

지형도와 연속지적도의 가구계 폴리곤 집합간의 N:M 대응쌍 탐색

(Determination of N:M Corresponding
Pairs between Block Polygon Sets
from a Topographical Map and a
Cadastral Map)

허 용* 김 정 옥** 유 기 윤***

(Yong Huh) (Jung Ok Kim) (Ki Yun Yu)

요약 본 논문은 상이한 추상화 및 일반화 기준에 의하여 지형도와 지적도 사이에서 발생하는 가구계 폴리곤 개체의 입도차로 인하여 발생하는 N:M 대응쌍을 탐색할 수 있는 기법을 제시한다. N:M 대응쌍을 도출하기 위하여 어떤 지도상의 가구계 폴리곤 집합에서 병합이 수행될 폴리곤들은 다른 지도상의 어떤 폴리곤과 유의미한 공간적 중복성을 가져야한다는 가정하에 반복적인 중첩 분석을 수행하였다. 반복 과정은 두 지도 사이에서 각각 병합이 수행된 폴리곤 사이의 공간적 중첩이 1:1 대응쌍의 조건을 만족시킬 경우 종료된다.

키워드 : N:M 대응쌍 탐색, 가구계 폴리곤 집합, 지형도, 지적도

Abstract In this paper, we proposed a new algorithm for determination of many-to-many corresponding pairs between block-polygon sets from the national topographical map and the cadastral map in Korea Land Information System, caused by different abstraction and generalization rules of the two maps. Our proposed algorithm starts from an assumption that a block-polygon for a N:M pair should significantly overlap at least one block polygon of the counterpart group, and determines N:M pairs using an iterative updating and searching with this overlapping analysis. This iteration process is terminated when the N:M corresponding pairs satisfy our predefined 1:1 corresponding condition.

Keywords : N:M Corresponding Pairs, Block Polygon Sets, Topographical Map, Cadastral Map

* 이 논문은 2009 GIS 공동추계학술대회에서 '지형도와 연속지적도의 가구계 폴리곤 집합간의 N:M 대응쌍 탐색에 관한 연구'의 제목으로 발표된 논문을 확장한 것임

** 본 연구는 국토해양부 첨단도시기술개발사업-지능형 국토정보기술혁신 사업과제의 연구비지원에 의하여 수행되었음 (07국토정보C04)

* 서울대학교 건설환경공학부 박사과정, hy7808@snu.ac.kr

** 서울대학교 건설환경공학부 박사과정, geostar1@snu.ac.kr

*** 서울대학교 건설환경공학부 교수, kiyun@snu.ac.kr(교신저자)

논문접수 : 2009.08.07

수정일 : 2009.09.10

심사완료 : 2009.09.25

1. 서론

실세계를 공간정보로 표현하는 것은 추상화 모형 및 일반화 기준에 의하여 결정된다[1]. 따라서 동일한 지표상황일지라도 서로 다른 형태 및 범주로 표현될 수 있다. 지형도와 지적도 상의 가구계는 도로에 의하여 둘러싸인 공간으로 정의되며, 도로의 표현에 대한 정의 및 일반화에 따라 표현 또는 생략되는 도로에 의하여 N:M 대응쌍이 발생하게 된다. 본 연구에서는 가구계 폴리곤 사이의 N:M 대응관계를 도출하기 위한 중첩분석 및 폴리곤 병합을 반복하는 알고리즘을 이용하여 복수의 가구계를 결합하여 가상의 단일 개체를 구성하고 이들 사이의 1:1 대응관계를 탐색하는 알고리즘을 제안한다.

2. N:M 대응쌍 탐색 알고리즘

본 연구에서 두 가구계 폴리곤 집합 사이에서 발생할 수 있는 모든 경우의 중첩면적비를 계산하고 그 수치가 임계값 θ_1 이상일 경우 일대일 대응쌍으로 판별한다. 편의상 아래첨자 1은 지적도를 2는 지형도를 의미하며, i, j 는 각각의 가구계 폴리곤 집합에서 해당 가구계의 인덱스를 나타낸다. 만약 가구계 폴리곤 i, j 가 일대일 관계를 가진다면 행렬 $C_{1 \rightarrow 2}[i, j]$ 는 1값을 가지게 된다.

$$C_{1 \rightarrow 2}[i, j] = \left\{ (P_{1i}, P_{2j}) \in P_1 \times P_2 \mid \frac{A_{P_{1i}} \cap A_{P_{2j}}}{A_{P_{1i}} \cup A_{P_{2j}}} > \theta_1 \right\}$$

복수의 가구계 폴리곤들을 하나의 개체로 병합하고 1:1 대응쌍을 찾아내는 과정은 아래와 같은 포함관계 행렬을 이용한다. 행렬 $R_{1 \rightarrow 2}[i, j]$ 은 $C_{1 \rightarrow 2}[i, j]$ 의 원소값이 1이 아니면서 어떤 지도상의 가구계 폴리곤 P_{1i} 가 다른 지도상의 가구계 폴리곤 P_{2j} 에 일정 비율 이상 중첩되어 있는 모든 조합을 표시한다. $R_{2 \rightarrow 1}[i, j]$ 도 유사한 방식으로 정의된다.

$$R_{1 \rightarrow 2}[i, j] = \left\{ (P_{1i}, P_{2j}) \in P_1 \times P_2 \mid \left(\frac{A_{P_{1i}} \cap A_{P_{2j}}}{A_{P_{1i}}} > \theta_2 \right) \wedge \left(\frac{A_{P_{1i}} \cap A_{P_{2j}}}{A_{P_{1i}} \cup A_{P_{2j}}} \leq \theta_1 \right) \right\}$$

$$R_{2 \rightarrow 1}[i, j] = \left\{ (P_{1i}, P_{2j}) \in P_1 \times P_2 \mid \left(\frac{A_{P_{1i}} \cap A_{P_{2j}}}{A_{P_{2j}}} > \theta_2 \right) \wedge \left(\frac{A_{P_{1i}} \cap A_{P_{2j}}}{A_{P_{1i}} \cup A_{P_{2j}}} \leq \theta_1 \right) \right\}$$

$R_{1 \rightarrow 2}[i, j]$ 과 $R_{2 \rightarrow 1}[i, j]$ 를 이용하여 중첩분석을 수행하였을 때, 어떤 지도상의 가구계 폴리곤들 사이에서 나머지 지도상의 어떤 가구계 폴리곤의 면적을 일정 부분 이상 분할하여 중첩되는 가구계 폴리곤들을 공간적으로 연결되는 가구계 폴리곤으로 정의할 경우 아래와 같이 표현된다.

$$L_{1 \rightarrow 1}[i, i'] = \left\{ (P_{1i}, P_{1i'}) \in P_1 \times P_1 \mid \exists j \text{ s.t. } R_{1 \rightarrow 2}[i, j] \wedge R_{1 \rightarrow 2}[i', j] \right\}$$

$$L_{2 \rightarrow 2}[j, j'] = \left\{ (P_{2j}, P_{2j'}) \in P_2 \times P_2 \mid \exists i \text{ s.t. } R_{2 \rightarrow 1}[i, j] \wedge R_{2 \rightarrow 1}[i, j'] \right\}$$

즉 각각의 지도상에서 가구계 폴리곤들에 대한 병합을 수행할 개별 가구계 폴리곤들 사이의 연결성을 $L_{1 \rightarrow 1}[i, i']$ 과 $L_{2 \rightarrow 2}[j, j']$ 이 표현하게 된다. 이러한 $C_{1 \rightarrow 2}[i, j]$, $R_{1 \rightarrow 2}[i, j]$, $R_{2 \rightarrow 1}[i, j]$, $L_{1 \rightarrow 1}[i, i']$ 그리고 $L_{2 \rightarrow 2}[j, j']$ 행렬을 이용하여 최종적으로 각각의 지도상에서 어떤 가구계 폴리곤들을 병합하여 일대일 결합 가구계 폴리곤 쌍을 탐색하는 과정은 그림 1과 같다. 여기서 $L_{1 \rightarrow 1}[i, i']$ 을 거듭 제곱하여 $L^{n \rightarrow 1}[i, i']$ 을 계산하는 것은 [2]에서 네트워크의 연결성을 표현하기 위하여 개별 노드사이의 연결성 행렬 C^1 을 거듭제곱하여 최종 네트워크 연결성 행렬 C^n 을 얻는 것과 유사하다. 단 본 연구의 연결성은 0과 1의 이진값으로 표현되는 차이가 있다.

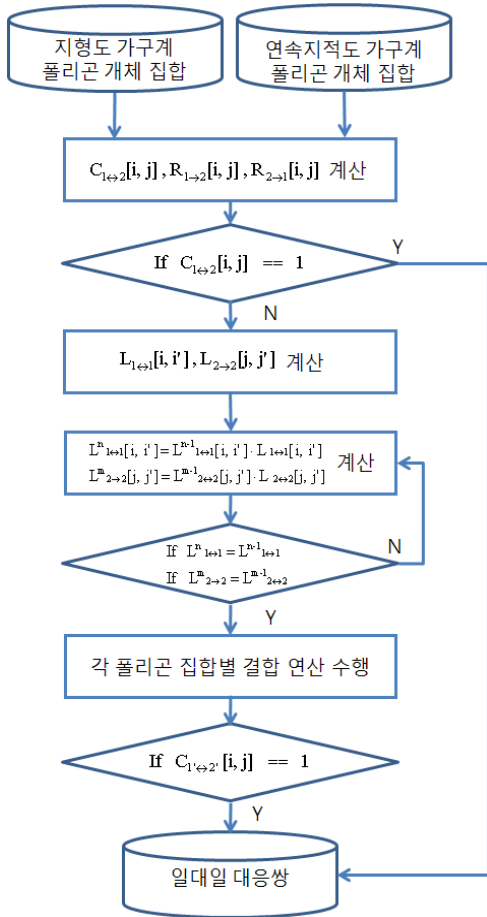


그림 1. N:M 대응 가구계쌍 도출을 위한 결합 가구계쌍 탐색 알고리즘

- (a) $C_{1 \rightarrow 2}[i, j]$ 을 이용, 일대일 대응쌍 추출
- (b) 일대일 대응쌍이 아닌 가구계 폴리곤 집합에 대해 $L_{1 \rightarrow 1}[i, i']$ 과 $L_{2 \rightarrow 2}[j, j']$ 을 각각 자신의 행렬과 원소값의 변화가 없을 때까지 거듭제곱
- (c) $L^{n \rightarrow 1}[i, i']$ 과 $L^{n \rightarrow 2}[j, j']$ 의 각 행별로 1값을 가지는 원소에 해당하는 가구계 폴리곤들을 결합 연산에 사용
- (d) 결합연산에 사용된 가구계 폴리곤을 이용하여 다시 $C_{1 \rightarrow 2}[i, j]$ 을 계산하여 일대일 결합 가구계 폴리곤 쌍 도출

3. 알고리즘 임계값 결정

제안된 알고리즘을 수행하기 위해서는 θ_1, θ_2 의 임계값이 정의되어야 한다. 본 연구에서는 모의 실험 자료를 이용하여 임계값을 결정하였다. 먼저 연구 대상지역을 수원시 팔달구 주변의 $1.5 \times 2 \text{ km}^2$ 로 가정하고 이 공간에 1,000개의 점을 무작위 생성한다. 이 점집합을 이용하여 델로니 삼각망을 생성한다. 다시 무작위로 전체 삼각망을 구성하는 선분 중 20%를 두 번 추출하여 남은 80% 선분들로 두 개의 다각형망을 생성한다. 이 때 지형도와 지적도의 통계적 위치 불일치의 RMSE가 1.53m인 점을 고려하여[3][4], 99% 유의수준인 $\pm 4.5\text{m}$ 사이의 값을 무작위 생성하여 한 개의 다각형망을 구성하는 각각의 점의 x, y 좌표값에 더한다.

이 과정을 통하여 초기에 생성된 1,000개의 점집합이 어떤 지형지물의 좌표가 지형도에서 표현된 좌표라면 조정된 좌표값은 그에 대응되는 지적도 상의 좌표가 된다. 또한 무작위로 추출된 선분들로 인하여 병합된 델로니 삼각형들로 인하여 가구계 폴리곤 개체의 입도차를 모의 생성할 수 있다.

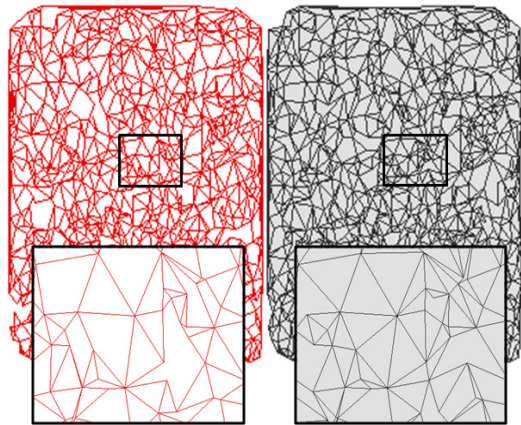


그림 2. $\pm 4.5\text{m}$ 의 위치 오차량 및 N:M 대응 조건을 반영한 모의 실험 자료

그 결과 $\pm 4.5m$ 의 위치 오차량이 존재하며 서로 다른 가구계 폴리곤 입도차를 가짐으로써 N:M 대응관계가 존재하는 두 개의 폴리곤 집합이 얻어진다. 그림 2는 이렇게 얻어지는 두 개의 다각형 집합을 보여준다.

이렇게 생성된 모의 실험 자료는 처음의 1,000개 꼭지점을 이용한 초기 델로니 삼각망과 그 이후의 몇몇 조작에 대한 이력을 간단한 인덱싱 처리를 통하여 쉽게 추적할 수 있으므로 θ_1, θ_2 의 최적값을 얻기위한 통계적 혼련 자료를 손쉽게 얻을 수 있다.

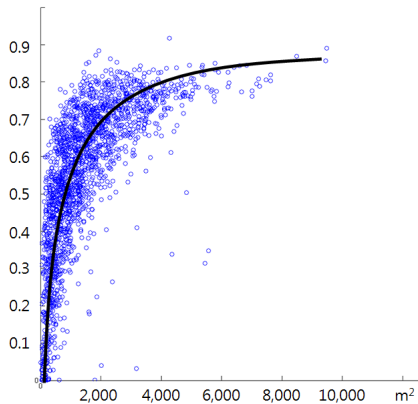


그림 3. 위치 오차량을 고려하였을 때
1:1 대응 다각형의
 $(A_{P_{1i}} \cap A_{P_{2i}}) / (A_{P_{1i}} \cup A_{P_{2i}})$ 과 $A_{P_{1i}}$ 의 산점도

그림 3은 모의 실험 자료에서 실제 1:1 대응 관계인 가구계 폴리곤 쌍의 평균 면적별 중첩면적비를 보여준다. 면적이 작을수록 실제 대응되더라도 중첩면적비가 매우 작게 측정될 수 있다는 것을 확인할 수 있다. 수원시 팔달구의 가구계의 면적이 대부분 $2,000m^2$ 이상임을 고려하여 θ_1 를 0.7로 결정하였다. 만약 적용 대상 지역의 가구계 폴리곤 면적이 평균적으로 이보다 작다면 θ_1 도 그에 맞추어 조정해야 한다. 하지만 이 값이 일정 수준 이하로 낮아질 경우 1:1 대응쌍 탐색의 신뢰도가 낮아지게 된다.

θ_2 의 경우 위치 오차량을 두 개의 다각형 집합에 반영하기 전에 중첩분석을 통하여 알 수 있는 참인 N:M 대응 관계를 개별 폴리곤의 인덱스쌍으로 저장한 후에 위치 오차량을 부여하고 두 다각형 집합의 모든 $(A_{P_{1i}} \cap A_{P_{2j}}) / A_{P_{1i}}$ 를 분석한다. 앞에서 저장된 참인 N:M 대응 관계는 그림 4에서와 같이 o로 표시된 산점도로 표현하고, 위치 오차량으로 인하여 원래 중첩 관계가 없던 폴리곤 사이의 중첩관계, 즉 잘못된 중첩에 의한 $(A_{P_{1i}} \cap A_{P_{2j}}) / A_{P_{1i}}$ 는 별도의 산점도(*)로 표시한다. 역시 연구 대상 지역의 가구계의 면적이 대부분 $2,000m^2$ 이상임을 고려하였을 때, 두 개의 산점도를 안정적으로 구분할 수 있는 임계값은 약 0.2로 판단하였고, 이를 바탕으로 θ_2 를 결정하였다.

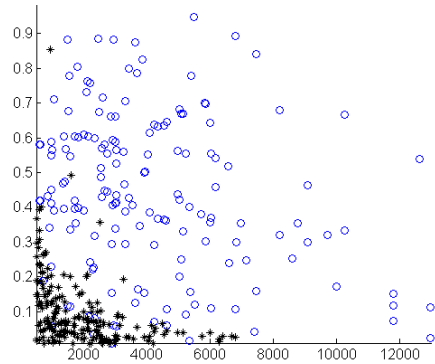


그림 4 위치 오차량을 고려했을 때
실제 대응관계를 가지는 N:M 대응(o)과
오차로 인한 오류 중첩(*)에 의한
 $(A_{P_{1i}} \cap A_{P_{2j}}) / A_{P_{1i}}$ 과 $A_{P_{1i}}$ 의 산점도

4. 결론

본 연구는 가구계 폴리곤 개체의 입도차로 인하여 발생하는 N:M 대응쌍을 탐색하기 위한 알고리즘을 제안하였다. 제안 알고리즘은 어떤 지도상의 가구계 폴리곤 집합에서 복수의 폴리곤이 다른 지도상의 어떤 폴리곤과 유의미한 공간적 중복성을 공유할 경우 병합이 발생할 연결성을 가지고 있다는 가정에서 출발하였다. 하지만 두 지도가 가지고 있는 위치 오차량으로 인하여 알고리즘상의 임계값들에 대한 보다 정밀한 검증이 필요하다. 특히 중첩 면적에 기반한 임계값의 결정은 개별 가구계의 면적이 작을수록 불안정한 단점을 가지고 있다. 따라서 일대일 대응을 판별하기 위한 중첩비에 대한 결정과 공간적 연결성을 판별하기 위한 임계값에 대한 보다 안정적인 분석 모형이 필요하며, 개별 가구계의 면적 및 형상 등을 고려하여 최적화시킬 수 있는 추가 연구가 필요하다.

참고 문헌

- [1] Uitermark, H., Ontology-based Geographic Data Set Integration, Ph.D Dissertation, Deventer, Netherlands, 2001.
- [2] Chou, Y.H., Exploring spatial analysis in geographical information systems. Onward press, 1997.
- [3] 홍성언, "행정구역 경계지역에서의 지적불부합지 실태분석", 한국지형공간정보학회지, 제16권 제1호, 2008, pp. 57-65.
- [4] 홍성언, "수치지정사자료를 이용한 지적불부합지 조사 방법의 개선", 한국지형공간정보학회지, 제14권 제2호, 2006, pp. 23-32.