네트워크 분할 기법을 이용한 도로 네트워크 데이터 정합

Road network data matching using the network division technique

허 용¹⁾ · 손화민²⁾ · 이재빈³⁾

Huh, Yong · Son, Whamin · Lee, Jeabin

Abstract

This study proposes a network matching method based on a network division technique. The proposed method generates polygons surrounded by links of the original network dataset, and detects corresponding polygon group pairs using an intersection-based graph clustering. Then corresponding sub-network pairs are obtained from the polygon group pairs. To perform the geometric correction between them, the Iterative Closest Points algorithm is applied to the nodes of each corresponding sub-networks pair. Finally, Hausdorff distance analysis is applied to find link pairs of networks. To assess the feasibility of the algorithm, we apply it to the networks from the KTDB center and commercial CNS company. In the experiments, several Hausdorff distance thresholds from 3m to 18m with 3m intervals are tested and, finally, we can get the F-measure of 0.99 when using the threshold of 15m.

Keywords : Network matching, Network division, Intersection-based graph clustering, Iterative Closest Point, Hausdorff Distance

초 록

본 연구는 도로 네트워크를 분할한 뒤 네트워크 정합을 수행하는 방법론을 제안한다. 네트워크의 링크에 의해 둘러싸인 공간으로부터 폴리곤 객체를 생성하고 중첩기반 그래프 군집화 기법을 적용하여 이들 폴리곤 객체들의 대응 집합 쌍을 탐색하고, 이로부터 독립적인 정합을 수행하는 하위 네트워크 쌍을 얻었다. 이후 분할된 네트워크 데이터 쌍의 노드 집합에 Iterative Closest Point 알고리즘을 적용하여 기하보정을 수행한 후 하우스도르프 거리를 이용하여 1:N 링크 정합을 수행하였다. 제안된 방법의 효용성을 평가하기 위해 국가교통 DB센터에서 관리하는 KTDB의 네트워크 데이터와 국내 상용 차량 내비게이션 업체의 네트워크 데이터를 대상으로 알고리즘을 적용하고 결과를 평가하였다. 3m에서 18m까지 3m 간격으로 하우스도르프 거리에 임계값을 적용하여 네트워크를 정합한 결과 15m의 임계값에서 F-measure 기준 0.99의 정확도를 얻을 수 있었다.

핵심어 : 네트워크 정합, 네트워크 분할, 중첩기반 그래프 군집화, Iterative Closest Point, 하우스도르프 거리

1. 서 론

서로 다른 기관 또는 서로 다른 시점에 구축된 공간정보를 융합하는 것은 공간정보의 효율적인 상호 갱신은 물론, 개별

공간정보로부터는 얻을 수 없었던 새로운 정보를 얻을 수 있다는 점에서 그 필요성이 점차 증가하고 있다. 네트워크 데이터는 차량용 내비게이션 서비스의 핵심 데이터로 공공분야를 중심으로 구축되고 있는 다른 공간정보와는 달리 민간분야

1) Regular Member, Engineering Research Institute, Seoul National Univ., Korea (E-mail: huhhyong78@gmail.com)

2) Student Member, Dept. of Civil and Environmental Engineering, Seoul National Univ., Korea (E-mail: shm0407@gmail.com)

3) Corresponding Author, Regular Member, Department of Civil Engineering, Mokpo National Univ., Korea (E-mail: lee2009@mokpo.ac.kr)

에서도 활발하게 구축되고 있는 공간정보이다. 그런데 공공 분야의 네트워크 데이터는 중앙 및 지방 정부가 관리하고 있는 도로와 시설물에 관련된 정보를 중심으로 구축된 반면, 민간분야의 네트워크 데이터는 POI(Point Of Interest)로 표현되는 주택과 상가 건축물 및 그 부속시설에 관련된 정보를 중심으로 구축되고 있다. 따라서 이들을 융합함으로써 다양한 서비스를 개발하는 것이 가능하다.

공간정보를 융합하기 위해서는 동일한 지형지물을 표현하는 대응객체를 탐색해야한다. 만약 고유 명칭과 같은 속성정보가 융합하고자하는 두 공간정보에 포함되어 있다면 손쉽게 대응객체를 탐색할 수 있지만, 그렇지 않다면 기하학적 특징을 이용한 정합을 수행해야 한다. 최초의 네트워크 데이터 정합 사례는 미국 통계청(U.S. Census Bureau)에 의해 수행된 미 지질조사국(U.S. Geological Survey)의 TIGER(Topographically Integrated Geographic Encoding and Reference system)와 통계청의 GBF(Geographic Base File)/DIME(Dual Independent Map Encoding) 통합으로 노드의 위치와 노드에 연결된 링크의 개수 및 방향을 고려하여 노드 정합을, 링크의 길이와 방향 등을 이용하여 링크 정합을 수행하였다(Rosen and Saalfeld, 1985; Saalfeld, 1988). 그런데 1:1 대응관계만이 존재하는 노드와는 달리 링크는 1:N, M:1, M:N 등의 다양한 대응관계가 존재한다. 이 문제를 해결하기 위하여 Walter와 Fritsch(1999)는 두 네트워크 데이터에서 교대로 대응 링크 집합을 확장하면서 정합을 수행하는 Buffer Growing 알고리즘을 제안하였다. 이 알고리즘은 M:N 대응관계를 효과적으로 탐색할 수 있어 Mantel과 Lipeck(2004), Stigmar(2005), Zhang과 Meng(2007) 등의 연구에서 지속적으로 개선되었다. 반면 Li와 Goodchild(2011)은 네트워크 데이터 정합에 선형계획법을 적용한 최적화 문제로 접근하는 방법을 제안하였다. 이후 Yang *et al.*(2013)은 초기에 측정된 링크와 노드의 유사도에 확률적 완화(probabilistic relaxation) 기법을 적용함으로써 유사도 사이의 양립성(compatibility)을 고려하여 정확도를 향상하였다.

이처럼 많은 연구가 수행되었지만 보다 정확한 네트워크 데이터 정합을 수행하기 위해서는 두 네트워크 데이터의 위치 및 형상의 불일치 현상을 효과적으로 해결하는 것이 필요하다(Rosen and Saalfeld, 1985; Saalfeld, 1988; Walter and Fritsch, 1999). 선행 연구에서의 정합은 일반적으로 기하학적 특징을 비교하여 대응객체를 탐색한다. 노드 정합은 노드의 위치와 노드에 연결된 링크들의 위상정보를 활용하며, 링크 정합은 링크 사이의 거리와 방향을 비교하거나 특정 링크에 버퍼연산을 수행한 후 다른 네트워크 데이터의 어떤 링크들이 해당 버

퍼에 어느 정도 포함되는 가를 이용한다. 하지만 상이한 기관들이 각자의 활용목적에 따라 구축되고 관리되기 때문에 동일한 지형지물이라도 위치와 형상에 차이가 발생한다. 이러한 차이가 증가할수록 기하학적 특징을 이용하는 정합의 결과는 불안정해질 수밖에 없다. 따라서 전처리 과정에서 분석자에 의하여 대응 노드를 탐색하거나 엄격한 탐색조건을 적용하여 노드 정합을 수행한 후, 기하보정을 통하여 기하학적 불일치 현상을 완화한 네트워크 데이터에서 나머지 대응 노드 또는 대응 링크를 탐색하는 방법이 적용되고 있다(Walter and Fritsch, 1999). 하지만 위와 같은 전처리를 수행하기 위해서는 지도변환을 적용할 공간적 범위를 결정해야한다. 정합을 수행할 지역의 공간정보를 일괄적으로 취득하고 동일한 지도 작성 기준에 의하여 구축된 네트워크 데이터라면 단일 기하보정 모형을 적용하여 계통적 오차만을 해소하여도 네트워크 정합의 정확도를 효과적으로 향상시킬 수 있다. 그러나 정확성을 우선시하는 공공분야의 공간정보와는 달리 민간분야의 공간정보는 사용자들이 요구하는 최신성을 충족시키기 위하여 3개월 내지 6개월 주기로 갱신을 수행하고 있다. 이 과정에서 다양한 외부 공간정보를 활용하는 것은 물론 내부의 현장조사 부서에 의하여 자체적으로 갱신자료를 수집하고 반영하고 있다. 그 결과 인접한 지역의 네트워크 데이터라 할지라도 위치와 형상 불일치 정도와 방향이 상이한 현상이 발생하고 있다. 특히 서로 다른 세밀도를 가지는 네트워크 데이터 간에 이러한 현상이 더 많이 보고되고 있다. 이 경우 올바른 대응 노드를 탐색하였다하더라도 단일 기하보정 모형으로 전체 네트워크 데이터를 보정하는 것은 한계를 가진다. 따라서 정합을 수행할 두 네트워크 데이터를 일정한 규칙을 적용하여 하위 네트워크 데이터 쌍으로 분할하고 각각의 데이터에 기하보정을 수행하는 분할 기법을 적용하는 것이 효과적인 해결방안이 될 수 있다.

본 연구에서는 네트워크 데이터의 링크에 둘러싸인 공간을 폴리곤 객체로 생성하고 이들 폴리곤 객체들의 대응 집합 쌍을 탐색함으로써 분할된 하위 네트워크 데이터별로 정합을 수행하는 방법을 제안하였다. 중첩기반 그래프 군집화 기법을 적용하여 대응 폴리곤 집합 쌍을 탐색한 후, 이들 폴리곤 집합을 생성하는데 사용된 두 네트워크 데이터의 노드 집합 쌍에 ICP(Iterative Closest Point) 알고리즘을 적용하여 대응 노드 쌍을 탐색하고, 이를 이용하여 기하보정을 수행한다. 기하보정 후 하우스도르프 거리를 이용하여 링크 정합을 수행하였다. 제안된 방법의 효용성을 평가하기 위해 국가교통DB센터(<http://www.ktdb.go.kr>)에서 관리하는 KTDB의 네트워크 데이터와 국내 상용 차량 내비게이션 업체의 네트워크 데이터를 대상으로 적용하고 결과를 분석하였다.

2. 연구 방법

제안된 네트워크 데이터 정합 방법은 Fig. 1과 같이 네트워크 데이터의 분할과 대응 링크 쌍 탐색으로 구성된다. 네트워크 데이터의 분할은 다시 네트워크 데이터의 폴리곤 데이터 변환, 폴리곤 중첩 관계의 이분 그래프 표현, 그리고 그래프 군집화를 이용한 M:N 대응 폴리곤 집합 및 이들의 네트워크 데이터 쌍 탐색으로 이루어진다. 대응 링크 탐색은 ICP 알고리즘을 이용한 노드 정합 및 기하보정과 하우스도르프 거리를 이용한 링크 정합으로 구성된다. 각각의 세부적인 내용은 다음과 같다.

Fig. 1. Workflow of this study

2.1 네트워크 데이터 분할

2.1.1 네트워크 데이터의 폴리곤 데이터 변환

Timpf(1998)은 지도를 구성하는 계층적 요소를 Trans-hydro network, Container, Area, Element로 구분하였다. Container란 도로 또는 하천(Trans-hydro network)으로 둘러싸인 공간으로 가구계(block)에 대응되는 개념이며, Area는 Container를 토지이용으로 구분한 필지에 대응되는 개념이다. 마지막으로 Element는 Area 내부에 위치하는 건물이나 시설물에 대응되는 개념이다. 지도의 주제나 축척에 따라 동일한 지역일지라도 Area와 Element의 표현에는 많은 차이가 발생할 수 있지만 상대적으로 Trans-hydro network는 서로 다른 지도에서도 유사하게 표현된다. 따라서 이 네트워크에 의하여

정의되는 Container는 지도 융합과정에서 정합의 기본단위로 활용될 수 있다(Timpf, 1998). 본 연구에서도 네트워크 정합을 수행할 하위 네트워크의 공간적 범위를 Container 기준으로 분석하고자 한다. 즉 Fig. 2(a)와 (b)와 같이 네트워크 데이터가 있을 때 링크로 둘러싸인 Container를 폴리곤 객체로 변환하면 Fig. 2(c)와 (d)와 같은 폴리곤 데이터를 얻게 된다. 얻어진 폴리곤 데이터로부터 대응 객체 집합 쌍 및 이들로부터 하위 네트워크 데이터 쌍을 구한 후, 네트워크 정합을 수행한다.

Fig. 2. Conversion of network data into polygon data; (a) KTDB network data, (b) CNS network data, (c) Conversion result of (a), (d) Conversion result of (b)

2.1.2 폴리곤 중첩 관계의 이분 그래프 표현

차량 내비게이션 네트워크 데이터는 차량의 이동 가능한 경로를 모두 포함해야 하기 때문에 동일한 축척으로 구축된 다른 데이터에 비하여 더 세밀한 수준으로 링크를 표현해야 한다. 따라서 Fig. 2(a)에 비하여 Fig. 2(b)는 더 세밀한 단계까지 Container가 표현된 것을 확인할 수 있다. 이 경우 대응 폴리곤을 탐색하는 과정에서 1:1 대응관계뿐만 아니라 1:N, M:1 그리고 M:N과 같은 다양한 대응관계가 발생한다. 본 연구에서는 중첩기반 그래프 군집화 기법(Huh *et al.*, 2011)을 적용하여 1:1 대응관계는 물론 다양한 대응관계를 가지는 폴리곤 집합 쌍을 탐색한다. 이 방법은 두 폴리곤 데이터에서 각각의 폴리곤 객체를 노드로, 두 데이터 사이에서 중첩관계가 발생하는 객체 쌍을 링크로 표현한다. 동일한 데이터 사이에서는 폴리곤 중첩관계가 발생하지 않으므로 중첩관계는 Fig. 3(c)와 같이 이분 그래프의 형태로 얻어진다. 그런데 모든 공간 정보는 공간적 불확실성(spatial uncertainty)에 의하여 객체와 그 객체가 표현하는 실제 지형지물과의 위치와 형상이 불일치하는 현상이 발생한다. 따라서 원시 데이터 사이의 폴리곤 중첩여부를 이용할 경우 잘못된 중첩관계가 발생할 수 있다. 이 문제를 해결하기 위하여 본 연구에서는 불확실한 경계(indeterminate boundary) 모형을 이용하였다. 본 모형에 의하면 어떤 폴리곤 x_i 의 경계에는 불확실성이 포함되어 있으므로 이 폴리곤이 표현하고자 하는 실제 지형지물의 경계는 x_i 의 경계를 ϵ 만큼 확장한 x_i^{ext} 와 축소한 x_i^{int} 의 사이에 존재

한다고 가정한다(Huh *et al.*, 2011). 공간적 불확실성이 발생하더라도 x_i^{int} 는 안정적으로 실제 지형지물이 존재하는 범위이기 때문에 Fig. 3(a)와 (b)처럼 각 폴리곤을 수축시킨 x_i^{int} 사이의 중첩여부만으로 Fig. 3(c)와 같이 이분 그래프로 표현하였다. 여기서 ϵ 는 훈련지역에서 대응 관계를 가지는 폴리곤 객체 쌍의 대응 꼭지점 사이의 거리들을 분석함으로써 얻을 수 있다. 본 연구의 실험지역에서 수작업에 의하여 대응 꼭지점 사이의 거리들을 측정하고, 95% 유의수준에서 평균에 표준편차의 1.96배를 더하면 8.70m가 된다. 그런데 두 지도에서 모두 경계를 축소하므로 2로 나누어 각 지도에서 폴리곤의 경계를 축소해야 하는 거리 ϵ 을 4.35m로 결정하였다.

2.1.3 대응 폴리곤 집합 쌍 및 하위 네트워크 데이터 쌍 탐색

2.1.2 절의 이분 그래프에서 링크는 Eq.(1)과 같은 연결성 행렬 C 로 표현할 수 있다. 만약 폴리곤 x_i 의 x_i^{int} 와 x_j 의 x_j^{int} 가 중첩되어 있다면 $C(i, j)$ 의 원소값은 1이며, 그렇지 않다면 0이 된다. 그런데 앞에서 설명한 바와 같이 동일한 폴리곤 데이터 내에서는 자기 자신 이외에는 중첩관계가 발생하지 않으므로 Eq.(2)와 같은 이분 그래프의 형태를 가지게 된다.

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

$$C' = \begin{bmatrix} I_{n \times n} & C \\ C^T & I_{m \times m} \end{bmatrix} \quad (2)$$

여기서, n 과 m 은 폴리곤 데이터 A 와 B 의 객체의 개수이며, C' 는 폴리곤 데이터 A 와 B 사이에서 Eq.(1)과 같은 방법으로 표현된 $n \times m$ 크기의 연결성 행렬이다. Eq.(2)의 행렬에서 링크로 연결된 노드 집합의 탐색은 네트워크 연결성 분석 기법과 유사한 방식으로 얻을 수 있다(Chou, 1997). Fig. 3(c)의 이분 그래프는 Fig. 4(a)의 연결성 행렬로 표현할 수 있으며, 그래프 군집화 이론(Huh *et al.*, 2011)에 따라 행렬을 거듭제곱하면 Fig. 4(b)와 같은 행렬을 얻게 된다. 이 행렬의 각 행에서 원소값이 1인 열에 대응되는 폴리곤들을 데이터 A 와 B 사이의 대응 집합 쌍으로 분할한다. 이러한 방법으로 Fig. 4(b)에서 두 번째 행에서 {689, 693} : {3798, 3752, 3815}인 2:3 대응쌍을 얻을 수 있다. 동일한 방법으로 첫 번째, 세 번째 그리고 다섯 번째 행에서는 각각 1:1 대응쌍을, 여섯 번째 행에서는 {715}와 {3904, 3781, 3823, 3835}의 1:4 대응쌍을 얻을 수 있다.

Fig. 3. Finding corresponding polygons based on intersection relation; (a) result of negative buffer operation of Fig. 2(c), (b) result of negative buffer operation of Fig. 2(d), (c) representation of a bipartite graph

Fig. 4. Graph clustering of adjacency matrix C; (a) Matrix C' of Eq. (2), (b) result of power matrix method to matrix C'

하지만 위와 같은 과정은 중첩여부만을 이용하기 때문에 두 폴리곤 데이터 사이에서 폴리곤 집합의 형상이 매우 상이한 경우가 발생할 수 있다. 따라서 탐색된 대응 군집 쌍의 형상이 일정 수준 이상 유사해야 한다는 기준을 추가해야 한다. 만약 Fig. 4(b)에서 k 번째 행에서 폴리곤 데이터 A와 B에 속한 폴리곤 집합을 $a^{(k)}, b^{(k)}$ 이라고 한다면 Eq.(3)과 같이 각각 병합한 후 무게 중심을 일치시켰을 때의 중첩 면적비로 유사도를 측정한다. 본 연구에서는 이 유사도가 0.90 이상일 경우에 $a^{(k)}, b^{(k)}$ 에 속한 폴리곤들을 둘러싼 링크들로부터 하위 네트워크 쌍을 얻게 된다.

$$Sim(a^{(k)}, b^{(k)}) = \frac{U(a^{(k)}) \cap F_c(U(b^{(k)}))}{U(a^{(k)}) \cup F_c(U(b^{(k)}))} \quad (3)$$

여기서, U 는 폴리곤 집합의 병합(aggregation) 함수이며, F_c 는 폴리곤 집합의 무게 중심을 이용한 위치 오차 조정 함수이다.

2.2 네트워크 링크 정합

2.2.1 ICP 알고리즘을 이용한 노드 정합 및 기하보정

ICP 알고리즘은 Besl(1992)에 의하여 제안된 점 정합 방법으로 두 점 집합에서 가장 가까운 거리에 있는 쌍을 탐색하고, 이들의 거리를 최소화 시키는 변환함수를 반복적으로 탐색한다. 구체적인 연산과정은 네트워크 데이터로부터 ICP 알고리즘을 적용할 점 집합 데이터 추출, 추출된 두 점 집합 데이터 사이에 초기 대응점 탐색, 변환함수를 이용하여 오탐색된 대응점 제거, 오차 측정 및 최적화의 단계를 대응점의 탐색과 변환함수의 결정을 종료조건을 만족할 때까지 반복하게 된다. ICP 알고리즘의 성능은 초기 변환함수를 구하는데 사용되는 첫 번째 대응점 탐색 결과와 이후 반복적인 탐색 및 변환 과정에서 대응점을 탐색하는 판별기준에 많은 영향을 받는다. 공간정보의 경우 위치정보를 가지고 있기 때문에 초기 탐색결과와 비교적 안정적으로 얻을 수 있다. 그런데 ICP 알고리즘은 사전에 정의된 변환함수를 이용하여 위치와 형상 불일치를 조정하게 된다. 하지만 공간정보의 경우 개별 공간정보의 구축 과정에 따라 지역적으로 비계통적인 위치와 형상 불일치가 발생할 수 있다. 따라서 하나의 변환함수를 통해 전체 공간정보에 발생하는 불일치를 해결하는 것은 적합하지 않다. 본 연구에서는 분할된 하위 네트워크에서 각각의 ICP 알고리즘을 적용, 불일치를 해소한 이후 정합을 수행함으로써 지역적으로 존재하는 불일치에 의한 정합 오류를 줄이는 것이 가능하다. 이 과정에서 ICP 알고리즘의 변환함수로는 어파인(affine) 함수를 적용하였으며, 반복적인 대응점 탐색에서 거리가 9m 이내인 가장 인접한 노드 쌍을 변환함수의 결정에 적용하였다.

2.2.2 하우스도르프 거리를 이용한 링크 정합

본 연구에서는 링크 정합을 위해 공간정보 분야에서 제안된 거리측정 방법으로 가장 대표적인 방법들 중 하우스도르프 거리 알고리즘을 활용하였다. 두 점집합 A와 B가 주어졌을 때, A와 B 사이의 하우스도르프 거리는 다음과 같이 정의된다(Rucklidge, 1996).

$$H(A, B) = \max\{h(A, B), h(B, A)\} \quad (4)$$

$$h(A, B) = \max_{p_a \in A} \{\min_{p_b \in B} \{d(p_a, p_b)\}\} \quad (5)$$

$$h(B, A) = \max_{p_b \in B} \{\min_{p_a \in A} \{d(p_a, p_b)\}\} \quad (6)$$

Eq. (4)는 먼저 A의 각 점으로부터 B까지의 최단 거리들을 찾은 다음, 그 최단 거리들 가운데서 가장 큰 것을 $h(A, B)$ 의 값으로 선택하는 과정을 나타낸다. 직관적으로, $h(A, B)=d$ 라면, A의 각 점들은 B의 점들 중 하나로부터 d만큼의 거리 이내

에 있게 되고, A 의 점들 중에서 B 의 가장 가까운 점으로부터 정확히 d 만큼 떨어져 있는 점이 하나이상 존재해야 한다. Eq.(6)은 거리 $h(B, A)$ 에 대하여 위와 같은 내용을 나타낸다. 즉, Eq.(7)과 같이 $h(A, B)$ 와 $h(B, A)$ 중에서 더 큰 쪽을 하우스도르프 거리 $H(A, B)$ 로 결정한다.

$$H(A, B) = \max \left\{ \max_{p_a \in A} \left\{ \min_{p_b \in B} \{d(p_a, p_b)\} \right\}, \max_{p_b \in B} \left\{ \min_{p_a \in A} \{d(p_a, p_b)\} \right\} \right\} \quad (7)$$

이러한 표현 방식으로 인하여, 하우스도르프 거리는 최대-최소 거리라고도 불린다. 만약 하우스도르프 거리 값이 d 라면, 객체 A 의 모든 점은 B 의 점들 중 하나 이상으로부터 d 만큼의 거리 이내에 있어야 하며, 그 반대의 경우도 마찬가지이다. 또한 A (또는 B)의 점들 중 적어도 하나는, 그와 가장 가까운 B 의 점으로부터 정확히 d 만큼 떨어져 있게 된다. 본 연구에서는 분할된 하위 네트워크들을 구성하는 vertex 쌍을 단위로 링크들을 구분하고 target 집합과 reference 집합에 속한 모든 링크들 간의 하우스도르프 거리를 결정하고 하우스도르프 거리가 최소화되는 정합 쌍들에 대해 정합을 실시하였다. 본 연구 대상지역에서 차량 내비게이션 네트워크는 KTDB 네트워크보다 상세한 도로 선형을 포함하고 있다. 따라서 두 네트워크 사이에서는 M:1 대응 관계가 일반적으로 발생하고 있으며, Eq.(7)을 적용하여 차량 내비게이션 네트워크의 개별 링크가 KTDB의 어떤 링크에 대응되는지를 탐색하였다. 비록 2.2.1절의 과정을 통하여 두 네트워크 사이의 위치 오차를 최소화하였지만 국지적으로는 여전히 많은 오차가 발생하게 된다. 따라서 $h(link(i)_{CNS}, link(j)_{KTDB})$ 가 얼마 이내일 경우 대응관계를 가지는 것으로 결정하기 위한 임계거리가 필요하다.

본 연구에서는 3m에서 18m까지 3m 간격으로 임계거리를 적용하여 최적 임계거리를 분석하였다.

3. 적용 및 평가

제안된 방법을 평가하기 위하여 수원지역의 국가교통DB 센터의 KTDB 네트워크 데이터와 국내 차량 내비게이션 업체의 네트워크 데이터에 적용하였다. 국가교통DB는 교통정책 및 계획 수립 등에 필요한 기초자료를 조사·분석하여 공동 활용하기 위한 국가 차원의 교통DB로서 도로·철도·항만·공항·물류기지 등 교통시설 및 수단의 운영실태, 기종점교통량, 교통주제도 및 네트워크 정보를 포함하고 있다. 차량 내비게이션 네트워크 데이터에는 차량운전과 관련되어 도로등급, 링크종별과 같은 일반적인 교통네트워크 정보와 함께 버스전용차로구분 및 운행시간, 유무료 사항, 정방향 근거리 방면 명칭, 정방향 중거리 방면 명칭 등과 같은 정보를 포함하고 있다. Table 1은 두 데이터의 원시자료, 위치정확도 그리고 갱신주기와 같이 정합 과정에 영향을 미칠 수 있는 사항을 비교하여 보여준다.

Table 1. Specifications of network data

	KTDB network data	CNS network data
Source data	· 1:5,000 National topographic map · ITC standard node/link data	· National topographic map · New address map · Field survey
Positional accuracy	5.0 m	unknown
Revision period	1 year	3 months

Fig. 5. (a) and (b) KTDB and commercial CNS network data, (c) and (d) Detected 525 corresponding polygon dataset, (e) corresponding node pairs, (f) result of geometric correction, (g) result of link matching

실험대상지역의 두 네트워크 데이터는 그림 5(a), (b)와 같다. 이들을 폴리곤 데이터로 변환한 뒤, 2.1절에 기술한 방법론을 적용하여 Fig. 5(c), (d)와 같이 525개의 대응 폴리곤 객체 집합 쌍을 얻을 수 있다. 대응 폴리곤 객체집합 쌍을 기준으로 하위 네트워크 집합 쌍들을 구성하고 이들 집합 쌍을 구성하는 링크와 노드를 대상으로 네트워크 정합을 실시하였다. Fig. 5(e)는 2:4 대응 폴리곤 집합 쌍을 구성하는 하위 네트워크 쌍의 노드들을 대상으로 ICP 알고리즘을 적용한 결과로 9개의 대응 노드 쌍을 탐색하였다(Fig. 5(e)). 어파인 변환을 수행하여 기하보정을 실시한 결과는 Fig. 5(f)와 같다. 이를 바탕으로 하우스도르프 거리를 이용하여 대응 링크 집합을 탐색하면 Fig. 5(g)와 같이 네트워크 링크 정합이 수행된다. 여기서 점선으로 표시된 부분은 탐색된 7개의 대응 링크 쌍들을 보여준다.

하우스도르프 거리를 활용하여 링크 정합을 수행하는 과정에서 3m에서 18m까지 3m 간격으로 임계거리를 적용하면서 정확도를 평가한 결과는 Table 2와 같다. 임계거리가 6m이 내일 경우 정확률(precision)은 높지만 재현율(recall)이 매우 낮았다. 반면 임계거리가 12m 이상일 경우 정확률과 재현율이 모두 0.97이상으로 높았으며, 15m에서 정확률 0.981, 재현율 1.000 그리고 Eq.(8)과 같이 측정되는 F-measure가 0.991로 가장 정확한 결과를 얻을 수 있었다.

$$Fmeasure = \frac{2 \times precision \times recall}{precision + recall} \quad (8)$$

여기서, *precision*은 탐색된 정합 쌍 중 참인 정합 쌍의 비율이며, *recall*은 실제 존재하는 참인 정합 쌍 중 탐색된 정합 쌍의 비율이다.

Table 2. Assessment of the proposed method

threshold(m)	F measure	Precision	Recall
3	0.562	1.000	0.390
6	0.825	0.993	0.705
9	0.943	0.990	0.900
12	0.981	0.986	0.976
15	0.991	0.981	1.000
18	0.977	0.955	1.000

하지만 교차로 지역에서는 길이가 15m보다 짧은 링크가 존재할 수 있으며, 이 경우 Fig. 6과 같은 문제가 발생할 수 있다. 실선으로 표현된 차량 내비게이션 네트워크의 링크는 대

응되는 KTDB 네트워크의 링크를 가지지 않지만, 길이가 약 10m로 실제로는 대응되지 않는 KTDB 네트워크의 링크와의 $h(link(i)_{CNS}, link(j)_{KTDB})$ 가 8.97m가 되어 대응쌍으로 탐색되었다. 이와 같은 문제를 해결하기 위해서는 링크의 방향각과 같은 거리 외의 추가적인 유사도를 적용하는 것이 필요하다고 판단된다.

Fig. 6. An example of false matching results

4. 결론

본 연구에서는 이중의 공간 데이터의 네트워크 정합에 있어서 정합을 수행할 각각의 네트워크 데이터를 균일한 품질을 가지는 하위 네트워크 데이터로 분할하고 독립된 정합 모형을 적용하는 분할정복 개념을 적용하였다. 이를 위해 네트워크 데이터의 링크에 둘러싸인 공간을 폴리곤 객체로 생성하고 이들 폴리곤 객체들의 대응 집합 쌍을 탐색한 후, 이들 폴리곤 집합을 생성하는데 사용된 네트워크 데이터를 하위 네트워크 쌍으로 분할하였다. 분할된 하위 네트워크 쌍의 노드에 ICP 알고리즘을 적용하여 기하보정을 실시한 후 하우스도르프 거리를 이용하여 링크 정합을 수행하였다. 이러한 과정을 통하여 공간 DB에 존재하는 기하학적 오류가 해당 하위 네트워크별로 관리 및 탐색될 수 있어 주변 지역의 정합결과를 왜곡하는 문제를 방지하고, 하위 네트워크별로 정합 과정을 적용함으로써 정확률을 향상시키는 것이 가능하였다. KTDB 네트워크와 차량 내비게이션 네트워크에 적용결과 하우스도르프 거리 임계값 15m에서 F measure 기준 0.99의 정확도를 얻을 수 있었다. 하지만 교차로 지역에서 길이가 짧은 링크의 경우 잘못된 대응 링크와의 하우스도르프 거리가 링크의 길이만큼만 발생하기 때문에 정합에 오류가 발생하였다. 이러한 문제를 해결하기 위해서는 거리 이외의 방향각과 같은 추가 정보를 링크 정합과정에 반영함으로써 정확도를 향상시키는 것이 필요하다고 판단된다.

Acknowledgement

이 논문은 2013년도 정부(미래창조과학부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임 (2012R1A1A1A1015203)

References

- Besl, P. and McKay, N. (1992), A method for registration of 3-D shapes, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 14, No. 2, pp. 239-256.
- Huh, Y., Yu, Y. and Heo, J. (2011), Detecting conjugate-point pairs for map alignment between two polygon datasets, *Computers, Environment and Urban Systems*, Vol. 35, No. 3, pp. 250-262.
- Li, L. and Goodchild, M. (2011), An optimisation model for linear feature matching in geographical data conflation. *International Journal of Image and Data Fusion*, Vol. 2, No. 4, pp. 309-328.
- 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.
- Min, D. Zhilin, L. and Xiaoyong, C. (2007), Extended Hausdorff distance for spatial objects in GIS, *International Journal of Geographical Information Science*, Vol. 21, No. 4, pp. 459-475.
- Rosen, B. and Saalfeld, A. (1985), Match criteria for automatic alignment, *Proceedings of Auto-Carto 7*, 11-14 March, Washington, D.C., pp. 1-20.
- Rucklidge, W. (1996), *Efficient visual recognition using the Hausdorff distance*, Springer, New York, N.Y., 95 p.
- Saalfeld, A. (1988), Conflation automated map compilation, *International Journal of Geographical Information Science*, Vol. 2, No. 3, pp. 217-228.
- Stigmar, H. (2005), Matching route data and topographic data in a real-time environment. *Proceeding of the Xth Scandinavian research conference on geographical information science*, Stockholm, Sweden, pp. 89-107.
- Timpf, S. (1998), *Hierarchical structures in map series*, Ph.D The technical university of vienna
- 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.
- Yang, B., Zhang, Y. and Luan, X. (2013), A probabilistic relaxation approach for matching road networks, *International Journal of Geographical Information Science*, Vol. 27, No, 2, pp. 319-338.

(Received 2013. 07. 18, Revised 2013. 08. 02, Accepted 2013. 08. 30)