

K-Means 와 GHA를 이용한 위성영상 구름영역 검출

*서석배 김종우 최해진
한국항공우주연구원 위성운영센터
e-mail : sbseo@kari.re.kr

Cloudy Area Detection in Satellite Image using K-Means & GHA

*Seok-Bae Seo, Jong-Woo Kim, and Hae-Jin Choi
Satellite Mission Operation Department
Korea Aerospace Research Institute

Abstract

This paper proposes a new algorithm for cloudy area detection using K-Means and GHA (Generalized Hebbian Algorithm). K-Means is one of simple classification algorithm, and GHA is unsupervised neural network for data compression and pattern classification. Proposed algorithm is based on block based image processing that size is 16×16 . Experimental results shows good performance of cloudy area detection except blur cloudy areas.

I. 서론

패턴인식에 관한 많은 알고리즘 연구와 시스템 개발이 있었지만 인간의 아직까지 인식성능에는 근접하지 못하고 있다^{[1][2]}. 위성영상에서 구름영역 검출도 경우도 마찬가지로 인간은 쉽게 구별해 낼 수 있지만 패턴인식 알고리즘으로 접근하기에는 많은 어려움이 따른다. 특히 위성영상의 경우 큰 사이즈를 가지며 비슷한 영역이 많이 존재하기 때문에 많은 양의 계산이 요구된다. 본 논문에서는 위성영상의 데이터 특성을 고려한 패턴인식방법을 제안하고, 구름영역추출에 적용하여 보았다.

패턴분류 알고리즘 중 널리 알려진 K-Means는 성능은 다소 떨어지지만 간단하고 수행속도가 빠른 것이 특징이다. GHA는 신경회로망의 한 종류로 비교사(unsupervised) 학습으로 데이터의 형태특징을 추출할 수 있고 그 성능이 우수하여 패턴인식에 널리 사용되어지고 있다^[2].

K-Means 알고리즘은 전체 데이터를 몇 개의 그룹으로 분류하여 적절한 표본데이터를 추출하

는데 사용하였고, GHA는 추출된 데이터를 이용하여 데이터 형태를 잘 표현할 수 있는 계수를 생성하는데 사용된다. 또한 전체적인 계산량을 줄이기 위해서 먼저 각 블록의 평균치를 계산하여 구름이 될 가능성이 희박한 어두운 부분을 후보영역에서 제외한 다음, 후보영역에 대해서만 GHA 알고리즘을 적용하는 방법을 도입하였다. 제안한 알고리즘의 성능을 검증하기 위해서 다목적위성 아리랑 1호의 대전 EOC (Electronic Optic Camera) 영상을 사용하였다.

II. 배경이론

2.1 K-Means Algorithms

알고리즘이 간단하고 수행속도가 빠른 특징을 가지는 K-Means 알고리즘은 간단한 패턴분류에 널리 사용된다. 알고리즘은 다음의 두 단계로 이루어진다.

[단계1] ①분류하고자하는 클러스터의 개수 K를 정하고 K개의 초기 클러스터에 데이터를 하나씩

대입한다. 남은 ⑥데이터를 클러스터 중 가장 작은 거리를 가지는 클러스터로 분류하고 중심 값을 갱신한다.

[단계2] 두 번째 계산부터는 전체 데이터에 대해서 [단계1]의 ⑥를 반복한다. 정지조건은 반복회수도달이나 분류결과 고정으로 설정한다.

2.2 Generalized Hebbian Algorithm

GHA는 PCA의 한 종류로 순방향 구조의 단층 신경망이며 자율학습의 특성이 있다^[2]. 학습된 GHA 웨이트는 가중치 특성이 있으며, 이를 이용하면 입력데이터를 보다 작은 차원으로 축소시킬 수 있다. 본 논문에서는 웨이트들이 가중치에 의해 내림차순 정렬되는 GHA 특성을 이용하여 데이터의 형태에 대한 특징을 추출한다. GHA를 다음의 3단계로 요약하여 설명한다.

Step 1. 신경회로망의 연결강도 w_{ji} 를 무작위의 작은 실수값으로 초기화하고 학습률 η 와 반복회수 n 을 설정한다. 여기서 $i = 0, 1, \dots, p-1$, $j = 0, 1, \dots, m-1$ 로 p 와 m 은 각각 입력노드와 출력노드의 수를 나타낸다.

Step 2. 연결강도 w_{ji} 를 반복회수 n 만큼 식 (1), 식 (2)를 이용하여 학습시킨다.

$$y_j(n) = \sum_{i=0}^{p-1} w_{ji}(n)x_i(n) \quad (1)$$

$$\Delta w_{ji}(n) = \eta[y_j(n)x_i(n) - y_j(n)v(n)] \quad (2)$$

여기서, $v(n) = \sum_{k=0}^m w_{kj}(n)y_k(n)$ 이다.

Step 3. Step 2에서 구한 연결강도 w_{ji} 를 이용하여 특징추출을 위한 계수 C 를 생성한다. C 는 식 (3)으로 나타내며 각각의 인자는 식 (4)로 구할 수 있다. 가중치는 j 가 적을수록 높으며, c_0 는 다른 계수에 비하여 매우 큰 값을 지닌다.

$$C = \{c_0, c_1, \dots, c_{m-1}\} \quad (4)$$

$$c_j = \sum_{i=0}^{p-1} w_{ji}(n)x_i(n) \quad (5)$$

최종 수정된 w_{ji} 는 학습입력 x_i 에 대한 고유벡

터의 특징이 있으며, 이를 이용하면 입력보다 낮은 차원으로 데이터를 표현할 수 있다. 여기서 m 은 p 보다 작은 정수로 입력벡터의 축소된 차수를 의미한다. 학습이 성공적으로 수행되면 p 차원으로 표현된 학습입력의 직교성(orthogonal)이 증가되어 낮은 차원으로 학습입력을 효과적으로 표현할 수 있다^[3].

III. 제안하는 알고리즘

3.1 Input Image

그림 1은 다목적위성 1호로 촬영한 대전영상으로 2592x3300의 크기를 가진다. 실험을 위하여 1024x768로 크기를 재조정했으며, 16x16의 블록으로 분할하였다. 따라서 16x16 블록의 개수는 5184(64x81)개이며 각각의 블록 영상은 256(16x16)개의 그레이 스케일 화소를 가진다.

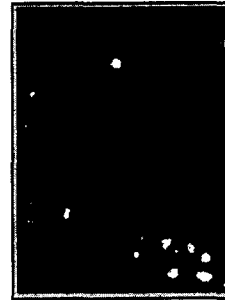


그림 1. 대전 EOC 영상

3.2 Bright Area Selection using Block Average

그림 1에서 구름영역은 주변에 비하여 매우 밝음을 볼 수 있다. 따라서 계산량을 줄이기 위해서 각 블록의 평균값을 이용하여 밝은 영역을 그룹후보로 설정하고 해당영역에 대해서만 그룹영역 검출 알고리즘을 적용한다.

3.3. Simple Area Selection using K-Means & GHA

위성영상은 큰 사이즈를 가지고 있으며 그 형태가 비슷한 부분이 많다. 중복되는 데이터가 계산되는 것을 방지하기 위해서 본 논문에서는 K-Means 알고리즘을 이용하여 데이터를 몇 개의 그룹으로 나누고, 각 그룹에서 일정한 개수의

데이터를 그룹영역 검출을 위한 데이터로 사용함으로써, 신경회로망의 계산시간을 단축시키고 고 큰 패턴의 특징을 추출할 수 있었다.

그림 2는 GHA를 이용한 영상의 복원을 설명한 것으로 학습으로 구한 베이스에 각각의 블록이 가지는 계수를 곱해서 누적하면 원래의 영상이 복원됨을 볼 수 있다¹⁴⁾. 각각의 계수의 크기는 해당 베이스에 대한 상관도로, 높을수록 그 베이스의 특징을 많이 가지고 있다. 그림 2에서 첫 번째 베이스는 블록 전체의 밝기를 나타내는 것으로 DCT의 DC 계수와 유사하다. 본 논문에서 그룹 영역을 검출하는 방법은 GHA의 학습에 의해 베이스를 구하고, 그 베이스를 이용해서 각각의 영역의 복잡도를 계산하는 것으로, 계수값이 높을수록, 그리고 계수 값들의 차이가 클수록 복잡한 영상임을 알 수 있다.

$$I = C_1 \times B_1 + C_2 \times B_2 + C_3 \times B_3 + \dots + C_n \times B_n$$

그림 2. GHA를 이용한 영상복원

IV. 실험결과

4.1 Bright Area Selection using Block Average

그림 3은 각각의 16x16의 블록에 대해서 그레이 스케일 평균을 구한 다음 140 이상이 되는 영역만 선별한 것이다. 그림에서 보면 그룹 영역뿐만 아니라 흰색 건물까지 선택된 것을 볼 수 있는데, 이에 대한 구분은 이후의 GHA를 이용한 복잡도 계산에 의해서 수행된다.



그림 3. Bright Areas by Block Average

4.2 Simple Areas Detection

a) Candidate data Selection using K-Means

1024x1296 크기의 영상은 5184개의 16x16 영상을 가진다. 그림 4는 클러스터 K의 개수에 의한 데이터 분류결과중 일부로 16개와 100개 일 때를 표시한 것이다. 본 실험에서는 16개의 클러스터를 사용했으며 그에 대한 데이터 개수 분포는 그림 5와 같다.

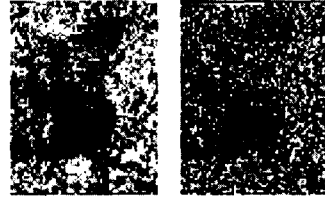


그림 4. 클러스터 개수에 대한 분류결과 (좌: 16개, 우: 100개)

K-Means 알고리즘을 이용하여 5184개의 데이터를 16개 그룹으로 나누고, 각 그룹에서 10개씩 임의로 선택한 총 160개의 데이터의 형태는 그림 6과 같으며, 이는 GHA 학습에 사용된다.

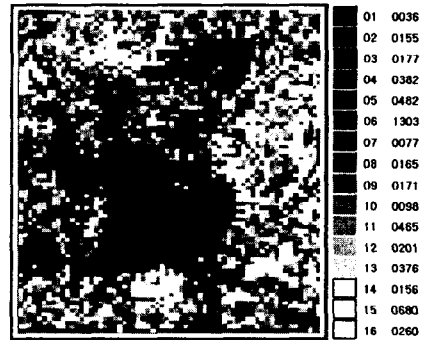


그림 5. 16개 클러스터에 대한 데이터 개수

b) Shape Detection using GHA

블록의 형태를 추출하고 복잡도를 계산하기 위해서 30개의 베이스를 가지는 GHA 시스템을 이용하였다. 학습률과 학습회수를 변경해가며 실험한 결과, 위성영상데이터는 상대적으로 큰 학습계수를 가지며 서로 독립적인(orthogonal) 베이스의 수가 적음을 알 수 있었다. 그림 7은 학습계수 0.2 반복회수 100회로 학습한 베이스를

시각적으로 표시한 결과이다.

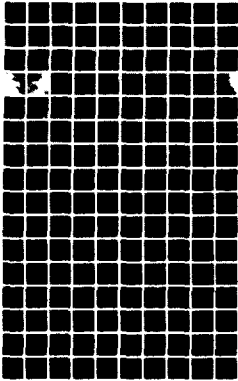


그림 6. K-Means에 의해 선택된 후보 데이터

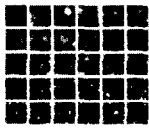


그림 7. GHA로 학습된 베이스

결과에서 10번째부터는 형태가 비슷함을 볼 수 있는데, 이는 위성영상의 데이터는 단순하며 10개 정도의 특징의 조합으로 표현할 수 있다는 것을 의미한다.

그림 8은 학습에 의해 생성된 베이스를 이용한 복잡도 계산을, 밝은 영역으로 선택된 부분에만 적용하여 건물영역을 제외시킨 결과이다. 영역선택은 각 후보블록에 GHA 베이스를 이용하여 계수를 구하고 그 계수의 차이가 적은 것들만 구름영역으로 간주한 것이다. 그림 8의 경우 하나의 영상에 대해서 몇 번의 계산에 의한 결과이지만, 많은 위성영상을 입력으로 하여 학습한다면 자동으로 구름을 추출할 수 있을 것으로 생각된다.

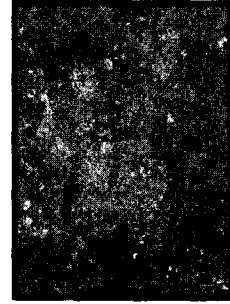


그림 8. 최종 구름영역 검출 결과

IV. 결 론

본 논문에서는 K-Means 알고리즘과 GHA를 이용하여 구름영역을 추출하였다. 먼저 각 블록에 대해서 그레이스케일 평균값을 이용해 후보영역을 설정함으로써 알고리즘이 적용되는 영역을 축소하였다. 다음으로 복잡도를 이용하여 구름영역을 검출하기 위해 적용된 GHA의 효과적인 학습을 위해서, K-Means를 이용하여 특징데이터 개수를 줄였으며, 마지막으로 GHA에 의해서 최종적인 구름영역을 검출하였다.

본 실험에서는 그레이 스케일이 큰 영역에서 복잡한 부분을 제외하여 구름영역을 찾는 방법을 위한 특징을 추출하는 것에 목적을 두었으며, 보다 상세하고 효율적인 구름영역 추출에 대해서 지속적인 연구를 수행할 예정이다.

참고문헌

- [1] Earl Gose, Ricahrd Johnsonbaugh, and Steve Jost 1996, Pattern Recognition and Image, Prentice Hall Pub., pp 213-214.
- [2] S. Haykin 1999, Nerual networks a comprehensive foundation, Prentice Hall Pub., (pp 413-499).
- [3] Oja. E. 1992, A Simplified erneron model as a principal component analyzer, Journal of mathematical biology, pp. 267-273.
- [4] Seok-Bae Seo, Daijin Kim, and Dea-Seung Kang 2001, Face Image Compression algorithm using Triangular Feature Extraction and GHA, Journal of IEEK, Vol. 38-SP, pp. 11-18.