라인 곡선 곡률 기반의 벡터 데이터 해싱

정회원 이 석 환*, 권 기 룡**°

Vector Data Hashing Using Line Curve Curvature

Suk-Hwan Lee*, Ki-Ryong Kwon**° *Regular Members*

요 약

최근 CAD 설계도면 및 GIS 디지털 맵과 같은 벡터 데이터 모델의 응용 분야가 확대되면서 이에 대한 보호 기술이 필요하게 되었다. 본 논문에서는 벡터 데이터 모델의 인증 또는 복사방지에 필요한 벡터 데이터 해싱 방법을 제안한다. 제안한 해싱에서는 벡터 데이터 모델 내 주요 레이어 상에 폴리라인들을 그룹화한 다음, 폴리라인의 1차 및 2차 곡선 곡률 분포에 따라 그룹 계수를 생성한다. 그리고 이들 그룹 계수를 랜덤 계수 키 패턴으로 투영기에 의하여 특징 계수를 얻은 다음, 이를 이진화 과정에 의하여 최종 이진 해쉬를 생성한다. 설계도면 및 디지털 맵을 이용한 실험 결과로부터 제안한 방법에 의하여 생성된 해쉬가 다양한 공격에 대한 강인성과 랜덤 키에 의한 보안성 및 유일성을 만족함을 확인하였다.

Key Words : Vector data model, Content hashing, CAD design drawing, GIS digital map, Line curve curvature

ABSTRACT

With the rapid expansion of application fields of vector data model such as CAD design drawing and GIS digital map, the security technique for vector data model has been issued. This paper presents the vector data hashing for the authentication and copy protection of vector data model. The proposed hashing groups polylines in main layers of a vector data model and generates the group coefficients by the line curve curvatures of the first and second type of all polylines. Then we calculate the feature coefficients by projecting the group coefficients onto the random pattern and generate finally the binary hash from the binarization of the feature coefficients. From experimental results using a number of CAD drawings and GIS digital maps, we verified that the proposed hashing has the robustness against various attacks and the uniqueness and security by the random key.

Ⅰ. 서 론

벡터 데이터 모델은 건축, 자동차, 조선, IT 하드웨어 분야 등의 산업 분야에서 활용되는 CAD 설계 도면[1]과 지리정보시스템의 디지털 맵[2] 등과 같은 데이터로 이 데이터를 기반으로 다양한 응용 콘텐츠 분야로 확대되어 사용되고 있다. 따라서 벡터 데이터 모델 기반의 다양한 콘텐츠가 개발되면서 이에 대한 보호 기술의 필요성이 제기되면서, CAD 설계도면[3-7]과 GIS 벡터맵[8-10]의 저작권 보호를 위한 다양한 워터마킹 기법이 연구되어져 왔다. 그러나 Delp[11]는 21세기 콘텐츠 보안 기술 방향으로 비즈니스 모델 변경과 함께 저작권 보호가 아닌 콘텐츠 인증에 대한 중요성을 제기하였다.

콘텐츠에 대한 인증 기술의 가장 일반적으로 방법으로 해싱이 있으며, 해싱에는 암호학적 기반의 해싱과 콘텐츠 기반 해싱이 있다. 암호학적 해싱은 비트 변화에 민감하므로 화질이 유지되면서 다양한 형태로

※ 본 논문은 2009년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(KRF-2009-0071269)
* 동명대학교 정보보호학과 (skylee@tu.ac.kr), ** 부경대학교 IT융합응용공학과 (krkwon@pknu.ac.kr), (°:교신저자)
논문번호 : KICS2010-08-407, 접수일자 : 2010년 8월 19일, 최종논문접수일자 : 2011년 1월 20일

변형이 가능한 멀티미디어 또는 콘텐츠에는 적합하지 못하다. 따라서 많은 연구자들은 강인성과 보안성을 만족하는 영상[12,13], 비디오[14,15]와 같은 영상 콘텐츠 해싱에 대하여 제안하였으며, 최근에는 3D 모델 해싱[16-19]에 대한 연구가 이루어지고 있으나, 이는 벡터 데이터 모델과는 다른 3D 메쉬 모델에 대해서만 적용되는 방법이다.

워터마킹에서와 같이 콘텐츠 해싱에서도 콘텐츠의 데이터 구조에 따라 다르게 적용되어야 한다. 예를 들어, 영상 및 비디오는 고정된 위치와 크기의 화소 배열과 같은 프레임 또는 이미지들로 구성되어 있으며, 화소들의 공간 영역 또는 DCT, DWT, FMT 변환 영역 기반으로 많은 기법들이 제안되었다. 3D 그래픽스 모델은 정형화 또는 고정되지 않은 3D 꼭지점 좌표에 의한 NURBS 곡면 또는 다각형 메쉬로 구성되며, 기하학적 형상 특징 분포 기반의 해싱이 제안되었다. 그러나 이들 해싱들은 다른 데이터 구조로 인하여 벡터 데이터 모델에 대해 적용되지 못한다. 즉, 벡터 데이터 모델은 여러 개의 레이어 구조로 되어 있으며, 각 레이어별로 점, 폴리라인, 폴리곤 및 타원 등의 원형 성분에 의하여 모델링된다. 여기서 3D 그래픽 모델과는 달리 각 원형 성분들은 독립적으로 다른 성분들과의 연관성이 거의 없다.

본 논문에서는 강인성과 보안성을 만족하는 CAD 설계도면 또는 디지털 맵과 같은 벡터 데이터 모델 기반 해싱을 제안한다. 제안한 해싱에서는 입력된 벡터 데이터 모델 내에 주요 레이어를 폴리라인 밀도에 따라 선택한 다음, 각 레이어별로 곡선 에너지에 따라 폴리라인들을 그룹화한다. 그리고 폴리라인의 1차 및 2차 곡률 계수에 따라 그룹 계수를 얻은 다음, 이를 랜덤 계수 벡터로 투영함으로써 2D 특징 계수 벡터를 얻는다. 마지막으로 이 계수 벡터를 이진화 과정에 의하여 최종 이진 해쉬를 생성한다. 따라서 제안한 해싱에서는 1차 및 2차 곡률 계수 분포를 이용함으로써 폴리라인의 재변수화 및 유클리드 변환과 이외의 벡터 데이터 공격에 강인성을 가지며, 랜덤 계수 벡터로의 투영에 의하여 보안성을 가지게 된다. 실험 결과로부터 제안한 해싱이 다양한 공격에 대하여 강인성을 가지며, 또한 보안성과 키에 대한 유일성을 가짐을 확인하였다.

본 논문의 구성을 살펴보면, 2장에서는 벡터 데이터 모델 해싱 프레임워크 및 1차 및 2차 곡률에 대한 기본적인 이론에 대하여 언급하고, 3장에서는 제안한 벡터 데이터 해싱에 대하여 자세히 설명한다. 4장에서는 강인성 및 보안성에 대한 실험 결과에 대하여 살펴보며 마지막으로 5장에서는 본 논문의 결론을 맺는다.

Ⅱ. 관련 이론

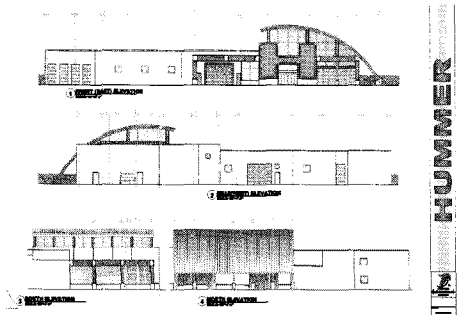
2.1 벡터 데이터 모델 해싱 구조

콘텐츠 해싱[12-19]은 콘텐츠 내에 특징 벡터와 랜덤 키와 결합하여 이를 이진 해쉬로 생성하는 것으로 강인성, 보안성 및 유일성 등의 조건을 만족하여야 한다. 이 때 특징 벡터 추출은 강인성 여부를 결정하는 주요한 단계로 콘텐츠 구조에 따라 다르게 제안되어져야 한다.

강인성은 원 모델의 화질이 유지되면서 수행되는 공격에 대하여 해쉬는 보존되어야 한다. 여기서 공격 형태는 영상, 3D 모델 및 벡터 데이터 모델에 따라 다르다. 예를 들어, 벡터 데이터 모델은 CAD/GIS 관련 편집 툴을 이용하여 레이어 공격, 객체 공격 등이 있으며, 해싱은 이들 공격에 강인하여야 한다. 보안성은 키를 알지 못할 때, 해쉬를 예측하기가 불가능하여야 한다는 것이다. 이는 특징 벡터 추출 과정 또는 이진 해쉬 생성 과정에서 랜덤변수 키 또는 치환 키를 이용하거나, 랜덤 양자화에 의하여 보안성을 향상시킬 수 있다. 유일성은 임의의 모델과 임의의 키에 의하여 생성된 해쉬는 유일하여야 한다는 것이다. 예를 들어, 동일한 모델에서 두 개의 키에 의하여 생성된 두 해쉬들은 서로 연관성이 없어야 한다. 또한 다른 모델에서 동일한 키에 의하여 생성된 두 해쉬들은 서로 연관성이 없어야 한다. 이는 특징 벡터 추출과 키에 의한 결합 과정에 의하여 유일성 유무가 결정된다.

벡터 데이터 모델 해싱과 기존의 영상 해싱[12-15] 및 3D 모델 해싱[16-19]과의 차이점은 강인한 특징 벡터 추출 과정이다. 그리고 보안성을 향상시키기 위한 키 생성과 특징 계수와 키와의 결합은 이전 영상 및 3D 해싱과는 유사하나, 특징 계수의 성질에 따라 다르게 적용되어야 한다. 일반적으로 벡터 데이터 모델은 여러 레이어들의 합성에 의하여 모델링되며, 각 레이어는 레이어의 특성을 표현하도록 점, 선, 면, 원 또는 텍스트 등의 기하학 성분에 의하여 구성된다. 여기서 각 레이어 및 레이어 내의 각 성분들은 각각 독립적으로 설계되어진다.

그림 1은 Hummer Elevation 설계도면과 1;1000 디지털 맵의 벡터 데이터 모델들을 보여주고 있다. 그림 1 (a)의 설계도면은 30개의 레이어로 구성되며, 이들 중 폴리라인에 해당되는 주요 설계도면 레이어는 1개이고, 그림 1 (b)의 디지털 맵은 72개 레이어로 구성되었으며, 이들 중 폴리라인에 해당되는 레이어는



(a)

(b)

그림 1. (a) Hummer Elevation 설계도면 및 (b) 1:1000 디지털 맵의 벡터 데이터 모델

58개이다. 이들 폴리라인을 포함하는 레이어들 중 가장 주요한 레이어들을 해쉬 추출에 사용된다.

이와 같은 벡터 데이터 모델에서 특징 벡터들은 레이어별로 기하학 성분들의 특징 분포에 의하여 추출될 수 있다. 여기서 레이어와 성분들은 각각 인덱스에 의하여 탐색되며, 이들 인덱스들은 쉽게 변경이 가능하다. 즉, 영상의 화소 배열과는 달리 레이어 및 성분들의 순서 및 인덱스는 고정적이 않는다. 따라서 벡터 데이터 모델 해싱에서는 레이어 및 성분 재배열과 다양한 벡터 공격에 강인한 특징 벡터 추출이 가장 주요하다. 제안한 해싱에서는 설계도면 및 디지털 맵 등에서 가장 주요하게 사용되는 폴리라인 성분들의 곡률 분포를 이용하여 특징 벡터를 추출하며, 이를 랜덤 계수 키 패턴으로 투영함으로써 최종 이진 해쉬를 생성한다.

2.2 라인 곡선 곡률

본 논문에서는 벡터 그래픽 모델의 특징 계수를 추출하기 위하여 곡선의 1차, 2차 곡률을 사용한다. 곡선 곡률은 라인의 재변수화 (re-parameterization) 및 유클리디안 (Euclidean) 변환에 불변인 특성을 가진다. 본 절에서는 라인 곡선의 일반화된 곡률 (Generalized curvature of line curve)[20,21]에 대하여 간략히 살펴보기로 한다.

n차 정칙한 (regular) R^n 상에 $n+1$번 미분 가능한 곡선 Υ이 주어졌을 때, Υ의 Frenet 프레임은 정규직교 벡터 (orthonormal vector) $e^{(1)}(t),\cdots e^{(n)}(t)$의 집합이며, 이 벡터들을 Frenet 벡터라 한다. 여기서 Frenet 프레임은 각 위치 $\Upsilon(t)$에서 곡률 또는 비틀림률 (torsion)과 같은 곡선의 국부적인 서술자 (descriptor)로 사용되는 정규직교벡터 $e^{(i)}(t)$들의 기준 프레임이다. 여기서 $e^{(i)}(t)$는 Gram-Schmidt orthogonalization 알고리즘 기반으로 $\Upsilon(t)$의 미분에 의하여 $e^{(1)}(t)=\frac{\Upsilon'(t)}{\|\Upsilon'(t)\|}$,

$e^{(j)}(t)=\frac{\overline{e^{(j)}}(t)}{\|\overline{e^{(j)}}(t)\|}$, $\quad \overline{e^{(j)}}(t)=\Upsilon^{(j)}(t)-\sum_{i=1}^{j-1}<\Upsilon^{(j)}(t),$ $e^{(j)}(t)>e^{(j)}(t)$와 같이 구하여진다. 이 때 일반화된 곡률은 실수 함수 $\chi^{(i)}(t)=\frac{<e'^{(i)}(t),e^{(i+1)}(t)>}{\|\Upsilon'(t)\|}$에 의하여 정의된다. 여기서 Frenet 프레임과 일반화된 곡률은 라인의 재변수화 상에서 불변이므로 곡선의 미분 기하학 성질을 나타낸다.

본 논문에서는 이상의 연속적인 곡선에 대한 Frenet 프레임과 일반화된 곡률을 이용하여 이산 구조의 폴리라인의 1차 및 2차 곡률 계수를 구하고자 한다. 예를 들어, $n+1$개의 꼭지점 $v_0,\cdots v_n$들로 구성된 폴리라인 p이 주어졌을 때, Frenet 벡터는

$$e^{(1)}[k]=\frac{\Upsilon'[k]}{\|\Upsilon'[k]\|},\ k\in[0,n]$$
$$e^{(j)}[k]=\frac{\overline{e}^{(j)}[k]}{\|\overline{e}^{(j)}[k]\|},\qquad (1)$$
$$\overline{e}^{(j)}[k]=\Upsilon^{(j)}[k]-\sum_{i=1}^{j-1}<\Upsilon^{(j)}[k],e^{(j)}[k]>e^{(j)}[k],$$

와 같이 정의되고, 곡률은

$$\chi^{(i)}[k]=\frac{<e'^{(i)}[k],e^{(i+1)}[k]>}{\|\Upsilon'[k]\|} \qquad (2)$$

와 같이 정의된다. 여기서 1차 Frenet 벡터 $e^{(1)}[k]$는 곡선 Υ의 각 점 v_k에서 정의된 단위 접선 벡터 (unit tangent vector)이고, 2차 Frenet 벡터 $e^{(2)}[k]$는 단위 곡률 벡터 (unit curvature vector) 또는 단위 법선 벡터 (unit normal vector)으로 한 직선에 대한 곡선의 편차를 나타낸다. 임의의 점 v_k에서 $e^{(1)}[k]$와 $e^{(2)}[k]$는 v_k에서 접촉 평면 (osculating plane)을 정의한다. 그리고 3차 Frenet 벡터 $e^{(3)}[k]$는 종법선 벡터 (binormal vector)로 v_k에서 $e^{(1)}[k]$와 $e^{(2)}[k]$와 항상

직교한다.

제안한 방법에서는 임의의 폴리라인 상의 각 꼭지점에서 1차, 2차 Frenet 벡터에 의하여 구하는 1차 곡률 $\chi^{\{1\}}[k]$과 2차, 3차 Frenet 벡터에 의하여 구하는 2차 곡률 $\chi^{\{2\}}[k]$를 구한 후, 이들 곡률 계수에 의하여 폴리라인 특징 계수를 구한다. 이에 대한 설명은 다음 장에 자세히 언급하기로 한다.

Ⅲ. 제안한 벡터 데이터 해싱

벡터 데이터는 CAD 설계도면, GIS 도면 등 다양한 산업 분야에서 활용되는 주요 데이터 또는 모델이다. 본 논문에서는 이들 벡터 데이터에 의하여 생성된 모델의 인증 및 복사방지를 위한 벡터 데이터 해싱을 제안한다. 이 때 해싱은 강인성과 보안성을 만족하여야 하고, 해쉬 추출 시 원 데이터가 필요 없도록 설계되어진다.

일반적인 벡터 데이터 모델은 주요 레이어들로 구성되며, 각 레이어 내에는 점, 선(폴리라인), 면(폴리곤), 원, 타원 및 텍스트 등의 원형 기하학 성분들을 포함한다. 이들 중 성분들 중 폴리라인과 폴리곤이 대부분을 차지한다. 제안한 벡터 데이터 해싱에서는 임의의 레이어를 선택한 후, 이들 레이어 내에 폴리라인과 폴리곤 성분들의 특징 계수를 추출한 후, 이를 해쉬 데이터로 생성한다. 제안한 해싱 구조는 기본적인 콘텐츠 해싱의 구조를 가지며, 그림 2에서와 같이 레이어 선택, 그룹화, 특징 계수 추출 및 해쉬 생성의 단계로 구성된다.

본 장에서 제시된 기호 정의는 다음과 같다. 임의의 벡터 데이터 모델 $\mathbf{M}=\{\mathbf{L}_i|i\in[1,N]\}$은 N개의 레이어 $\mathbf{L}$들로 구성되며, 각 레이어 $\mathbf{L}_i=\{\mathbf{P}_i,\mathbf{C}_i,\mathbf{Ac}_i,\mathbf{T}_i\}$은 폴리라인 성분 $\mathbf{P}_i=\{\mathbf{p}_{ij}|j\in[1,N(P_i)]\}$, 원 성분 $\mathbf{C}_i=\{\mathbf{c}_{ij}|j\in[1,N(C_i)]\}$, 타원 성분 $\mathbf{Ac}_i=\{\mathbf{ac}_{ij}|j\in[1,N(Ac_i)]\}$ 및 텍스트 성분 $\mathbf{T}_i=\{\mathbf{t}_{ij}|j\in[1,N(T_i)]\}$들로 구성된다. 각 성분들에 대한 자세한 설명은 AutoCAD[22]에 제시되어 있으므로, 본 논문에서는 이에 대한 설명은 제외하기로 한다. 여기서 $N(P_i)$, $N(C_i)$, $N(Ac_i)$, 및 $N(T_i)$는 레이어 $\mathbf{L}_i$ 내의 폴리라인, 원, 타원 및 텍스트 성분의 개수들을 나타낸다.

임의의 폴리라인 $\mathbf{p}_{ij}=\{\boldsymbol{v}_{i0},\boldsymbol{v}_{i1},\cdots,\boldsymbol{v}_{N_{ij}}\}$는 N_{ij}개의 꼭지점들로 구성된다. 이 때 시작점과 끝점이 $\boldsymbol{v}_{i0}=\boldsymbol{v}_{N_{ij}}$일 경우, 이를 닫혀진 폴리라인 또는 폴리곤이라 하며 본 논문에서는 폴리곤을 폴리라인과 같은 성분으로 간주한다. 최종 생성된 해쉬 $\mathbf{H}=\{h_{n_1,n_2}|n_1,n_2\in[1,N_H]\}$는 $N_H\times N_H$ 비트로 구성된다.

3.1 레이어 선택

벡터 데이터 모델은 여러 레이어를 가지며 각 레이어들은 모델의 특성에 따라 고유한 기하 특징을 가진다. 예를 들어, GIS 도면에서 도로 레이어 같은 경우에서는 직선 위주의 폴리라인과 사각형 형태의 폴리곤들이 많이 분포되어 있으며, 시계와 같은 설계 도면에서는 타원이나 곡선 형태의 폴리라인들이 많이 분포되어 있다. 제안한 방법에서는 도면에서 가장 주요한 레이어들만 선택하여 이 레이어 내에 해쉬를 추출하고자 한다. 기존 3D CAD 설계도면[7]과 3D GIS 워터마킹[10]에서는 주요 성분들의 밀도에 따라 레이어를 선택하였다. 제안한 방법에서는 폴리라인 성분은 주요 해쉬 성분으로 하므로, 레이어 $\mathbf{L}_i$ 내의 타성분에 대한 폴리라인 분포 비율 r_i과 폴리라인 밀도 ρ_i

$$r_i=\frac{N(P_i)}{N(P_i)+N(C_i)+N(Ac_i)+N(T_i)}=\frac{N(P_i)}{N_i},\quad \rho_i=\frac{N(P_i)}{\sum_{k=1}^{N_L}N_k} \tag{3}$$

와의 곱 $r_i\rho_i$을

$$r_i\rho_i=\frac{N(P_i)^2}{N_i\sum_{k=1}^{N_L}N_k} \tag{4}$$

이용하여 주요 레이어를 선택한다. 즉, 제안한 방법에서는 모든 레이어들을 $r_i\rho_i$에 따라 내림차순으로 정려한 다음, $r_i\rho_i$가 높은 레이어 순으로 전체 폴리라인

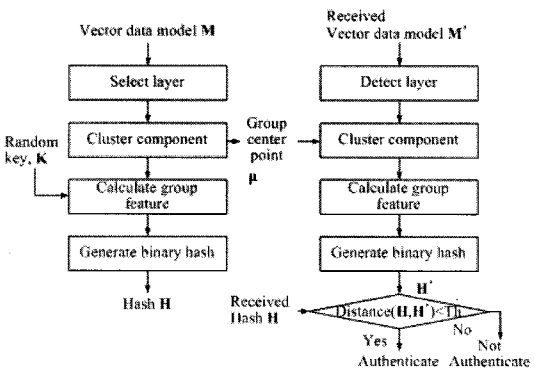


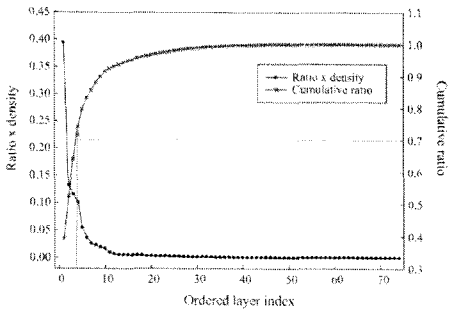
그림 2. 벡터 데이터 모델 기반 (a) 해쉬 생성 과정 및 (b) 해쉬 인증 과정

개수 $\sum_{k=1}^{N} N(P_k)$에 대한 누적 개수 $\sum_{k=1}^{N_L} N(P_k)$ 비율 $\gamma_{N_L} = \sum_{k=1}^{N_L} N(P_k) / \sum_{k=1}^{N} N(P_k)$이 Th보다 클 때까지 레이어들을 선택한다. 즉, 해쉬 추출로 선택된 레이어 $\mathbf{L}^*$는

$$\begin{aligned} \mathbf{L}^* &= \{\mathbf{L}_1, \cdots, \mathbf{L}_i, \cdots \mathbf{L}_{N_L}\}, \\ r_i\rho_i &> r_{i+1}\rho_{i+1}, \forall i \in [1, N_L], \\ \gamma_i &= \sum_{k=1}^{i} N(P_k) / \sum_{k=1}^{N} N(P_k) > Th \end{aligned} \tag{5}$$

와 같이, 벡터 데이터 모델 내에 전체 폴리라인 개수의 $Th \times 100\%$ 이상의 폴리라인을 가진다. 여기서 Th가 작으면 선택된 레이어 개수가 증가하므로 해쉬의 길이가 증가하나, 중요하지 않은 레이어가 선택될 수 있으므로 강인성이 약해질 수 있다. 반대로 Th가 크면 중요한 일부 레이어만 선택되어지므로 강인성이 증가되나, 해쉬의 길이가 작아지므로 보안성이 약해질 수 있다. 따라서 본 논문에서는 선택된 레이어 개수 N_L가 최소 2개 이상이 되도록 Th를 실험적으로 0.7로 선택하였다. 그림 3은 1:5,000 축적 비율의 디지털 맵 상에서 폴리라인 분포 비율과 폴리라인 밀도의 곱 $r_i\rho_i$에 의하여 정렬된 레이어별 폴리라인 개수 누적 비율 γ_i을 보여준다. 이 그림을 살펴보면, 일부 주요

(a)



(b)

그림 3. (a) 1:5,000 디지털맵과 (b) $r\rho$에 의하여 내림차순으로 정렬된 레이어별 폴리라인 개수 누적 비율

레이어에 폴리라인 밀도가 높으며, 약 70% 이상이 4개의 레이어에 집중되어 있음을 볼 수 있다. 또한 대부분의 벡터 데이터 모델에서도 주요 레이어에 폴리라인인 집중되어 있음을 확인하였다. 그러나 그림 1 (a)의 Hummer Elevation 모델과 같은 일부 벡터 데이터 모델에서는 폴리라인을 포함하는 레이어가 단일한 경우이다. 이와 같은 모델에서는 위의 레이어 선택 과정 없이 폴리라인 단일 레이어가 선택된다.

3.2 레이어별 그룹화

벡터 데이터 내의 레이어들은 레이어 인덱스 정보에 의하여 탐색되고, 레이어 내의 모든 폴리라인과 꼭지점들 또한 인덱스 정보에 의하여 탐색된다. 그러나 이들 인덱스 정보들은 변경이 용이하므로 레이어, 폴리라인, 꼭지점들의 순서가 쉽게 변경된다. 따라서 하나의 폴리라인에 하나의 해쉬 비트를 생성할 때 많은 비트수를 생성할 수 있지만, 원 데이터 변형없이 인덱스 정보 변경으로도 해쉬는 쉽게 변경될 수 있다. 제안한 방법에서는 폴리라인의 특정 성질에 의하여 그룹화한 다음, 그룹별로 해쉬 비트를 생성한다. 이는 해쉬 비트수는 작아질 수 있지만, 공격에 대한 강인성이 매우 우수한 장점을 가진다.

그룹화 과정에서는 먼저 선택된 레이어 $\mathbf{L}_i = \{\mathbf{p}_j | j \in [1, N_{L_i}]\}$ 내에 모든 폴리라인 $\mathbf{p}_j$들의 곡선 활동량 (Curve action)인 에너지 $E[\mathbf{p}_j]$을 구한다. 여기서 $\mathbf{p}_j = \{v_0, v_1, \cdots, v_{N_{P_j}}\}$의 곡선 에너지 $E[\mathbf{p}_j]$는

$$E[\mathbf{p}_j] = \frac{1}{2} \sum_{k=1}^{N} \| v_k - v_{k-1} \| \Delta k \tag{6}$$

이며, Δk는 이산 구간 거리로 본 논문에서는 Δk=1인 단위 구간으로 간주한다. $\mathbf{p}_j$의 정규화된 곡선 에너지 $\overline{E}[\mathbf{p}_j]$을

$$\overline{E}[\mathbf{p}_j] = \frac{E[\mathbf{p}_j]}{\sum_{j=1}^{N_L} E[\mathbf{p}_j]} \tag{7}$$

구한 다음, $\overline{E}[\mathbf{p}_j]$에 대한 k-means 알고리즘 기반으로 모든 폴리라인들을 해쉬 비트 N_H 길이와 동일한 개수만큼 그룹화 $\mathbf{G}_i = \{G_{i,n} | n \in [1, N_H]\}$한다.

여기서 k-means 알고리즘은 초기값에 의존하므로, 제안한 방법에서는 $\overline{E}[\mathbf{p}_j]$의 최대값과 최소값의 간격을 1D 해쉬 비트 길이인 N_H개의 등간격으로 나눈 후, 이들 간

격의 중간점을 초기 중심 벡터 $\mu_i^{(0)} = \{\mu_{i,n}^{(0)} | n \in [1, N_H]\}$으로 놓는다. 그리고 모든 폴리라인 p_j들을 $\bar{E}[\mathrm{p}_j]$이 중심 벡터 $\mu_i^{(t)}$ $(t \geq 0)$와 간격이 최소가 되는 그룹 $G_{i,n}$으로

$$G_{i,n} = \left\{ \mathrm{p}_j : |\bar{E}[\mathrm{p}_j] - \mu_{i,n}^{(t)}| \leq |\bar{E}[\mathrm{p}_j] - \mu_{i,n^*}^{(t)}|, \atop \text{for all } n^* = 1, \cdots n, n^* \neq n \right\} \quad (8)$$

와 같이 할당한다. 그런 다음, 각 그룹 내에 $\bar{E}[\mathrm{p}_j]$의 평균을 다시 계산한 후 이를 중심 벡터 $\mu_i^{(t+1)}$로 놓는다. 이전 중심 벡터와의 차이가 $|\mu_i^{(t)} - \mu_i^{(t+1)}| < 10^{-5}$ $(\forall i \in [1, N_L])$와 같이 거의 없을 경우, 그룹화 과정을 종료하고 그렇지 않을 경우 위의 단계를 반복 수행한다.

이상의 반복 과정에 의하여 선택된 모든 레이어 L_i $(i \in [1, N_L])$ 내의 폴리라인 그룹을 생성한다.

$$G_{i,n} = \left\{ \mathrm{p}_j : \frac{\mu_{i,n-1} + \mu_{i,n}}{2} < \bar{E}[\mathrm{p}_j] < \frac{\mu_{i,n} + \mu_{i,n+1}}{2}, \atop j \in [1, N_{G_{i,n}}] \right\} \quad (9)$$

여기서 $\mu_{i,n} = \mu_i^{(t)}$으로 최종 반복에 의하여 생성된 그룹 중심값이고, $N_{G_{i,n}}$는 그룹 $G_{i,n}$내의 폴리라인의 개수이다.

3.3 레이어별 특징 벡터

3.3.1 폴리라인 곡률 계수

각 레이어의 그룹 내의 모든 폴리라인 $\mathrm{p}_j = \{v_0, v_1, \cdots, v_{N_j}\}$들의 평균 1차 곡률 $\bar{\chi}^{\{1\}}(\mathrm{p}_j)$과 평균 2차 곡률 $\bar{\chi}^{\{2\}}(\mathrm{p}_j)$를 수식 2로부터

$$\bar{\chi}^{\{1\}}(\mathrm{p}_j) = \sum_{k=3}^{N_j} \chi^{\{1\}}[k] / (N_j - 3), \quad (10)$$
$$\chi^{\{1\}}[k] = \frac{< \mathrm{e}'^{\{1\}}[k], \mathrm{e}^{\{2\}}[k] >}{\|\Upsilon'[k]\|},$$

$$\bar{\chi}^{\{2\}}(\mathrm{p}_j) = \sum_{k=3}^{N_j} \chi^{\{2\}}[k] / (N_j - 3), \quad (11)$$
$$\chi^{\{2\}}[k] = \frac{< \mathrm{e}'^{\{2\}}[k], \mathrm{e}^{\{3\}}[k] >}{\|\Upsilon'[k]\|}$$

와 같이 각각 구한다. 여기서 2차 곡률 계산에 필요한 3차 Frenet 벡터 $\mathrm{e}^{\{3\}}[k]$는 3차 미분이 가능하여야 하므로, 그림 4에서와 같이 폴리라인의 v_0, v_1, v_2들은 초기점들로 사용되고, 꼭지점 v_3부터 곡률 계산이 가능하다. 꼭지점 $v_k (k \in [3, N_j])$에서의 2차 및 3차 Frenet

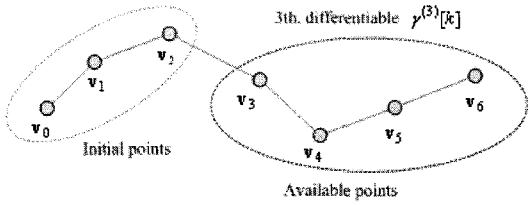


그림 4. 임의의 폴리라인 $\mathrm{p}_j = \{v_0, v_1, \cdots, v_6\}$ 상의 초기점과 3차 미분이 가능한 점

벡터인 $\mathrm{e}^{\{2\}}[k]$ 및 $\mathrm{e}^{\{3\}}[k]$은 수식 (1)로부터

$$\mathrm{e}^{\{2\}}[k] = \frac{\bar{\mathrm{e}}^{\{2\}}[k]}{\|\bar{\mathrm{e}}^{\{2\}}[k]\|}, \quad (12)$$
$$\bar{\mathrm{e}}^{\{2\}}[k] = \Upsilon^{\{2\}}[k] - < \Upsilon^{\{1\}}[k], \mathrm{e}^{\{1\}}[k] > \mathrm{e}^{\{1\}}[k]$$

$$\mathrm{e}^{\{3\}}[k] = \frac{\bar{\mathrm{e}}^{\{3\}}[k]}{\|\bar{\mathrm{e}}^{\{3\}}[k]\|}, \quad (13)$$
$$\bar{\mathrm{e}}^{\{3\}}[k] = \Upsilon^{\{3\}}[k] - \sum_{i=1}^{2} < \Upsilon^{\{i\}}[k], \mathrm{e}^{\{i\}}[k] > \mathrm{e}^{\{i\}}[k]$$

와 같이 구하며, 1차 및 2차 Frenet 벡터의 1차 미분 $\mathrm{e}'^{\{1\}}[k]$ 및 $\mathrm{e}'^{\{2\}}[k]$은 Frenet-Serret 공식[21]에 의하여

$$\begin{bmatrix} \mathrm{e}'^{\{1\}}[k] \\ \mathrm{e}'^{\{2\}}[k] \end{bmatrix} = \begin{bmatrix} 0 & -\Upsilon^{\{1\}}[k] \\ \Upsilon^{\{1\}}[k] & 0 \end{bmatrix} \begin{bmatrix} \mathrm{e}^{\{1\}}[k] \\ \mathrm{e}^{\{2\}}[k] \end{bmatrix} \quad (14)$$

와 같이 구한다. 그리고 꼭지점 v_k에서의 n차 미분 $\Upsilon^{\{n\}}[k]$은

$$\Upsilon^{\{n\}}[k] = \sum_{i=0}^{n} (-1)^i \binom{n}{i} v_{k-i} \quad (15)$$

와 같다. 이상의 방법에 의하여 모든 그룹 내에 폴리라인의 1차 및 2차 곡률 계수 쌍 $(\bar{\chi}^{\{1\}}(\mathrm{p}_j), \bar{\chi}^{\{2\}}(\mathrm{p}_j))$을 구한다.

3.3.2 그룹 특징 계수

제안한 방법에서는 그림 5 (a)에서와 같이 레이어별 폴리라인 그룹 상에서 1차 및 2차 그룹 곡률 계수 쌍 $(x_{i,n}^{\{1\}}, x_{i,n}^{\{2\}})$를

$$x_{i,n}^{\{1\}} = \sum_{j=1}^{N_{i,n}} \bar{\chi}^{\{1\}}(\mathrm{p}_j) / N_{i,n}, \quad (16)$$
$$x_{i,n}^{\{2\}} = \sum_{j=1}^{N_{i,n}} \bar{\chi}^{\{2\}}(\mathrm{p}_j) / N_{i,n}$$

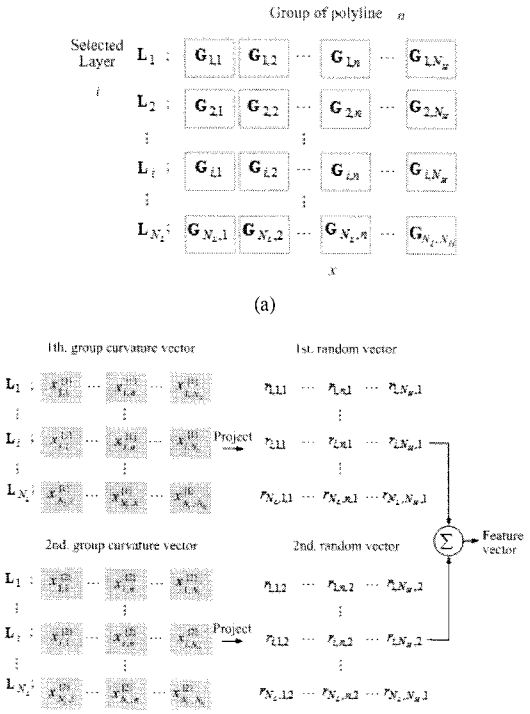


그림 5. 레이어별 (a) 폴리라인 그룹 및 (b) 곡률 벡터와 랜덤 벡터에 의한 특징 벡터 추출

와 같이 위에서 구한 그룹 내의 모든 폴리라인의 1차 및 2차 곡률 계수의 평균치로 사용한다. 즉, 전체 레이별 그룹 계수는 그림 5 (b)에서와 같이 $N_L \times N_H$ 크기의 1차 그룹 곡률 벡터 $\mathbf{X}^{\{1\}}$와 2차 그룹 곡률 벡터 $\mathbf{X}^{\{2\}}$로 구성된다.

$$\mathbf{X}^{\{1\}} = \left\{x_{i,n}^{\{1\}} | i \in [1, N_L], n \in [1, N_H]\right\}, \quad \mathbf{X}^{\{2\}} = \left\{x_{i,n}^{\{2\}} | i \in [1, N_L], n \in [1, N_H]\right\} \tag{17}$$

여기서 제안한 방법에서는 평균 및 분산이 (m_R, σ_R^2)인 $N_L \times N_H$ 크기의 가우시안 랜덤 키 벡터를 2개 $\mathbf{R}_1$,$\mathbf{R}_2$ 생성한 다음, $[\mathbf{X}^{\{1\}}]^T$와 $\mathbf{R}_1$ 및 $[\mathbf{X}^{\{2\}}]^T$와 $\mathbf{R}_2$의 벡터곱에 의하여 $N_H \times N_H$크기의 그룹별 특징 계수 벡터 $\mathbf{F}$를

$$\mathbf{F} = [\mathbf{X}^{\{1\}}]^T \mathbf{R}_1 + \alpha [\mathbf{X}^{\{2\}}]^T \mathbf{R}_2 \tag{18}$$

와 같이 구한다. 이를 원소별로 표현하면

$$f_{n_1,n_2} = \sum_{i=1}^{N_L} x_{n_1,i}^{\{1\}} r_{1,i,n_2} + \alpha \sum_{i=1}^{N_L} x_{n_1,i}^{\{2\}} r_{2,i,n_2}, \quad \forall n_1, n_2 \in [1, N_H] \tag{19}$$

와 같다. 여기서 $[\mathbf{X}]^T$는 X의 전치(transpose) 행렬이고, α는 1차 곡률 계수 $x_{i,n}^{\{1\}}$와 2차 곡률 계수 $x_{i,n}^{\{2\}}$의 크기 범위를 동일하게 하는 것으로

$$\alpha = \max\left\{x_{i,n}^{\{1\}}\right\} / \max\left\{x_{i,n}^{\{2\}}\right\} \tag{20}$$

와 같다.

3.4 해쉬 생성

최종 이진 해쉬 H는 위에서 구한 그룹 특징 계수 벡터 F의 이진화 과정에 의하여

$$h_{n_1,n_2} = \begin{cases} 1, \text{if} f_{n_1,n_2} > Th \\ 0, \text{if} f_{n_1,n_2} \leq Th \end{cases} \tag{21}$$

와 같이 간단히 구하여진다. 여기서 문턱치 Th는 h_{n_1,n_2}이 베르누이(Bernoulli) 분포 $P[h_{n_1,n_2}=1] \simeq P[h_{n_1,n_2}=0] \simeq 1/2$를 가지도록 다음과 같은 과정에 의하여 결정된다.

1) 초기 문턱치 Th_0를 f_{n_1,n_2}의 DR (Dynamic range) 중 임의의 값을 선택한다.
2) 그룹 특징 계수 f_{n_1,n_2}들을 문턱치 Th에
$$Th = \begin{cases} Th_t, \text{if} t > 0 \\ Th_0, \text{if} t = 0 \end{cases}$$
의하여 두 집합 $\mathbf{G}_1$, $\mathbf{G}_2$로 분류한다.
$\mathbf{G}_1 = \left\{f_{n_1,n_2} | f_{n_1,n_2} > Th\right\}$, $\mathbf{G}_2 = \left\{f_{n_1,n_2} | f_{n_1,n_2} \leq Th\right\}$
3) 두 집합 $\mathbf{G}_1$, $\mathbf{G}_2$ 내의 평균 그룹 계수 m_1, m_2을 구한 다음 이의 평균을 문턱치 Th_t로 놓는다.
4) 수렴조건인 $Th_t \approx Th_{t-1}$을 만족할 때까지 2)-4) 단계를 반복 수행한다.

3.5 해쉬 인증

전송된 벡터 데이터 모델 M′에 대한 인증은 랜덤 계수 키 K와 레이어별 그룹 중점 μ에 의하여 그림 2 (b)에서와 같이 해쉬 생성 과정과 동일하게 해쉬 H′를 생성한다. 그리고 제안한 방법에서는 생성된 해쉬 H′와 원 해쉬 H와의 정규화된 해밍 거리차 $d(\mathbf{H},\mathbf{H}')$를 이용하여

$$d(\mathbf{H},\mathbf{H}') = \frac{1}{N_H \times N_H} \sum_{n_1=1}^{N_H} \sum_{n_2=1}^{N_H} |h_{n_1,n_2} - h'_{n_1,n_2}| \tag{22}$$

표 1. 대표적인 벡터 데이터 모델의 레이어 및 폴리라인 개수

테스트 모델		전체 개수		선택 개수		폴리라인 비율 (N/T×100)[%]
		레이어	폴리라인 (T)	레이어	폴리라인(N)	
설계 도면	Blocks and Tables	16	76	2	55	72.4
	Db	29	1,737	2	1,351	77.8
	Hummer Elevation	30	1,492	1	1,492	100.0
	Stadium North Elevation	52	5,442	2	3,916	71.9
	Tablet	5	1,458	2	1201	82.4
	Taisei Detail Plan	37	2,288	1	2,288	100.0
디지털맵	1:1000 map	72	633	2	481	75.9
	1:5000 map	74	2,489	3	1851	74.4
	1:25000 map	116	18,911	6	16,361	86.5
	1:250000 map	29	2,.001	4	1,530	76.5

$$M' \rightarrow \begin{cases} Auth., & \text{if} d(H,H') < e \\ NoAuth. & \text{if} d(H,H') \geq e \end{cases} \quad (23)$$

와 같이 벡터 데이터 모델 M′의 인증 여부를 결정한다. 여기서 정규화된 해밍 거리차 $d(H,H')$는 두 모델이 인지적 화질이 유사할 경우 0에 가깝고, 두 모델이 다른 모델일 경우 0.5에 가깝다.

본 논문에서는 해쉬 인증 문턱치인 e를 결정하기 위하여 ROC(Receiver operation characteristic) 기반의 추정 (hypothesis) 실험을 수행하였다. 추정 실험에서는 50개의 벡터 데이터 모델 상에서 100개의 키로 생성된 해쉬들과 공격받은 모델에서 추출된 해쉬들과의 정규화된 해밍 거리차를 구한 후, [0, 0.5] 범위의 문턱치 e를 0.05 단위로 가변하면서 TP(True positive rate)와 FP(False positive rate)를 각각 측정하였다. 여기서 TP 확률은 $p_{TP} = \Pr[d(H,H') < e]$으로 공격받은 모델 M′의 해쉬 H′가 원 모델 M로 인증될 확률이다.

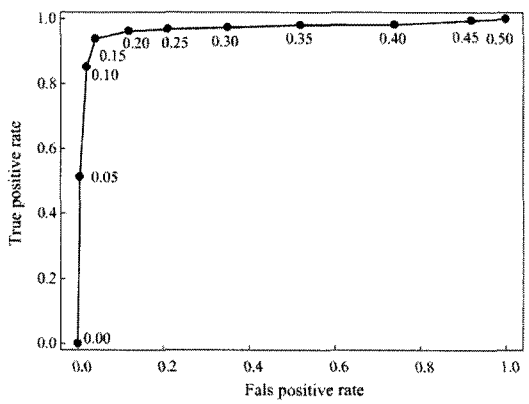


그림 6. 해쉬 인증 문턱치에 대한 ROC 곡선

그리고 FP 확률은 $p_{FP} = \Pr[d(H_k,H') < e]$으로 공격받은 모델 M′의 해쉬 H′가 다른 모델 M_k로 인증될 확률이다. 실험 결과인 ROC는 그림 6에서와 같으며, 이 곡선으로부터 문턱치 e가 0.15일 때 TP는 약 0.95이고 FP는 약 0.04으로 낮은 FP에서 가장 높은 TP을 얻을 수 있음을 확인하였다. 따라서 본 논문에서는 해쉬 인증 문턱치를 0.15로 결정하였다.

Ⅳ. 실험 결과

본 실험에서는 제안한 벡터 데이터 모델 해싱의 강인성 및 유일성과 보안성을 평가하기 위하여 autodesk 사의 AutoCAD 소프트웨어[22]에서 제공한 테스트 파일과 국토정보지리원[23]에서 제공한 디지털 맵 등 50여개의 다양한 벡터 데이터 모델들을 사용하였다. 이들 모델들은 다양한 레이어 개수와 폴리라인 개수를 가지므로, 모델마다 선택된 레이어의 개수가 다르다. 레이어별 그룹화 과정에서 폴리라인의 그룹 개수는 1D 해쉬의 비트수 N_H와 동일하다. 본 실험에서는 레이어별 그룹 개수 N_H=15으로 놓은 다음, $N_H \times N_H$ =15×15=225 비트의 해쉬를 생성하였다. 표 1에서는 대표적인 벡터 데이터 모델의 레이어 개수와 폴리라인 개수를 보여준다. 이 표를 살펴보면, 일반적으로 설계도면 모델인 경우 폴리라인이 많은 단일 또는 2개의 레이어에 집중적으로 분포되어 있으며 디지털 맵인 경우에도 일부 레이어에서 폴리라인의 밀도가 높은 것을 알 수 있다. 그리고 선택된 레이어 내의 폴리라인이 전체 레이어의 폴리라인의 70% 이상 차지함을 볼 수 있다.

4.1 강인성 평가

강인성 평가를 위한 실험에서는 AutoCAD SW 상의 각종 편집 기능을 공격으로 수행하였다. 이 때 강인성 평가 도구로 오류 검출 확률 p_{fd}

$$p_{fd}(\alpha,\mathbf{H},\mathbf{H}') = 1 - \Pr[d(\mathbf{H},\mathbf{H}') < e|\alpha] \qquad (24)$$

을 사용하였다. 여기서 p_{fd}는 임의의 공격 세기 α에서 원 벡터 데이터 모델에서 생성된 해쉬 $\mathbf{H}$와 공격 변형된 벡터 데이터 모델에서 생성된 $\mathbf{H}'$와의 정규화된 해밍 거리차가 문턱치보다 클 확률이다. 본 실험에서는 각종 공격의 세기 s에서 $d(\mathbf{H},\mathbf{H}') > e, (e = 0.15)$일 때까지 랜덤 키 생성 및 해쉬 추출 과정을 반복 수행함으로써 p_{fd}를 계산하였다. 강인성 평가에 대한 실험 결과는 그림 7에서와 같다.

4.1.1 RST 및 재배열

제안한 해싱의 특징 벡터인 곡선 곡률은 회전 및 이동과 같은 변환에 전혀 변함이 없다. 그러나 임의의 스케일링 인자에 따라 곡선 곡률이 선형적으로 변경된다. 따라서 제안한 해싱에서는 해쉬 생성 과정에서 전체 폴리라인의 곡선 에너지 E을 저장한 다음, 스케일링된 벡터 데이터 모델 상의 전체 곡선 에너지 E′와 비교하여 스케일링 인자 s를 예측한다. 그리고 모든 폴리라인들의 곡선 에너지가 $1/s$이 되도록 정규화를 수행한다. 여기서 N은 전체 레이어 개수이고, N_{L_i}은 레이어 내의 폴리라인의 개수를 나타낸다. 이상의 정규화 과정을 통하여 스케일링된 공격에 대하여 해쉬는 오류없이 검출됨을 확인하였다. 즉, 오류 검출 확률 p_{fd}은 0이다.

레이어와 기하학 성분들은 고유의 인덱스를 가지며, 이 인덱스들은 재배열에 의하여 변경이 가능하다. 이 때 벡터 데이터 모델의 화질은 전혀 변함이 없다. 그러나 제안한 해싱에서는 폴리라인 밀도에 따른 레이어 선택과 곡선 에너지에 따른 그룹화 과정을 수행하므로 레이어와 기하학 성분들이 재배열되더라도 해쉬 추출에는 전혀 영향을 받지 않는다. 따라서 재배열 공격에 대하여 오류 검출 확률 p_{fd}은 0이다.

4.1.2 객체 삭제 및 복사

벡터 데이터 모델의 기하학 성분들은 서로 연관이 없이 독립적으로 삭제 및 복사가 가능하다. 본 실험에서는 임의의 객체 성분들을 전체 성분의 약 10-50%을 삭제 및 복사를 수행하였다. 그림 8은 Hummer Elevation 설계도에서 1층 앞면도 도면과 30% 객체가 삭제된 도면과 복사된 도면을 보여주고 있다. 먼저 삭제 도면을 살펴보면, 주요 도면만 남기고 나머지 부분은 삭제됨을 볼 수 있다. 그리고 복사 도면을 살펴보면, 주요 도면인 오른쪽 부분이 반복해서 복사됨을 볼 수 있다.

객체 삭제 및 복사에 대한 오류 검출 확률은 그림 7 (a) 및 (b)에서와 같다. 먼저 객체 삭제 결과인 그림

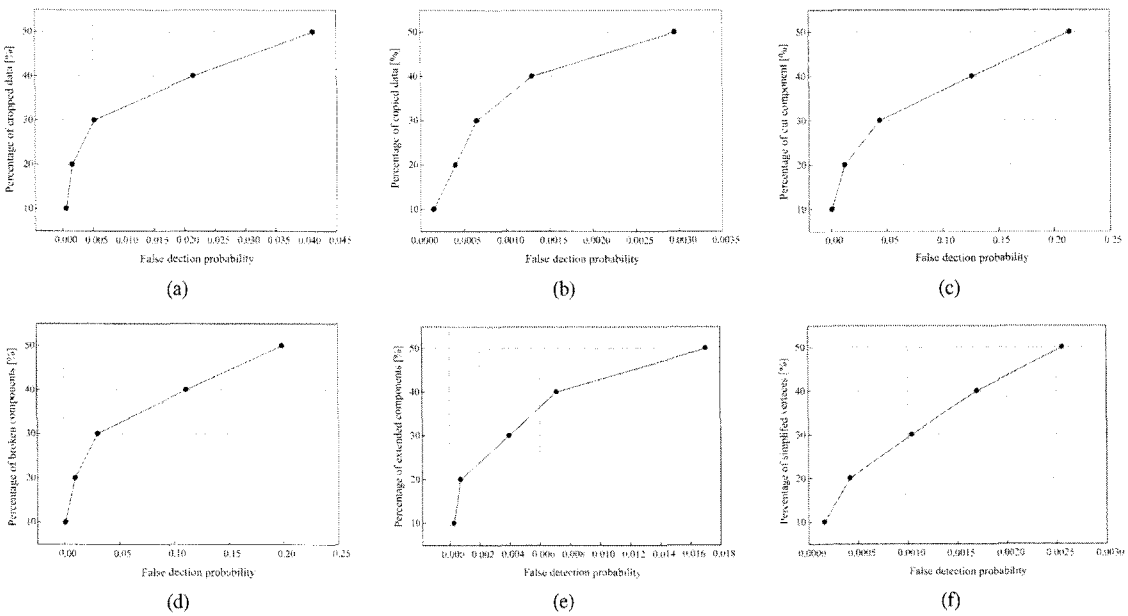
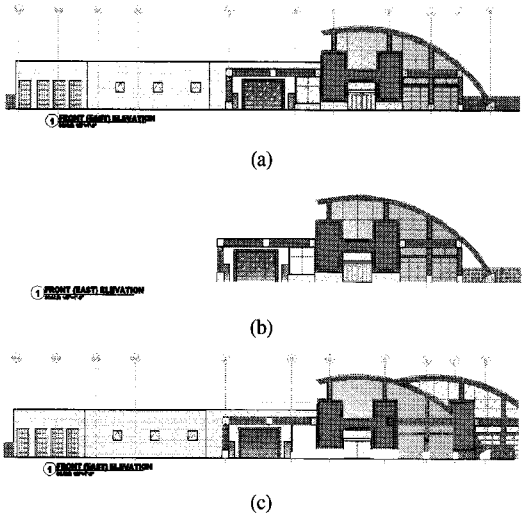


그림 7. 객체에 대한 공격인 (a) 삭제, (b) 복사, (c) 자르기, (d) 끊기, (e) 연장 및 (f) 꼭지점 삭제에 대한 오류 검출 확률



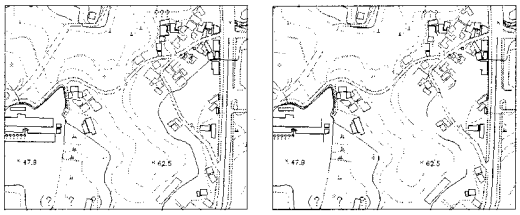
(a) (b) (c)

그림 8. (a) Hummer Elevation의 1층 앞면도와 (b) 30%정도 데이터가 삭제된 1층 앞면도 및 (c) 30%정도 주요 도면이 복사된 1층 앞면도

7 (a)를 살펴보면, 삭제량이 10-30%일 때 p_{fd}는 6×10-4-5.2×10-3으로 낮으나, 삭제량이 40-50%일 때 p_{fd}는 2.1×10-2-4.1×10-2으로 높게 나타났다. 데이터 복사 결과인 그림 7 (b)를 살펴보면, 복사량이 50%일 때 까지 p_{fd}는 1.5×10-4-2.9×10-3으로 데이터 삭제보다 매우 낮게 나타났다. 즉, 해쉬 특징 벡터는 객체 삭제에 의하여 영향을 미치나, 객체 복사에 대하여 거의 영향을 미치지 않는다.

4.1.3 객체 자르기 (Trim)

객체 자르기는 폴리곤의 모서리를 절단하거나 연속적인 폴리라인 또는 폴리곤을 임의의 점에서 자르는 것으로 자를 모서리와 자를 객체를 사용자 좌표계의 XY 평면에 투영한다. 본 실험에서는 전체 객체들 중 10%-50% 객체를 임의의 선택한 후 이들 객체 라인들에 대하여 자르기를 수행하였다. 그림 9 (a)인 1:5000 디지털 맵 상에서 50% 정도 객체들에 대하여 잘라진 디지털 맵은 그림 9 (b)에서와 같다. 이 그림으로부터 여러 개의 폴리라인들이 임의의 점에 자려진 것을 볼 수 있다. 객체 자르기에 대한 오류 검출 확률은 그림 7 (c)에서와 같다. 이 결과를 살펴보면, 잘려진 객체량이 10-30%일 때 p_{fd}은 6.0×10-4-4.3×10-2 이나 40-50%일 때 1.2×10-1-2.1×10-1으로 매우 높게 나타났다. 이는 많은 폴리라인들이 작은 라인들로 잘려지며, 잘려진 라인들은 다른 그룹으로 할당되며 또한 곡선 곡률이 변경되기 때문이다. 그러나 제안한 해싱에서는 그 이외의 폴리라인들의 그룹 계수 분포에 의하여 약 30%정도 객체 자르기에 대하여 4%의 낮은 에러 검출 확률을 가짐을 확인하였다.

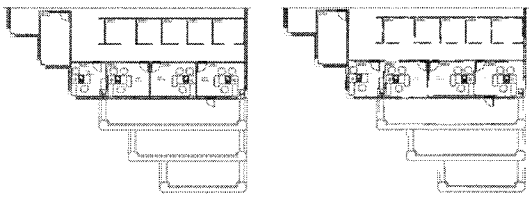


(a) (b)

그림 9. (a) 1:5000 디지털 맵 상의 한 지역과 (b) 50% 객체에 자르기가 수행된 디지털 맵

4.1.4 객체 끊기 (Break)

객체 끊기는 폴리라인을 두 개의 폴리라인으로 분할하거나 한쪽 끝을 제거하는 것으로, 사용자의 첫 번째 지점과 두 번째 지점의 사이의 객체 부분을 지운다. 본 실험에서는 객체 자르기 실험에서와 같이 전체 객체의 10-50% 정도를 임의로 선택하여 객체 끊기를 수행하였다. Db 설계 도면의 아래 부분과 30% 정도 객체가 끊겨진 도면은 그림 10 (a)와 (b)에서와 같다. 이 그림을 살펴보면, 폴리라인들이 끊겨진 것을 볼 수 있다. 객체 끊기에 대한 오류 검출 확률은 그림 7 (d)에서와 같이 객체량이 10-30%일 때 p_{fd}은 7.5×10-4-2.9×10-2으로 비교적 낮게 나타났으나, 객체량이 40-50%일 때 p_{fd}은 1.1×10-1-1.9×10-21으로 다소 높게 나타났다. 이는 객체 자르기 실험 결과와 유사하게 끊기에 의하여 폴리라인들이 여러 개로 나뉘고, 나누어진 폴리라인들은 다른 그룹으로 할당되어 곡선 곡률에 영향을 주기 때문이다.



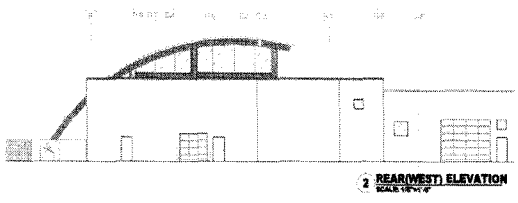
(a) (b)

그림 10. (a) Db 설계 도면의 아래 부분 및 (b) 30% 정도 객체가 끊겨진 Db 설계 도면

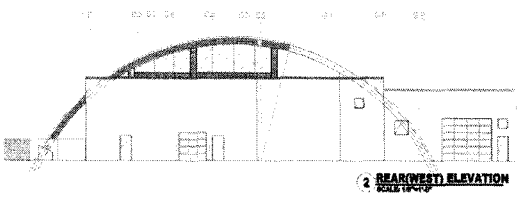
4.1.5 객체 연장 (Extend)

객체 연장은 선택된 객체를 기준의 정의된 객체 모서리에 연결되도록 늘이는 것으로 원 객체의 형상은 유지되면서 객체 길이가 연장된다. 본 실험에서는 임의의 객체를 10-50% 정도를 선택하여 이를 임의의 기준 객체 모서리에 연결되도록 연장하였다. Hummer

Elevation 설계도면의 뒷면 부분은 그림 11 (a)에서와 같고, 20% 정도 객체들이 연장된 도면은 그림 11 (b)에서와 같다. 이 그림으로부터 폴리라인, 타원 등의 객체들이 연장됨을 볼 수 있다. 객체 연장에 대한 오류 검출 확률은 그림 7 (e)에서와 같이, 객체량이 10-40%일 때 p_{fd}은 2.8×10-4-7.1×10-3으로 낮게 나타났으나, 객체량이 50%일 때 p_{fd}은 1.7×10-2으로 다소 높게 나타났다.



(a)

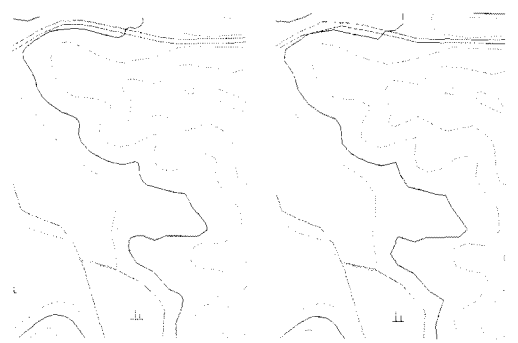


(b)

그림 11. Hummer Elevation의 (a) 뒷부분 도면과 (b) 20% 정도 객체들이 연장된 도면

4.1.6 객체 간단화

객체 간단화는 객체를 구성하는 꼭지점들을 형상을 유지하면서 줄여나가는 것으로 데이터 압축에 사용된다. 본 실험에서는 부표본화(subsample)를 이용하여 모든 폴리라인을 구성하는 꼭지점들을 10-50% 정도 줄였다. 그림 12 (a)는 1:25000 디지털 맵의 일부분을 보여주며, 그림 12 (b)는 모든 폴리라인 상에 50% 정도 꼭지점들이 줄여진 맵을 보여준다. 이 그림을 살펴보면 간단화된 폴리라인의 전체 형상은 유지되나, 곡선이 매끄럽지 못함을 볼 수 있다. 객체 간단화에 대한 오류 검출 확률은 그림 7 (f)에서와 같이, 간단화된 꼭지점 비율이 10-50%일 때 p_{fd}은 1.6×10-4-2.5×10-3으로 낮게 나타났다. 실험 결과로부터 제안한 해싱은 50% 정도까지 간단화된 공격에 대하여 강인함을 확인하였다.



(a) (b)

그림 12. 1:25000 디지털 맵의 (a) 아래 부분과 (b) 50% 꼭지점이 삭제된 부분

4.2 해쉬 유일성

해쉬의 유일성은 모델에 대한 유일성과 키에 대한 유일성으로 나눌 수 있다. 모델에 대한 유일성은 다른 모델에 의하여 생성된 해쉬들이

$$\Pr(\mathbf{H}_1 \neq \mathbf{H}_2) \approx 1 \quad \text{where } \mathbf{H}_1 = \text{hash}(\mathbf{M}_1,\mathbf{K}),\ \mathbf{H}_2 = \text{hash}(\mathbf{M}_2,\mathbf{K}),\ \mathbf{M}_1 \neq \mathbf{M}_2 \tag{25}$$

와 같이 서로 다를 확률이 매우 높아야 하는 것이고, 키에 대한 유일성은 동일 모델 상에서 생성된 해쉬들이

$$\Pr(\mathbf{H}_1 \neq \mathbf{H}_2) \approx 1 \quad \text{where } \mathbf{H}_1 = \text{hash}(\mathbf{M},\mathbf{K}_1),\ \mathbf{H}_2 = \text{hash}(\mathbf{M},\mathbf{K}_2),\ \mathbf{K}_1 \neq \mathbf{K}_2 \tag{26}$$

와 같이 서로 확률이 매우 높아야 하는 것이다. 즉, 모델 및 키에 대한 유일성 실험에서 두 해쉬 간의 정규화된 해밍 거리차가 거의 0.5에 가까워야 한다는 것이다.

본 실험에서는 모델에 대한 유일성 평가를 위하여 50개의 모델에 의하여 생성된 해쉬들 간의 정규화된 해밍 거리차를 구하였고, 키에 대한 유일성 평가를 위하여 각 모델에서 1,000개의 키를 생성한 후, 이 키에 의하여 생성된 해쉬들 간의 정규화된 해밍 거리차를 구한 후, 이들 결과에 대한 분포를 표 2에서와 같이 3단계로 나누어 구하였다. 여기서 $\Pr[D(\mathbf{H}_1,\mathbf{H}_2) \geq 0.5-e]$은 두 해쉬 간이 서로 다르다는 것을 나타내며, 이와 반대로 $\Pr[D(\mathbf{H}_1,\mathbf{H}_2) < e]$는 두 해쉬 간의 유사성이 매우 높음 나타낸다. 모델 유일성 평가에서는 $\Pr[D(\mathbf{H}_1,\mathbf{H}_2) \geq 0.5-e]$이 약 0.95정도이고, $\Pr[D(\mathbf{H}_1,\mathbf{H}_2) < e]$이 10-6이다. 이는 동일 키 상에서 다른 모델에 의하여 생성된 해쉬들이 서로 다를 확률이 거의 95% 이상임을 나타낸다. 키 유일성 평가에서는 $\Pr[D(\mathbf{H}_1,\mathbf{H}_2) \geq 0.5-e]$이 0.98이고, $\Pr[D(\mathbf{H}_1,\mathbf{H}_2) < e]$이 0이다. 이는 동일 모델에서 다른 키에 의하여 생성된 해쉬들이 서로 다를 확률이 거의 98%임을 나타낸다. 이 결과로부터 제안

구분	모델 유일성 평가 $H_1 = hash(M_1,K), H_2 = hash(M_2,K)$	키 유일성 평가 $H_1 = hash(M,K_1)\, H_2 = hash(M,K_2)$
$Pr[D(H_1,H_2) \geq 0.5-e]$	0.956987	0.98
$Pr[e \leq D(H_1,H_2) < 0.5-e]$	0.043012	0.02
$Pr[D(H_1,H_2) < e]$	0.000001	0.00

표 2. 해쉬 유일성에 대한 평가 결과

한 해싱에 의하여 생성된 해쉬가 95% 유일함을 확인하였다.

V. 결 론

본 논문에서는 벡터 데이터 모델의 인증 및 복사방지를 위한 폴리라인의 곡선 곡률 분포 기반의 벡터 데이터 해싱을 제안하였다. 제안한 해싱에서는 벡터 데이터 모델 내의 레이어들 중 폴리라인 분포가 가장 많은 주요 레이어들을 선택하고, 이들 레이어 내의 폴리라인들을 곡선 에너지에 따라 그룹화한다. 그리고 그룹별로 폴리라인들의 1차 및 2차 곡선 곡률 분포를 이용하여 그룹 계수를 구한 다음, 이들을 랜덤 계수 패턴으로 투영하여 특징 계수들을 얻는다. 마지막으로 특징 계수들의 이진화 과정으로부터 최종 이진 해쉬를 생성한다. 실험 결과로부터 제안한 벡터 데이터 해싱은 RST 및 객체 삭제/복사, 자르기, 끊기, 연장 및 간단화 등의 공격에 대하여 강인하며, 랜덤 계수 키에 의한 보안성 및 유일성을 가지게 됨을 확인하였다. 본 논문에서 제안한 벡터 데이터 해싱은 CAD 설계도면 및 GIS 디지털 맵과 같은 벡터 데이터 기반의 여러 콘텐츠들에 대한 인증 과정에 사용될 수 있을 것이다. 향후 본 연구진들은 제안한 해싱의 미분 엔트로피 기반의 보안성을 평가하고자 하며, 벡터 데이터 콘텐츠 뿐만 아니라 3D 영상에 대한 해싱 기반의 인증 시스템으로 확대하여 연구하고자 한다.

참 고 문 헌

[1] G. Farin, J. Hoschek and M.-S. Kim, *Handbook of Computer Aided Geometric Design*, Elsevier science, 2002.

[2] K. Chang, *Introduction to Geographic Information System*, 4th Edition. McGraw Hill, 2007.

[3] R. Ohbuchi, H. Ueda, and S. Endoh, "Robust Watermarking of Vector Digital Maps," *Proc. of IEEE International Conference on Multimedia and Expo (ICME)*, Vol.1, pp.577-580, 2002.

[4] M. Vogit and C. Busch, "Watermarking 2D-Vector data for geographical information systems," *Proc. of the SPIE, Security and Watermarking of Multimedia content*, San Jose, USA, Vol.4675, pp.621-628, 2002.

[5] K.R. Kwon, H.J. Chang, Gwang S. Jung, K.S. Moon, S.H Lee, "3D CAD Drawing Watermarking Based On Three Components", *IEEE International Conference on Image Processing (ICIP)*, pp.1385-1388, Oct. 2006.

[6] K.R. Kwon, J.-S. Sohn, Y. Huh, S.-H. Lee, "The Watermarking For 3D Cad Drawing Using Line, Arc,3Dface Components," *IEEE International Conference on Multimedia & Expo (ICME)*, pp.1361-1364, 2006.

[7] 이석환, 권기룡, "k-means++ 기반의 설계도면 워터마킹 기법," *대한전자공학회논문지*, 제46권 CI편 제5호, pp.57-70, 2009년 9월.

[8] C. Y. Shao, H. L. Wang, X. M. Niu, X. T. Wang, "A Shape-Preserving Method for Watermarking 2D Vector Maps Based on Statistic Detection," *IEICE Transactions on Information and Systems*, Vol.E89-D, No.3, pp.1290-1293, March 2006.

[9] 김정엽, 박수홍, "최근점 쌍을 이용한 벡터 맵 디지털 워터마킹," *정보과학회논문지, 정보통신*, 제36권 제6호, pp.536-544, 2009년 23월.

[10] 김준희, 이석환, 권성근, 박승섭, 권기룡, "GIS 벡터맵 폴리라인 워터마킹 방법," *한국멀티미디어학회논문지*, 제13권 제4호, pp.582-593, 2010년 4월.

[11] E. J. Delp, "Multimedia security: the 22nd century approach," *Multimedia Systems*, Vol. 11, No.2. pp.95-97, Oct. 2005.

[12] A. Swaminathan, Y. Mao, and M. Wu, "Robust

and secure image hashing," *IEEE Trans. on Information Forensics and Security*, Vol.1, Issue 2, pp.215-230, June 2006.

[13] V. Monga and M.K. Mhcak, "Robust and Secure Image Hashing via Non-Negative Matrix Factorizations," *IEEE Trans. on Information Forensics and Security*, Vol.2, Issue 3, Part 1, pp.376-390, Sept. 2007.

[14] B. Coskun, B. Sankur, and N. Memon, "Spatio -Temporal Transform Based Video Hashing," *IEEE Trans. on Multimedia*, Vol.8, Issue 6, pp.1190-1208, Dec. 2006.

[15] C. De Roover, C. De Vleeschouwer, F. Lefebvre, B. Macq, "Robust video hashing based on radial projections of key frames," *IEEE Trans. on Signal Processing*, Vol.53, Issue 10, Part 2, pp.4020-4037, Oct. 2005.

[16] 이석환, 권기룡, "키 기반 블록 표면 계수를 이용한 강인한 3D 모델 해싱," *대한전자공학회논문지*, 제47권 CI편 제1호, pp.1~14, 2010년 1월.

[17] 이성주, 문대성, 김학재, 정용화, 이옥연, "3차원 기하학적 해싱을 이용한 퍼지볼트에서의 지문 정합," *정보보호학회논문지*, 제18권 제1호, pp.11-21, 2008년 2월.

[18] 이석환, 권기룡, "객체별 특징 벡터 기반 3D 콘텐츠 모델 해싱," *대한전자공학회논문지*, 2010년 11월., 제47권 CI편 제6호, pp.75-85, 2010년 11월.

[19] 이석환, 권기룡, "3D 모델 해싱의 미분 엔트로피 기반 보안성 분석," *한국통신학회논문지*, 2010년 12월.

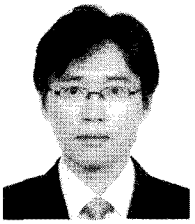
[20] E. Kreyszig, *Differential Geometry*, Dover Publications, New York, 1991,

[21] Differential geometry of curves, http://en.wikipedia.org/wiki/Differential_geometry_of_curves

[22] AutoCAD, http://www.autodesk.com

[23] 국토정보지리원, http://www.ngi.go.kr

이 석 환 (Suk-Hwan Lee) 정회원



1999년 2월 경북대학교 전자공학과 공학사
2001년 2월 경북대학교 전자공학과 공학석사
2004년 8월 경북대학교 전자공학과 공학박사
2005년 3월~현재 동명대학교 정보보호학과 조교수
<관심분야> 워터마킹, DRM, 영상신호처리

권 기 룡 (Ki-Ryong Kwon) 정회원

1986년 2월 경북대학교 전자공학과 공학사
1990년 2월 경북대학교 전자공학과 공학석사
1994년 8월 경북대학교 전자공학과 공학박사
2000년 7월~2001년 8월 Univ. of Minnesota, Post-Doc.
1996년 3월~2006년 2월 부산외국어대학교 컴퓨터전자공학부 부교수
2006년 3월~현재 부경대학교 전자컴퓨터정보통신공학부 교수
<관심분야> 멀티미디어 정보보호, 멀티미디어 통신 및 신호처리