

표준편차를 이용한 K-means 알고리즘의 최적의 클러스터 수 추정

Estimation of optimal number of K-means algorithm using standard deviation

김용민¹⁾ · 김용일²⁾

Kim, Yong Min · Kim, Yong Il

¹⁾ 서울대학교 공과대학 건설환경공학부 박사과정(E-mail:kym0210@snu.ac.kr)

²⁾ 서울대학교 공과대학 건설환경공학부 정교수(E-mail:yik@snu.ac.kr)

1. 연구 배경 및 목적

클러스터링 수행 시 최적의 클러스터 수를 결정하는 것은 현재까지 연구되고 있는 본질적이고 실질적인 문제이다(Tibshirani et al., 2001). 본 연구에서는 일반적으로 사용되는 통계치인 표준편차를 이용하여 위성영상의 클러스터링 수행 시 최적의 클러스터 수를 추정한다.

2. 실험영상

최적의 클러스터 수를 추정하기 위해 본 연구에서 사용된 데이터는 IKONOS, Landsat 5 TM 영상이며, 각 영상별로 도심 지역 및 산림 지역을 포함함으로써 센서의 특성과 영상에 포함된 피처에 따라 결과가 어떤 양상으로 변화하는지에 대해 네 가지의 경우를 선택하였다.

3. 실험방법 및 실험결과

실험영상의 클러스터링을 위해 K-means 알고리즘이 사용되었으며, 각 영상마다 1~20개의 클러스터 수로 설정된 클러스터링 결과를 취득하였다. 최적의 클러스터 수를 추정하기 위해 매클러스터링마다 각 클러스터의 밴드별 표준편차를 구하고, 그 표준편차들을 밴드별로 평균하였다. 그 후, 클러스터가 증가함에 따라 각 밴드별 표준편차의 변화량을 산정하였다. 최종적으로 클러스터가 증가함에 따라 각 밴드별 표준편차의 변화량이 모두 1 이하가 된 시점의 클러스터 수를 최적의 클러스터 수로 추정하였다.

IKONS, Landsat 5 TM 영상에 대해 실험한 결과, [그림 1]~[그림 4]와 같이 각 실험영상에 대해 클러스터 수에 따른 표준편차 변화량을 산출하였으며, 다음 [표 1]과 같이 실험영상에 대한 최적의 클러스터 수를 결정하였다. 일반적으로 다수의 피처를 포함한 도심 지역이 산림 지역에 비해 더 많은 최적의 클러스터 수가 추정된 것을 확인할 수 있었다. 또한 공간해상도가 높은 IKONOS 영상의 경우 Landsat 5 TM 영상보다 상대적으로 많은 최적의 클러스터 수를 추정하였으며, Landsat 5 TM 영상의 경우 공간해상도는 상대적으로 낮음에도 불구하고 분광해상도가 높아 IKONOS 영상에서와 비슷한 수의 클러스터를 추정했음을

확인할 수 있었다.

[표 1] 제안된 방법에 의해 추정된 최적의 클러스터 수

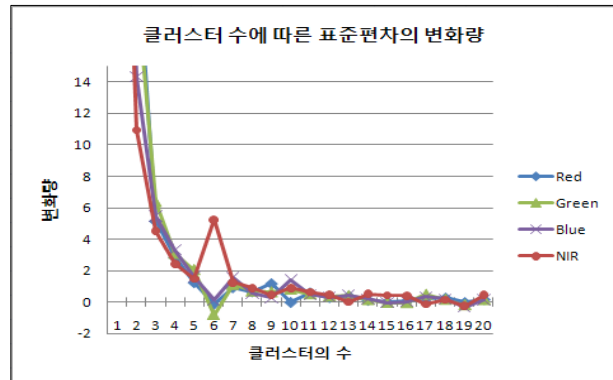
	IKONOS		Landsat 5 TM	
	도심지	산악지	도심지	산악지
최적의 클러스터 수	11	7	9	6

4. 결론

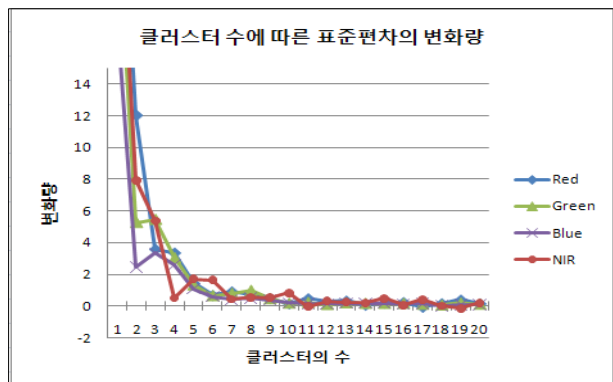
본 논문에서는 위성영상의 클러스터링을 수행함에 있어 최적의 클러스터 수를 추정하는 연구를 수행하였다. 최적의 클러스터 수를 추정하기 위해 가장 일반적인 통계치인 표준편차를 이용하였으며, 네 가지 경우에 대해 그 결과를 도출하였다. 향후, 도출된 결과에 대해 이론적으로 검증함과 동시에 기존의 알고리즘과의 비교를 통해 제안 알고리즘의 성능을 평가할 것이다.

감사의 글

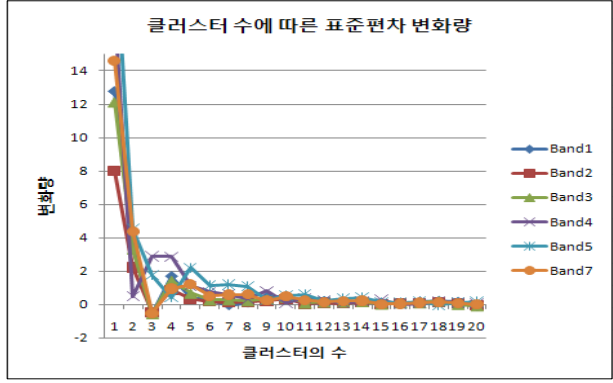
본 연구는 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 20090085392).



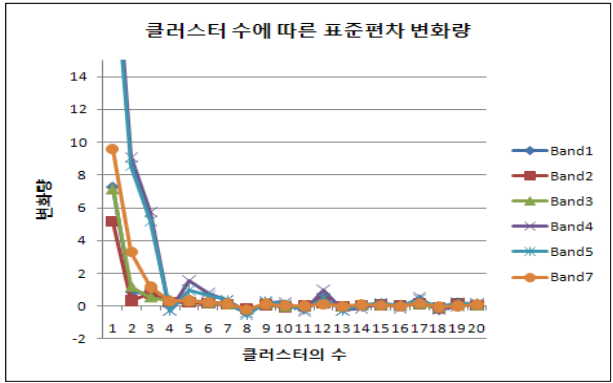
[그림 1] IKONOS 영상의 도심 지역 영상과 클러스터 수에 따른 표준편차의 변화량



[그림 2] IKONOS 영상의 산악 지역 영상과 클러스터 수에 따른 표준편차의 변화량



[그림 3] Landsat 5 TM 영상의 도심 지역 영상과 클러스터 수에 따른 표준편차의 변화량



[그림 4] Landsat 5 TM 영상의 산림 지역 영상과 클러스터 수에 따른 표준편차의 변화량

참고문헌

Tibshirani, R., Walther, G. and Hastie, T. (2001), Estimating the number of clusters in a data set via the gap statistics, Journal of Royal Statistical Society, B, Vol. 63, Part 2, pp. 411-423