

人工衛星 遠隔探査 데이터의 分析 正確度 向上에 關한 研究
—分類過程에서의 Bayesian MLC 適用을 中心으로—
Improving Correctness in the Satellite Remote
Sensing Data Analysis

—Laying Stress on the Application of Bayesian MLC
in the Classification Stage—

안 철 호 김 용 일
Ahn, Chul-Ho Kim Yong-Il

요 旨

본 연구에서는 원격탐사 수치화상 데이터의 분류단계에 가중치를 고려한 Bayesian MLC를 적용하여 그 분석 정확도를 향상시키고자 하였다.

우선, Bayesian 결정법칙을 원격탐사분야 측면에서 분석해 보고 정규확률 밀도함수를 이용하여 n 차원으로 확장시켰다.

이 유도과정에서 정의되는 사전확률 항에, 평행육면체 분류결과를 가중치로 적용하여 분류를 실행하였다. 그리고 최종적 분류정확도는 확률함수데이터에 χ^2 분포를 가정한 임계치 처리를 하므로써 오분류확률이 높은 화소를 추출하여 그 양을 기준으로 평가하였다.

연구의 전체 처리과정에 사용한 인공위성 데이터는 LANDSAT TM(1985년 10월 21일; 116-34)이며 연구 대상지역은 서울시 행정구역 내이다.

가중치를 적용해 본 결과 5.21%의 분석정확도 향상을 이루었으며, 따라서 본기법은 도시 지역과 같이 복잡한 분포특성을 가지는 지형에 효과적으로 활용될 수 있다고 생각된다.

ABSTRACT

This thesis aims to improve the analysis accuracy of remotely sensed digital imagery, and the improvement is achieved by considering the weight factors(a priori probabilities) of Bayesian MLC in the classification stage.

To be concrete, Bayesian decision theory is studied from remote sensing field of view, and the equations in the n -dimensional form are derived from normal probability density functions.

The amount of the misclassified pixels is extracted from probability function data using the thresholding, and this is a basis of evaluating the classification accuracy.

The results indicate that 5.21% of accuracy improvement was carried out. The data used in this study is LANDSAT TM(1985.10.21; 116-34), and the study area is within the administrative boundary of Seoul.

1. 서 론

1970年代以後로 현재까지, Maximum Lik-

* 서울대학교 공과대학 명예교수

** 서울대학교 공과대학 시간강사

elihood Classification(MLC)은 分類枝法의 하나로 遠隔探査 分野에 널리 사용되어 왔다.^{5),11)}

遠隔探査 分野에서는, 航空機나 人工衛星에서 多波長帶의 디지털 映像을 얻어 내는 다중분광 스캐닝 기술의 발달에 따라, 이들 數值映像으로 土地利用/土地被覆과 같은 主題分類圖(them-

atic classification maps)를 제작하기 위해 MLC를 이용하는 境遇가 많아졌다.¹²⁾

1972年 7月 地球資源探查衛星 LANDSAT-1號가 發射된 이후, 遠隔探查 디지털 데이터의 패턴認識이나 分類 過程에서 이 MLC 技法이 주된 역할을 하여왔다.⁷⁾

그 대표적인 예로, 1978年 美航空宇宙局(NASA)에서 LANDSAT MSS 데이터로 Denver市의 空間的 土地利用 모델을 제작함에 있어서 이 MLC를 주된 分類技法으로 이용하였고(NASA Technical Memorandum 79710), 1980年 NATO Advanced Study Institute에서 발간한 報告書의 數値映像處理(Digital Image Processing)分野를 보면 遠隔探查의 映像分析에 있어 決定理論(decision theory)으로서 Bayesian ML C를 이용하고 있다.¹³⁾

그런데 最近의 例를 들어, '91年 1月 PE & RS에 掲載된 "Rapid MLC"라는 研究論文(T. M. Lillesand)에서 Parallelipiped 分類와 LUT 接近 方法을 이용하여 處理速度가 향상된 MLC를 제안하고 있을 만큼 MLC는 아직도 理論的으로 발전되고 있고, 改善의 餘地를 남기고 있는 理論으로 여겨진다.²⁾

따라서 本 연구에서는 遠隔探查 데이터를 MLC로 처리함에 있어서, 分類正確度를 향상시키기 위하여 事前確率을 導入하고, 이 加重值를 적용한 分類를 실행하고자 한다.

또한, 分類 決定을 함에 있어 核心的 役割을 하는 Bayes 이론을 定義하고, 最大尤度를 測定하는 기준인 確率密度函數를 1次元에서 n-次元으로 확장하는 誘導를 해 보고자 한다.

이러한 統計的 接近方法이 遠隔探查 데이터 처리에 반드시 도입되어야 하는 理由는 다음과 같다.

① 遠隔探查는 어떤 특정한 人工物만을 대상으로 하는 것이 아니므로, 그 데이터는 모든 自然物이 지나는 任意性(Randomness)에 의해, 분석하고자 하는 分類項目間의 特徵的 差異를 흐리게 하는 偶發的인 變動(Variation)을 포함하고 있다.

統計的 接近方法은 이러한 變動要素를 분석하

여 分類正確度에 영향을 미칠 수 있는 可能性을 감소시키는 데 도움이 된다.

② Training 패턴을 選定하여 判別函數를 결정함에 있어 비록 사소한 정도일지라도 간혹 不確實性(Uncertainty)이 존재하게 된다.

이러한 不確實性이 너무 심하지 않다면, 統計的 理論을 도입하여 이에 따른 誤差發生 可能性을 감소시킬 수 있다.

③ 分類하고자 하는 分類項目의 패턴은 測定空間(measurement space)上에서 실제 중복되어 나타날 수 있으며, 이러한 分類項目 間의 重複部分은 식별하기 어려워진다.

이 境遇, 統計的 패턴認識 技法을 적용하여 分類項目 間의 分離度를 높여 分類正確度를 향상시킬 수 있다.

2. 본 론

2.1 Bayesian 최대우도분류의 적용개념

원격탐사 데이터의 分類段階는 前處理, 強調段階와 함께 수치영상 처리분야의 한 영역을 차지하고 있다.¹⁸⁾²⁰⁾

畫像의 分類段階에서는 意思決定過程이 全的으로 컴퓨터에 의해 이루어지며, 그 目的은 判讀者의 主觀的이고 애매모호한 解析 대신 컴퓨터에 의한 定量的이고 反復的인 解析을 함으로써 보다 信賴할 수 있고 正確한 結果를 얻고자 하는 것이다. 또한 컴퓨터를 이용한 人工衛星 데이터의 分類方法은 廣域의 地表面을 效率的이고 一貫性 있게 도면화하고 관리할 수 있는 역할을 한다.²³⁾

數値映像 分類는 상당한 統計的 可變要素를 지니고 있는 데이터를 다루게 되므로, 統計的 決定理論과 같은 수학적 기법에 의존할 수 밖에 없다.

分類項目 群集內에서 또는 分類項目 間의 重複과 같은 데이터의 可變性을 나타내게 하는 要素로는 여러가지가 존재한다.¹⁵⁾

그런데, 한 畫素를 어떤 特定 項目으로 分類하기 위한 결정 과정은, 어느 정도 誤差 發生의

確率을 지니는, 統計學的인 「推測」에 지나지 않는다.

따라서, 수많은 畫素를 개별적으로 分類함에 있어 각 畫素가 최소의 誤差 發生可能性을 가지도록 결정하는 것이 論理的으로 要求된다.

本質的으로 이러한 특성을 만족시키고, 數學的으로 해결할 수 있는 分類理論이 바로 최대우도 혹은 Bayes 최적화 분류이다.¹⁴⁾

1) Bayes 결정법칙

遠隔探査 데이터로 구성된 어떤 數值映像의 特性分布(예 : 各 畫素의 gray level)를 測定하고, 각 畫素가 두 分類項目(예 : 植物, 土壤) 중 어디에 속하느냐를 결정해야 한다면, 이 문제는 1次元 2-分類項目 分類의 境遇가 된다.

만일 각 分類項目을 대표할 수 있는 畫素 集合을 抽出한다면(예 : Training 데이터), 相對的 週期 히스토그램을 構成할 수 있으며, 이들은 무한의 데이터 집합에서의 連續的 確率密度 函數로 近似시킬 수 있다.

어떤 畫素에 대해 分類 決定을 하기 위해서는 事後 確率(畫素가 특성치 X를 가질 때 ω_i 에 속할 確率)을 알아야 하며 이 確率 $P(\omega_i|X)$ 는 Bayes 定理에서 誘導된다.¹⁹⁾ (1次元 2-分類項目 分類의 境遇)

$$P(\omega_i|X) = P(X|\omega_i)P(\omega_i)/P(X) \dots\dots(2-1)$$

$$\text{단, } P(X) = \sum_{i=1}^2 P(X|\omega_i)P(\omega_i)$$

식 (2-1)에서 $P(X)$ 는 각 分類項目에서 동일

하므로 다음과 같은 Bayesian 결정법칙이 유도된다.

a pixel belongs to class 1 if

$$P(X|\omega_1)P(\omega_1) > P(X|\omega_2)P(\omega_2)$$

a pixel belongs to class 2 if

$$P(X|\omega_2)P(\omega_2) > P(X|\omega_1)P(\omega_1) \dots\dots(2-2)$$

Bayesian 決定法則에서, $P(X|\omega_i)P(\omega_i)$ 를 判別函數라고 하며 D_i 로 표시한다.

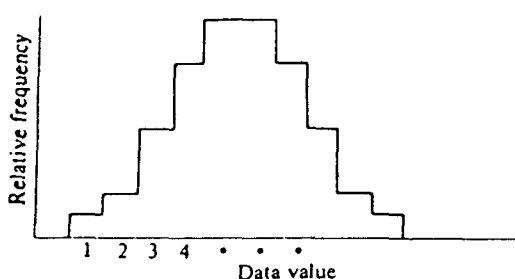
判別函數는 $D_i(X)$ 의 값이 가장 큰 境遇 i 分類項目으로 X 가 分類되도록 하는 역할을 하게 된다.

2) 정규확률밀도함수를 이용한 n차원에서의 확장

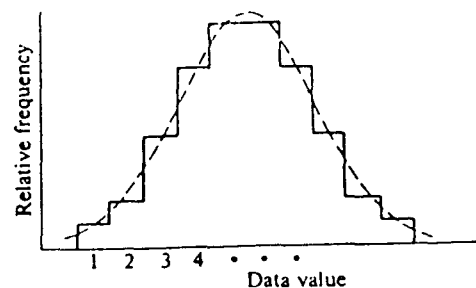
遠隔探査 데이터는 그 volume이 매우 크기 때문에 分類項目에서의 確率函數에 따른 히스토그램 표현을 컴퓨터에 貯藏하기 위해서는 많은 量의 메모리가 必要하다.²¹⁾ (예 : LANDSAT TM의 subscene 境遇 : line×colume=2,982×3,460(pixels)=1.03×10⁷(pixels), 1pixel=8bit=1byte이므로 1 밴드 當約 10Mbyte의 容量이 必要. TM에서 7 밴드가 있으므로 총 所要容量은 約 70 Mbyte)

더우기 여러 波長帶(Band)를 同時에 사용하여 MLC를 실행하고자 한다면 메모리의 量은 거의 감당할 수 없다.

이러한 問題를 극복하기 위한 方法들 中 하나는 간단한 函數 形態의 分布曲線으로 각 히스토



(a) 1 채널의 데이터 分布(히스토그램)



(b) 正規確率密度函數에 의한 데이터 分布

그림 2-1 1次元 데이터 分布와 그 推定曲線

그림이나 確率函數를 추정하는 것이다. 일반적으로 MLC에서는, 각 分類項目에 대한 確率函數를 正規確率密度函數로 定義하여 표현하고 있다.

그 이유는 다음과 같다.

① 遠隔探査의 여러 應用分野에서, 確率論的 처리를 할 때 「多變量 正規分布」로써 데이터를 模型化하는 것이 適合하였다.³⁾

② 만일 데이터 分布가 이러한 假定에서 어느 정도 벗어나는 境遇가 있더라도, 分類正確度에는 큰 영향을 주지 않기 때문에 상당히 安定的이다.⁶⁾

③ 실제로, 여러 알고리즘의 分類方式을 적용하였을 때 「多變量 正規分布」의 가정은 分類作業의 遂行能力(正確度)이나 費用側面(速度, 알고리즘의 複雜性 등)에서 가장 적절하다.³⁾

1次元 또는 單一變量(univariate)의 境遇, 分類項目 i에 대한 正規確率密度函數는 다음 식과 같다.¹⁴⁾

$$P(X|\omega_i) = \frac{1}{(2\pi)^{1/2}\sigma_i} \exp\left[-\frac{1}{2} \frac{(X-\mu_i)^2}{\sigma_i^2}\right] \quad \dots\dots(2-3)$$

여기서 X=全體 測定空間에서의 測定값

ω_i =分類項目 i를 의미

exp=自然代數의 밑

$\mu_i=E[X|\omega_i]$ 分類項目 i內的 測定값의 平均(기대값)

$\sigma_i^2=E[(X-\mu_i)^2|\omega_i]$ 分類項目 i內的 測定값의 分散

실제로 μ_i 와 σ_i^2 은 未知값이며 分類實行前段階의 Training 샘플에서 推定되어야 한다.

2채널 다중분광 데이터의 境遇, 각 分類項目에 대한 2變量(bivariate) 確率函數는 각 채널의 모든 가능한 쌍의 데이터 값의 發生頻度數를 測定함으로써 推定할 수 있다.

2次元(2變量) 正規密度函數는 (2-4)와 같이 계산된다.

$$P(x_1, x_2|\omega_i) = \frac{1}{2\pi(\sigma_{i11}\sigma_{i22}-\sigma_{i12}^2)^{1/2}}$$

$$\times \exp\left\{-\frac{1}{2} \frac{\frac{(x_1-\mu_{i1})^2}{\sigma_{i11}} - \frac{2\sigma_{i12}(x_1-\mu_{i1})(x_2-\mu_{i2})}{\sigma_{i11}\sigma_{i22}} + \frac{(x_2-\mu_{i2})^2}{\sigma_{i22}}}{1 - \frac{\sigma_{i12}^2}{\sigma_{i11}\sigma_{i22}}}\right\} \quad \dots\dots(2-4)$$

여기서, $\mu_{ij}=E[x_j|\omega_i]$ 즉, 分類項目 i에 대한 채널 j의 데이터 평균값

$\sigma_{ijk}=E[(x_j-\mu_{ij})(x_k-\mu_{ik})|\omega_i]$ 즉, 分類項目 i에 대한 채널 j와 k간의 공분산(covariance)

그런데 (2-4)와 같이 n次元으로 確率密度函數를 전개해 나가자면 무척 복잡하므로, 벡터 및 행렬의 개념을 도입하여 식을 구성하는 것이 편리하다.

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad U_i = \begin{bmatrix} \mu_{i1} \\ \mu_{i2} \\ \vdots \\ \mu_{in} \end{bmatrix} \quad \Sigma_i = \begin{bmatrix} \sigma_{i11} & \sigma_{i12} & \dots & \sigma_{i1n} \\ \sigma_{i21} & \sigma_{i22} & \dots & \sigma_{i2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{in1} & \sigma_{in2} & \dots & \sigma_{inn} \end{bmatrix} \quad \dots\dots(2-5)$$

여기서, x=데이터 벡터

U_i =分類項目 i의 평균벡터

Σ_i =分類項目 i의 共分散行列

(2-5)로서 n次元 多變量(multivariate) 正規確率密度函數를 표현하면,

$$P(X|\omega_i) = \frac{1}{(2\pi)^{n/2} |\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2} (X-U_i)^T \Sigma_i^{-1} (X-U_i)\right\} \quad \dots\dots(2-6)$$

여기서, $|\Sigma_i|$ =共分散行列 Σ_i 의 determinant

Σ_i^{-1} = Σ_i 의 inverse matrix

$(X-U_i)^T$ =벡터 $(X-U_i)$ 의 transpose

(2-6)과 같이 平均 벡터 U_i 와 共分散行列 Σ_i 로 구성된 正規確率密度函數 $P(X|\omega_i)$ 는 $P(X|\omega_i) \sim N(U_i, \Sigma_i)$ 로 표시한다.

실제로, 각 分類項目의 平均 벡터와 共分散行列은 未知數이므로 training 패턴에서 추정되어야 하며, \hat{U}_i 와 $\hat{\Sigma}_i$ 를 U_i 와 Σ_i 의 不偏推定量이라고 하면 다음과 같다.

$$\hat{\mu}_{ij} = \frac{1}{q_i} \sum_{l=1}^{q_i} x_{jl} \quad j=1, 2, \dots, n$$

$$\hat{\sigma}_{ijk} = \frac{1}{q_i - 1} \sum_{l=1}^{q_i} (X_{jl} - \hat{\mu}_{ij})(X_{kl} - \hat{\mu}_{ik})$$

$$k=1, 2, \dots, n \quad \dots\dots(2-7)$$

지금까지, 正規確率密度函數를 定義하고 各 데이터를 벡터와 행렬의 형태로 표현하여 n-밴드 다중분광 데이터의 境遇에 바로 適用할 수 있게 誘導하였다.

여기서의 假定과 誘導過程에서 주의할 점은 다음과 같다.

① 各 分類項目에 대한 平均 벡터와 共分散行列을 推定하기 위해서는 적절한 테스트 샘플의 選定이 가능해야 한다.⁴⁾

確率抽出에서 추정량의 企待값과 추정하는 母數와의 차이를 偏倚(bias)라고 하는데, training 샘플의 선정이 적절치 못하면 偏倚推定量이 되어버리므로 이는 앞의 가정에 위배된다.²²⁾

理論적으로, n개의 波長 밴드가 分類에 이용된다면, 分類項目 當 필요한 최소의 training 샘플은 n+1이며, 더 적은 境遇에는 공분산 행렬의 determinant가 0가 되며 逆行行列도 계산할 수 없으므로 (2-6)의 계산이 불가능해진다.

실제로는, training 샘플의 갯수가 10n~100n개 정도일 때 媒介變數의 正確한 추정이 가능하다.^{16), 17)}

② 특이한 境遇, 分類項目 內의 데이터 分布가 分명한 multimodal 確率函數(두개 이상의 최대점을 가지는 函數)의 성질을 띠는 境遇가 있는데 이는 正規確率密度函數의 가정에 위배된다.

이러한 성향이 두드러질 境遇 해결방법은, MLC 分類實行에 들어가기 전에 群集化 分類 실행을 하여 multimodal 分類項目을 識別하고 이를 여러 subclass로 나누어 각각이 正規分布의 성질을 가지도록 만드는 것이다.

3) 判別函數의 誘導

분류가 실행될 전체집합에 대해 평균손실을 최소화하기 위해 Bayes Optimal 원칙에 근거하여 식을 유도하면 다음과 같다.

$$D_i(X) = P(X|\omega_i)P(\omega_i)$$

$$= \frac{P(\omega_i)}{(2\pi)^{n/2} |\Sigma_i|^{1/2}} \exp\left[-\frac{1}{2} (X-U_i)^T \Sigma_i^{-1} (X-U_i)\right] \quad \dots\dots(2-8)$$

事前 確率 P(ω_i)에 default 값 1.0을 주는 境遇,

$$D_i(X) = -\frac{1}{2} \ln |\Sigma_i| - \frac{1}{2} (X-U_i)^T \Sigma_i^{-1} (X-U_i) \quad \dots\dots(2-9)$$

일반적으로 이 값이 최대화 되는 分類項目으로 畫素벡터 X를 分類하거나, (2-10)이 최소화 되는 分類項目으로 分類하는 과정이 바로 Bayesian 最大尤度分類이다.

$$D_i(X) = \ln |\Sigma_i| + (X-U_i)^T \Sigma_i^{-1} (X-U_i) \quad \dots\dots(2-10)$$

P(ω_i)에 default 를 1.0으로 주지 않는 境遇는 가중치를 주어 다루게 된다.

2.2 誘導式의 分析

분석 1) 지금까지의 誘導過程에서 보면, 모든 데이터들은 分類過程을 거치면서 判別函數가 판단하는 어느 하나의 分類項目으로 결정지어지게 된다.

그런데, 遠隔探査 데이터 分類에 있어서 어느 分類項目에도 명확히 속하지 않는 데이터들이 發生하는 境遇가 있으며, 이는 주로 媒介變數를 추정하는 training 샘플 選定이 不適合하였거나 그 데이터가 주로 分布하는 分類項目을 무시했기 때문이다. 이러한 데이터는 判別函數 자체가 존재하지 않기 때문에 正確한 分類를 실행해 줄 수는 없지만, 그 分布特性으로부터 檢出해 낼 수 있다.

그 檢出方法으로는 ‘임계치처리’ 技法이 있는데, 이것은 分類된 分類項目과 데이터 벡터에 따른 函數값 P(X|ω_i)를 인위적으로 지정한 限界값과 비교하여, 確率값이 限界값 아래이면 그 데이터는 ‘reject class’로 分類가 되도록 하는 것이다.

분석 2) 本 誘導에서 사용한 決定原則인 Bayes Optimal에서, ‘zero-one loss function’을 적용하여 논리를 전개하였다.

이 境遇를 ‘최대우도 決定法則’이라고 하며, 遠隔探査 데이터 분석에 패턴인식을 적용할 때 가장 널리 쓰이고 있다.

대부분의 境遇 “全體 分類 데이터의 誤差發生

確率의 최소화”를 目的으로 하는 이 법칙은 가장 正確한 分類 技法으로 평가된다.

그러나, 이 법칙이 發生確率이 낮은 ($P(\omega_i)$ 값이 작은) 分類項目에 대해서는 열등한 分類性向을 나타내고 있으므로, 특별한 分類 目的의 境遇, 즉 드물게 나타내는 分類項目 特性을 分類하고자 할 때는 그 힘이 약하다고 볼 수 있다

이러한 境遇에는 특별한 分類目的의 特性을 반영하도록 zero-one loss function 보다 복잡한 형태의 損失函數를 사용하면 문제가 해결된다.

2.3 事前確率에 의한 加重值의 適用

Bayesian 最大尤度分類 理論에서 유도된 式 (2-8)의 $P(\omega_i)$, 즉 事前確率을 1.0으로 주지 않고 기존 分類結果로부터 예측되는 分類項目의 分布로부터 事前情報를 얻어 $P(\omega_i)$ 에 대입함으로써 分類正確度를 높이는 方法을 생각해 보았다.

式 (2-8)을 다시 보면, 判別函數 $D_i(X)$ 는

$$D_i(X) = \ln P(\omega_i) - \frac{n}{2} \ln 2\pi - \frac{1}{2} \ln |\Sigma_i| - \frac{1}{2} (X - U_i)^T \Sigma_i^{-1} (X - U_i)$$

여기서, $D_i(X) \leq D_j(X) = \ln D_i(X)$ 이며 k 개의 分類項目으로 分類를 하고자 할 때 $P(\omega_i)$ 는 k 개의 分類項目 중 i 分類項目이 發生할 事前確率을 의미한다.

이때, $P(\omega_i)$ 에 대한 정보는 자체적으로 미리 찾아낼 수는 없고, 이미 分類가 되어 있는 경우로부터 얻어야 한다.⁶⁾ (예: 土地利用圖, 土地被覆圖, 統計資料, supervised 및 unsupervised 分類結果 등)

本 研究에서는 平行六面體 分類를 실행하여 그 分類結果로부터 事前確率 $P(\omega_i)$ 를 抽出하여 各 分類項目 別 加重值를 지정하였다.

그리고 이 $P(\omega_i)$ 를 고려한 Bayesian 最大尤度分類를 실행하여 그 分類結果를 $P(\omega_i)$ 를 고려하지 않은 境遇와 比較評價하였다.

分類의 흐름도는 그림 (2-2)와 같다.

이 事前確率에서 抽出한 加重值를 이용하여 分類를 실행하는 境遇, 그 適用方式과 分野는

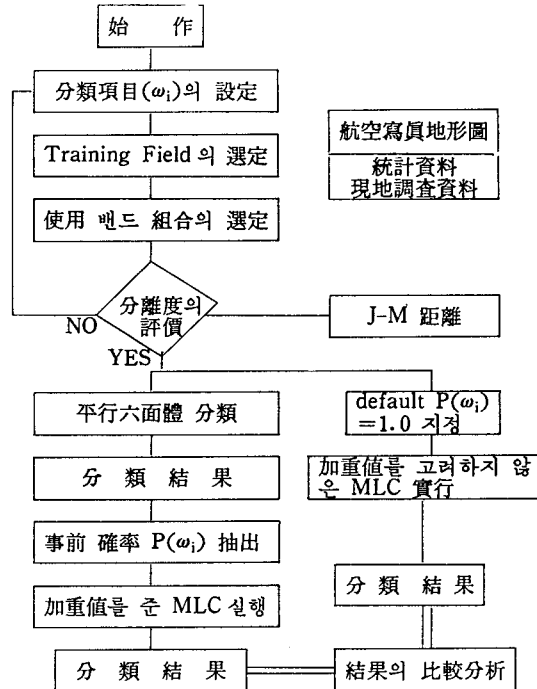


그림 2-2. 本 研究에서 實行한 分類過程의 흐름도

크게 나누어 다음 세 가지로 생각해 볼 수 있다
적용 1) 土地利用分類나 土地被覆情報 抽出에 있어, 이미 알고 있는 情報(土地利用圖, 統計資料 등)를 이용하거나, 비교적 간단한 分類方法(群集化, 平行六面體分類 등)을 미리 실행하여, 그 結果資料로부터 各 分類項目의 豫想되는 分布에 따라 加重值를 決定한다.⁹⁾

적용 2) 連續的 形態로 測定되는 情報(遠隔探査 數値映像이나 廣域 航空寫眞 등)와 不連續的으로 測定된 附加的 變數(岩石分布나 土壤分布 등의 情報)를 결합시켜 分類決定 過程에 도움을 줄 수 있다.

적용 3) 時間的(또는 季節的) 차이를 가지는 遠隔探査 데이터의 分類에 있어서, 附加變數 結合에 의한 時期的 加重值를 적용하기 위해 事前確率을 적용할 수 있다.

이러한 時間 順序的(time-sequential) 分類의 方法은 土地利用變化를 抽出해 내는 境遇에도 이용될 수 있으며, 이 때 Markov-type의 豫測 模型을 이용하여 변화패턴에 따른 事前確率을 抽出하는 것이 효과적이다.¹⁴⁾

本 研究에서는 적용 1)의 境遇, 즉 土地 被覆 情報를 이용한 土地利用圖 作成에 있어서 加重 值를 고려한 最大尤度分類의 分類正確度에 대한 效果를 찾아보려고 한다.

表 2-1. 入力加重值

CLASS	區 分	PIXEL 수	면 적 (km ²)	가중치
1	URBAN(LOW)	277,642	244.7	37
2	URBAN(HIGH)	135,784	110.2	18
3	GRASS	2,327	1.9	1
4	CROP	42,127	34.2	5
5	VINYL HOUSE	23,411	19.0	3
6	FOREST	219,346	178.1	29
7	BARE SOIL	20,433	16.5	3
8	WATER	29,573	24.0	4
합 계		750,463	608.6	100

2.4 分類正確度 評價

本 研究에서는, 統計的인 根據가 뒷받침되고, 個人的 判讀誤差의 可能性이 없을 뿐 아니라 所要時間, 經費面에서 월등히 效率的인 臨界值 處理技法을 適用하여 分類正確度를 評價하였다.¹⁰⁾

分類技法에 따른 分光의 距離測定에 의해 誤分類될 確率을 抽出하여 그 結果를 確率數值畫像의 形態로 나타낸다.⁸⁾

誤分類된 畫素는 높은 距離값을 가지므로, 確率數值畫像에서 밝게 나타낸다.

最大尤度分類의 境遇 距離測定값은 Mahalan-

obis 距離에 의한다.

$$M = (U_i - U_j)^T \Sigma^{-1} (U_i - U_j)$$

M은 χ^2 分布를 하며 n의 自由度를 가진다.¹⁹⁾

이 분포의 距離 히스토그램을 보면 오른쪽 끝 부분이 誤分類될 確率이 높은 畫素로 구성되므로, 끝부분의 어떤 地點(臨界值)을 定義하여 그 基準에 따라 誤分類確率이 높은 畫素를 區別하도록 한다.¹¹⁾

즉, T=어떤 分類項目의 臨界值

$$C\% = \text{誤分類되었다고 여겨지는 畫素의 百分率}$$

라고 할 때 T는 C%의 畫素가 T와 같거나 큰 距離값을 가지게 되는 地點의 境界값이다. T는 χ^2 分布에 의해 표현되며 이때 χ^2 分布는 自由度(使用 밴드 數)와 信賴水準의 函數이다.

本 연구에서는 分類時 사용밴드 數에 따라 自由度 n=3이며, 信賴水準(confidence level)은 95%($\alpha=0.05$)로 주어 χ^2 分布로부터 7.815의 값을 얻었다.

이 값에 따라 各 分類項目別 데이터 分布에서 臨界值를 주어 誤分類되었을 確率이 높은 畫素를 抽出해 본 結果는 表 (2-2)와 같다.

分類正確度의 比較는

—未補正 畫像의 最大尤度分類(結果 file : SEOTHS)

—地形補正(數值標高 모 델) 畫像의 最大尤度分類(結課 file : DEMTHS)

表 2-2 臨界值 處理 結果

번호	분류 항목	SEOTHS		DEMTHS		DWTMS	
		畫素數	%	畫素數	%	畫素數	%
1	FOREST	245835	32.76	272923	36.37	274374	36.56
2	URBAN(HIGH)	160352	21.37	187612	25.00	194113	25.87
3	URBAN(LOW)	114895	15.32	126973	16.92	131717	17.55
4	CROP	33716	4.49	35117	4.68	53108	7.08
5	WATER	29134	3.88	29256	3.89	29860	3.98
6	GRASS	930	0.12	1132	0.15	1594	0.21
7	BARE SOIL	3634	0.48	3741	0.50	6563	0.87
8	VINYL HOUSE	3963	0.53	6454	0.86	10962	1.46
9	MISCLASSIFIED	158004	21.05	87255	11.63	48152	6.42
합 계		750463	100.00	750463	100.00	750463	100.00

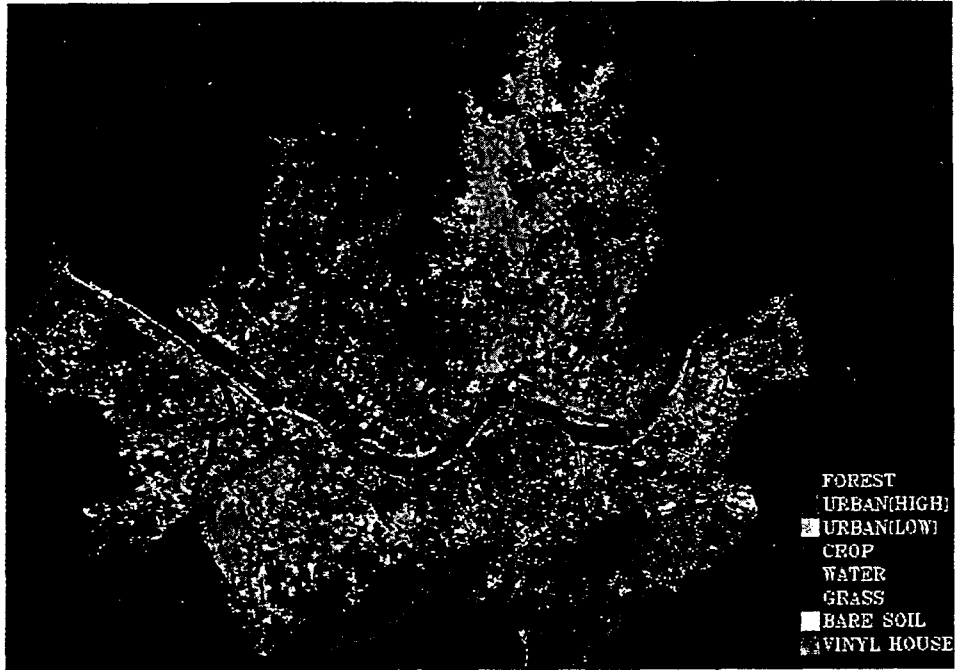


사진 1 未補正 最大尤度法 分類結果 數值畫像

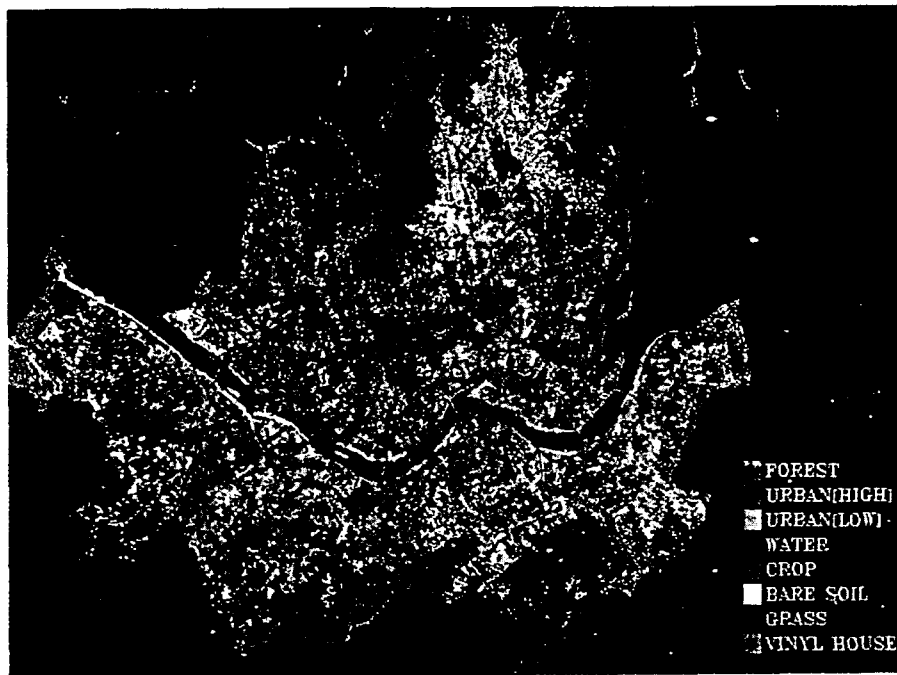


사진 2 數值標高모델 補正 後 事前確率을 考慮한 最大尤度法 分類結果 數值畫像



사진 3. 未補正 最大尤度分類 正確度評價 數值畫像

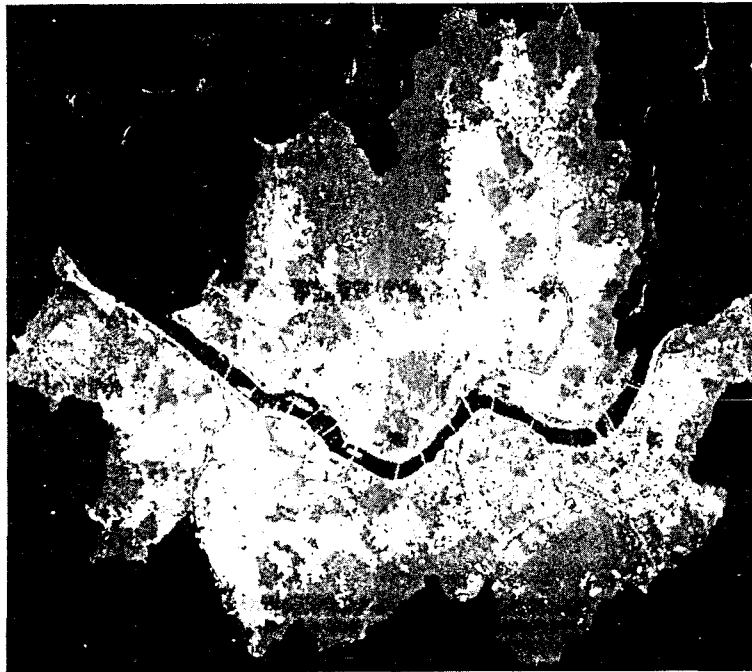


사진 4. 數值標高모델 補正 後 事前確率을 考慮한 最大尤度分類 正確度評價 數值畫像

—地形補正(數值標高 모델)後 加重值를 考慮한 最大尤度分類(結果 file : DWTHS)의 세 가지로 나누어 實行하였으며, 이는 未補正畫像에 대해 地形補正과 加重值의 效果를 各各 알아보기 위한 것이다.

分類正確度 評價 結果數值畫像은 寫眞 3, 4이며, 最終 分類數值畫像은 寫眞 1, 2이다.

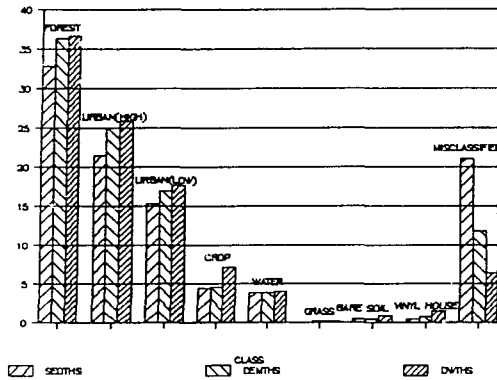


그림 2-3 臨界值 處理結果 比較 그래프(分類正確度)

3. 結 論

本 研究에서 얻어진 結論은 다음과 같다.

1) 平行六面體 分類의 結果로 이끌어 낸 加重值를 事前確率로 주어, Bayesian 最大尤度分類를 實行해 본 結果, 都市의 土地利用分類 正確度가 向上됨을 알 수 있었다.

信賴水準 95% ($\alpha=0.05$)의 臨界值 處理結果 加重值 適用效果는 5.21%로서, 지형 보정 效果를 합하면 全體의으로 14.63%의 分類正確度 向上을 이루었다.

2) χ^2 分布에서 α 의 값은 一般的으로 0.05(95%) 또는 0.01(99%)을 사용하나, 0.01의 경우 誤分類畫素量이 너무 적어져서 地形補正이나 加重值適用의 效果를 評價하기에 어려움이 따르므로 本 研究에서는 $\alpha=0.05$ 를 사용하였다.

3) 가중치에 의해 분류정확도를 향상시키는 기법은 우리 나라와 같이 複雜하고 多樣하게 分布하는 地形에 遠隔探查 技法을 適用하는 경우 반드시 필요한 過程으로 여겨지

며, 여러가지 다른 應用分野에서도 活用될 수 있을 것으로 豫想된다.

4) 以上の 結果에 의해 地形補正이나 加重值適用의 效果는 확실히 立證되었으나, 그 正確度 向上 程度는 客觀的으로 適用할 수 없으며, 分類對象地域에 따라 다를 수 있다.

參考文獻

1. Aronoff, S., 1985, "The Minimum Accuracy Value as an Index of Classification Accuracy", PE & RS, Vol. 51, No. 1, pp. 99-111.
2. Bolstad, Paul V., and T.M. Lillesand, 1991, "Rapid Maximum Likelihood Classification", PE&RS, Vol. 57, No. 1, pp. 67-74.
3. Congalton, Russell G., 1983, "Assessing Landsat Classification Accuracy Using Discrete Multivariate Analysis Statistical Techniques", PE & RS, Vol. 49, No. 12, pp. 1671-1678.
4. Hepner, George F., and Thomas Logan, 1990, "Artificial Neural Network Classification Using a Minimal Training Set: Comparison to Conventional Supervised Classification", PE & RS, Vol. 56, No. 4, pp. 469-473.
5. Hixson, Marilyn et al., 1980, "Evaluation of Several Schemes for Classification of Remotely Sensed Data", PE & RS, Vol. 46, No. 12, pp. 1547-1553.
6. Ince, Fuat, 1987, "Maximum Likelihood Classification, Optimal or Problematic? A Comparison with the Nearest Neighbour Classification", International Journal of Remote Sensing, Vol. 8, No. 12, pp. 1829-1838.
7. Jampoler, S., and B. Haack, 1988, "A Comparison of Classification Techniques for Landsat Thematic Mapper Data", 1988 ACSM-ASPRS Fall Convention Technical Papers, pp. 82-91.
8. Mead, R. A., and John Szajgin, 1982, "Landsat Classification Accuracy Assessment Procedures", PE & RS, Vol. 48, No. 1, pp. 139-141.
9. Prisley, Stephen P., and James L. Smith, 1987, "Using Classification Error Matrices to Improve the Accuracy of Weighted Land-cover Models", PE & RS, Vol. 53, No. 9, pp. 1259-1263.
10. Scarpace, F. L., and B. K. Quirk, 1980, "A Method of Assessing Accuracy of a Digital Classification", PE & RS, Vol. 46, No. 11, pp.

- 1427—1431.
11. Schowengerdt, R. A., 1983, *Techniques for Image Processing and Classification in Remote Sensing*, Academic Press.
 12. Siegal, Barry S., and Michael J. Abrams, 1976, "Evaluation of Classification Techniques Applied to Geologic Mapping Using LANDSAT Data", 42nd Annual Meeting of ASP Technical Papers.
 13. Simon, J. C., and R. M. Haralick, 1980, *Digital Image Processing Proceedings of the NATO Advanced Study Institute*, D. Reidel Publishing Company.
 14. Therrien, Charles W., 1989, *Decision, Estimation, and Classification*, John Wiley & Sons.
 15. 安哲浩, 1983, "Remote Sensing Data의處理와解析", 韓國測地學會誌, 제 1 권, 제 2 호, pp. 1—15.
 16. 安哲浩, 1984, "人工衛星(LANDSAT) 데이터에 의한 서울시에 있어서의 都市域의 變化의 抽出" 韓國測地學會誌, 제 2 권, 제 1 호, pp. 5—16.
 17. 安哲浩, 尹定燮, 安基元, 1985, "都市 및 地域研究를 위한 人工衛星 데이터의 分析 및 그 活用方案에 관한 研究", 韓國測地學會誌, 제 3 권, 제 1 호, pp. 1—14.
 18. 金勇日, 1988, *FILTERING* 技法에 의한 LANDSAT TM DATA의 分析正確度 向上에 관한 研究, 서울大學校 大學院, 碩士學位論文.
 19. 金宇哲 外., 1985, 現代 統計學, 영지문화사.
 20. 柳福模, 1986, 遠隔探測, 開文社.
 21. 上野秀夫, 1985, "ランドサット TMデータの土地被覆分類について", 日本リモートセンシング學會, 第 5 回學術講演會 論文集, pp. 53—56.
 22. 沼田實, 出口近士, 1983, "クラスター分析の應用による MSSデータの土地被覆分類にあける分類項目の設定", 日本リモートセンシング學會, 第 3 回學術講演會論文集, pp. 151—152.
 23. 新技術調査センター, 1986, 海外研究開發レポート 都市域モニタ及び土地利用調査へのリモートセンシングデータの應用研究, 新技術調査センタ.