

GIS DB 갱신을 위한 공간 객체간 Partial Hausdorff Distance 측정 방법 연구

Study on Method for Measuring Partial Hausdorff Distance between Spatial Objects for Updating GIS DB

방윤식*, 허용, 최재완, 유기윤

Yoonsik Bang*, Yong Huh, Jaewan Choi, Kiyun Yu

서울대학교 건설환경공학부

{bangys1004*, hy7808, iz2003, kiyun}@snu.ac.kr

요약

본 논문에서는, 건설도면의 건물 경계선을 이용하여 기존 지형도의 건물 레이어를 자동 갱신하는 과정에서 건물 객체간 유사를 측정하는 과정에서 Hausdorff 거리 측정 방법을 적용해 보았다. 서로 다른 축척에 의하여 발생하는 형상의 세밀도 차이에서 발생하는 오차를 최소화하기 위하여 partial Hausdorff 거리를 적용하였다. 이 때 partial fraction을 정의하기 위한 기준으로 폴리곤 절점, 경계선 길이, 폴리곤 면적을 이용한 세 가지 방법을 적용하였다. 이 기준을 실제 데이터에 적용한 결과, 경계선 기준과 폴리곤 면적 기준 방법이 절점 기준에 비하여 안정적이라는 것과, 적용 기준에 따라 적절한 partial fraction 값을 사용하면 잡음에 의한 영향을 최소화할 수 있다는 것을 확인할 수 있었다.

1. 서론

최근 지리정보체계의 최신성을 확보하기 위한 방안으로, 건물 또는 도로 등 항공측량 등을 통하여 갱신되는 주기보다 빠르게 변화하는 공간 객체를 건설도면을 이용하여 갱신하기 위한 연구가 활발히 진행되고 있다. 건설도면을 이용하여 기존 지형공간정보를 갱신하기 위해서는 좌표 변환의 문제는 물론 기존 공간객체와 갱신과정에서 입력된 건설도면상의 객체와의 정합과정이 필요하다. 하지만 지리정보체계의 객체와 건설도면상의 객체를 연계

할 수 있는 식별자 체계의 미비와 건설도면상의 객체에 부여되는 실좌표 정확도의 문제로 인하여, 좌표를 기준으로 지리정보체계상의 객체와 건설도면상의 객체를 중첩하여 비교하는 과정이 필요하다. 이 때 두 자료의 객체 사이의 거리는 정합에 중요한 결정 인자가 된다. 그런데 건설도면과 지리정보체계의 공간 객체의 기하정보는 그 축척이나 세밀도(level of detail) 등에서 서로 다른 특성을 나타낸다. 따라서 속성정보 없이 기하정보만으로 이를 판단하기 위해서는 이러한 특성을 고려한 거리 측정방법이 필요하다.

공간 분석에서 객체들 사이의 거리, 즉 '공간적 거리'를 측정하는 방법은 두 점 사이의 거리로 단순화할 수 없는 문제점이 있다. 2차원 벡터 공간 데이터에서 점 객체는 단순히 두 점 사이의 유클리드 거리를 측정하는 것만으로 공간적 거리를 대표하는 값으로 사용할 수 있다. 그러나 대부분의 공간 객체는 점들의 집합으로 표현되는 선 객체 또는 면 객체의 형태이다. 이러한 경우에는 유클리드 거리 이외에, 객체간 거리를 측정할 새로운 방식이 필요하다 이 때 가장 많이 사용되는 방법이 바로 하우스도르프 거리(Hausdorff Distance, HD)이다.

HD는 원래 두 개의 점 집합들 사이의 거리를 측정하기 위하여 제시되었다. 이 거리는 스칼라량 H 으로 표현되며, 한 집합 내의 임의의 점에서 거리 H 이내에 다른 집합 내의 점이 적어도 하나 이상 존재한다는 것을 의미한다. 기존의 HD는 측정하고자 하는 객체 형상의 잡음에 민감한 문제점을 가지고 있다. 이를 보완하기 위하여 여러 종류의 개선된 HD가 제안되었다. 대표적으로 Huttenlocher(1993)의 Partial HD(PHD), Azencott(1996)의 censored HD(CHD), Dubuisson과 Jain(1994)의 modified HD(MHD) 등이 있다.

하지만 이들이 제안한 기법들은 기본적으로 점 집합사이의 거리를 측정하기 위한 기법들로서 폴리곤의 형상을 구성하는 절점들의 분포 형태에 영향을 받게 된다. 즉, 잡음이 발생하지 않은 지역이 세밀한 형상을 표현하기 위하여 다수의 절점들이 분포하고 있다면 이들 절점들과 비교될 공간객체와의 거리들이 최종 거리 판단에 큰 영향을 미치게 된다. 본 연구에서 이를 같은 문제점을 해결하기 위하여 HD 측정이 발생하는 공간객체 상의 지점들을

균일하게 분포하기 위한 기법으로 두 공간객체의 경계선 길이를 정규화하고 등간격으로 분포시킨 절점을 이용한 방법과 두 공간객체의 면적을 정규화하고 내부를 동일한 개수의 격자로 분할하고 각각의 격자들을 거리 계산의 기준점들로 사용하였다. 또한 잡음 및 세밀도 차이에 의하여 발생할 수 있는 오차를 완화하기 위하여 기존 연구들에서 사용된 partial fraction을 적용하였으며, partial fraction을 변화시키면서 제안된 기법을 통하여 측정된 HD의 변화를 비교해보았다.

2. Hausdorff Distance

2.1. HD의 정의

좌표를 가지고 있는 점들이 순차적으로 나열된 형태로 정의되는 폴리곤 객체 A 와 B 사이의 HD는 다음과 같이 정의된다.

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

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

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

여기서 $h(A, B)$ 는 directed HD라고 부르며, 임의의 점 p_a 에 대하여 가장 가까운 B 의 점과의 거리($\min_{p_b \in B} d(p_a, p_b)$)를 구하고 모든 p_a 에서 측정된 거리 중 가장 큰 값($\max_{p_a \in A} \{\cdot\}$)으로 정의된다. 다시 말하면, 객체 B 에 대해 버퍼를 생성할 때에 그 버퍼가 A 의 모든 점을 포함하도록 하는 버퍼의 크기가 곧 $h(A, B)$ 의 값이 된다. 그림 1은 이렇게 구한 $h(A, B)$ 와 $h(B, A)$ 를 보여준다. 이 중 큰 값이 곧 A 와 B 사이의 HD로 정의된다.

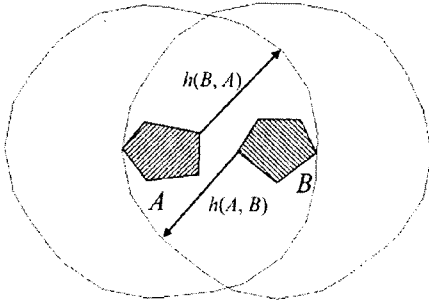


그림 2 HD의 개념(Min, 2007)

2.2. Partial HD의 정의 및 종류

HD는 점의 집합으로 이루어진 일반적인 형태의 객체 사이의 거리를 측정하는 기준을 제시해 준다. 그러나 이러한 HD는 만약 객체를 이루는 점들 가운데 일부의 좌표값에 손상이 생겼을 경우, 그 손상에 의한 영향을 크게 받는다. 이는 directed HD들 중 하나의 값이 비정상적으로 커지게 되고, HD의 정의상 극값을 취하기 때문이다. 이러한 문제점을 해결하기 위하여 다양한 기법이 제안되었다.

Huttenlocher(1993) 등은 안정적인 HD를 얻을 수 있는 partial HD(PHD) 기법을 제안하였다. 이 기법은 손상된 기하학적 형상의 영향을 최소화하기 위하여,

$$h_k(A, B) = K_{a \in A}^{th} d_B(a)$$

와 같은 방법을 이용하였다. $K_{a \in A}^{th}$ 는 순차적으로 정렬된 $d_B(a)$ 값들 중 K 번째에 해당하는 순서의 값을 의미한다. 이 기법에서 잡음에 의하여 영향을 받은 정도에 따른 HD를 얻기 위하여 partial fraction f 를 정의하였는데, f 는 0부터 1사이의 값

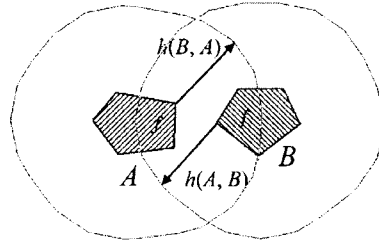


그림 3 Partial Hausdorff 거리

을 가지며, $K = fN_A$ 로 결정된다. 따라서 객체 B 에 대해 버퍼를 생성할 때에 그 버퍼가 A 의 일정 비율(f)을 포함하도록 하는 버퍼의 크기가 곧 $h(A, B)$ 의 값이 된다.

이외에 Azencott (1996) 등은 잡음문제를 해결하기 위하여 censored HD(CHD)를 제안하였다.

$$h_{CHD}(A, B) = K_{a \in A}^{th} L_{b \in B}^{th} \|a - b\|$$

여기서 $K_{a \in A}^{th}$ 는 순차적으로 나열된 $L_{b \in B}^{th} \|a - b\|$ 값들 중에서 K 번째에 해당하는 순서의 값을 나타내고 $L_{b \in B}^{th}$ 는 순차적으로 나열된 유클리드 거리값 $\|a - b\|$ 의 L 번째 값을 나타낸다. 잡음에 의한 훼손정도에 따른 유사도 판단을 위하여 β, γ 를 파라미터로 정의하였는데, β 는 0부터 1사이의 값을 갖고, $K = \beta N_A$ 로 결정되며, γ 는 0부터 1사이의 값을 갖고, $L = \gamma N_B$ 로 결정된다.

Dubuisson과 Jain(1994)은 $d_B(A, B)$ 의 거리값들을 평균한 값을 택하는 modified HD(MHD)를 제안하였다. 이 방법은 PHD나 CHD와 같이 사용자가 직접 파라미터를 지정할 필요가 없다.

$$h_{MHD}(A, B) = \frac{1}{N_A} \sum_{a \in A} d_B(a)$$

3. Partial Fraction 계산 기준

3.1. 폴리곤 절점 기준

Partial HD를 계산할 때에 가장 중요한 작업은 partial fraction (f)값을 계산하는 것이다. 이 때, f 의 값을 어떤 요소를 기준으로 결정하느냐가 결과값에 큰 영향을 미친다. 가장 간단한 방법으로, 폴리곤을 구성하는 절점을 기준으로 계산할 수 있다. 즉, A 의 절점들로부터 B 를 구성하는 절점까지의 최소거리들을 짧은 순서대로 늘어놓았을 때, f 만큼의 비율의 순위를 차지하는 거리값을 $h(A, B)$ 의 값으로 택하는 것이다. 따라서 위의 식을 그대로 적용할 수 있다.

그러나 일반적인 경우에 폴리곤의 점밀도가 일정하지 않으므로, 지역에 따라 결과값에 미치는 영향이 서로 대등하지 못하게 된다. 점 밀도가 높은 부분(자세히 표현된 부분)은 f 값의 결정에 영향을 많이 미치는 반면, 점 밀도가 낮은 부분(간직선으로 간략히 표현된 부분)은 상대적으로 적은 영향을 미친다.

3.2. 경계선 기준

비율 f 의 값을 결정할 때 절점의 개수 대신 폴리곤의 경계선 경계선을 기준으로 계산할 경우 위와 같은 문제점을 어느 정도 해결할 수 있다. 이 경우에는, A 의 둘레(변) 위의 모든 점들로부터 B 까지의 최소거리들을 짧은 순서대로 늘어놓았을 때, f 만큼의 비율의 순위를 차지하는 거리 값을 $h(A, B)$ 의 값으로 택하는 것이다. 이

는 다음과 같은 식으로 표현할 수 있다.

$$h(A, B) = \min \left\{ \epsilon : f \leq \frac{s_\epsilon(A)}{s(A)} \right\}$$

여기에서, $s(A)$ 는 A 의 둘레이고, $s_\epsilon(A)$ 는 A 의 둘레 중 B 로부터의 거리가 ϵ 이 내인 부분의 길이이다.

3.3. 면적 기준

위와 같은 계산을 폴리곤의 면적을 기준으로 수행하는 방법도 생각할 수 있다. 이는 다음과 같이 표현할 수 있다.

$$h(A, B) = \min \left\{ \epsilon : f \leq \frac{a_\epsilon(A)}{a(A)} \right\}$$

여기에서, $a(A)$ 는 A 의 면적이고, $a_\epsilon(A)$ 는 A 의 면적 중 B 로부터의 거리가 ϵ 이 내인 부분의 면적이다.

4. 실험 및 평가

실험은 서울대학교 공과대학 38동 건물에 대한 수치지형도상의 벡터 데이터와, CAD 도면으로부터 추출한 건물 외곽선 데이터를 이용하여 진행하였다(그림 3).

절점 기준의 PHD를 계산하는 것은 단순한 두 점 사이의 유클리드 거리를 계산하는 것만으로도 정확한 결과를 얻을 수 있다. 그러나 경계선 기준 또는 면적 기준의 경우, 연속적으로 변하는 비율 f 값에 대해 둘레 길이와 면적을 연속적으로 계산하는 것은 매우 복잡한 연산 과정이 필요하다. 때문에 구현 과정에서 근사법을 쓰는 것이 합리적이고, 이에 따라 약간의 오차가 생길 수 있다.

경계선 기준 PHD의 실제 구현 과정에

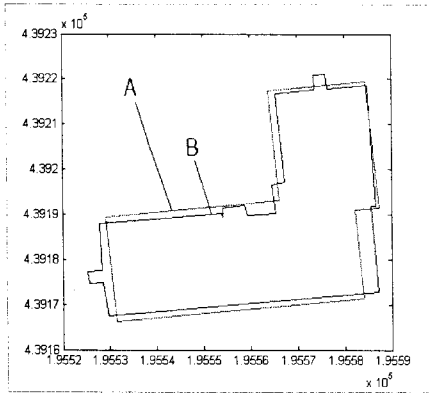


그림 4 실험 대상 데이터. A는 수치지형도, B는 CAD 도면에서 얻은 건물 외곽선 데이터이다(단위:m).

서는, 대상 폴리곤의 각 변(edge) 위에 작고 일정한 간격으로 새로운 점들을 찍은 다음 그 점들로 이루어진 폴리곤을 대상으로 질점 기준의 PHD를 계산하였다. 면적 기준 PHD의 경우에는 폴리곤 내부 영역 내에 일정한 간격으로 새로운 점들을 찍은 다음 마찬가지로 방법으로 계산하였다(그림 4). 새로운 점들을 찍는 간격은 폴리곤의 크기와 관계없이 두 폴리곤에

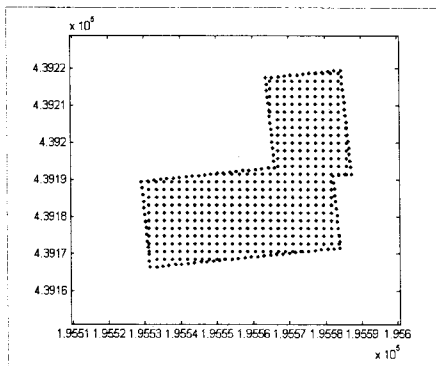


그림 5 수치지형도 데이터의 폴리곤 내부 영역에 새로운 점들을 찍은 모습.

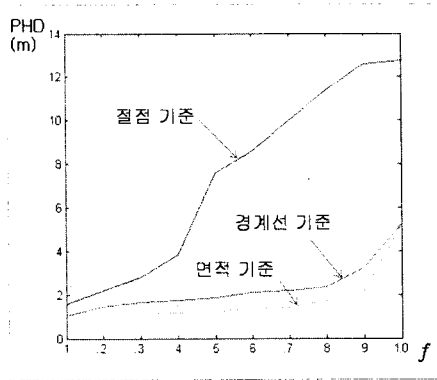


그림 6 f 값에 따른 PHD 계산 결과.

대해 같은 개수의 점이 찍히도록 하였다. 이는 거리를 측정하는 두 객체가 실세계에서 서로 같은 물체를 나타내는지 여부를 알고자 하는 것이 계산의 목적이기 때문이다. 그렇기 때문에 두 객체 폴리곤의 크기가 서로 다르더라도 거리 측정 결과에 동등한 영향을 주도록 해야 한다.

이러한 방법을 통해 두 폴리곤 사이의 PHD를 세 가지 기준에 대해서 계산한 결과는 표 1과 같다. 비율 f 값을 0.1부터 1.0까지 변화시켜 가면서 각각의 경우에 PHD를 계산하였다. 이론적 배경으로부터, 비율 f 값이 1.0일 경우의 PHD 값이 곧 기존의 HD 값과 같음을 유추할 수 있다.

그림 5에 나타난 결과에서 볼 수 있듯이, 질점 기준으로 계산할 경우 점-선분 거리가 아닌 점-점 거리가 계산되므로 경계선 기준 또는 면적 기준의 경우에 비해 상대적으로 큰 값이 나오게 된다. 이 실험에서 폴리곤이 의미하고 있는 건물 객체의 경우, 폴리곤의 꼭지점 뿐 아니라 각 변들과 내부 공간도 건물 객체를 구성하는 중요한 요소이다. 따라서 점 기준보다는 경계선 또는 면적 기준의 계산 방법이 좀 더 타당하다.

위의 주장을 좀 더 명확히 뒷받침하기

표 1 랜덤 잡음을 가한 데이터의 PHD 계산 결과의 RMSE 값. (단위:m)

f 값	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
점 기준	0.3923	0.4271	0.4843	0.5911	0.5142	0.3454	0.4701	0.6055	0.6224	0.7081
경계선 기준	0.3201	0.2689	0.2472	0.1979	0.2425	0.2738	0.4572	0.6184	0.4381	0.6047
면적 기준	0.3129	0.3503	0.3827	0.3718	0.3408	0.3239	0.3101	0.2258	0.2136	0.6047

위하여 데이터의 손상 및 오차에 대한 강건함을 시험해볼 수 있다. 앞에서 사용한 원본 데이터 중, CAD 도면으로부터 추출한 건물 외곽선 데이터에 $\pm 1m$ 이내의 랜덤 잡음(오차)을 발생시켰다(그림 6). 그런 다음, 잡음을 가한 데이터를 이용하여 앞의 계산 과정을 반복하였다(그림 7). 이를 10회 반복 실험하였고, 원본 데이터의 결과를 기준으로 잡음을 가한 데이터의 결과들의 RMSE를 취하였다. 그 결과는 표 2와 같다. 그림 8은 이 결과를 그래프로 나타낸 것이다.

그림의 결과에서 볼 수 있듯이 대상 데이터에 의미 없는 잡음이 발생하였을 경우, f 값의 변화에 따른 RMSE 값은 PHD 계산 방법에 따라 다르게 나타난다. 전반

적으로 점 기준 PHD는 경계선 또는 면적 기준 PHD에 비하여 RMSE 값이 높으므로 상대적으로 덜 강건하다. 경계선 기준 PHD는 낮은 f 값(0.2~0.5)에 대하여 비교적 강건한 모습을 보이지만, 0.7 이상의 f 값에 대하여 RMSE 값이 급격히 증가하고 있다. 이에 반해 면적 기준 PHD는 높은 f 값(0.7~0.9)에 대하여 강건한 모습을 보인다. 이것은 폴리곤 좌표값 데이터에 잡음이 생길 경우, 경계선 길이는 일반적으로 증가하는 반면, 면적은 변화를 나타내지 않기 때문이다. 이 효과는 f 값이 작을 경우에는 잘 드러나지 않다가, f 값이 커지면 폴리곤 전체의 특성이 반영되면서 위의 결과와 같은 특징이 나타나게 된다. 결과적으로, 계산에 사용하는 PHD 계산 방법에 따라 적합한 f 값의 크기가 달라진다는 것을 알 수 있다.

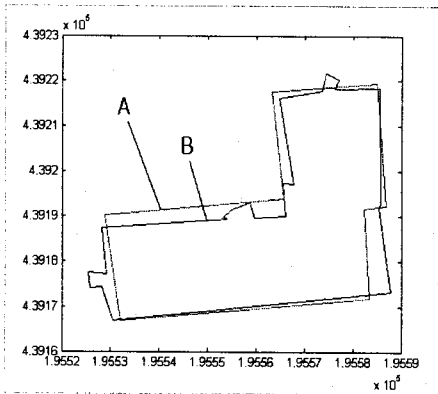


그림 7 원본 데이터에 $\pm 1m$ 이내의 랜덤 잡음을 가한 데이터. A는 수치지형도, B는 CAD 도면에서 얻은 건물 외곽선 데이터이다(단위:m).

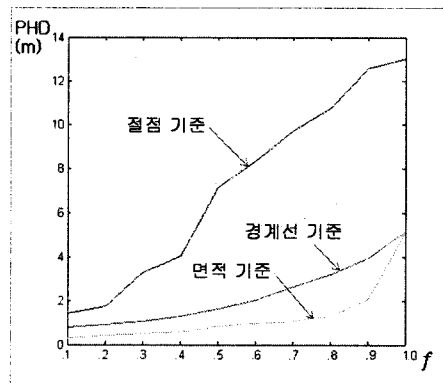


그림 8 랜덤 잡음을 가한 데이터의 PHD 계산 결과.

표 2 원본 데이터의 PHD 계산 결과 (단위:m)

f 값	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
점 기준	1.5684	2.2020	2.8018	3.8406	7.5842	8.5735	10.013	11.397	12.597	12.732
경계선 기준	1.0557	1.4498	1.6428	1.7411	1.8481	2.0906	2.1712	2.3681	3.2465	5.2398
면적 기준	0.8641	1.0049	1.0976	1.1710	1.2444	1.3194	1.4130	1.7457	2.1802	5.2398

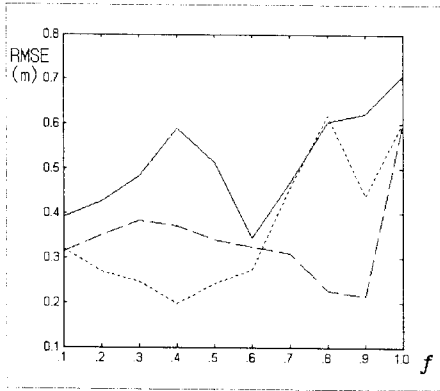


그림 9 랜덤 잡음을 가한 데이터의 PHD 계산 결과의 RMSE 값. 실선이 점 기준, 점선이 경계선 기준, 파선이 면적 기준의 PHD이다.

5. 결론

본 논문에서는 Hausdorff 거리(HD)를 이용하여 점집합으로 이루어진 공간 객체 사이의 공간적 거리를 측정하는 방법에 대하여 연구하고, 실제 건물 폴리곤의 경우에 이 방법을 적용하여 그 결과를 분석하였다.

데이터의 손상에 대해 강건함을 유지하기 위하여 기 개발된 여러 종류의 수정된 기법들 중, Partial HD(PHD)를 이용하여 거리를 측정하였다. 이 때 가장 중요한 파라미터인 partial fraction (f)값을 결정하는 기준 요소의 형태에 따라 PHD를 점 기준, 경계선 기준, 면적 기준으로 세분화

하고 각 기준의 계산 방법을 제안하였다. 그리고 이 방법을 실제 수치지형도와 CAD 도면에서 추출한 데이터에 적용한 결과, 점 기준보다는 경계선과 면적 기준의 PHD가 더 타당한 방식임을 확인할 수 있었다.

또한 데이터 손상에 대한 강건함을 시험하기 위해 원본 데이터에 랜덤 잡음을 가한 다음 세 가지 방식으로 PHD를 측정하였다. 그 결과 낮은 f 값(0.2~0.5)의 경우에는 경계선 기준 PHD가, 높은 f 값(0.7~0.9)의 경우에는 면적 기준 PHD가 타당한 것으로 나타났다. 따라서 PHD의 계산 기준에 따라 적절한 크기의 f 값을 사용하면 잡음 오차의 영향을 최소화하여, 공간 객체간 유사도 측정을 위한 거리 측정의 정확도를 향상시킬 수 있다.

감사의 글

본 연구는 국토해양부 첨단도시기술개발사업 - 지능형국토정보기술혁신사업과 제의 연구비지원(07국토정보C04)에 의해 수행되었습니다.

참고문헌

1. 고중호, 이철우, Hausdorff Distance를 이용한 영역 정합 기반 자동 영상 모자이크. 전자통신기술 논문지, 제3권 제 1호, pp.73-79, 2000.

2. 박정환, 손홍규, 컬러영상의 경계정보와 색상정보를 활용한 동일건물인식. 대한토목학회논문집, 제 26권 제 3D호, pp.519-525, 2006.
3. AZENCOTT, R., DURBIN, F., and PAUMARD, J., Multiscale identification of building in compressed large aerial scenes. *Proc. of 13th International Conference on Pattern Recognition(ICPR '96)*, Vol. 3, pp. 974-978, 1996.
4. DUBUISSON, M.-P. and JAIN, A. K., A modified Hausdorff distance for object matching. *Proc. of International Conference on Pattern Recognition(ICPR '94)*, Vol. 1, pp. 566-568, 1994.
5. HUTTENLOCHER, D.P., KLANDERMAN, G.A. and RUCKLIDGE, W.A., Comparing Images Using the Hausdorff Distance. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 9, pp. 850-863, 1993.
6. MIN, D., ZHILIN, L. and XIAOYANG, C., Extended Hausdorff Distance for Spatial Objects in GIS. *International Journal of Geographical Information Science*, Vol. 21, No. 4, pp. 459-475, 2007.
7. PENG, X., CHEN, W. and MA, Q., Feature-based Nonrigid Image Registration Using a Hausdorff Distance Matching Measure. *Optical Engineering*, Vol. 46, No. 5, pp. 057201-1-16, 2007.
8. TIAN, L., Kamata, S.i., UESHIGE, Y. and KUROKI, Y., An Automatic Image-Map Registration Algorithm Using Modified Partial Hausdorff Distance. *Proc. of IGARSS2005*, pp. 3534- 3537, 2005.