

# 형상 유사도 기반의 유전 알고리즘을 활용한 이종 수치지도 간의 면 객체 집합 정합 알고리즘 개발

## Development of polygon object set matching algorithm between heterogeneous digital maps - using the genetic algorithm based on the shape similarities

허 용<sup>1)</sup> · 이재빈<sup>2)</sup>

Huh, Yong · Lee, Jeabin

### Abstract

This paper proposes a matching algorithm to find corresponding polygon feature sets between heterogeneous digital maps. The algorithm finds corresponding sets in terms of optimizing their shape similarities based on the assumption that the feature sets describing the same entities in the real world are represented in similar shapes. Then, by using a binary code, it is represented that a polygon feature is chosen for constituting a corresponding set or not. These codes are combined into a binary string as a candidate solution of the matching problem. Starting from initial candidate solutions, a genetic algorithm iteratively optimizes the candidate solutions until it meets a termination condition. Finally, it presents the solution with the highest similarity. The proposed method is applied for the topographical and cadastral maps of an urban region in Suwon, Korea to find corresponding polygon feature sets for block areas, and the results show its feasibility. The results were assessed with manual detection results, and showed overall accuracy of 0.946.

Keywords : Polygon feature set matching, Genetic algorithm, Topographic map, Cadastral map

### 초 록

본 연구는 유전 알고리즘을 이용하여 다대다 면 객체 정합을 수행하는 방법을 제안한다. 동일한 지형·지물을 표현하는 객체 집합의 형상은 서로 동일하다는 가정 하에 형상 유사도를 최적화하는 객체 집합을 두 지도 사이에서 탐색함으로써 정합을 수행한다. 이 때 어떤 객체가 객체 집합에 포함되는지의 여부를 이진 부호로 표현하고, 이진 부호들을 결합한 이진 문자열로 후보해를 표현한다. 초기 후보해들로 해집단을 생성한 뒤, 유전 알고리즘에 의하여 점진적으로 해집단의 품질을 개선함으로써 최적해를 탐색하였다. 제안된 방법을 평가하기 위하여 수원시 도심지역의 수치지형도와 지적도에서 가구계 대응 면 객체 집합을 탐색하였으며 제안된 알고리즘의 효율성을 확인할 수 있었다. 또한 수작업에 의한 탐색결과를 이용하여 평가한 결과 0.946의 정확도를 얻었다.

핵심어 : 면 객체 정합, 유전 알고리즘, 수치지형도, 연속지적도

## 1. 서론

지도 융합(map conflation) 과정에서 객체 정합은 기하학적 특징을 비교하여 두 지도 사이의 대응 객체를 탐색한다. 일반

적으로 한 지도의 기준 객체를 중심으로 일정한 탐색 공간을 정의하고 여기에 포함되는 다른 지도의 객체 중에서 가장 높은 유사도를 가지는 객체를 대응 객체로 결정한다. 하지만 동일한 지형·지물을 표현하더라도 상이한 데이터 모델링과 묘

1) 정희원 · 서울대학교 공학연구소 선임연구원(E-mail : huhjong78@gmail.com)

2) 교신저자 · 정희원, 목포대학교 토목공학과 조교수(E-mail : lee2009@mokpo.ac.kr)

화 기준 때문에 어떤 지도의 한 객체가 다른 지도의 객체 집합에 대응하거나 객체 집합이 서로 대응하는 현상이 발생할 수 있다. 이처럼 일대다 또는 다대다 대응 관계가 발생하는 경우 지형·지물의 일부만을 표현하는 개별 공간 객체의 기하학적 특징을 비교해서는 객체 집합의 전체(holistic) 특징을 비교할 수 없다. 따라서 정합 여부를 판단할 객체 집합들을 각 지도에서 병합한 상위 객체(super feature)를 생성한 후 이들의 기하학적 특징을 비교하여 정합 여부를 판단해야 한다. 그런데 개별 지도에서 조합 가능한 상위 객체의 개수는 원시 공간 객체의 개수에 비례하여 기하급수적으로 증가하는 것은 물론, 어떤 상위 객체와 대응되는 다른 지도의 상위 객체를 탐색하는 과정에서 수많은 비교가 필요하기 때문에 효율적인 탐색 알고리즘이 필요하다.

공간정보공학 분야에서 이 문제는 네트워크 정합을 중심으로 많은 연구가 수행되었다. 대표적으로 Walter and Fritsch(1999)가 제안한 buffer growing 알고리즘은 한 지도의 객체 집합을 기준으로 다른 지도에서 가장 유사한 객체 집합을 탐색한 후 탐색된 객체 집합을 기준으로 처음 지도에서 가장 유사한 객체 집합을 다시 탐색한다. 한 개의 객체로부터 시작하여 위 과정을 반복함으로써 각 지도의 객체 집합은 주변 객체들을 병합하면서 정합 기준을 만족할 때까지 서로 확장된다. 이 같은 상호 확장 탐색은 대응 객체 집합의 탐색을 효과적으로 해결할 수 있으며, Mantel and Lipeck (2004), Stigmar (2005) 그리고 Zhang and Meng (2007) 등의 연구에 적용되었다. 그러나 네트워크에서는 선 객체의 연결 구조를 따라 상위 객체를 구성할 객체 집합을 확장하는 과정이 비교적 단순하지만 면 객체의 경우 주변에 인접한 다수의 객체 중에서 병합할 객체의 순위를 결정해야 한다. 특히 건물처럼 일정 거리 이상 떨어져 있는 객체들에서는 일정 범위 이내에 존재하는 모든 객체들을 고려해야 하기 때문에 수많은 확장 가능한 경우의 수를 고려해야 한다. 이 문제를 해결하기 위하여 Bel Hadj Ali (2000)와 Huh et al. (2011)는 객체 중첩 관계에 의한 그래프 군집화 방법을 이용하였다. 객체를 노드 그리고 두 지도 사이의 객체 중첩 관계를 에지로 표현하는 그래프에서 에지 연결성을 기준으로 노드를 군집화한 후 각 군집 내에서 대응 객체 집합을 탐색하였다. 하지만 선분의 길이, 방향, 위치 그리고 연결성 등과 같은 다양한 특징을 탐색 과정에 이용할 수 있는 네트워크와 달리 단순한 중첩 관계만을 이용하므로 군집화 결과가 위치 오차에 민감한 문제를 가지고 있다. 특히 두 지도 사이의 위치 오차에 비하여 작은 면적을 가지는 객체가 실제 대응되어야 하는 객체가 아닌 주변의 다른 객체와 중첩될 경우 군집화 결과가 왜곡될 수밖에 없다(Huh et al., 2011). 따

라서 ICP (iterative closest point) 알고리즘처럼 위치 오차를 보정하기 위한 지도 변환(map transformation)과 객체의 대응 관계 분석(correspondence analysis)을 결합할 수 있는 방법이 필요하다. 그런데 점 객체 집합을 대상으로 개발된 ICP 알고리즘에서 객체의 대응 관계는 기본적으로 일대일 대응을 가정한다. 따라서 지도 변환에 필요한 대응점 쌍의 탐색과 탐색된 대응점 쌍을 이용한 지도 변환이 서로 반복적으로 연결되어 점진적으로 최적해를 탐색하는 것이 가능하다. Butenuth et al.(2007)은 면 객체의 꼭지점을 추출하여 점 객체 집합을 두 지도에서 추출한 다음 꼭지점에 접합 선분의 방향각과 같은 정보를 위치 정보와 결합하여 보다 안정적으로 대응점 쌍을 탐색하고 위치 오차를 해결할 수 있는 방법을 제안하였다. 하지만 이 경우 한 지도에서 선분으로 연결된 두 꼭지점이 다른 지도에서 선분으로 연결된 두 꼭지점에 서로 대응해야 한다는 것과 같은 위상과 관련된 조건을 탐색 과정에 반영할 수 없다는 한계를 가지고 있다(Huh et al., 2011). 또한 작은 객체들이 밀집한 지역일수록 많은 꼭지점들이 추출되기 때문에 위치 편차에 의한 왜곡된 탐색 결과가 초기 대응 관계 탐색과정에서 발생할 경우 지역적 최대값(local maxima) 문제가 발생할 가능성이 높아진다.

위와 같은 문제를 해결하기 위하여 본 연구는 두 지도의 지역적 위치 오차는 회전이나 축척 변환을 생략한 평행 이동만으로 설명할 수 있다는 가정 하에 무게 중심을 일치시킨 상태에서 가장 유사한 형상을 가지는 대응 객체 집합을 탐색함으로써 위치 오차에 영향을 받지 않는 다대다 객체 정합 방법을 제안한다. 즉 평행 이동에 불변(invariant)인 형상 유사도로 목적 함수를 정의하였을 때, 가장 높은 형상 유사도를 가지는 객체 집합 쌍을 탐색하는 최적화 과정을 통하여 정합 문제를 해결하게 된다. 그런데 일반적인 최적화 과정은 목적 함수의 편미분 벡터를 이용하여 초기 해로부터 점진적으로 개선된 해를 탐색하지만 본 연구의 경우 어떤 객체가 대응 객체 집합에 포함되는지 여부, 즉 이산적인 해공간에서 탐색이 이루어지므로 연속 함수를 가정한 편미분 벡터를 이용하는 것은 적합하지 않다. 따라서 본 연구에서는 추계적(stochastic) 최적화 기법의 하나인 유전 알고리즘을 이용한다. 이 알고리즘은 임의로 생성한 복수의 후보해로 초기 해집단을 구성한 뒤, 후보해 한 쌍을 선택하여 교차 연산과 변이 연산을 수행하여 새로운 후보해 한 쌍을 생성, 이들 네 개의 후보해 중에서 우수한 두 개를 기존 후보해 대신 해집단에 삽입한다. 이 과정을 반복함으로써 점차 최적해에 가까운 후보해들로 해집단이 구성되며 종료 조건에 도달한 시점의 가장 우수한 후보해를 최적해로 얻게 된다. 따라서 유전 알고리즘은 목적 함

수의 편미분이나 초기해가 필요하지 않으며, 최적해를 얻고자 하는 함수의 후보해가 선택, 교차, 변이 연산이 가능한 형태로 표현될 수 있다면 적용이 가능한 장점을 가지고 있다 (Moon, 2008).

제안된 방법의 구체적인 내용은 다음 장에서 설명한다. 3장에서는 경기도 수원시 도심지역의 수치지형도 2.0과 연속지적도의 면 객체를 대상으로 제안된 방법을 적용한 결과를 분석하며, 마지막으로 4장에서 결론과 함께 향후 연구방향에 대하여 논한다.

## 2. 방법론

본 연구에서 제안한 다세대 면 객체 정합은 Fig. 1과 같은 단계로 수행되며, 각 단계의 세부적인 내용은 다음과 같다.

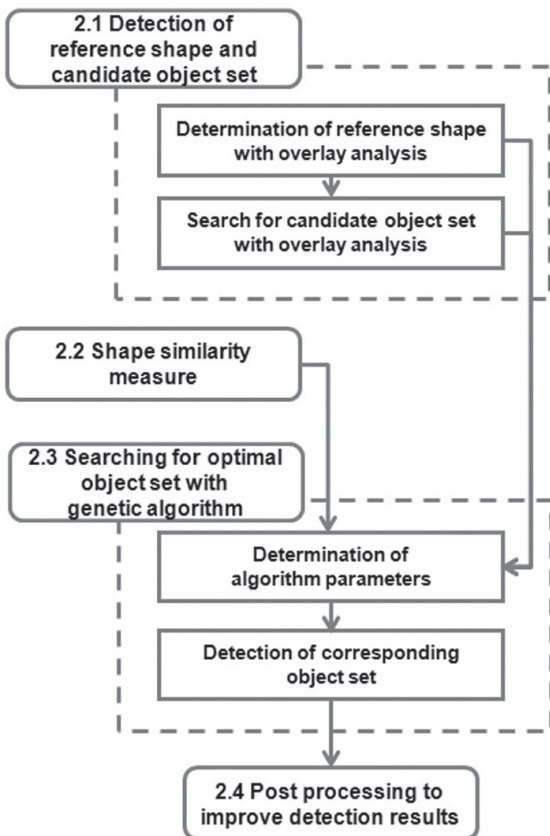


Figure 1. Workflow of the proposed method

### 2.1 기준 형상 및 후보 객체 집합 탐색

유전 알고리즘을 적용하여 최적해를 탐색하는 방법은 복수의 후보해를 대상으로 종료 조건에 도달할 때까지의 반복적

인 연산을 필요로 한다. 따라서 탐색할 해공간, 즉 대응 객체 집합을 탐색할 후보 객체 집합의 크기를 최대한 줄임으로써 탐색의 효율을 높이는 것이 필요하다. Bel Hadj Ali (2000)와 Huh et al. (2011) 역시 객체의 중첩관계를 이용하여 군집화를 수행함으로써 대응 객체 집합을 탐색할 후보 객체들의 범위를 효과적으로 제한할 수 있었다. 단, 위치 오차에 의한 왜곡된 중첩 관계를 해결하기 위하여 Bel Hadj Ali (2000)는 객체 군집을 객체가 포함된 지도를 기준으로 객체 집합 쌍으로 분할한 뒤 하나의 객체를 제외한 나머지 객체들로 대응 객체 집합을 구성하였을 때의 유사도를 모두 비교하면서 가장 높은 유사도를 가지는 대응 객체 집합을 탐색하는 후처리를 수행하였다. 하지만 이 방법은 처음 탐색된 군집에 포함되지 않는 객체를 대응 객체 집합에 추가할 수 없는 한계를 가지고 있다. Huh et al. (2011) 역시 Fig. 2와 같이 불확실한 경계(Burrough and Frank, 1996) 모형을 적용하여 위치 오차에 영향을 받지 않는 객체  $x_i$ 의 내부 영역(interior region)  $x_i^{int}$ 들의 중첩 관계를 이용하여 군집화를 수행하였지만 Bel Hadj Ali (2000)와 동일한 한계를 가지고 있다.

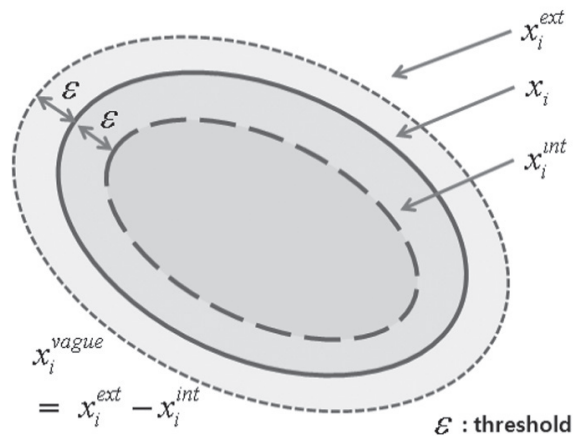


Figure 2. Indeterminate boundary model of object  $x_i$

이 문제를 해결하기 위하여 본 연구에서는 객체의 외부 영역(exterior region)  $x_i^{ext}$ 까지도 고려하여 탐색 공간을 확장한다. 불확실한 경계 모형에 의하면 위치 오차에 의하여 객체  $x_i$ 의 경계에는 공간 불확실성이 존재하기 때문에 실제 경계는  $x_i$ 의 모호 영역(vague region)  $x_i^{vague}$  내에 존재한다. 이 영역은 Fig. 2와 같이  $x_i$ 의 경계를 기준으로  $\epsilon$ 만큼  $x_i$ 을 확장한 외부 영역  $x_i^{ext}$ 와 축소된 내부 영역  $x_i^{int}$ 의 공간적 차 집합으로 정의된다. 따라서 내부 영역을 이용하여 초기 대응 객체 집합 쌍을 얻은 뒤, 이들 집합의 외부 영역에 포함되는 객체들까지 확장한 해공간에서 가장 유사도가 높은 대응 객

제 집합의 조합을 탐색할 경우 보다 정확한 탐색이 가능하다. 하지만 내부 영역을 구하는 과정에서 크기가 작은 객체들은  $\epsilon$ 의 길이에 따라 제거(collapse)되는 경우가 발생할 수 있다. 이 문제를 해결하기 위하여 한 지도의 면 객체들의 크기는 내부 영역을 얻는 과정에서 제거되지 않을 정도로 충분히 크다고 가정한다. 만약 지도  $A$ 의 객체들이 이 조건을 만족한다면 두 지도 사이의 객체 중첩 관계는 식(1)과 같은 인접성 행렬(adjacency matrix)  $M$ 으로 표현된다. 이 행렬에서 인접성을 분석하여 얻게 되는  $k$ 번째 군집  $C^{(k)}$ 에서 지도  $A$ 에 포함된 객체들의 집합  $A^{(k)} = \{a_i^{(k)}\}$ 를 병합하여 기준 형상을 생성한 뒤, 기준 형상의 외부 영역에 포함되는 지도  $B$ 의 객체들을 후보 객체 집합  $B^{(k)} = \{b_j^{(k)}\}$ 으로 탐색한다. 이렇게 기준 형상과 후보 객체 집합을 얻게 되면  $B^{(k)}$ 를 구성하는 객체 집합이  $A^{(k)}$ 와 대응되는 객체 집합을 탐색할 해공간을 정의하게 된다.

$$M_c(i, j) = \begin{cases} 1 & \text{if } a_i^{int} \cap b_j \neq \emptyset \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

여기서  $\epsilon$ 은 두 지도 사이에서 발생할 수 있는 최대 위치오차로 결정되며 훈련 지역에서 추출된 대응 꼭지점 사이의 거리를 이용하여 추정할 수 있다.

### 2.2 형상 유사도 평가 함수

면 객체 집합의 형상 유사도( $S_l$ )를 평가하기 위하여 본 연구는 식(2)와 같이 지도융합 분야에서 일반적으로 사용되고 있는 중첩 면적비를 이용한다. 지도  $B$ 의 후보 객체 집합  $B^{(k)} = \{b_j^{(k)}\}$ 에서  $l$ 번째 임의의 부분집합을  $B_l^{(k)}$ 로 표현한다면 기준 형상과  $B_l^{(k)}$ 의 유사도는 병합된 객체들의 무게 중심을 이용하여 위치 오차를 조정할 후에 측정된다.

$$S_l = S(A^{(k)}, B_l^{(k)}) = \frac{\text{aggr}(A^{(k)}) \cap F_c(\text{aggr}(B_l^{(k)}))}{\text{aggr}(A^{(k)}) \cup F_c(\text{aggr}(B_l^{(k)}))} \quad (2)$$

여기서,  $\text{aggr}$ 는 객체 집합의 병합 함수이며,  $F_c$ 는 무게 중심을 이용한 위치 오차 조정 함수이다.

### 2.3 유전 알고리즘을 이용한 최적 집합 탐색

2.1절의 방법을 이용하여 기준 형상을 구성하는 객체 집합  $A^{(k)}$ 와 후보 객체 집합  $B^{(k)}$ 를 얻게 되면  $B^{(k)}$ 의 어떤 객체가  $A^{(k)}$ 에 대응하는 객체 집합에 포함될지의 여부는 Fig. 3(a)과 같이 1 또는 0의 이진 부호로 표현할 수 있다. 유전 알고리즘의 관점에서 개별 이진 부호는 유전자, 이들을 결합한 문자열은 후

보해를 표현하는 염색체가 된다. 만약 후보해가 101100이라면 후보 객체 집합에서 1값이 할당된 첫 번째, 세 번째, 네 번째 객체는 기준 형상에 대응되는 후보해를 구성하는 조합에 포함되고, 0값이 할당된 객체들은 포함되지 않게 된다. 유전 알고리즘은 Fig. 3(b)와 같이 후보해 한 쌍을 해집단에서 부모 염색체로 선택한 뒤, 이진 부호 문자열로 표현되는 이들 염색체의 일부를 교환하는 교차 연산과 낮은 확률로 하나의 이진 부호를 다른 부호로 변환하는 변이 연산을 수행하여 새로운 이진 부호 문자열 한 쌍, 즉 후보해 한 쌍을 자식 염색체로 생성한다. 이들 네 개의 후보해 중에서 기준 형상과의 유사도가 높은 두 개의 후보해를 앞에서 선택된 부모 염색체 한 쌍 대신 해집단에 삽입한다.

선택 연산에서 어떤 염색체가 부모 염색체로 선택될 확률은 식(3)과 같이 해의 품질에 비례하는 룰렛휠 선택(roulette wheel selection) 방법으로 결정하였다. 다양한 부모 염색체를 이용하면 다양한 자식 염색체를 탐색할 수 있으며, 이를 통하여 국소 최대치 문제를 해결할 수 있다. 따라서 낮은 유사도를 가지는 염색체일지라도 일정 수준 이상 선택되는 것이 바람직하다(Moon, 2008). 룰렛휠 선택 방법은 최대 유사도를 가지는 염색체와 최소 유사도를 가지는 염색체 사이의 상대적인 확률을 식 (3)의  $k$ 값을 조절함으로써 제어할 수 있다. 일반적으로  $k$ 값은 3 내지 4가 적용되며(Moon, 2008), 본 연구에서는 4를 적용하였다.

$$p_l = (S_l - S_w) + (S_b - S_w) / (k - 1), k > 1 \quad (3)$$

여기서,  $p_l$ 은  $l$ 번째 염색체의 상대적 선택확률,  $S_l$ 는  $l$ 번째 염색체의 유사도,  $S_w$ 와  $S_b$ 는 각각 해집단에서 가장 낮은 유사도와 높은 유사도이다.

유전 알고리즘을 실제 적용하기 위해서는 교차 연산과 변이 연산의 확률과 함께 해집단을 구성하는 염색체의 개수와 종료조건을 결정해야 한다. 본 연구에서는 선택된 두 부모 염색체가 교차할 확률과 자식 염색체에서 임의로 선택된 하나의 유전자가 변이될 확률을 각각 95%와 5%로 결정하였다. 또한 후보 객체 집합의 면 객체 개수( $n$ )로 해집단을 구성하는 염색체의 개수를 정하였으며, 탐색 과정에서 식(2)의 유사도가 0.97 이상인 염색체가 발생하거나 반복과정의 회수가  $2n^2$ 에 도달하면 탐색 과정을 종료한다. 최적해는 종료 시점의 해집단에서 가장 유사도가 높은 염색체에 대응되는 객체 집합으로 결정된다. Fig. 4는 위와 같은 유전 알고리즘의 전체적인 의사코드이다.



Binary code representation of candidate object set

| Each object of candidate object set                        | Obj1   | Obj2   | Obj3   | Obj4   | ... | Obj77  | Obj78  | Obj79  | Obj80  |
|--|--------|--------|--------|--------|-----|--------|--------|--------|--------|
| Binary representation of an object to be used(1) or not(0) | 0 or 1 | 0 or 1 | 0 or 1 | 0 or 1 | ... | 0 or 1 | 0 or 1 | 0 or 1 | 0 or 1 |

(a)

Parent chromosomes



Crossover operation



Offspring chromosomes



Mutation operation

100100110110...

100100110110...

100100100110...

100100100110...

100000100110...



111000100110...

111000100110...

111000110110...

111000110110...

(b)

Figure 3. Detection of candidate corresponding object set within tolerance region of reference shape(a) and generation of two offspring chromosomes from two parent chromosomes with crossover and mutation operations(b)

```

generate  $n$  initial chromosome  $c_i$ ;
repeat {
    select two chromosomes  $c_1$  and  $c_2$ 
     $offspring_1, offspring_2 = crossover(c_1, c_2)$ 
     $offspring'_1, offspring'_2 = mutation(offspring_1, offspring_2)$ 
    compare( $c_1, c_2, offspring'_1, offspring'_2$ )
    insert optimal two chromosomes into population instead of  $c_1$  and  $c_2$ 
} until terminal condition
    
```

Figure 4. Pseudocode of genetic algorithm

2.4 정확도 개선을 위한 탐색 객체 집합의 후처리

본 연구의 유전 알고리즘은 식(2)의 유사도가 0.97 이상인 자식 염색체가 발생하거나 반복회수가  $2m^2$ 에 도달하면 탐색은 중지된다. 이 과정에서 면적이 작은 객체들은 식(2)의 유사도에 미치는 영향 역시 작기 때문에 포함되어야 할 객체가 포함되지 않거나 반대로 포함되지 않아야 하는 객체가 포함되어도 0.97 이상의 유사도를 얻을 수 있다. 이 문제를 해결하기 위하여 유전 알고리즘으로 탐색된 객체 집합의 무게 중심과 기준 형상의 무게 중심을 이용하여 두 지도의 위치 오차를 조정 한 후 식(4)를 이용하여 각 후보 객체들의 중첩 면적비를 이용하여 보다 높은 유사도의 객체 집합을 탐색한다. 만약 유전 알고리즘에 의하여 탐색된 객체 집합에 포함되지 않는 객체 중에서 중첩 비율이 0.80 이상인 객체는 대응 객체 집합에 추가하며, 반대로 포함된 객체 중에서 0.20 이하인 객체는 제거한다. 이 비율들은 훈련지역의 실험 결과를 바탕으로 결정되었다.

$$S(A^{(k)}, b_l^{(k)}) = \frac{aggr(A^{(k)}) \cap F_c(b_j^{(k)})}{F_c(b_j^{(k)})} \quad (4)$$

3. 실험 자료 및 결과 분석

3.1 실험 자료 분석

경기도 수원시 도심지역의 1:5,000 수치지형도2.0과 연속지적도를 대상으로 제안된 방법을 적용 및 평가하였다. 2.1절의 기준 형상과 후보 객체 집합을 탐색하기 위해서는 객체의 내부와 외부 영역을 결정하기 위한  $\epsilon$ 가 결정되어야 한다. 본 연구에서는 실험지역에 인접한 훈련지역에서 두 지도의 대응 꼭지점 쌍 200개를 추출한 후 거리 오차를 분석하였다. 그 결과

Fig. 5와 같이 정규분포와는 다른 형태를 가지고 있음을 확인할 수 있었다. 이 경우 평균과 표준편차를 이용하여 임계값을 추정할 경우 왜곡된 값을 얻게 되므로 Reimann et al.(2005)

이 제안한 Box-Plot 방법을 이용한다. 이 방법은 정규분포와 같은 분포 모형을 고려하지 않고 분포 자체의 제1사분위수(LH, lower hinge), 중앙값(Median), 제3사분위수(UH, upper hinge)를 이용하여 이상치(outlier)를 제외한 분포의 극대값(UW, upper whisker)을 식(5)와 같이 결정한다. 즉 전체 관측값의 가운데 절반이 분포하는 구간을 정의하는 LH와 UH를 계산한 후, 이 두 값의 차의 1.5배를 UH에 더한 값(UIF, upper inner fence)을 계산한다. 이후 관측값들 중에서 이 값보다 작은 최대값을 임계값(UW)으로 결정한다. 훈련 지역에 적용한 결과 5.7m가 두 지도상 대응 꼭지점이 가질 수 있는 최대 거리, 즉 두 지도사이에서 발생할 수 있는 최대 위치 오차로 추정하고, 이 거리를  $\epsilon$ 로 이용하였다.

$$UW = \max(x|x < UIF) \quad (5)$$

여기서  $UIF = UH + 1.5(UH - LH)$ 이며,  $x$ 는 관측된 대응 꼭지점 쌍의 거리

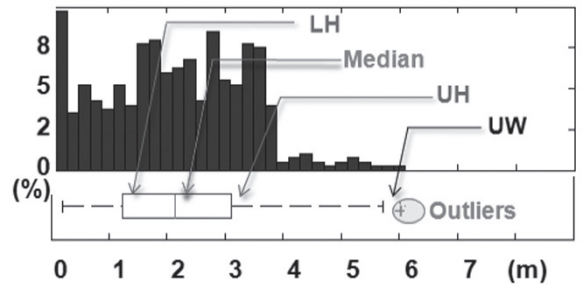


Figure 5. Histogram and boxplot analysis on lengths of corresponding point pairs in training site

3.2 결과 분석

제안된 방법을 적용 및 평가하기 위하여 수치지형도2.0과 연속지적도를 대상으로 동일한 지형·지물을 표현하는 대응 객체 집합을 탐색하였다. 연속지적도는 토지를 분할하는 경계를 표현하는 반면, 수치지형도2.0은 지표의 형태 또는 지표상의 사물을 표현한다. 따라서 형상이 유사한 대응 객체 집합을 탐색하기 위해서는 두 지도에서 유사한 표현 기준에 의하여 측량 및 묘화된 공통 지형·지물이 존재해야 한다. 두 지도를 구축 명세서와 함께 분석한 결과 도로에 대응되는 수치지형도2.0 레이어의 면 객체들과 필지의 지목이 도로에 대응되는 면 객체들이 동일한 지형·지물을 기하학적으로 유사한 형

상으로 표현하고 있음을 확인할 수 있었다. 하지만 도로 필지의 경우 수치지도의 객체들에 비하여 일부 지역에서는 과대하게 분할되어 있는 반면, 다른 지역에서는 반대로 병합되어 있어 중첩 관계를 이용한 기준 형상이 단순 면 객체(simple polygon)가 아닌 매우 복잡하고 거대한 네트워크 형상의 면 객체(complex polygon with holes)가 된다. 따라서 도로의 공간적 여집합, 즉 도로에 둘러싸인 가구체에 대응되는 객체들을 이용하여 다대다 객체 집합의 탐색 결과를 분석하였다. 정확도 평가를 위하여 수치지형도의 가구체 면 객체에 대응되는 연속적지도의 필지 객체 집합을 탐색한 뒤, 필지의 지목을 이용하여 개별 필지 별로 결과를 평가하였다.

Table 1. Accuracy test

|   | Number of true parcels corresponding to block | Number of true parcels corresponding to road |
|---|---|--|
| Number of detected parcels corresponding to block | 745   | 42   |
| Number of detected parcels corresponding to road  | 11  | 184  |
| Overall accuracy                                  | $(745+184)/(745+42+184+11) = 0.946$           |  |

Table 1은 총 40개의 수치지형도 가구체 면 객체와 982개의 필지를 대상으로 평가한 결과로 다음의 table 1과 같은 0.946의 정확도를 얻을 수 있었다. 정확도 평가를 가구체와 도로로 구분하여 분석할 경우 정확도(precision)와 재현율(recall)을 적용할 수 있다. 가구체의 경우 정확도는 0.947 ( $745/(745+42)$ ), 재현율은 0.985 ( $745/(745+11)$ )로 매우 높은 정확도를 얻은 반면, 도로의 경우 정확도는 0.944 ( $184/(184+11)$ ), 재현율은 0.841 ( $184/(184+42)$ )로 특히 재현율에서 낮은 결과를 얻었다. 이는 도로에 대응되는 필지들이 가구체 대응 필지로 오탐색된 결과로 상대적으로 도로 대응 필지들의 크기가 가구체 대응 필지들의 크기보다 작아서 발생한 결과로 추정된다. 앞의 2.4 정확도 개선을 위한 탐색 객체 집합의 후처리에서 중첩 면적비를 이용하는데 크기가 작은 도로 필지들은 수치지형도와 연속적지도에서 가구체에 대응되는 면 객체의 형상에 약간의 차이가 발생하여도 상대적으로 크기가 큰 가구체 대응 필지들에 비하여 중첩 면적비에 큰 영향을 받게 된다. 이러한 특징을 분석하기 위하여 Fig. 3(a)의 기준 형상에 대응하는 필지 객체 집합을 탐색한 결과(Fig. 6)를 지목이 표시된 지적도(Fig. 7)와 비교하였다. 2.3절까지의 유전 알고리즘만을

적용할 경우  $A_1, A_2, A_3$  그리고  $A_5$ 와 같은 작은 객체들이 대응 객체 집합에 포함되지 않은 상태에서도 식(2)의 유사도가 종료 조건을 만족할 수 있다. 따라서 이들 객체가 포함되지 않은 상태로 최적해를 얻는 경우가 발생하였다. 하지만 탐색된 대응 객체 집합의 무게중심을 이용하여 위치 오차를 조정된 후 식(4)를 이용하여 개별 필지별로 탐색 결과를 조정된 결과보다 안정적인 탐색 결과를 얻을 수 있음을 확인할 수 있었다.

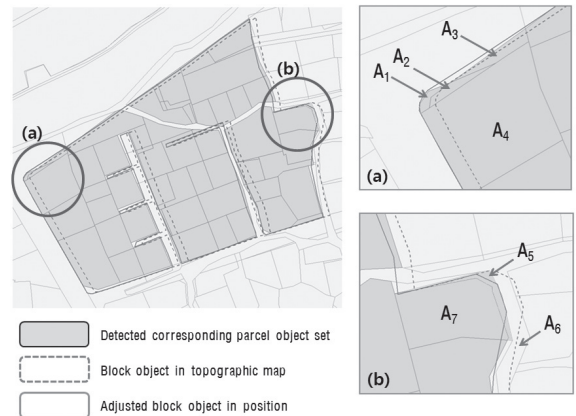


Figure 6. Matching result in Fig. 3(a)

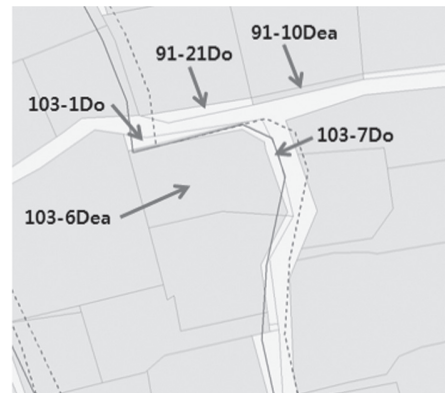


Figure 7. Land use categories of parcels in Fig. 6(b)

하지만 식(2)와 (4)는 두 지도의 무게 중심을 이용한 평행 이동만으로 위치 오차를 해소할 수 있다고 가정한다. 따라서 대응 객체 집합의 형상에 일정 수준 이상의 오차가 존재할 경우 무게 중심을 이용하여 오차를 조정된 이후에도 작은 객체들의 중첩 관계가 여전히 왜곡될 수 있다. 이 문제는 Fig. 6(b)의 탐색 결과와 필지의 실제 지목을 보여주는 Fig. 7를 비교함으로써 확인할 수 있다. 2.4절의 후처리를 수행하면 면 객체  $A_5$ 는 가구체에 대응되는 필지로 결정된다. 하지만 이 필지의 지

목은 도로이며, 103-6Dea(대지)번지의 형상과 수치지형도 가구계 면 객체의 형상을 비교하여도 만이 가구계에 대응되는 것이 타당하다. 하지만 Fig. 6(b)의 수치지형도 가구계 면 객체의 형상이 우측으로 다소 넓게 묘화되어 식(4)에 의한  $A_3$ 의 중첩 면적비가 0.8 이상으로 측정되었고 그 결과 가구계에 대응되는 필지로 탐색되었다. 이 문제는 Fig. 7의 91-21Do(도로), 91-10Dea, 103-1Do, 103-7Do 번지의 필지들과 같이 가구계와 도로의 경계 지역에 위치하는 크기가 십여  $m^2$ 이하의 작은 필지들에서 주로 발생하였다. 하지만 대부분의 실험에서 Fig. 8과 같이 필지 면 객체들의 크기는 백여  $m^2$  이상이었기 때문에 table 1과 같은 높은 정확도의 탐색 결과를 얻을 수 있었다. 두 그림에서 점선으로 표현된 면 객체는 수치지형도의 가구계로 연속지적도의 대응 필지 집합에 비하여 2~3 m 정도의 위치 편차를 가지고 있다. 위치 편차가 보정되지 않은 상황에서 단순 중첩관계만을 고려할 경우 Fig. 8(a)과 같이 골목길에 대응되는 필지들의 크기에 비하여 위치 편차가 상대적으로 작을 경우에만 중첩관계 또는 중첩비율에 큰 외곡이 발생하지 않기 때문에 참인 대응관계를 탐색할 수 있다. 하지만 Fig.8(b)의 경우 골목길에 대응되는 필지가 수치지형도 가구계와 50% 이상 중첩되어 있기 때문에 참인 아닌 대응관계가 탐색된다. 하지만 필지들의 크기가 일정 수준 이상일 경우 제안된 방법은 두 경우 모두에서 안정적인 대응 객체 집합을 탐색할 수 있었다.

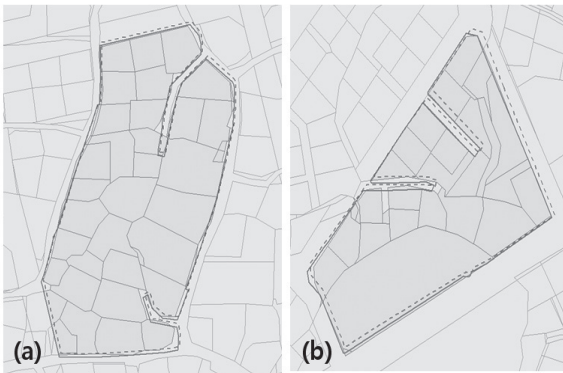


Figure 8. Detection of corresponding object sets between topographic map and cadastral map

#### 4. 결론

본 연구는 두 지도사이에서 다대다 대응 면 객체 집합을 탐색하기 위하여 객체 중첩 관계를 이용하여 한 지도의 객체 집합을 병합하여 기준 형상을 얻은 뒤, 이 형상에 가장 유사한 다른 지도의 객체 집합을 유전 알고리즘을 이용하여 탐색하

는 방법을 제안하였다. 동일한 지형·지물을 표현하는 객체 집합의 전체 형상은 서로 동일하다는 가정 하에 객체 집합을 병합한 상위 객체의 형상 유사도를 최적화하는 객체 집합을 두 지도 사이에서 탐색함으로써 다대다 면 객체 정합을 수행한다. 어떤 객체가 대응 객체 집합에 포함되는지 여부는 이산적인 해공간에서 결정되므로 목적 함수의 편미분 벡터를 이용하는 일반적인 최적화 방법 대신 추계적 최적화를 수행하는 유전 알고리즘을 이용하였다. 제안된 방법을 평가하기 위하여 경기도 수원지역의 1:5,000 수치지형도 2.0과 연속지적도를 대상으로 가구계를 구성하는 두 지도의 대응 객체 집합을 탐색하고 연속지적도의 지목을 이용하여 평가한 결과 0.946의 정확도를 얻을 수 있었다. 대부분의 잘못된 탐색은 대응 객체 집합의 무게 중심을 이용하여 위치 오차를 조정한 이후에도 남아있는 위치 및 형상 불일치에 의하여 중첩 관계가 왜곡되는 작은 면 객체에서 발생하였다.

제안된 다대다 객체 정합 방법은 공간정보 융합에 관련된 다양한 분야에 적용될 수 있다. 탐색하고자 하는 객체 집합을 유전 알고리즘에 적용이 가능한 형태의 후보해로 표현하고 정합 기준을 적절한 목적 함수로 표현하는 방법론을 개발할 경우 일반적인 주제도와 영상 기반 분류 결과와의 비교는 물론, Lidar자료로부터 얻어진 TIN자료와 이들 자료 사이의 비교에서도 활용될 수 있다. 또한 네트워크 정합에서도 활용 가능하리라 기대된다.

#### 감사의 글

이 논문은 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임 (2012R1A1A1A1015203)

#### References

- Bel Hadj Ali, A. (2002), Moment representation of polygons for the assessment of their shape quality, *Journal of Geographical Systems*, Vol. 4 No. 3, pp. 209-232.
- Burrough, P.A. and Frank, A.U. (1996), Geographic Objects with Indeterminate Boundaries, *GISDATA Series* vol. 2, Taylor & Francis, London, UK
- Butenuth, M., von Gosseln, G., Tiedge, M., Heipke, C., Lipeck, U. and Sester, M. (2007), Integration of heterogeneous geospatial data in a federated database,



- ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 62, No. 5, pp. 328–346.
- Huh, Y., Yu, K. and Heo, J. (2011), Detecting conjugate-point pairs for map alignment between two polygon datasets, *Environment and Urban Systems*, Vol. 35, No. 3, pp. 250–262.
- Mantel, D. and Lipeck, U. (2004), Matching cartographic objects in spatial databases, *Proceedings of the XXth Congress of the ISPRS*, Istanbul, Turkey, pp. 172–176.
- Moon, B. (2008), *IT Cookbook 81: Genetic algorithm with evolutionary approach*, Hanbit Media Inc., Seoul, Korea
- Reimann, C., Filzmoser, P. and Garrett, R. G. (2005), Background and threshold: Critical comparison of methods of determination, *Science of the Total Environment*, Vol. 345, No. 1-3, pp. 1–16.
- Stigmar, H. (2005), Matching route data and topographic data in a real-time environment, *Processing of the Xth Scandinavian research conference on geographical information science*, Stockholm, Sweden, pp. 89–107.
- Walter, V. and Fritsch, D. (1999), Matching spatial data sets: a statistical approach, *International Journal of Geographical Information Science*, Vol. 13, No. 5, pp. 445–473.
- Zhang, M. and Meng, L. (2007), An iterative road-matching approach for the integration of postal data. *Computers, Environment and Urban Systems*, Vol. 31, No.5, pp. 597–615.

---

(접수일 2012. 12. 10, 심사일 2013. 01. 23, 심사완료일 2013. 02. 25)