

위성영상의 통계적 분류를 위한 유효 트레이닝 기법에 관한 연구 The Effective Training Method for the Statistical Classification of Remotely Sensed Imagery

이병길* · 김용일** · 어양담***

Lee, Byoung-Kil · Kim, Yong-Il · Eo, Yang-Dam

要 旨

위성영상의 통계적 분석에서는 트레이닝을 통하여 추정된 각 클래스의 평균과 분산을 통계적 유사성 결정의 근거로 사용한다. 따라서, 트레이닝 작업의 결과는 전체적인 분류정확도에 큰 영향을 주게 된다. 이상적인 위성영상 데이터의 분포는 정규분포에 근사한 것으로 가정되고 있으나, 실제로는 약간의 밀집도와 편의를 보이며 이러한 분포의 이상이 트레이닝 결과 추정되는 분산뿐만 아니라 분류 결과 자체에도 많은 영향을 준다. 본 연구에서는 트레이닝을 통해 추정된 분산의 특성과 이러한 특성이 화소값의 분포와 어떤 관련이 있는지를 규명하고, 그에 따른 분류결과의 변화에 대해 검토하였고, 분산 과소추정의 영향을 최소화할 수 있는 트레이닝 기법을 제시하였다.

ABSTRACT

In statistical analysis of remotely sensed data, means and variances of each classes are used as the basis of statistical similarity determination. Therefore, the overall accuracy of classification is affected by the training results. It is assumed that the ideal distributions of pixel values follow normal distributions, but practically they have some aggregations and biases. These anomalies of distribution can affect the classification results greatly as well as the variances of training results. In this study, relationships between the inferential variances of the training sets and the distributions of pixel values are examined, and the resulting changes of classification results are studied. Furthermore, the training method which minimizes the effect of underestimation of variances is proposed.

1. 연구의 배경 및 목적

위성영상에 대한 통계적 분석은 화소값(DN : Digital Number)을 이용하여 영상의 각 화소를 통계적으로 유사한 분류항목으로 분류하는 과정으로 분류된 결과는 사용 목적에 적합한 정확도를 가져야 한다.

통계적 영상분류 작업에서, 분류에 필요한 매개변수(parameter)는 트레이닝 작업으로부터 얻어지는데, 트레이닝 작업은 매우 주관적으로 이루어지기 때문에, 동일영상이라 하더라도 작업자에 따라 분류작업결과가 다를 수 있다.⁴⁾ 다시 말하면, 트레이닝 작업은 모집단에 대한 모수를 추정하기 위한 표본추출(sampling)로 이해할 수 있

으며, 트레이닝 작업의 성과에 따르는 모집단 추정의 정확도가 최종적인 분류 결과에도 많은 영향을 미치게 된다.

실제로 많은 위성영상의 분석 작업에서 트레이닝에 의해 산출된 결과가 실제 모집단의 분포를 정확하게 추정하지 못하며, 특히 트레이닝 영역의 크기에 따라 분류결과가 달라짐을 경험할 수 있다. 이렇게 트레이닝이 통계적 분류작업에 매우 중요함에도 불구하고 트레이닝 기법이나 트레이닝의 기제에 관한 연구는 많지 않다.

트레이닝에 관한 연구 사례로는 Hixon 등이 분류 알고리즘 선정보다 트레이닝 샘플의 대표성이 분류 정확도의 향상에 더 중요하다고 주장하였고,³⁾ Van Deusen은 트레이닝 데이터의 대표성을 갖추기 위한 각 클래스별 샘플화소비율을 확률이론에 근거하여 연구하였으며,⁵⁾ 또한 어양담은 유전자 알고리즘을 적용하여 트레이닝 데이터의 정규성을 향상시켜 모집단에 근사하게 할 수 있음을 제시하였다.²⁾

*서울대학교 도시공학과 박사과정

**서울대학교 도시공학과 조교수

***서울대학교 도시공학과 박사후과정

본 연구에서는 기준 데이터를 이용하여 추출된 모집단의 기술통계를 구하고, 여러 번의 실험 트레이닝을 통해 추정된 통계치를 모집단의 통계치와 비교 분석하여, 트레이닝에서 구해지는 분산이 과소추정됨을 예시하고, 위성영상 데이터에서 트레이닝 영역 추출에 의한 분산의 추정이 작아지는 원인을 규명하였으며, 분산의 과소추정에 따라 통계적 영상분류기법을 적용한 분류결과에 일어나는 변화에 대해 검토하였고, 최종적으로 분산의 과소추정에 의해 발생하는 분류결과와 변화를 최소화할 수 있는 트레이닝 기법을 제안하고, 검증하였다.

2. 트레이닝 모의 실험

본 연구에서 사용된 영상은 1994년 5월에 획득한 LANDSAT TM(row471×column380) 영상이며, 기준 데이터는 영상취득과 가까운 시기에 얻어진 항공사진을 수치지도화한 수치지도를 이용하였다. 논, 밭 및 도시와 같이 수치지도로부터 폐합된 폴리곤을 얻을 수 있으며, 적정수 이상의 화소수를 가진 클래스를 선택하여 각 클래스에 대한 모집단을 영상으로부터 추출하였다.

2.1 모집단의 기술통계

영상으로부터 논, 밭, 도시 3개 클래스에 대한 화소값을 추출하여 영상분류에 많이 사용되는 6개 밴드(밴드 1, 2, 3, 4, 5 및 7)에 대하여 평균, 분산, 첨도²⁾, 왜도³⁾ 및 정규분포와의 상관성 등을 구하였다(표 1).

클래스 논은 전체적으로 첨도와 왜도가 높으며, 클래스 밭은 밴드 1과 밴드 2에서 첨도와 왜도가 높고, 클래스 도시는 상대적으로 첨도와 왜도가 낮다. 첨도가 높다는 것은 데이터가 밀집되어 분산이 나타내는 것에 비해 데이터의 밀집이 더 심함을 의미한다. 또한 클래스 도시의 히스토그램은 형태가 매끄럽지 못한 다첨두(multi-model)형태를 보이고 있다. 그러나, 정규분포와의 상관성을 구한 정규상관성에서 볼 수 있는 것과 같이 모든 클래스의 모든 밴드에서 정규분포와의 상관성이 높은 것으로 나타나고 있다. 이는 위성영상 데이터가 전체적으로는 정규분포에 근사한 분포를 가지나 부분적으로 밀집, 다첨두 등 정규분포의 특성을 저해하는 요인을 약간씩 가지고 있음을 의미한다.

$$2) \text{ 첨도는 } \frac{n(n-1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{x_i - \bar{x}}{s} \right)^4 \frac{3(n-1)^3}{(n-2)(n-3)} \text{ 로}$$

표 1. 클래스의 기술 통계치

논	밴드1	밴드2	밴드3	밴드4	밴드 5	밴드 7
평균	88.09	40.36	51.19	57.16	92.00	45.91
분산	23.46	11.45	35.90	70.83	224.38	78.24
첨도	16.20	6.81	5.13	5.06	1.36	2.26
왜도	2.44	1.77	1.46	0.95	-0.33	0.19
정규상관성	0.93	0.94	0.96	0.99	0.98	0.98
밭	밴드1	밴드2	밴드3	밴드4	밴드 5	밴드 7
평균	88.84	41.37	53.99	63.05	103.31	52.38
분산	34.24	20.97	69.76	70.99	286.61	109.76
첨도	4.21	11.36	1.42	1.78	0.73	1.44
왜도	1.25	1.53	0.80	0.47	-0.11	0.33
정규상관성	0.96	0.99	0.95	0.97	0.93	0.94
도시	밴드1	밴드2	밴드3	밴드4	밴드 5	밴드 7
평균	102.38	49.05	64.63	64.38	106.38	60.54
분산	93.61	34.71	88.05	43.48	221.96	118.76
첨도	0.92	1.28	1.62	.55	1.97	2.44
왜도	0.34	0.34	0.28	0.36	1.97	0.48
정규상관성	0.95	0.96	0.97	0.99	0.98	0.98

정의되며, 분포의 뾰족한 정도를 정규 분포와 비교하여 나타내는 양으로 양의 첨도는 상대적으로 더 뾰족하고 음의 첨도는 덜 뾰족함을 의미한다.

3) 왜도는 $\frac{n}{(n-1)(n-2)} \sum \left(\frac{x_i - \bar{x}}{s} \right)^3$ 으로 정의되며, 평균에 대한 분포의 비대칭 정도를 나타내는 양으로 왜도가 양수 이면 분포의 비대칭 꼬리가 양의 값 쪽으로 치우쳐 있으며, 왜도가 음수이면 이와 반대가 된다.

2.2 트레이닝 및 분석

본 연구에서는 트레이닝 기법간의 차이를 비교하기 위해 점표본을 임의 추출하여 트레이닝 데이터로 사용하는 방법과 여러 개의 트레이닝 영역을 추출하여 트레이닝 데이터로 사용하는 방법의 두가지 방법을 사용하였다.

2.2.1 점표본 추출

임의 표본추출은 통계적으로 가장 유의성을 가지고 표본추출 방법이나, 표본의 수를 많이 해야 하는 위성영상 데이터의 경우에는 사용이 용이하지 않다. 그러나, 본 연구에서는 통계적으로 유의한 트레이닝 데이터를 얻기 위

표 2. 각 클래스이 추론 통계치(입의 점표본 추출기법)

논화소수	밴드1	밴드2	밴드3	밴드4	밴드 5	밴드 7
308	평균 88.12	40.44	51.50	57.06	93.01	46.65
0.83%	분산 17.12	9.33	27.44	68.29	210.01	65.84
369	평균 87.98	40.22	50.87	56.71	91.66	45.77
1.00%	분산 18.63	9.46	29.19	72.97	219.70	73.25
모집단	평균 88.09	40.36	51.19	57.16	92.00	45.91
36990	분산 23.46	11.45	35.90	70.83	224.38	78.24
발화소수	밴드1	밴드2	밴드3	밴드4	밴드 5	밴드 7
116	평균 89.81	42.04	55.28	63.43	103.42	52.86
0.80%	분산 30.79	19.97	65.58	62.64	232.36	84.25
154	평균 88.96	41.19	54.11	65.42	102.45	52.06
1.07%	분산 28.53	17.87	59.55	61.73	228.23	80.82
모집단	평균 88.84	41.37	53.99	63.05	103.31	52.38
14414	분산 34.24	20.97	69.79	70.99	286.61	109.76
도시화소수	밴드1	밴드2	밴드3	밴드4	밴드 5	밴드 7
10	평균 100.40	47.90	62.50	64.60	106.20	60.20
1.39%	분산 91.84	28.29	63.85	10.44	137.76	97.56
22	평균 104.36	50.09	65.31	69.50	108.18	62.18
3.05%	분산 104.41	36.62	84.21	19.25	143.87	83.33
모집단	평균 102.38	49.05	64.63	64.38	106.38	60.54
721	분산 93.61	34.71	88.05	43.48	221.96	118.79

해 모집단 데이터를 얻기 위해 모집단 데이터를 ASCII 파일로 변환한 후, 필요한 수의 표본을 입의 추출하여 트레이닝 데이터로 사용하였다(표 2).

2.2.2 트레이닝 영역 표본 추출

트레이닝 영역을 표본으로 추출하여 트레이닝 데이터를 얻는 트레이닝 방법은 실제 원격탐사 분류 작업에서 많이 사용되는 방법으로 작업이 용이하며, 결과를 신속하게 얻을 수 있는 장점이 있다. 실험을 위한 많은 횟수의 트레이닝이 기술적으로 어렵기 때문에, 본 연구에서는 먼저 논과 밭에서는 27개의 트레이닝 영역을, 도시에서는 5개의 트레이닝 영역을 추출하여 표본으로 사용하였다. 논과 밭에 대해서는 표본으로부터 난수 발생에 의해 임의로 선택된 4개, 5개, 6개, 7개, 8개 및 9개의 트레이닝 영역을 트레이닝 데이터로 사용하는 실험을 21회씩 반복하였고, 도시에 대해서는 표본의 수가 작아, ${}_3C_1$, ${}_3C_2$, ${}_3C_3$ 및 ${}_3C_4$ 의 조합으로 트레이닝 영역을 트레이닝 데이터로 사용하였다.

각각의 트레이닝 데이터에 대하여, 평균과 분산을 구

표 3. 사용된 트레이닝 영역 수에 따른 트레이닝 영역 표본 추출의 검정 결과(추정 성공 횟수/검정횟수)

트레이닝 영역수	4*	5**	6***	7****	8	9
논	평균 53/126	44/126	39/126	38/126	37/126	36/126
	분산 28/126	20/126	22/126	18/126	14/126	12/126
밭	평균 72/126	68/126	55/126	59/126	55/126	52/126
	분산 20/126	26/126	28/126	14/126	15/126	8/126
도시	평균 21/30	9/60	9/60	10/30		
	분산 7/30	15/60	19/60	24/30		

*,**,***,****: 사용된 도시의 트레이닝 영역수, 각각 1, 2, 3, 4개.

표 4. 밴드에 따른 트레이닝 영역 표본 추출의 검정 결과(추정 성공 횟수/검정횟수)

밴드	1	2	3	4	5	7
논	평균 26/126	26/126	14/126	28/126	74/126	79/126
	분산 3/126	21/126	5/126	47/126	17/126	21/126
밭	평균 84/126	53/126	42/126	69/126	47/126	50/126
	분산 0/126	3/126	19/126	34/126	28/126	13/126
도시	평균 4/30	3/30	3/30	26/30	8/30	5/30
	분산 15/30	5/30	16/30	15/30	9/30	5/30

하고, 트레이닝 데이터가 모집단과 같음을 유의성 5%에서 Z-검정과 χ^2 -검정을 각각 수행하였다.

검정 결과는 표 3 및 표 4와 같으며, 분사의 경우 사용된 트레이닝 영역 수가 증가됨에 따라 검정결과 기각되는 트레이닝 데이터가 많아짐을 알 수 있다. 이는 트레이닝 영역 수의 증가와 무관하게 트레이닝 영역 표본 추출에 의한 분산은 일정하게 유지되는 반면에 χ^2 -검정의 자유도는 증가하여, 통계치의 기각이 많이 일어나기 때문이다.

표 3, 4와 앞서 살펴본 모집단의 기술통계를 비교하면, 첨도가 높게 나온 밴드에서 대체적으로 분산의 추정 에 많이 실패하고 있음을 알 수 있다.

3. 트레이닝 데이터 분산의 과소추정

위성영상 데이터가 정규분포를 따른다고 가정하면, 표본분산 S^2 는 $\chi^2(n-1)$ 을 따르며, 이 때, n 은 표본의 수이다. χ^2 분포의 평균은 $n-1$ 이므로, 표본의 수가 많아지면, 표본분산을 모분산과 같아진다.

점표본 추출에 의한 트레이닝 결과는 표본 크기가 모

집단 크기의 약 1%이상인 경우에 모분산과 거의 동일한 통계치를 제시하고 있다. 그런데, 원격탐사용 위성영상 데이터는 정규분포에 근사하기는 하나, 한쪽으로 편이 되어 있고, 약간 밀집되어 있어 표본추출 시표본의 크기가 작으면 평균과 가까운 화소가 선택될 확률이 높아져 분산이 작아지는 경향을 가지게 되므로 모분산의 추정에서 실패하게 된다.

점표본 추출의 경우와 달리 트레이닝 영역 표본 추출의 경우는, 사용하는 트레이닝 영역 수의 증가, 즉 화소 수의 증가에 의한 자유도의 증가가 분산의 확대와 무관하게 나타났다. 트레이닝 영역 표본 추출은 작은 크기의 트레이닝 영역을 여러 개 사용하여 분산을 구하는 방법을 사용하였으며, 이를 통계적으로 표현하면 하나의 정규모집단에서 여러 개의 표본을 추출하여 모집단의 분산을 추정하는 것이 된다. 따라서, 공통분산 σ^2 에 대한 합동추정량은 식 (3-1)과 같이 주어진다.¹⁾

$$S_p^2 = \frac{\sum_1^{n_1} (X_1 - \bar{X})^2 + \sum_1^{n_2} (X_1 - \bar{X})^2}{n_1 + n_2 - 2} = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \quad (3-1)$$

정규모집단에서 추출된 확률변수의 경우, $\frac{(n_1 - 1)S_1^2}{\sigma^2}$ 과 $\frac{(n_2 - 1)S_2^2}{\sigma^2}$ 이 각각 χ^2 -분포를 따른다면, S_1^2 과 S_2^2 이 σ^2 보다 클 확률과 작을 확률이 모두 0.5로 동일하기 때문에, 이의 가중 평균에 해당하는 합동추정량이 모분산에 가까워지는 것은 자명하다.

그러나, 정규모집단에 비해 밀집된 분포를 보이는 위성영상의 경우에는 각 트레이닝 영역의 물리적 한계(트

레이닝 영역의 크기)에 의해 분산이 과소추정될 확률이 높기 때문에 이를 이용하는 추정되는 합동추정량 S_p^2 도 과소추정될 확률이 높아진다. 따라서, 트레이닝 영역 표본 추출에 의한 트레이닝 데이터에서는 사용된 표본인 화소수의 증가와 무관하게 분산이 과소추정될 확률이 높다.

4. 분산의 과소추정이 분류에 미치는 영향

분산이 과소추정되는 정도는 클래스 모집단의 척도에 따라, 영상에 따라, 작업자에 따라 심지어는 트레이닝 시행마다 변한다. 분산의 과소추정과 그 정도의 차이에 따라 통계적 영상분류에서 분류경계(decision boundary)의 이동이 발생된다.

그림 1에서 클래스 1과 클래스 2의 분산은 각각 4와 9였으나 모두 2.25로 분산이 감소하였다. 그러나 분산이 감소된 정도는 클래스 1과 클래스 2에 대해 각각 0.5625와 0.25로 다르다. 이와 같이, 분류경계는 분산의 과소추정 정도가 심한 쪽으로 이동하며, 분류결과도 클래스 1이 더 많은 비중을 차지하게 된다.

5. 분산 변화에 유효한 트레이닝 기법의 제안

가장 이상적인 트레이닝 데이터 수집방법은 점표본의 임의추출에 의한 트레이닝 데이터 수집 방법이나, 충분한 표본의 크기를 충분히 크게 하기 위해서는 시간상, 비용상의 한계가 존재하게 된다. 따라서, 실제적으로는 트레이닝 영역 추출에 의한 트레이닝 데이터 수집방법을 사용하게 되며, 이 방법은 분산을 과소추정하는 경향이

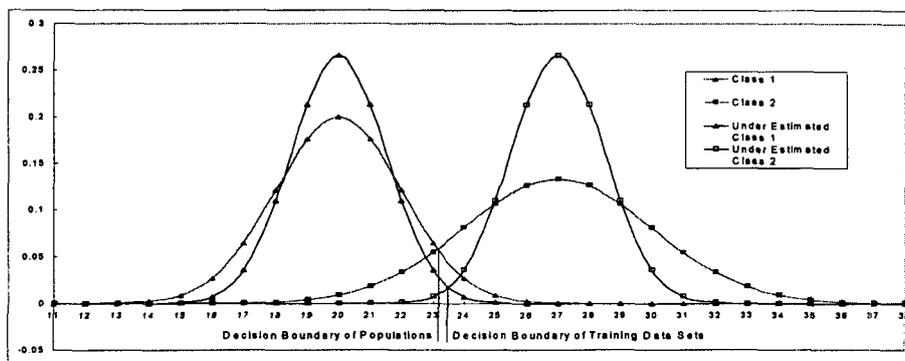


그림 1. 분산 추정이 클래스 확률밀도 함수와 분류 경계면 결정에 미치는 영향

있으며, 그 결과로 분류에 안 좋은 영향을 줄 수 있다는 것을 알아보았다.

그러므로, 분류정확도의 향상과 분류결과의 신뢰도 제고를 위해서는 클래스 및 밴드별로 차이 나는 분산의 과소추정을 최소화하여야 하고, 이를 위한 방법으로 단순히 트레이닝 영역의 수만을 늘리는 것 외에, 모분산에 대한 표본분산의 비가 일정하도록 트레이닝 영역의 크기를 조정하여야 한다.

위성영상의 한 밴드에서 각 클래스에 대한 모집단이 정규분포를 따른다고 가정하면, 각 클래스에 대한 모분산(σ_i^2)과 그 클래스에 대한 트레이닝 영역 한 개의 표본분산 (S_{ij}^2)은 모두 $\frac{(n_{ij}-1)S_{ij}^2}{\sigma_i^2} \sim \chi^2(n_{ij}-1)$ 가 성립한다.

그런데, 표본분산이 과소추정된다면 통계적으로 식 (5-1)이 성립한다.

$$\frac{(n_{ij}-1)S_{ij}^2}{\sigma_i^2} \sim k\chi^2(n_{ij}-1) \quad (5-1)$$

여기서, k 는 1보다 작은 임의의 실수.

여기서, $k\chi^2(n_{ij}-1)$ 의 기대값은 $kE\{\chi^2(n_{ij}-1)\}$ 이고, 이것은 $k(n_{ij}-1)$ 과 같다. 이 때, $\chi^2(k(n_{ij}-1))$ 의 기대값은 $k(n_{ij}-1)$ 이므로, $k\chi^2(n_{ij}-1)$ 과 동일한 분포임을 알 수 있다. 이는 과소추정된 표본분산이 실제로 비해 작은 자유도를 갖는 χ^2 -분포를 따름을 의미한다.

따라서, 모분산에 대해 과소추정된 표본분산의 비를 일정하게 만들기 위해서는 식(5-2)이 성립해야 하며, 이 식은 일정하게 정해진 상수여야 한다.

$$\frac{S_{ij}^2}{\sigma_i^2} = \frac{\chi^2(k(n_{ij}-1))}{n_{ij}-1} = C \quad (5-2)$$

여기서, C 는 1보다 작은 상수.

식 (5-2)으로부터 표본의 크기 n_{ij} 는 다음과 같이 결정된다.

$$\begin{aligned} E\left\{\frac{\chi^2(k(n_{ij}-1))}{n_{ij}-1}\right\} &= \frac{E\{\chi^2(k(n_{ij}-1))\}}{E\{n_{ij}-1\}} \\ &= \frac{k(n_{ij}-1)}{E(n_{ij})-1} = C \end{aligned} \quad (5-3)$$

식 (5-3)은 $k(n_{ij}-1) = C(E(n_{ij})-1)$ 인 경우에 성립한다.

그러므로, χ^2 -분포로부터 일정한 값을 기대하기 위해서는 자유도, 즉, 표본의 크기가 고정되어야 한다. 따라서, $n_{ij}-1$ 이 모든 클래스의 모든 트레이닝 영역에 대해 일정하다면, 모분산에 대한 표본분산의 비율이 일정하게 유지될 것을 기대할 수 있다.

이와 같은 가설을 검증하기 위해 실제로 하나의 트레이닝 영역에 25개의 화소가 포함되도록 30개의 트레이닝 영역을 논과 밭 클래스에 대해 실시하여 결과를 비교하였다.

표 2에 제시된 두 집단의 각 밴드가 동일한 집단인지 여부를 Z-검정을 이용하여 유의수준 5%에서 검정하였다.

귀무가설 $H_0 = \mu_n - \mu_m$ 이며, 유의수준 5%의 기각역은

$$|Z| > 1.96 \text{이고, 검정통계량은 } Z = \frac{(\bar{X} - \bar{Y})}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} \text{ 로부}$$

터 밴드 별로 각각 1.26, 1.26, 0.59, 0.76, -0.20 및 -0.29이므로, 귀무가설을 채택한다. 따라서, 모든 밴드에서 논과 밭 두개 클래스에 대한 표본분산 대 모분산의 비율이 일정하다고 할 수 있다.

결론적으로 모든 클래스 및 밴드에 대해 표본분산의 비율이 일정하게 되는 것을 기대하기 위해서는 각 클래스 별 트레이닝 데이터의 수는 대상지역의 특성에 따라 비례하도록 결정하되, 각 트레이닝 영역의 크기를 일정하게 하여 다수의 트레이닝 영역을 선정하여야 한다. 이런 트레이닝 기법을 통하여 각 클래스에 대하여 평균은 모평균에 수렴하고, 분산이 과소추정되는 비율은 일정한 값에 수렴하도록 할 수 있다.

6. 결 론

위성영상의 통계적 분류는 화소값이 가지는 통계치를 이용하는 것으로 이러한 통계치를 추정할 때 사용되는 트레이닝 데이터의 정확성은 전체적인 분류 결과에 큰 영향을 미치게 된다. 본 논문에서는 트레이닝 데이터에 의해 추정되는 통계치가 어떠한 특성을 가지며, 분류결과에 어떠한 방법으로 영향을 미치는지에 대하여 실험, 분석 및 연구하였다. 본 연구를 통하여 얻어진 결론은 다음과 같다.

첫째, 위성영상 데이터를 개별 클래스에 대해 분할한 모집단이 실제로 정규분포에 근사하나, 약산의 밀집도와 다침두를 가진다.

둘째, 일반적으로는 χ^2 -분포를 특성에 따라 트레이닝

표 2. 트레이닝 영역에 대한 표본분산 대 모분산의 비율의 비교

클래스 시행회수	밭						논					
	밴드 1	밴드 2	밴드 3	밴드 4	밴드 5	밴드 7	밴드 1	밴드 2	밴드 3	밴드 4	밴드 5	밴드 7
1	0.18	0.12	0.09	0.32	0.17	0.15	0.27	0.22	0.47	1.59	0.13	0.32
2	0.18	0.37	0.70	1.12	1.08	0.63	0.58	0.50	0.27	0.23	0.28	0.26
3	0.17	0.11	0.11	0.34	0.21	0.18	0.38	0.35	0.47	0.13	0.51	0.49
4	0.25	0.17	0.22	0.20	0.19	0.20	0.10	0.07	0.09	0.46	0.07	0.03
5	0.20	0.12	0.15	0.25	0.35	0.24	0.11	0.13	0.09	0.14	0.31	0.33
6	0.09	0.17	0.20	0.31	0.46	0.38	0.58	0.46	0.62	0.25	0.73	0.54
7	0.21	0.14	0.16	0.16	0.15	0.12	0.12	0.11	0.05	0.09	0.18	0.20
8	0.11	0.21	0.25	0.31	0.21	0.21	0.17	0.15	0.12	0.12	0.49	0.44
9	0.18	0.14	0.21	0.18	0.40	0.89	0.03	0.13	0.09	0.04	0.14	0.17
10	0.09	0.15	0.11	0.27	0.28	0.19	0.21	0.20	0.25	0.41	0.23	0.24
11	0.13	0.23	0.32	0.41	0.42	0.37	0.12	0.09	0.12	0.11	0.17	0.27
12	0.23	0.23	0.22	0.16	0.12	0.19	0.08	0.03	0.03	0.14	0.24	0.21
13	0.44	0.25	0.15	0.29	0.14	0.16	0.15	0.07	0.05	0.03	0.07	0.06
14	1.51	1.01	0.99	0.35	0.33	0.60	0.06	0.09	0.08	0.19	0.18	0.14
15	1.64	1.19	1.24	0.51	1.12	1.38	0.14	0.25	0.13	0.65	0.18	0.18
16	0.13	0.12	0.13	0.09	0.08	0.18	0.38	0.22	0.10	1.09	0.16	0.24
17	0.47	0.43	0.50	0.52	1.16	1.02	2.15	1.26	0.87	0.38	0.18	0.19
18	0.42	0.30	0.34	0.24	0.25	0.28	0.25	0.20	0.36	0.27	0.27	0.17
19	0.09	0.15	0.24	0.06	0.10	0.12	4.43	3.69	2.57	0.90	0.64	0.70
20	0.09	0.14	0.09	0.32	0.18	0.10	0.11	0.15	0.12	0.26	0.26	0.19
21	0.15	0.21	0.40	0.20	0.14	0.26	0.23	0.32	0.19	0.25	0.26	0.17
22	0.25	0.21	0.29	0.08	0.34	0.37	0.15	0.13	0.13	0.21	0.27	0.36
23	0.20	0.27	0.47	0.20	0.44	0.33	1.11	1.61	1.97	1.00	1.61	2.26
24	0.12	0.27	0.50	0.21	0.30	0.18	2.66	1.77	1.18	0.35	0.30	0.32
25	0.59	0.46	0.45	0.19	0.27	0.58	0.17	0.15	0.16	0.14	0.22	0.17
26	0.18	0.21	0.32	0.14	0.54	1.16	0.11	0.11	0.09	0.14	0.13	0.19
27	0.15	0.19	0.18	0.32	0.22	0.12	0.94	0.68	0.64	0.23	0.42	0.56
28	0.13	0.07	0.14	0.10	0.25	0.27	0.14	0.13	0.17	0.24	1.06	0.80
29	0.52	0.37	0.36	0.54	0.32	0.26	0.11	0.20	0.10	0.23	0.13	0.09
30	0.13	0.09	0.16	0.23	0.22	0.22	0.11	0.07	0.07	0.07	0.15	0.20
평균	0.31	0.27	0.32	0.29	0.35	0.38	0.54	0.45	0.39	0.34	0.33	0.35
분산	0.13	0.06	0.07	0.04	0.08	0.11	0.87	0.55	0.33	0.12	0.10	0.16

영역의 수가 늘면 분산 추정의 정확도도 높아지나, 첩도가 높은 클래스 및 밴드에서는 사용된 트레이닝 영역의 수와 무관하게 분산의 과소추정이 일어날 확률이 높다.

셋째, 분산이 과소추정되는 정도는 상황에 따라 다르기 때문에, 통계적 화소분류 시 분산이 과소추정된 클래스 방향으로 분류경계가 이동하는 영향을 주게 된다.

넷째, 분산 과소추정의 영향을 최소화하기 위해서는 개

별 트레이닝 영역의 크기를 일정하게 하고, 클래스 당 트레이닝 영역의 수를 조절하여 모집단의 비율에 유사하도록 트레이닝하여야 한다.

향후 연구과제는 SPOT XS; LANDSAT MSS, TM 등과 같은 여러 종류의 영상과 논, 밭 도시 외의 여러 항목에 대해 이상의 이론을 적용하여 실제 분류 정확도의 향상을 기대할 수 있는지를 입증하는 것이다.

참고문헌

1. 김우철 외 7명, “현대통계학”, 1988, 영지문화사.
2. 어양담, “위성영상 분류를 위한 트레이닝 정규화 알고리즘과 클래스 분리도 측정기법 개발”, 서울대학교 대학원 공학박사 학위논문, 1999.
3. Hixon, M. Schoez, D. and Fuhs, N., “Evaluation of Several Schemes for Classification of Remotely Sensed Data”, *Photogrammetric Engineering and Remote Sensing*, Vol. 46, No. 12, 1980, pp. 1547-1553.
4. Swain, P. H. and Davis, S.M. “Remote Sensing : The Quantitative Approach”, McGraw-Hill, 1978, pp. 159-164.
5. Van Deusen, P. C., “Unbiased Estimates of Class Proportions from Thematic Maps”, *Photogrammetric Engineering and Remote Sensing*, Vol. 62, No. 4, 1996, pp. 409-412.