

# GIS 벡터맵 데이터 압축을 위한 혼합형 폴리라인 단순화

임대엽<sup>†</sup>, 장봉주<sup>\*\*</sup>, 이석환<sup>\*\*\*</sup>, 권성근<sup>\*\*\*\*</sup>, 권기룡<sup>\*\*\*\*\*</sup>

## 요 약

본 논문은 혼합형 폴리라인 단순화와 공간 에너지 집중을 이용한 GIS 벡터맵 데이터 압축을 제안한다. 제안한 방법에서는 GIS 벡터맵에서 폴리라인 속성을 가지는 레이어들에 대해, 최소 면적 오차 기반의 혼합형 폴리라인 단순화 및 공간 에너지 집중에 의하여 선택된 레이어 내의 폴리라인 데이터들을 압축한다. 제안한 단순화 및 공간 에너지 집중은 폴리라인의 형상을 유지하면서 데이터 압축 효과를 향상시킨다. 가시적 화질과 압축률을 통한 평가 결과로부터 제안한 방법이 기존 방법보다 우수한 화질과 높은 압축률을 보임을 확인하였다.

## Hybrid Polyline Simplification for GIS Vector Map Data Compression

Dae-Yeop Im<sup>†</sup>, Bong-Joo Jang<sup>\*\*</sup>, Suk-Hwan Lee<sup>\*\*\*</sup>,  
Seong-Geun Kwon<sup>\*\*\*\*</sup>, Ki-Ryong Kwon<sup>\*\*\*\*\*</sup>

## ABSTRACT

This paper presents a GIS vector map data compression scheme based on hybrid polyline simplification method and SEC(spatial energy compaction). The proposed method extracts all layers which contain polylines in the GIS vector map and compress all polylines in extracted layers by the hybrid polyline simplification and SEC based on MAE(minimum area error) for each segment in the line. The proposed simplification and SEC increase the compression ratio while preserving the shape quality. We analyze the visual aspects and compression efficiency between the original GIS vector map and the compressed map. From experimental results, we verify that our method has the higher compression efficiency and visual quality than conventional methods.

**Key words:** Vector map compression(벡터맵 압축), Polyline Simplification(폴리라인 단순화), Spatial Energy compaction(공간 에너지 집중), Hybrid Simplification(혼합형 단순화)

## 1. 서 론

최근 Web GIS, 모바일 GIS, 및 3차원 GIS에 대한 GIS 서비스 시장이 형성되면서, 그에 따른 일반 사용자들의 요구가 점차 증가하고 있다[1-3]. GIS 기술의

이런 빠른 발전과 컴퓨터 및 지도제작 기술의 발달로 표현할 수 있는 지리정보가 더욱 정밀·다양해짐에 따라 GIS 맵 데이터 또한 방대해지고 있는 추세이며, 그로 인하여 GIS 맵 데이터의 고용량 문제가 대두되고 있는 실정이다. 특히, 이동성과 실시간 특성 기반

※ 교신저자(Corresponding Author) : 권기룡, 주소 : 부산광역시 남구 대연 3동 599-1, 부경대학교 대연캠퍼스 웁비관 1316호(608-739), 전화 : 051) 629-6257, FAX : 051) 629-6230, E-mail : krkwon@pknu.ac.kr

접수일 : 2012년 11월 13일, 수정일 : 2013년 1월 21일  
완료일 : 2013년 2월 12일

<sup>†</sup> 준회원, 부경대학교 IT융합응용공학과  
(E-mail : dev\_idy@naver.com)

<sup>\*\*</sup> 준회원, 부경대학교 정보보호협동과정  
(E-mail : roachbj@korea.com)

<sup>\*\*\*</sup> 정회원동명대학교 정보보호학과  
(E-mail : skylee@tu.ac.kr)

<sup>\*\*\*\*</sup> 정회원, 경일대학교 전자공학과  
(E-mail : sgkwon@kiu.ac.kr)

<sup>\*\*\*\*\*</sup> 중신회원, 부경대학교 IT융합응용공학과

※ 본 연구는 2012학년도 부경대학교의 지원을 받아 수행된 연구결과임(PK-2012-0471).

의 Web GIS와 모바일 GIS 서비스에서는 유·무선 네트워크를 통해 전송되는 공간 데이터의 양, 지리정보를 출력, 저장, 처리, 분석하는 모바일 단말기 자원, 유/무선 네트워크의 대역폭 등의 제한적인 서비스 환경으로 인해 GIS 맵 데이터의 방대한 용량에 상당히 민감하게 작용한다. 현재, 이들 GIS 서비스에서는 GIS 서비스를 제공받는 일반 사용자들의 요구를 충족시키면서, 데이터양을 줄이는 기술을 요구하고 있는 실정이다.

최근 방대한 GIS 벡터맵 데이터를 줄이기 위하여 폴리라인 단순화[4-9]와 데이터 압축 기법[10-12]들이 제안되어지고 있다. 폴리라인 단순화는 폴리라인의 형상을 유지하면서 꼭지점 개수를 줄이는 것으로 손실 압축에 해당된다. 데이터 압축은 꼭지점 데이터의 정밀도를 낮추는 손실 압축과 중복성 제거를 통한 무손실 압축으로 나누어진다. 그러나 기존 방법들은 Web GIS 및 모바일 GIS와 같은 제한된 환경에서의 압축 효과에 문제점을 가진다.

따라서 본 논문에서는 제한적인 GIS 서비스 환경과 GIS 맵 데이터의 고용량 문제를 해소하기 위하여 지리 형상을 보존하면서 데이터양을 줄이는 GIS 맵 데이터 압축 방법을 제안하고자 한다. 제안 기법은 먼저, GIS 벡터맵(vector map) 데이터에서 시각적으로 가장 많은 정보를 표현하는 도형 요소(spatial feature)들을 추출하고, 추출된 모든 도형 요소들에 대하여 제안 기법에 의한 압축을 수행한다. 압축 기법은 폴리라인 단순화(polyline simplification) 과정과 공간 데이터 압축(spatial data compression) 과정으로 구성된다. 혼합형 라인 단순화 과정을 통하여 불필요하게 복잡한 형태의 도형 요소들을 단순화시키고, 단순화 된 도형 요소들을 구성하고 있는 모든 좌표데이터에 대하여 공간 영역에서의 에너지 집중(spatial energy compaction)과정 기반의 데이터 압축을 수행함으로써, 최종적으로 원본 GIS 벡터맵의 전체 형상을 최대한 보존하면서, 데이터양이 줄어든 압축된 GIS 벡터맵 데이터가 생성하게 된다. 또한, 단순화 범위를 조절하여 다양한 정밀도 및 압축률을 가변함으로써, 다양한 GIS 서비스 환경에서 일반 사용자의 요구를 충족시킬 수 있게 한다. 제안한 기법의 성능 분석을 위하여 원본 GIS 벡터맵 데이터와 기존의 라인 단순화 기법 및 GIS 벡터맵 데이터 압축 기법들과의 비교 실험을 통하여 시각적 측면과 데이

터 압축률 측면에 대하여 제안 기법의 우수함을 확인한다.

본 논문의 구성으로 2장에서는 본 논문에 대한 관련 연구와 기존의 기법들의 대해 살펴보고, 3장에서는 제안하는 압축 기법에 대해 상세히 알아본다. 4장에서는 제안 기법의 성능평가를 위해 다양한 실험과 그에 대해 고찰한다. 마지막으로 5장에서 개선점 및 향후 연구 방향에 대한 결론을 맺는다.

## 2. GIS 벡터 데이터 압축

본 장에서는 GIS 벡터 데이터에 대한 특징을 살펴본 다음, 기존의 폴리라인 단순화 기법과 벡터 데이터 압축에 대하여 살펴보기로 한다.

### 2.1 GIS 벡터맵 데이터

GIS는 실세계의 다양한 지리정보를 디지털 수치지도 형태로 제작하고, 이에 해당하는 속성정보를 연결하여 처리하는 정보 시스템으로, 가장 중요한 요소는 GIS DB이다. GIS DB는 실세계의 지리적 공간 정보를 담고 있는 GIS 데이터(geo-spatial data)들이다[13]. GIS 데이터는 크게 공간 데이터와 속성 데이터로 구성된다. 공간 데이터는 수치지도 내에서 지형의 형상 및 위치 등 지리적인 정보를 가지고 있는 데이터로 점, 선, 면 등의 도형 요소들로 표현되어 있다. 속성 데이터는 공간 데이터의 도형 요소들의 본질을 정의하는 데이터이다. 공간 데이터, 즉 GIS 맵 데이터는 데이터 모델(data model)에 따라 공간 특징요소(feature)를 격자(grid)나 셀(cell)로 표현하는 래스터 데이터(raster data)와  $x, y$  좌표로 이루어진 선형 도형 요소들로 표현하는 벡터 데이터(vector data)로 분류된다[13].

특히, 벡터 데이터로 이루어진 GIS 벡터맵 데이터의 구조는 여러 개의 레이어(layer)들로 구성되어 있으며, 하나의 레이어는 지형, 하천, 도로 등 실세계의 지리정보들 중 하나의 주제에 대한 정보를 점, 선, 면 등의 선형 도형 요소들을 이용하여 표현한다.

### 2.2 폴리라인 단순화

지도 일반화에서 단순화(simplification) 과정은 중요한 개념으로, 지리적 공간 정보에서 중요한 특성

만을 살리며, 불필요한 정보를 버리는 과정이다.

Douglas-Peucker(DP) 알고리즘[4]은 1973년 제안된 라인 단순화 기법으로, 우수한 성능으로 인해 널리 활용되고 있다. 기본 원리는 라인의 시작점과 끝점을 연결한 선분과 중간점(middle point)들 간의 수선의 길이(perpendicular distance)를 이용하여 단순화시킨다. 오래전에 제안된 기법으로 라인 내에서 급격한 곡선부분에 대해 형상 왜곡이 많은 문제점을 가지며, 이와 관련하여 많은 연구가 이루어지고 있다 [4,5]. Zhao 등[6]이 제안한 Sleeve-fitting(SF) 알고리즘은 폴리라인의 시작점에서 연속적인 두 점에서의 주어진 임계치를 이용하여 계산된 공통 구간 즉, 임의의 너비를 가진 소매(sleeve)를 이용하여 단순화시킨다. SF 기법은 지역적인 단순화 과정으로 인하여 폴리라인에서 작은 홀(hole) 등의 구간이 단순화됨으로서 폴리라인의 전체 형상에 대한 왜곡이 심하게 발생한다. 선회함수(turning function, TF)는 두 형상의 유사도(similarity)를 비교하기 위하여 선분들의 방향각의 차이와 길이를 이용한 방향각의 변화량을 모델링하는 함수이다. 이런 특성을 라인 단순화에 적용한 것이 Carvalho 등[7]이 제안한 방법이며, 국내에서는 국내 수치지도 작성 작업 내규에 맞게 단순화 되는 조건을 변형시켜 폴리곤을 단순화 시키는 방법을 제안하였다[8]. 하지만, 해당 기법 역시 지역적인 단순화 과정으로 인하여 전체 형상의 왜곡 및 소멸되는 문제점을 보인다.

Park 등[9]은 DP, SF, 및 TP의 문제점을 보완하기 위하여 정량적 분석(quantitative characterization) 과정과 세분화 및 단순화(segmentation and simplification) 과정으로 이루어진 단순화 기법을 제안하였다. 해당 기법은 기존의 기법에 비하여 위치 오차와 면적오차에서 좋은 성능을 보이지만, 기존의 라인 단순화 기법의 중복 사용으로 인한 많은 연산량과 다량의 구간 생성으로 인한 단순화 생략, 혼란 데이터에 의해 세분화가 결정되기 때문에 잘못된 데이터 생성 시 단순화에 큰 영향을 미친다. 또한, 지역적 연산 과정으로 인하여 전체 면적오차에 있어서, 연산 복잡성에 비하여 기존의 알고리즘과 비슷한 성능을 보이는 단점을 가진다.

따라서 제안하는 방법에서는 원본 라인의 전체 형상을 보존하기 위하여, 기존의 라인 단순화 기법들의 특성들을 고려하고, 전체 면적오차를 최소화하는 라

인 단순화 기법을 제안한다. 또한, 하나의 가중치를 이용하여 단순화 범위를 조절한다. 기존의 라인 단순화 기법과의 성능 평가를 통하여 이를 입증한다.

### 2.3 벡터 데이터 압축

Chen 등[10]은 기존의 GIS 벡터 데이터 압축 알고리즘들을 크게 무손실 압축(lossless compression)과 손실 압축(lossy compression)으로 분류하였고, 파일 포맷 변환, 좌표계의 변환(mapping), 캐스트(cast) 방식, 리샘플링(re-sampling)방식 등이 있다. 대표적인 리샘플링 손실 압축 방식에는 앞 절에서 설명한 라인 단순화 기법을 이용한 압축 기법이 있다. 본 논문에서 제안하는 압축 기법은 파일 포맷 변환, 캐스트 방식, 라인 단순화 기법을 이용한 리샘플링 방식을 이용한 손실 압축이다. Jang 등[11,12]은 웹 GIS 서비스 환경에서의 효율적인 전송과 저장을 위한 GIS 벡터맵 데이터 압축 기법을 제안하였고, 제안 기법은 폴리라인을 구성하고 있는 모든 64비트 좌표데이터를 이용한 공간영역에서의 에너지 집중 과정(SEC, spatial energy compaction)과 zlib[16] 압축 라이브러리를 이용하여 압축한다. 해당 기법은 좌표 데이터들에 대하여, 두 번의 차 연산에 의해 생성된 결과 데이터들의 오프셋 연산(offset operation)을 이용한 파일 포맷의 변환 압축으로 간단하고, 다양한 정밀도 제공한다는 장점이 있다. 하지만, 모바일 GIS 및 Web GIS 서비스 환경에서는 GIS 맵을 보여주는 디스플레이 환경으로 인하여, 실수 형태의 좌표 데이터에서 정수 부분의 좌표 데이터들만을 이용하여 GIS 맵을 그리기 때문에, 이들 GIS 환경에서는 다양한 정밀도 제공이 힘들며, 데이터 압축률도 낮은 문제점을 가진다. 따라서 본 논문에서는 기존의 라인 단순화 알고리즘의 최소면적오차 구간을 이용한 혼합형 단순화 기법과 좌표데이터에 대한 공간영역에서의 에너지 집중 과정을 결합한 GIS 벡터맵 데이터 압축 기법을 제안하며, 성능 평가를 위하여 기존의 GIS 벡터맵 데이터 압축 기법과 비교 실험을 실시한다.

### 3. 제안한 GIS 벡터맵 압축

그림 1은 본 논문에서 제안하는 압축 기법의 전체 수행 과정을 나타낸 것이다. 제안 기법은 크게 압축

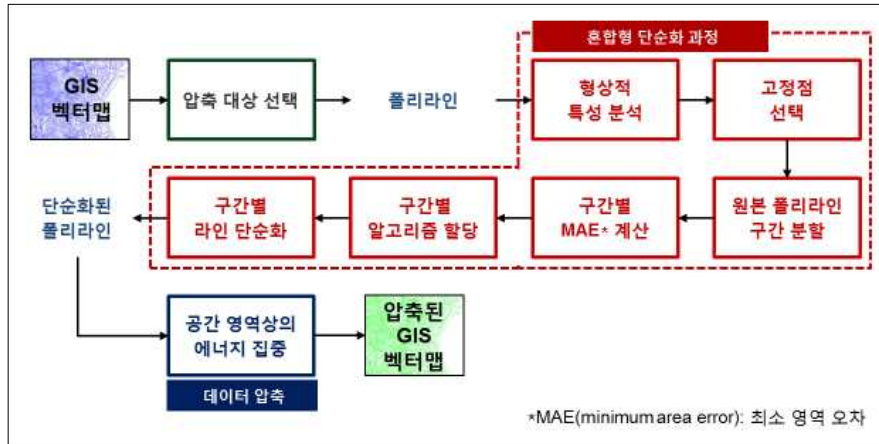


그림 1. 제안하는 압축 기법 과정도

대상 선택, 혼합형 단순화 과정, 데이터 압축 과정으로 이루어진다. 압축 대상 선택에서는 GIS 벡터맵 데이터에서 시각적으로 중요한 정보를 표현하는 폴리라인 요소들을 추출한다. 추출된 모든 폴리라인들에 대하여 혼합형 단순화 과정과 데이터 압축 과정이 수행된다.

3.1 혼합형 단순화 과정

3.1.1 형상적 특성 분석

먼저, 추출된 폴리라인의 형상적 특성을 분석을 위하여, 앞장에서 설명한 기존의 DP, SF, 및 TF 알고리즘들을 이용한다. 원본 폴리라인  $l$ 은  $x, y$  좌표 데이터로 이루어진 여러 꼭짓점( $v_{N-1}$ )으로 구성된다. 이때  $N$ 은 한 개의 폴리라인을 구성하는 모든 꼭짓점의 개수이다. 원본 폴리라인은 세 단순화 기법들을 수행하여 단순화 된 폴리라인  $l_{DP}, l_{SF}, l_{TF}$ 들을

$$l_{DP} = DP(l), l_{SF} = SF(l), l_{TF} = TF(l) \quad (1)$$

와 같이 각각 생성한다. 이때 생성된  $l_{DP}, l_{SF}, l_{TF}$ 들은 각각의 단순화 된 폴리라인들을 구성하는 시작점과 끝점들은 원본 폴리라인의 시작점( $v_0$ )과 끝점( $v_{N-1}$ )과 동일할 뿐, 각각의 단순화 된 폴리라인을 구성하는 꼭짓점의 개수와 구성하는 꼭짓점들의 좌표 데이터는 다르다.

3.1.2 고정점 선택

세 가지 라인 단순화 기법에서 모두 살아남는 꼭짓점들이 존재하며, 이들 점들은 원본 폴리라인의 형

상적 특성을 고려한 중요한 위치에 있는 꼭짓점들이 된다. 이들 꼭짓점들을 선택하기 위하여, 앞서 생성된 단순화 된 폴리라인  $l_{DP}, l_{SF}, l_{TF}$ 들을 구성하는 꼭짓점들 중에서 공통으로 포함되어 있는 꼭짓점들은 원본 폴리라인을 분할하기 위한 고정점들로 선택된다. 그림 2와 같이, 각각의 단순화 된 폴리라인에 대한 공통된 꼭짓점 검색을 위한 원본 폴리라인을 구성하는 꼭짓점의 개수  $N$ 와 같은  $N \times 2$  인덱스 테이블(index table)을 각각 만든다. 첫 번째 열은 순서를 나타내는 인덱스 번호이며, 두 번째 열은 인덱스 번호에 해당하는 꼭짓점의 존재 유무를 나타내는 값으로써, 값이 0이면 제거되어 존재하지 않는 상태이며, 1은 존재하는 상태를 나타낸다. 원본의 폴리라인의 경우, 두 번째 열의 모든 값이 1이 된다. 이렇게 생성된 각 단순화 된 폴리라인에 대한 인덱스 테이블들을 비교하여,

$$\begin{aligned} &\text{if equal}(T_{DPi}, T_{SF_i}, T_{TF_i}), \\ &\text{then } v_i : \text{고정점으로 선택}(i = 0, \dots, N-1) \\ &\text{otherwise, then } v_i : \text{고정점에서 제외} \end{aligned} \quad (2)$$

와 같이 고정점들을 탐색한다. 탐색된 고정점들은 원본 폴리라인을 분할하는데 사용한다.

3.1.3 원본 폴리라인 구간 분할

선택된 고정점들은 원본 폴리라인을 분할하는 중단점으로 사용된다. 그림 3처럼, 중단점들을 이용하여 원본 폴리라인을 여러 개의 구간  $S = \{S_k | k \in [0, K-1]\}$ 으로 분할한다. 각 구간  $S_k$ 은 최소 두 개 이상의 꼭짓점들로 구성된다. 이때 첫 번째

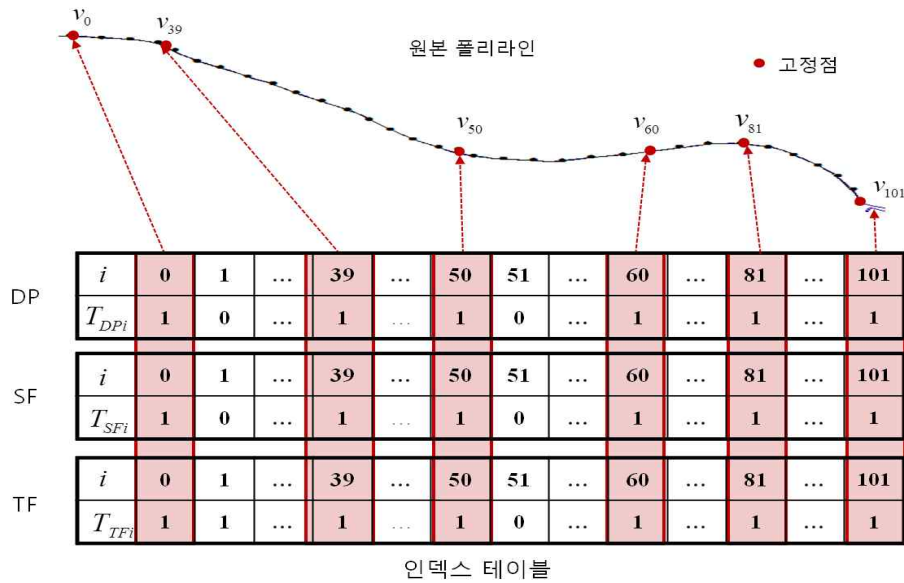


그림 2. 고정점 선택 과정의 예

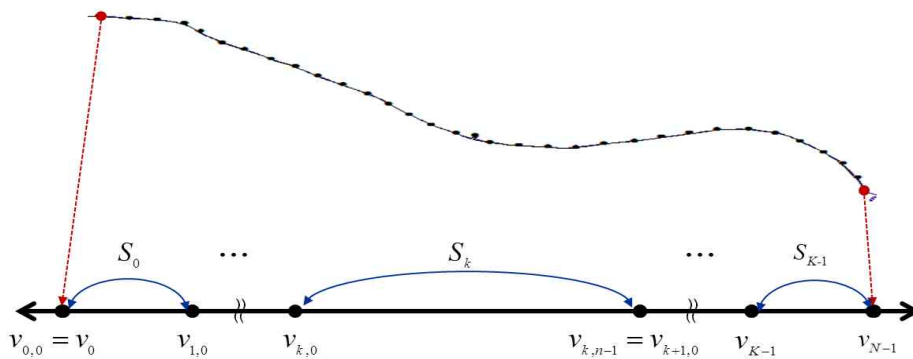


그림 3. 원본 폴리라인 구간 분할

구간  $S_0$ 의 시작점은 원본 폴리라인의 시작점  $v_0$ , 마지막 구간  $S_{K-1}$ 의 끝점은 원본 폴리라인의 끝점  $v_{N-1}$ 이 되며, 이외의 중간구간들은 인접한 두 구간  $S_k, S_{k+1}$ 에서 앞의 구간의 끝점과 뒤의 구간의 시작점은 같다. 제안 기법은 원본 폴리라인의 분할된 구간 단위로 단순화한다.

### 3.1.4 구간별 MAE 계산

제안하는 혼합형 단순화 기법은 원본 폴리라인의 형상 보존을 위하여, 분할된 구간  $S_k$  별로 최소 면적 오차  $MAE_k$ 을 계산한다. 우선, 하나의 구간  $S_k = [v_{k,0}, v_{k,n-1}]$ 는 시작점과 끝점이 다른 하나의 폴리라인으로 볼 수 있다. 이때  $n$ 은 하나의 구간  $S_k$ 에 포함된 꼭짓점들의 개수이다.  $MAE_k$ 는 아래의 단계를 거쳐 계산된다.

[단계 1] 폴리라인 형태의 각 구간  $S_k$ 의 시작점과 끝점을 선분으로 연결함으로써,  $S_k$ 을 시작점과 끝점이 동일한 폴리곤으로 간주한다.

[단계 2] 원본 폴리라인과 각각의 라인 단순화 기법(DP, SF, TF)에 의해 생성된 단순화된 폴리라인  $l_{DP}, l_{SF}, l_{TF}$ 들에 대하여 해당 구간  $S_k$ 의 면적  $A_{l,k}, A_{l_{DP,k}}, A_{l_{SF,k}}, A_{l_{TF,k}}$ 들을 식 (3)을 이용하여 계산한다.  $S_k$ 는 앞 절에서 설명한  $l_{DP}, l_{SF}, l_{TF}$ 들을 이용하여 생성된 고정점을 중단점으로 하기 때문에,  $l_{DP}, l_{SF}, l_{TF}$ 들도 원본 폴리라인  $l$ 과 동일한 개수의 구간으로 분할된다. 하지만,  $l_{DP}, l_{SF}, l_{TF}$ 안에 포함되는 꼭짓점들의 개수와 내용은 서로 다르다. 여기서,  $m, p, q$ 는 각각의 단순화 된 폴리라인  $l_{DP}, l_{SF}, l_{TF}$ 에서 하나의 구간을 구성하는 꼭짓점의 개수이다.

$$\begin{aligned}
 A_{l,k} &= \frac{1}{2} \sum_{j=0}^{n-1} (x_{l,k,j} \cdot y_{l,k,j+1} - x_{l,k,j+1} \cdot y_{l,k,j}) \\
 A_{l_{DP},k} &= \frac{1}{2} \sum_{j=0}^{m-1} (x_{l_{DP},k,j} \cdot y_{l_{DP},k,j+1} - x_{l_{DP},k,j+1} \cdot y_{l_{DP},k,j}) \\
 A_{l_{SF},k} &= \frac{1}{2} \sum_{j=0}^{p-1} (x_{l_{SF},k,j} \cdot y_{l_{SF},k,j+1} - x_{l_{SF},k,j+1} \cdot y_{l_{SF},k,j}) \\
 A_{l_{TF},k} &= \frac{1}{2} \sum_{j=0}^{q-1} (x_{l_{TF},k,j} \cdot y_{l_{TF},k,j+1} - x_{l_{TF},k,j+1} \cdot y_{l_{TF},k,j})
 \end{aligned} \tag{3}$$

단계 3 계산된 각각의 폴리라인에 대한 면적을 이용하여 각 구간  $S_k$  들에서의 DP, SF, TF 단순화 기법에 대한 면적오차  $s_{l_{DP},k}, s_{l_{SF},k}, s_{l_{TF},k}$  들을 식 (4)을 이용하여 계산한다.

$$\begin{aligned}
 s_{l_{DP},k} &= A_{l,k} - A_{l_{DP},k}, k=0, \dots, K-1 \\
 s_{l_{SF},k} &= A_{l,k} - A_{l_{SF},k}, k=0, \dots, K-1 \\
 s_{l_{TF},k} &= A_{l,k} - A_{l_{TF},k}, k=0, \dots, K-1
 \end{aligned} \tag{4}$$

[단계 4] 이렇게 계산된 각각의  $s_{l_{DP},k}, s_{l_{SF},k}, s_{l_{TF},k}$  들을 비교하여, 식 (5)에 의해 구간  $S_k$  에 대한 최소면적 오차  $MAE_k$  을 구한다.

$$MAE_k = \min(s_{l_{DP},k}, s_{l_{SF},k}, s_{l_{TF},k}) \tag{5}$$

### 3.1.5 단순화된 폴리라인 생성

원본 폴리라인의 각 구간  $S_k$  에서 최소면적오차를 보이는 라인 단순화 알고리즘을 해당 구간에 할당한다. 원본 폴리라인의 모든 구간들은 할당된 기법에 의해 생성된 단순화 된 라인의 해당 구간으로 대체 또는 단순화한다.

제안 기법은 다양한 정밀도의 벡터맵 생성과 동시에 압축률을 조절을 위하여, 각각의 라인 단순화 알고리즘들의 단순화 범위를 조절할 수 있는 하나의 가중치 요소(weighting factor)를 준다. 가중치 요소에 의하여, 각각의 라인 단순화 알고리즘의 임계값들은

$$\begin{aligned}
 DP: & d_1 \times (1+\alpha) \\
 SF: & d_2 \times (1+\alpha) \\
 TF: & (a,h) \times (1+\alpha)
 \end{aligned} \tag{6}$$

와 같이 조정되게 된다. 여기서,  $d_1, d_2$  는 각각 DP, SF 단순화 알고리즘에서의 사용하는 초기 임계거리로, 항상  $d_1, d_2 > 0$  의 범위를 가진다.  $a, h$  는 TF 단순화 알고리즘에서 사용하는 임계각도와 임계 길이로,  $a, h > 0$  의 범위를 가진다.  $\alpha$  는 압축률 및 단순화 조절

가중치로, 항상  $\alpha \geq -1$  의 범위를 가진다.  $\alpha = -1$  이면 원본의 폴리라인에 대하여 단순화가 수행되지 않은 폴리라인을 생성하는 것으로, 원본의 형상 그대로 제공한다. 가중치  $\alpha$  가 커질수록 단순화 되는 범위가 커지며, 그로 인하여 제거되는 꼭짓점 개수와 공간데이터량은 증가한다. 반면에, 형상 왜곡도 심하게 된다. 그림 4와 같이, 원본 폴리라인의 모든 구간에 대하여 할당된 단순화 알고리즘으로 단순화시킴으로써, 최종 단순화 된 폴리라인이 생성된다.

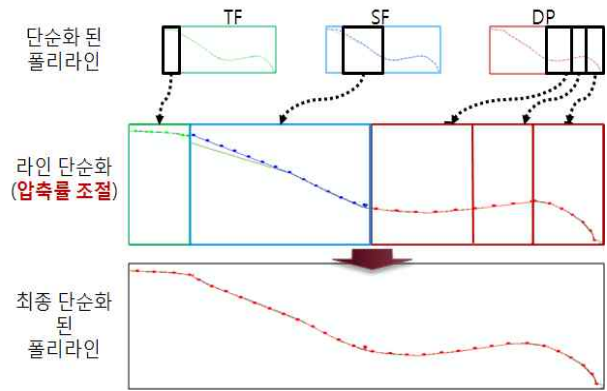


그림 4. 구간별 라인 단순화 알고리즘 할당 및 최종 단순화 된 폴리라인 생성

### 3.2 공간영역에서 에너지 집중 기반의 폴리라인 데이터 압축 과정

제안한 방법에서는 압축 효율을 더욱 높이기 위하여, 혼합형 단순화 기법이 적용된 각각의 폴리라인들에 대해 [12]에서 제안된 에너지 집중 과정을 수행한다. [12]의 공간에너지 집중을 이용한 압축 기법은 다음의 단계를 거쳐 수행된다.

[단계 1] 단순화 된 폴리라인을 구성하고 있는 64 비트 부동소수점 형태의  $x, y$  좌표 데이터를 가진 꼭짓점  $v'_j$  을  $V_j = \text{floor}(v'_j)$  와 같이 정수부 데이터  $V_j = (X_j, Y_j)$  만 추출한다.

[단계 2] 추출된 모든  $V_j$  에 대하여, 각각 해당 폴리라인을 포함하는 최소 사각형(minimum bounding box: MBR)의  $x, y$  좌표의 최소값  $MBR_{\min} = (x_{\min}, y_{\min})$  로 뺀다.

[단계 3] 모든 정수부 데이터  $V_j$  에 대한 각각의  $V_j^*$  는 MBR에 대한 벡터 변위 값이다. 이 벡터 변위의 평균  $AV^*$  를 계산한다.

[단계 4] 평균 벡터 변위  $AV^*$ 를 MBR에 대한 벡터 변위  $V_j^*$ 에 각각 뺀다.

[단계 5] 최종적으로, 단순화 된 폴리라인의 모든 꼭짓점 데이터  $v_j'$ 는 부호를 가진 간소화 된 데이터  $V_j^{**}$ 로 변하게 된다.  $V_j^{**}$ 의 부호값  $m_j$ 는 양수는 1, 음수는 0으로 저장한다.

위의 일련 과정을 통해 생성된 최종 데이터  $V_j^{**}$ 들과 각각에 해당하는 부호값  $m_j$ 들은 파라미터(parameter) 재배열을 통하여, 최종 압축될 스트림으로 저장되게 된다. 이렇게 생성된 최종 압축될 스트림에 범용적인 압축 알고리즘인 zlib을 사용하여 최종적으로 압축된 GIS 벡터맵이 생성된다. 압축된 GIS 벡터맵 파일은 기존의 SHP 파일의 구조와 유사한 형태를 이루게 된다.

#### 4. 실험 결과

제안한 압축 기법의 성능 평가를 위하여, 실험 데이터로 '마산시'와 '거제시'에 대한 GIS 벡터맵을 사용하였다. 실험에 사용된 GIS 벡터맵은 SHP 파일 포맷이다. SHP 파일은 대표적인 GIS 벡터맵 파일 포맷으로, main file, index file, dBase file로 구성되어 있다. 제안한 기법의 성능 평가는 지리 형상을 나타내는 데이터가 주로 포함되어 있는 main file에만 적용하며, 인덱스와 속성정보를 나타내는 index file과 dBase file은 제외한다. 하지만, main file의 폴리라인을 구성하는 꼭짓점의 개수에 대한 정보가 변경

되기 때문에, index file과 dBase file에서도 해당 값을 나타내는 부분의 값을 변경해줘야 한다. 실험 도구로는 Visual Studio 2008, ArcGIS 9.3과 Matlab 2011을 사용하여, 단순화율 및 압축율에 의하여 제안하는 방법의 성능을 평가하였다.

그림 5(a),(b)는 실험에 사용된 GIS 벡터맵으로, 1:1000 축적을 가진 '마산시'와 '거제시'에 대한 벡터맵이다. '마산시'은 약 3만개 이상의 도형 요소와 약 80만개 이상의 꼭짓점으로 이루어져 있다. '거제시'는 약 11만개의 도형 요소와 200만개의 꼭짓점들을 사용하여 지리적 공간 정보를 표현하고 있다. 그리고 폴리라인 요소가 차지하는 비중은 각각 '마산시'는 97%, '거제시'의 경우는 78%를 차지한다.

화질 측면에서의 성능 평가를 위하여, 원본 GIS 벡터맵과 비교 실험을 한 결과, 각각의 도시에 대한 전체 GIS 벡터맵에서 29%, 20%의 꼭짓점들이 제거되었음에도 원본 벡터맵과 시각적인 차이가 없었음을 확인하였다. 그림 6은 그 중 마산시에 대한 실험 결과를 나타낸 것이다.

또한, 이에 대한 화질 측면에서의 정량적인 평가를 위하여 기존의 단순화 기법[9]과 벡터와 영역 오차 측면에서 비교 실험을 실시하였다. 이는 동일 축적에서의 GIS 벡터맵을 대상으로 한 벡터 데이터 편집에 대한 기준이 마련되어 있지 않고, 일반적으로 시각적인 비교에 의한 주관적인 평가에 의존하고 있는 실정이다. 이를 위한 정밀도 측면에서의 정량적인 평가는 벡터맵을 구성하고 있는 폴리라인 단위로 원본 대비 평균 벡터 오차와 평균 면적 오차 측면에서

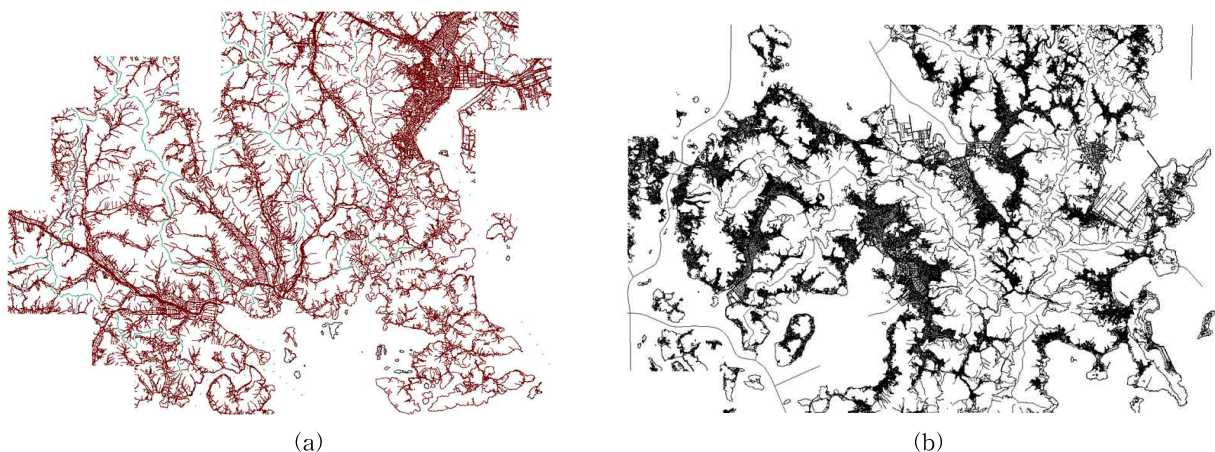


그림 5. 실험에 사용된 1:1000 GIS 벡터맵; (a) '마산시'(도형 요소 수: 36,696개, 꼭짓점 수: 825,188개) (b) '거제시'(도형 요소 수: 110,336개, 꼭짓점 수: 2,316,655개)

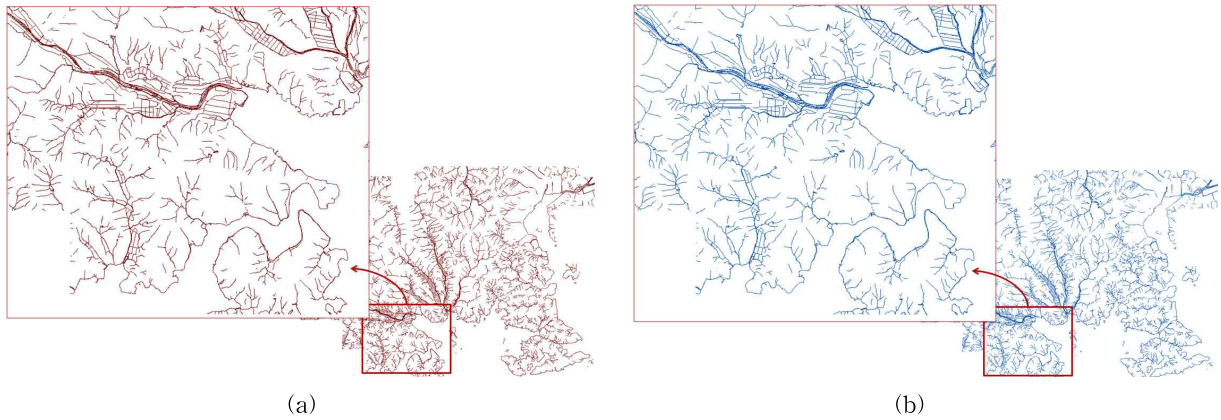


그림 6. '마산시' 벡터맵에 대한 화질 실험 (a) 원본 하천 레이어(폴리라인 수: 10,622개, 꼭짓점 수: 291,954개) (b) 결과 하천 레이어(폴리라인 수: 10,622개, 꼭짓점 수: 184,350개)

비교를 근거로 하고 있다. 그림 7은 기존의 라인 단순화 기법과의 성능 비교 실험 결과 데이터를 나타낸 그래프 그림으로 평균 벡터오차와 평균 면적오차에 대한 그래프이다. 여기서, 평균 벡터오차와 평균 면적오차는 각각 식 (7)-(9)에 의해 구할 수 있다.

$$E = \sum_i^{N-1} e_{ij} \tag{7}$$

$$\text{평균 벡터오차} = \frac{1}{C_L} \sum_{c=0}^{C_L-1} E_c \tag{8}$$

$$\text{평균 면적오차} = \frac{1}{C_L} \sum_{c=0}^{C_L-1} s_c \tag{9}$$

여기서,  $E$ 는 하나의 폴리라인에 대한 벡터오차로, 원본 폴리라인을 구성하는 하나의 꼭짓점의 벡터 오차  $e_{ij}$ 의 합이다.  $e_{ij}$ 는 원본 폴리라인의  $i$ 번째 꼭짓점에서 단순화 된 폴리라인  $j$ 번째 선분으로 내린 수선의 발이다.  $s_c$ 는 원본 폴리라인의 면적에서 단순화 된 폴리라인의 면적을 뺀 면적오차이다.  $C_L$ 은 단순

화 된 폴리라인의 총 개수이다.

이를 통해, 제안 기법이 기존 단순화 기법에 비하여 평균 벡터오차와 평균 면적오차 측면에서 나은 성능을 보이는 것을 확인할 수 있다.

#### 4.1 단순화율

원본 GIS 벡터맵 데이터와 혼합형 단순화 수행 결과는 표 1에 나타내었으며, 공간데이터양이 최소 2%에서 최대 37%의 제거되었음을 볼 수 있다. 여기서 단순한 구조의 폴리라인들로 구성된 시설물과 행정 경계의 경우는 제거되는 공간 데이터의 양이 적었다. 특히, 시설물의 경우는 97~98%의 단순화율을 보였다. 반면에 하천과 도로와 같이 복잡한 구조의 폴리라인 레이어일수록 제거되는 꼭짓점들의 수와 공간 데이터양이 많았다. 특히, 하천의 경우는 63~65%의 단순화율을 보였다. 실험 결과를 통해 복잡한 형태를 가진 맵일수록 제거되는 꼭짓점 및 공간 데이터양이 많음을 확인할 수 있다.

제안한 방법과 기존 방법[9]은 모두 DP, SF, TF 알고리즘을 사용하므로, DP 알고리즘의 임계 길이는 1m, SF는 1m, TF는 선분의 길이는 1m, 각도는 5°로 동일한 값으로 설정하여 비교 실험을 수행하였다. 그림 8(a)에서와 같이, 기존 방법이 동일한 파라미터 값에서 단순화율이 다소 낮지만, 그로 인해 제안 방법보다 형상 왜곡이 심한 것을 확인할 수 있다. 제안 방법이 전체 벡터맵에 포함된 폴리라인의 수를 봤을 때 5%내의 차이는 근소하다고 말할 수 있으며, 일부의 폴리라인에서는 제안 기법이 더 나은 결과를 보였다.

제안 기법과 기존 단순화 기법간의 오차 비교

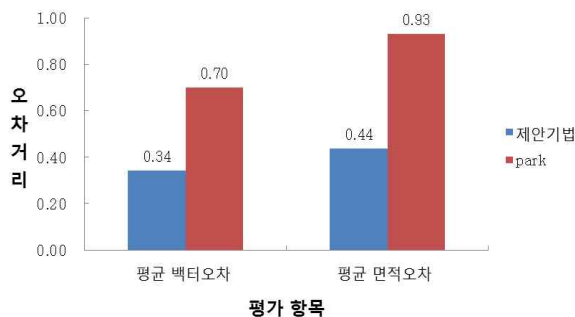


그림 7. 제안 기법과 기존 단순화 기법간의 성능 비교 그래프



표 1. 원본 GIS 벡터맵에 대한 혼합형 단순화 수행 결과 비교

벡터맵	레이어	폴리라인 수(개)	꼭짓점의 수(개)		공간 데이터 양(kB)	
			원본	결과	원본	결과
마산	철도	147	2,244	1,745	35.1	27.3
	하천	10,622	291,954	184,350	4,563.8	2,881.8
	도로	18,472	468,195	347,546	7,318.8	5,432.8
	시설물	6,136	22,966	22,694	359.0	354.8
	행정경계	138	6,868	6,595	107.4	103.1
	합계	35,515	792,227	562,930	12,384.1	8,799.7
거제	하천	5,433	293,483	192,374	4,587.7	3,007.2
	도로	9,325	386,375	263,424	6,039.8	4,117.8
	지류	67,749	816,701	749,886	12,766.7	11,722.2
	시설물	2,994	13,477	13,158	210.7	205.7
	행정경계	61	4,071	3,837	63.6	60.0
	합계	85,562	1,514,107	1,222,679	23,668.5	19,112.9

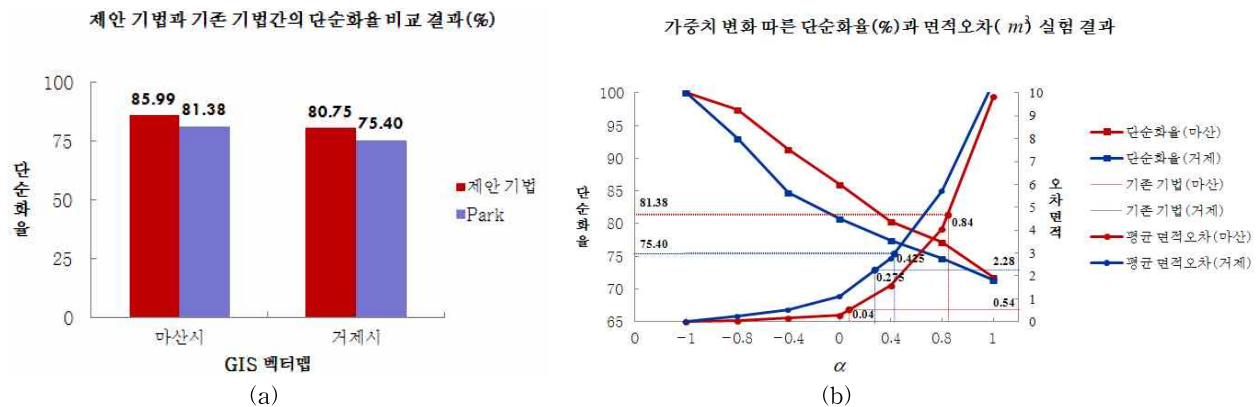


그림 8. (a) 단순화율(%) 및 (b) 가중치 변화에 따른 단순화율과 면적오차(m<sup>2</sup>)

제안 기법은 가중치  $\alpha$ 를 이용하여 원본 폴리라인에 대한 단순화 범위를 조절함으로써, 다양한 정밀도를 가진 GIS 벡터맵 생성 및 압축률 조절이 가능하다. 이에 대한 성능 평가는 그림 8(b)와 같다. 실험은 다양한 정밀도 및 형상 보존에 대한 성능 평가로 단순화율과 평균 면적오차를 측정하였다. 여기서, 원본 GIS 벡터맵에 대한 단순화율과 평균 면적오차는 각각 100%, 0 m<sup>2</sup>이다. 가중치  $\alpha$ 는 항상  $\alpha \geq -1$ 를 만족하여야 하며,  $\alpha = -1$ 이면 폴리라인 단순화가 되지 않은 원본 벡터맵의 정밀도를 그대로 보존하게 된다.  $\alpha = 0$ 이면, 각각의 기존 단순화 알고리즘의 초기에 설정된 임계값들에 의해 단순화되는 것을 의미한다. 실험 결과 가중치  $\alpha$ 가 증가할수록 제거되는 꼭짓점들은 많아지고, 면적오차가 커지는 것을 확인할 수 있

다. 면적오차가 커지는 것은 벡터맵의 형상 왜곡이 심해지는 것을 의미하는 부분이다. 또한, 각각의 벡터맵에 대해서 가중치  $\alpha$ 가 마산시는 약 0.4일 때, 거제시는 약 0.8일 때 동일한 단순화율을 보이는 것을 확인할 수 있다. 이것은 제안 기법의 가중치 변화를 통하여 기존의 단순화기법의 성능 또한 제공할 수 있는 것을 의미한다. 실험을 통해 제안 기법에서 가중치를 이용하여 가변적으로 정밀도 조절이 가능하며, 이것은 다양한 GIS 서비스 환경 및 일반 사용자들의 요구를 만족시킬 수 있는 다양한 정밀도를 가진 GIS 벡터맵 제공이 가능함을 확인할 수 있다.

#### 4.2 압축률

제안한 방법과 기존의 GIS 벡터맵 데이터 압축 방

표 2. 제안한 압축 기법과 다른 GIS 벡터맵 압축 기법간의 압축 데이터 비교(kB)

GIS 벡터맵	레이어 종류	원본 데이터	제안 방법(%)	Jang 방법(%)
마산	철도	43	23(54.06)	28(64.97)
	하천	5,145	2,117(41.15)	2,970(57.73)
	도로	8,326	3,913(47.00)	4,861(58.38)
	시설물	694	486(70.03)	488(70.32)
	행정경계	115	65(56.52)	67(58.26)
	합계	14,323	6,604(46.11)	8,414(58.74)
거제	하천	4,883	1,885(38.60)	2,664(54.56)
	도로	6,548	2,702(41.26)	3,652(55.77)
	지류	16,467	4,616(28.03)	4,920(29.88)
	시설물	375	252(67.20)	255(68.00)
	행정경계	68	37(54.41)	39(57.35)
	합계	28,341	9,492(33.49)	11,530(40.68)

법들 [11] 및 [12] 간의 데이터 압축률에 대해 실험을 실시하였다. 실험 결과는 그림 9과 표 2에서와 같이 제안한 방법이 마산시와 거제시 맵에서 각각 46%, 33%의 압축률을 보임을 알 수 있다. 특히, 마산과 거제에서 각각 하천과 지류 레이어에서 압축이 많이 되었으며, 시설물에서는 압축이 적게 되는 것을 확인하였다. 그리고 기존의 GIS 벡터맵 압축 기법에 비하여 압축률이 우수함을 확인하였다.

앞서 제안하는 GIS 벡터맵 압축 기법의 성능 평가를 위하여 다각도로 실험을 하였다. 실험 결과 제안하는 압축 기법은 적은 공간 데이터로 원본 GIS 맵에 가까운 형상을 보존하였다. 기존의 단순화 기법에 비하여, 형상 보존이 뛰어났으며, 다른 GIS 벡터맵 압축 기법보다 압축률이 뛰어남을 확인할 수 있었다. 최종적인 압축된 GIS 벡터맵은 원본 GIS 벡터맵 데

이터의 최대 28% 데이터만으로도 원본에 가까운 지리적 공간 정보를 제공할 수 있음을 확인할 수 있었다. 또한, 다양한 정밀도를 가진 GIS 벡터맵 제공 및 압축률 조절을 통해 다양한 환경 및 일반 사용자들의 요구를 만족하는 다양한 정밀도를 가진 GIS 벡터맵 생성이 가능함을 확인할 수 있었다.

### 5. 결론

본 논문에서는 모바일 GIS 및 Web GIS와 같이 제한적인 서비스 환경과 일반사용자들의 요구를 만족시킬 수 있는 혼합형 단순화 과정에 의한 GIS 벡터맵 데이터 압축 기법을 제안하였다. 혼합형 단순화 과정은 분할된 구간별로 면적오차가 최소가 되게 단순화시킴으로써, 단순화 된 폴리라인은 전체 면적오차가 최소가 되는 원본 폴리라인에 가까운 형상을 가진다. 또한, 불필요한 꼭짓점들을 제거함으로써, 데이터양도 줄일 수 있었으며, 단순화 범위 조절을 이용한 다양한 정밀도를 가진 GIS 벡터맵을 생성할 수 있다. 이렇게 단순화 된 폴리라인에 대하여 공간 영역에서의 에너지 집중 기반의 데이터 압축이 수행된다. 공간 영역에서의 에너지 집중 기반의 데이터 압축은 단순화된 폴리라인을 구성하고 있는 64비트 부동소수점 형태의 꼭짓점 데이터에 대해 압축을 수행함으로써 압축 효율을 향상시킨다. 실험 결과로부터 제안 기법에 의해 생성된 압축된 GIS 벡터맵은 원본 GIS 벡터맵에 비해 적은 데이터 용량을 가지면

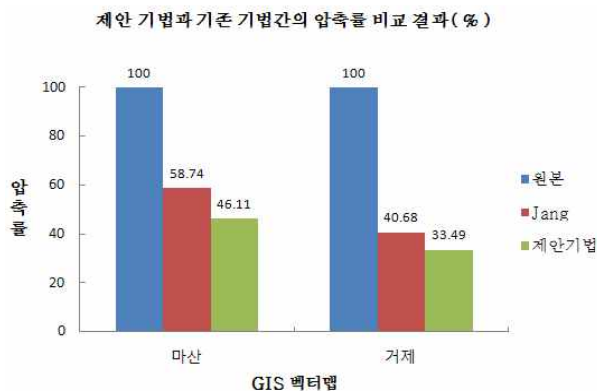


그림 9. 제안 기법과 기존 GIS 벡터맵 압축 기법 성능 비교 그래프

서도 시각적으로 차이를 구별하기 어려웠으며, 기존의 방법들에 비해 압축률이 우수함을 확인할 수 있었다.

### 참 고 문 헌

- [ 1 ] A. Alesheikh, H. Helali, and H. Behroz, "Web GIS: Technology and Its Applications," *Symposium on Geospatial Theory, Processing and Application 2002*, pp.1-9, 2002.
- [ 2 ] 스마트폰 기반 GIS, [http://gis.seoul.go.kr/GisWebDataStore/Gis\\_Edu/html/S1101/SGIS-HTML.jsp?sgis=1101](http://gis.seoul.go.kr/GisWebDataStore/Gis_Edu/html/S1101/SGIS-HTML.jsp?sgis=1101), 2012.
- [ 3 ] 박세환, "3차원 GIS 및 지도 데이터 기술 동향," *정보통신산업진흥원 주간기술동향*, 제1497호, pp.1-10, 2011.
- [ 4 ] D. Douglas and T. Peucker, "Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or Its Caricature," *The Canadian Cartographer*, Vol. 10, No. 2, pp.112-122, 1973.
- [ 5 ] Z. Xie, H. Wang, and L. Wu, "The Improved Douglas-Peucker Algorithm Based on the Contour Character," *19th International Conference on Geoinformatics*, pp.1-5, 2011.
- [ 6 ] Z. Zhao and A. Saalfeld, "Linear-Time Sleeve-Fitting Polyline Simplification Algorithms," *Autocarto 13, ACSM/ASPRS'97 Technical Papers, Seattle, Washington*, Vol. 5, pp.214-223, 1997.
- [ 7 ] J.D. Carvalho, D. Guliato, D.A. Santiago, and R.M. Rangayyan, "Polygonal Modeling of Contours using the Turning Angle Function," *Proceedings of Canadian Conference on Electrical and Computer Engineering*, pp. 1090-1093, 2007.
- [ 8 ] 박우진, 박승용, 유기윤, "건설도면의 GIS 데이터 변환을 위한 건물외곽선 단순화기법 비교," *한국지형공간정보학회지*, 제16권, 제3호, pp. 35- 41, 2008.
- [ 9 ] W.J. Park and K.Y. Yu, "Hybrid Line Simplification for Cartographic Generalization," *Pattern Recognition Letters*, Vol. 32, No. 9, pp.1267-1273, 2011.
- [ 10 ] F. Chen and H. Ren, "Comparison of Vector Data Compression Algorithms in Mobile GIS," *3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT)*, Vol. 1, pp.613-617, 2010.
- [ 11 ] B.J. Jang, S.H. Lee, K.S. Moon, and K.R. Kwon, "Effective Vector Map Image Compression Technique for Secure Encryption and Transmission," *Proc. FCV2010*, pp.299-303, 2010.
- [ 12 ] 장봉주, 문광석, 이석환, 권기룡, "GIS 디지털 맵의 안전한 전송 및 저장을 위한 효율적인 압축 기법," *멀티미디어학회논문지*, 제14권, 제5호, pp.201-218, 2011.
- [ 13 ] Geographic information system, <http://en.wikipedia.org/wiki/GIS>, 2012.
- [ 14 ] zlib, <http://zlib.net/>, 2012.



**임 대 엽**

2010년 부경대학교 컴퓨터멀티  
미디어공학과 학사 졸업  
(공학사)  
2012년 부경대학교 IT융합응용  
공학과 석사 졸업(공학석  
사)

관심분야 : 영상처리, GIS 데이터 압축, 차량 인포메이션  
시스템



**권 성 근**

1996년 경북대학교 전자공학과  
학사  
1998년 경북대학교 전자공학과  
석사  
2002년 경북대학교 전자공학과  
박사

2002년~2011년 삼성전자 무선사업부 책임연구원  
2011년~현재 경일대학교 전자공학과 조교수  
관심분야 : 멀티미디어 암호, 모바일 방송, 워터마킹



**장 봉 주**

2002년 부산외국어대학교 전자공  
학과 학사 졸업(공학사)  
2004년 부산외국어대학교 전자  
컴퓨터공학과 석사 졸업  
(공학석사)  
2004~2006년 (주)아크로엠 연구  
원

2007년~현재 부경대학교 정보보호협동과정 박사과정  
관심분야 : 영상압축, 멀티미디어 정보보호, 임베디드 시  
스템



**권 기 룡**

1986년 경북대학교 전자공학과  
학사 졸업(공학사)  
1990년 경북대학교 전자공학과  
석사 졸업(공학석사)  
1994년 경북대학교 전자공학과  
박사 졸업(공학박사)

2000년~2002년 Univ. of Minnesota, Post-Doc.  
1996년~2005년 부산외국어대학교 디지털정보공학부  
부교수  
2011년~2012년 Colorado State University, Visiting  
Scholar  
2006년~현재 부경대학교 IT융합응용공학과 교수  
2013년~현재 한국멀티미디어학회 정책자문위원  
관심분야 : 멀티미디어 정보보호, 멀티미디어 통신 및 신  
호처리



**이 석 환**

1999년 경북대학교 전자공학과  
졸업(공학사)  
2001년 경북대학교 전자공학과  
졸업(공학석사)  
2004년 경북대학교 전자공학과  
졸업(공학박사)

2005년~현재 동명대학교 정보보호학과 조교수  
관심분야 : 워터마킹, DRM, 영상신호처리