

Context Probability를 Weighting Function으로 사용한 Contextual Classifier

노준경 · 박규호 · 김명환
한국과학기술원, 전기 및 전자공학과
(1985년 10월 1일 받음)

Contextual Classifier with the Context Probability as a Weighting Function

Jung Kyung Rho · Kyu Ho Park · Myunghwan Kim
Korea Advanced Institute of Science and Technology
(Received October 1, 1985)

Abstract

The current methods of estimating context distribution function in contextual classifier are to "classify and count," GTGM (ground-truth-guided-method) and unbiased estimator.³⁾ In this paper we propose a new contextual classifier whose context distribution is replaced by context probability that is estimated from transition probability. The classification accuracy increases considerably compared with the classical one.

요 약

contextual classifier의 context distribution function을 구하는 방법으로서 "classify and count" GTGM(ground-truth-guided-method), unbiased estimator³⁾ 방법 등이 현재 사용되고 있다. 본 논문에서는 지금까지 사용된 context distribution 대신에 context probability라는 새로운 weighting function을 사용하는 contextual classifier를 제안한다. 이

와 같은 방법을 사용함으로써 classification의 정확도를 종전보다 상당히 높일 수 있음을 실험을 통해 보였다.

I. 서 론

일반적으로 원격탐사에 있어서 classification에 대한 노력은 각각의 ground resolution elements에 대한 multispectral pixel value에 pattern recognition 기법을 어떻게 적용시킬 것인가에 모아졌다. 근래에 와서 높은 classification의 정확도가 요구되는 부문에서 일반적인 spectral 한 정보뿐만 아니라 공간적인 정보도 포함시키려는 노력이 계속되고 있다. 공간적 정보라 하면 예를 들어 형태, 질감, 구조적 관계 등이 있다.

이러한 공간적인 정보를 포함시키는 classifier로서 contextual classifier는 구분하려고 하는 pixel의 주변 pixels의 정보를 이용하는 방법이다. 이와 같은 contextual classifier의 정확도는 분류 결정 함수의 weighting 함수인 context distribution을 어떻게 구하느냐에 크게 좌우된다. 본 논문에서는 이 context distribution을 context probability라는 새로운 weighting function으로 대체함으로써 classification의 정확도를 종전보다 상당히 높일 수 있는 방법을 제시한다. context probability는 transition probability로부터 구할 수 있으며, 이것은 공간적 영역에서 non-stationary한 특징을 가진다.

확률적인 분류의 관점에서 본다면, spectral한 특징에다 공간적 정보가 더해진 것이 정확하게 분류될 기회가 많은 것은 당연하다. 그러나, 이와 같은 경우 정보의 양이 많아지고 이것을 다루는 방법 역시 복잡해짐에 따라 계산량은 많아지게 된다. 따라서, 이러한 경우 multiprocessor-system 등의 고속처리 장치가 요구되며, 이와 관련된 연구가 필요하다.⁵⁾ II-1에서는 contextual classifier의 이론적 배경에 대한 설명을 하고 II-2에서 본 논문이 제시한 non-stationary contextual classifier(NSCC)에 대한 것과 이전의 contextual classifier들이 사용한 context distribution을 구하는 방법들에 대한 설명이 있다. II-3는 실험결과를 나타내 주고 있다.

II. 본 론

1. Contextual Classifier

contextual classifier의 이론적 배경을 볼 것 같으면 다음과 같다. multispectral image data는 $N = N_1 \times N_2$ 의 multivariate pixels의 행렬로 이루어진다. L 을 인접 pixel의 갯수, m 을 구분의 수, N_L 을 인접 pixels의 집합으로 정의하면,

image data set $X = \{X(S), S \in \Omega\}$,
 공간 집합 $\Omega = \{S = (i, j), 1 < i < N_1, 1 < j < N_2\}$,
 구분 집합 $K = \{W_1, W_2, \dots, W_m\}$

으로 정의할 수 있다. 위치 S에서의 인접 pixels의 집합은 X(S) pixel로부터 가까이 있는(반드시 이웃은 아니라도) L-1개의 pixels로서 정의될 수 있으며 그림 1이 두 가지 예를 보인다.

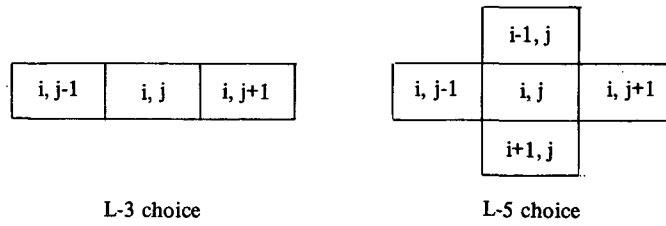


그림 1. Examples of neighborhood set

위치 S의 pixel을 분류하는 것을 C_s 라고 표시하면, X(S)의 올바른 분류를 θ_s 로 할 때 C_s 에 의해서 생기는 loss를 어떤 정해진 양의 함수 $L(\cdot, \cdot)$ 로부터 $loss = L(\theta_s, C_s)$ 로 표시할 수 있다. 따라서, 전체 image에 대한 loss는 전체 N개의 pixels에 대한 평균으로 나타낼 수 있다.

$$Loss = \frac{1}{N} \sum_s L(\theta_s, C_s)$$

위의 식에다 인접 pixels의 영향을 고려한다면 전체 image에 대한 기대 loss는 $R(X) = \frac{1}{N} \sum_s E[L(\theta_s, C_s(\bar{X}_s))]$ 과 같이 된다. 이때 \bar{X}_s 는 인접 pixels을 포함한 L-vector이다. 만일, 인접 pixels들끼리 class conditionally independent하다면, $L(\cdot, \cdot)$ 함수를 0-1 손실함수로 취할 때 다음과 같은 결정함수 $g_k(\bar{X}_s)$ 를 최대화시키는 분류 W_k 로 X_s 가 분류될 때 기대손실은 최소가 되고, 이것이 contextual classifier의 분류 방법이 된다.

$$g_k(\bar{X}_s) = \sum_{\substack{V_P \in K_P \\ V_P = K}} G(V_P) \prod_{k=1}^L f(X_s / W_k)$$

V_P : 가능한 분류의 방법의 vector (인접 pixels 포함)

$G(V_P)$: 분류하려는 image에서의 V_P 와 같은 형태의 구성이 나타나는 상대빈도

$f(X_s / W_k)$: class conditional densities

이때, 모든 pixels은 class conditionally independent하다고 가정하고, $f(X_s / W_k)$ 는 다

변수의 Gaussian 함수로부터 구할 수 있다. 또한, 일반적으로 $G(\underline{V}_P)$ 는 noncontextual classifier 의 결과로부터 얻을 수 있다.

2. Non-stationary contextual classifier

contextual classifier 가 처음 발표될 당시에는 그것의 우수성을 보이기 위해 simulated data set 을 사용했다. 즉, simulation 을 위해 마련된 template 로부터 context distribution 을 작성했다. 그러나, 실제의 data 를 다루는 경우, 이러한 template 를 얻을 수 없으며, 따라서 일반적인 방법으로 training sample 에 대하여 non contextual classifier 를 거쳐 이것의 결과로부터 context distribution 함수를 구하는 “classify and count” 방법이 많이 쓰인다. 그밖의 GTGM, unbiased estimator 방법 등이 있는데, GTGM은 ground truth data 를 구해, 그것으로부터 context distribution 을 구하는 것이고, unbiased estimator 는 $MSE_{VP} = E[T_{VP}(X) - G(\underline{V}_P)] = Var[T_{VP}(X)] + b_{VP}$ 로 주어지는 mean squared error 를 최소화시키는 estimator $T_{VP}(X)$ 를 찾는 방법이다. 자세한 것은 3) 을 참조하기 바란다.

이와 같이 지금까지의 방법에서는 $G(\underline{V}_P)$ 는 전체 data set 또는 block data set 로부터 구했으며, 이것은 context distribution 이 image 전체 내지는 한 block 안에서 일정하다는 가정이다. 그러나, 실제에 있어서 context distribution 은 place to place, 더 나아가 pixel to pixel 변하리라는 것을 예상할 수 있다. 본 논문에서는 pixel to pixel 변화하는 즉, non-stationary 하는 context probability $P_c(\underline{V}_P(S))$ 를 제시한다. $P_c(\underline{V}_P(S))$ 는 위치 S에서 \underline{V}_P 의 분류 구성이 나타날 확률을 의미하며, 4 개의 인접 pixel 을 그림 2 와 같이 고려했을 때 아래와 같이 된다.

$$P_c(\underline{V}_P(S)) = P(W_b(S)/W_a(S_1)) * P(W_b(S)/W_a(S_2)) \\ * P(W_b(S)/W_b(S_3)) * P(W_b(S)/W_c(S_4)) \\ \underline{V}_P = (b, a, a, b, c)$$

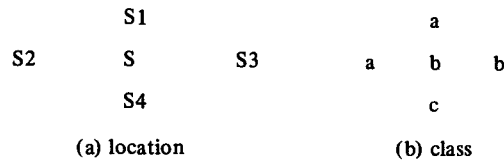


그림 2. Context formation

이때 $P(W_b(S)/W_a(S_1))$ 등은 transition probability 로서, 일반적으로 $P(W_k(S)/W_l(S+(i, j)))$ 는 $S+(i, j)$ 위치 $\{(i, j) \in N_L\}$ 의 pixel 이 W_l 이라고 가정했을 때 S 위치의 pixel

이 W_k 일 확률을 나타내 주는 것이다. 위의 식에서 볼 수 있듯이 context probability $P_c(V_P(S))$ 의 의미는 위치 S 에서 V_P 와 같은 구성을 가질 수 있는 확률이 되겠다. 이 context probability는 NSCC의 분류결정 함수의 weighting function 역할을 하게 된다.

Transition probability는 다음과 같이 구할 수 있다.” 원격탐사에서 사용되고 있는 multispectral image data는 Gaussian distribution을 가진다는 가정이 보편화되므로 class conditional densities $P(X(S)/W_k)$ 는 multivariate Gaussian function으로 표시할 수 있다.

$$P(X(S)/W_k) = \frac{1}{(\sqrt{2/\Sigma_k/\pi})^{q/2}} \exp\left\{-\frac{1}{2}(X(S)-M_k)^T \Sigma_k^{-1}(X(S)-M_k)\right\}$$

q : dimension of data

M_k : class means

Σ_k : class covariance matrices

Landsat data processing에 있어 일반적으로 Landsat II MSSdata는 4 band이나 processing time 등의 문제로 band를 2 band로 줄여서 사용한다. 본 논문에서는 Karhunen-Loeve transformation을 사용하였으며, 따라서 q 는 2가 된다. 또한, 평균 M_k 와 covariance matrices Σ_k 는 data set에서 먼저 training sample을 선정하고 이 sample에 대한 mean과 covariance matrices를 구해 그 class를 대표한다. 이와 같이 구한 class conditional densities $P(X(S)/W_k)$ 와 각 class의 a priori weight $P(W_k)$ 로 initial pixel label probability 또는 a posteriori 확률을 다음과 같이 구한다.

$$P_o(W_k(S)) = P(W_k/X(S)) = \frac{P(X(S)/W_k) P(W_k)}{\sum_{e \in c} P(X(S)/W_e) P(W_e)}$$

이때 a priori weight $P(W_k)$ 는 일반적으로 uniform a priori weight를 사용한다. 위의 initial pixel label 확률로부터 transition probability는 아래의 식으로서 구할 수 있다.

$$P(W_k(S) | W_e(S + (i, j))) = 1/36 \sum_{S_1 \in S_9} \sum_{S_2 \in S_4} P_o(W_k(S + S_1)) P_o(W_e(S + S_1 + S_2)) \\ 1/25 \sum_{S_3 \in N_{25}} P_o(W_e(S + (i, j) + S_3))$$

N_{25} : $S + (i, j)$ pixel에 중심을 가진 25개의 인접 pixels

N_9 : S 에 중심을 둔 9개의 인접 pixels

N_4 : $S + S_1$ 에 중심을 둔 4개의 인접 pixels

따라서, context probability는 아래와 같이 transition probability의 곱으로 표시할 수 있으며, 이때 계산에 포함된 pixel의 window는 그림 3과 같다.

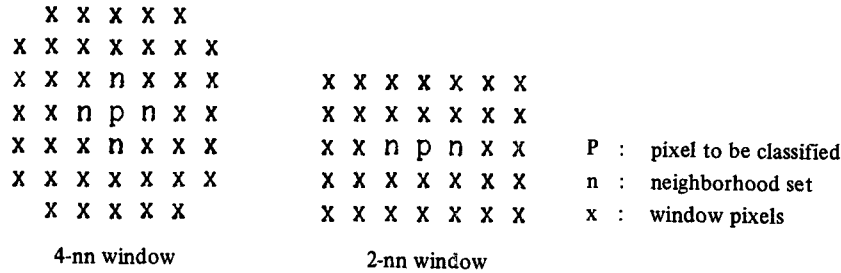


그림 3. Nonstationary window

$$P_c(\underline{V}_P(S)) = \prod_{i,j \in NL}^{L-1} P(W_k(S) | W_\ell(S+(i, j)))$$

그림과 같이 window의 크기는 인접 pixel의 개수를 달리함으로써 변하며 일반적으로 4-nearest neighborhood가 많이 쓰인다. 또한, window가 설정됨에 따라 분류에 있어서 adaptive한 분류가 가능함은 물론이다.

3. 실험 결과

실험은 KAISMIPS(Korea Advanced Institute of Science and Technology Micro-computer Image Processing System)에서 이루어졌는데, KAISMIPS는 MC 68000 CPU를 가진 Cromemco system과 Ramtek 9300 graphic display system, Trilog color plotter 등으로 이루어져 있다.

i. Real data

실제 data에 대한 실험대상으로 1981년에 찍혀진 Seoul 지방의 Landsat MSS data를 선택했다. 이 scene은 50×50 pixel로 이루어져 있으며, 4가지의 분류 가능지역으로 이루어져 있다. 이 4가지 지역은 강, 도심, 작물지대, 맨 땅 등이며 이 data에 대한 Bayes classifier, contextual classifier(2-nn, 4-nn), Nonstationary contextual classifier에 대한 결과가 Table 1 및 그림 4에 있다. Table 1에서 볼 수 있듯이 NSCC(4-nn)의 경우, Bayes classifier보다 10% 이상 contextual classifier(classify and count)보다 8% 이상 정확도가 증가됨을 알 수 있다. 이때 overall accuracy란,

$$\text{정확히 분류된 pixel 수} / \text{전체 pixel의 수} \times 100(\%)$$

이고 class by average 는 각 class 의 % 정확도의 평균값이다.

Table 1. Comparison of classifier accuracy (Real data)

DATA SET	CLASSIFIER	OVERALL ACCURACY	CLASS BY AVERAGE
Landsat 2	BAYES (uniform a priory)	67.1%	75.4%
	CONTEXTUAL (2-nn) (classify count)	72.8%	78.2%
Seoul, 1981 50x50	CONTEXTUAL (4-nn) (classify count)	74.1%	79.3%
	NSCC (2-nn)	79.3%	85.1%
	NSCC(4-nn)	83.1%	87.1%

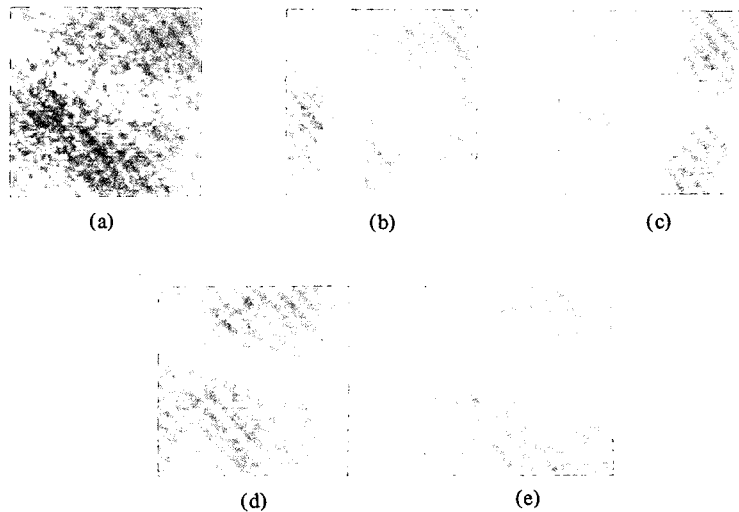


그림 4. Visual comparison of real data classification results

a) Bayes b) contextual (classify count) L=3 c) Contextual L=5 d) NSCC L=3 e) NSCC L=5

ii. Computer generated data

실제 data 에 있어서 정확도는 template 를 만들 수 없으므로, 100 % 정확하다고 생각할 수 없으므로 computer generated data set 를 만들어 실험을 해 보았다. Data generation 의 방법은 실제 data 의 training sample 로부터 mean 과 variance 를 구하고, 이것을 가지고 Gaussian random number generator 로부터 image data 를 만든다. 따라서, computer generated data set 도 실제 data 의 class 별 mean, variance 와 비슷하며 Gaussian 분포를 가진다. Table 2 는 이 data 의 Bayes classifier (uniform a priori, estimated a priori), con-

textual classifier(2-nn), NSCC(2-nn)에 대한 실험 결과를 나타내 주고 있다. 이 경우도 실제 data와 마찬가지로 NSCC의 정확도가 Bayes classifier나 Contextual classifier의 그것보다 상당히 높음을 알 수 있다. 실제 실험에 있어서 50×50 pixel의 image를 분류하기 위해서는 4-nn의 경우 window 때문에 56×56 의 image가 필요하다.

Table 2. Comparison of classifier accuracy (Simulated data)

DATA SET	CLASSIFIER	OVERALL ACCURACY	CLASS BY AVERAGE
SIMULATED 50x50	BAYES (uniform a priory)	90.6%	81.3%
	BAYES (estimated a priory)	91.1%	81.3%
	CONTEXTUAL (2-nn) (classify count)	96.9%	93.4%
	NSCC (2-nn)	99.3%	98.4%

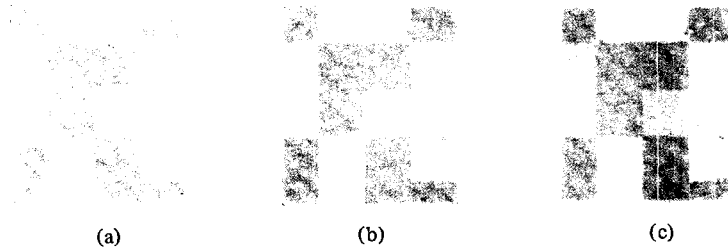


그림 5. Visual comparison of simulated data classification results

a) Bayes b) contextual (classify count) L=3 c) NSCC L=3

Ⅲ. 결 론

본 논문에서는 Non-stationary contextual classifier로 불리는 새로운 contextual classifier가 제시됐다. 이 방법은 종전의 contextual classifier의 context distribution function 대신에 context probability를 weighting function으로서 사용하였으며, 이것은 각각의 위치에서 변화하는 즉, 공간적 의미에서 non-stationary한 특성을 갖는다. 위의 방법은 adaptive한 적용이 가능하고 computer generated data 및 서울지방의 Landsat II MSS data를 가지고 실험해 본 결과, 종래의 contextual classifier("classify and count")나 Bayes classifier보다 상당히 좋은 결과를 얻을 수 있었다. 따라서, non-stationary contextual classifier는 특별히 높은 정확도가 요구되는 분야에 유용하게 쓰일 수 있다. 단지, 이 방법은 다

루는 data의 양이 많고 복잡한 계산식을 사용함에 따라 processing time이 많이 걸리나 computer component의 발전과 multiprocessor 등의 출현으로 그다지 큰 문제가 되지는 않으리라 여겨진다.

Reference

- 1) 노준경, 공간 및 주파수 대의 특성을 이용한 영상 분할, 전기 및 전자공학과, KAIST, 1985.
- 2) H. M. Kalaych, and D. A. Landgrebe, "Adaptive relaxation labeling," IEEE Trans, *Pattern Analysis and Machine Intelligence*, Vol. PAMI 3, pp.369-372, May 1984.
- 3) J. C. Tilton, S. B. Vardeman, and P. H. Swain, "Estimation of context for statistical classification of multispectral image data," IEEE Trans. *Geosci. and Remote Sensing*, Vol. GE-20, pp.445-451, Oct. 1981.
- 4) P. H. Swain, S. D. Vardeman, and J. C. Tilton, "Contextual classification of multispectral image data," *Pattern Recognition*, Vol. 13, No. 6, pp.429-441, 1981.
- 5) P. H. Swain, H. J. Siegel, and B. W. Smith, "Contextual classification of multispectral remote sensing data using a multiprocessor system," IEEE Trans. *Geosci. and Remote Sensing*, Vol. GE-18, pp.197-203, Apr. 1980.
- 6) D. A. Landgrebe, "The development of a spectral-spatial classifier for a earth observation data," *Pattern Recognition*, Vol.12, No.3, pp.165-175, 1980.