Outliers Using Star Coordinates. Proc. ACM SIGKDD '01, pp. 107-116
3. A. Inselberg, 1996, Parallel Coordinates: A Guide for the Perplexed, in Hot Topics Proc. of IEEE Conference on Visualization, pp. 35-38, IEEE Computer Society, Los Alamitos, CA, 1996
4. S. Feiner and C. Beshers, 1990, Worlds within Worlds: Metaphors for Exploring n-Dimensional Virtual Worlds, Proc. UIST '90, pp. 76-83
5. Duda. R. D. and P. E. Hart, 1973, Pattern Classification and Scene Analysis, New York, NY: John Wiley & Sons, pp. 482
6. John R. Jensen, 2000, Remote Sensing of the Environment-An Earth resource Perspective, Prentice Hall, pp. 181
7. Keke Chen, Ling Liu, 2004, VISTA: Validating and Refining Clusters via Visualization, Information Visualization, Vol. 3, No. 4, pp. 257-270
8. Soon. T. T., Kwan-Liu. M., 2003, StarClass: Interactive Visual Classification Using Star Coordinates. Proc. 3rd SIAM Intl. Conf. on Data Mining (SDM '03).
9. Research Systems, 2002, ENVI Tutorials, pp. 371-381