

Douglas Peucker 근사화 알고리즘과 빈 분류 기반 벡터 맵 데이터 압축

박진혁[†], 장봉주^{**}, 권오준^{***}, 정재진^{****}, 이석환^{*****}, 권기룡^{*****}

Vector Map Data compression based on Douglas Peucker Simplification Algorithm and Bin Classification

Jin-Hyeok Park[†], Bong Joo Jang^{**}, Oh Jun Kwon^{***}, Jae-Jin Jeong^{****},
Suk-Hwan Lee^{*****}, Ki-Ryong Kwon^{*****}

ABSTRACT

Vector data represents a map by its coordinate and Raster data represents a map by its pixel. Since these data types have very large data size, data compression procedure is a compulsory process. This paper compare the results from three different methodologies; GIS (Geographic Information System) vector map data compression using DP(Douglas-Peucker) Simplification algorithm, vector data compression based on Bin classification and the combination between two previous methods. The results shows that the combination between the two methods have the best performance among the three tested methods. The proposed method can achieve 4-9% compression ratio while the other methods show a lower performance.

Key words: Vector Map Compression, Douglas Peucker Simplification Algorithm, Bin Classification, Lossy Compression, GIS

1. 서 론

현대 GIS에서 사용되는 맵 데이터는 매우 높은 정밀도와 다양한 정보를 담고 있다. 이렇게 맵 데이터가 담고 있는 정보의 양이 증가하고, 고정밀화 된 데

에는 여러 가지 이유가 있겠지만, 몇 가지 중요한 이유는 최근의 발전된 컴퓨팅 및 스마트폰의 보급, 그로인해 그 어느 때 보다도 일반 사용자들의 GIS를 활용한 다양한 서비스 요구 및 사용 빈도의 증대를 꾀할 수 있을 것이다[1-3]. 이런 이유들로 맵이 표현

* Corresponding Author : Ki-Ryong Kwon, Address: (608-737) 599-1, 45 Yongso-ro, Namgu, Busan, Korea, TEL : +82-51-629-6257, FAX : +82-51-629-6230, E-mail : krkwon@pknu.ac.kr

Receipt date : Sep. 15, 2014, Revision date : Nov. 4, 2014
Approval date : Dec. 3, 2014

[†] Dept. of IT Convergence and Application Engineering, Pukyong National University
(E-mail : detdetplus@naver.com)

^{**} KICT, Korea Institute of Construction Technology
(E-mail : roachbj@kict.re.kr)

^{***} Dept. of Computer Software Engineering, Donggeui University (E-mail : ojkwon@deu.ac.kr)

^{****} Dept. of Multimedia Engineering, Dankook University
(E-mail : dothan@dankook.ac.kr)

^{*****} Dept. of Information Security, Tongmyong University (E-mail : skylee@tu.ac.kr)

^{*****} Dept. of IT Convergence and Application Engineering, Pukyong National University

* This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2014R1A1A4 A01006663), IT/SW Creative research program supervised by the NIPA (National IT Industry Promotion Agency) (NIPA-2014-1195), and ICT R&D program of MSIP / IITP(B0101-14-0059, Human Resource Development Program for Future Internet).

하는 정보가 질적, 양적인 수준이 올라감에 따라 자연스레 GIS 맵 데이터의 용량 역시 방대해졌으며, 실시간성이 요구되는 최근 GIS 분야에서 대용량의 데이터는 문제가 될 수밖에 없었다.

맵 데이터가 대용량화 되어 가는 것에서 파생된 여러 가지 문제의 해결을 위한 시도로 벡터 단순화 기법[4]과 데이터의 압축 기법[5,6]이 활발히 연구 되어져 왔다. 이 기법들 중 벡터 단순화 기법은 GIS 맵 데이터 중 벡터 데이터 맵에 사용되는 방법으로, 특정 프로세스를 통해 벡터 데이터의 좌표 점 개수를 줄임으로써 용량의 압축을 수행하는 기법이다. 가장 대표적인 기법으로는 DP(Douglas-Peucker) 벡터 단순화 알고리즘[4]이 있는데, DP 벡터 단순화 알고리즘을 포함한 벡터 단순화 기법에는 몇 가지의 중요한 단점이 존재하는데, 오브젝트 형상 표현에 불필요한 좌표 점을 제거하는 데 있어 정밀도의 허용오차에 의존하기 때문에 벡터 데이터의 정밀도에 따라 제거할 수 있는 좌표 점 개수가 제한적이게 되고 그에 따라 압축 효율 또한 경우에 따라서 불안정성을 띄게 된다[7]. 또한, 좌표 점을 직접 제거하기 때문에 허용오차 범위 내에서 제거 되었다고 해도 결과적으로는 지도의 정밀도가 저하 되는 문제점이 있다. 이런 점 때문에 고 정밀 지도 데이터를 요구하는 분야에는 적합하지 않지만, 주어진 구현 및 알고리즘이 간단하고 허용오차 범위 내에서는 좌표 점 제거를 통한 압축률이 보장되기 때문에 벡터 기반 GIS 기술 분야 중 고정밀도를 요구하는 분야를 제외한 여러 분야에 현재까지 널리 사용되고 있다. 이런 벡터 단순화 기법과는 달리 좌표점들을 클러스터링 하여 압축을 시도하는 압축 기법들도 있다[5,6]. 이 압축 기법은 좌표점들을 심벌 화 된 값으로 치환하는 것으로 원래의 좌표들이 가지는 값 대신 심벌 화 된 값들로 치환을 하여 표현하는 방법으로 좌표 값에 할당되는 비트를 줄임으로써 압축을 하는 기법이다. 하지만 이때에 문제점은 원 좌표들이 가지고 있던 값과 심벌 값으로 치환된 좌표들이 가지는 값에 따라 정밀도 면에서 오차가 발생하게 되고 이로 인해 오차가 누적 되어 원본과 압축된 데이터의 정밀도에 차이가 발생한다는 단점이 있다[7,8].

본 논문에서는 DP 단순화 알고리즘과 빈 분류 기반의 벡터 맵 데이터 압축 기법을 제안한다. 벡터 맵 데이터의 용량을 압축하는 기법으로는 DP 단순화

알고리즘으로 m 단위의 임계치를 설정하여 레이어 내에 포함된 오브젝트의 좌표 점을 단순화 하게 된다. 이에 따라 단순화 과정 시 좌표 점의 제거 개수가 달라지게 되며, 최종적으로 단순화 된 맵의 정밀도와 원본 맵의 정밀도 차이가 발생하게 되는데, 제안한 방법의 벡터 맵 압축 기법을 사용한 실험 결과로는 비교적 적은 용량을 가진 벡터 맵 데이터의 오브젝트 개수는 $1/2$ 가량이며, 꼭지점 개수는 $1/3-1/4$ 가량이 단순화되었음을 확인하였다. DP 단순화 알고리즘과 빈 분류 기반의 압축 기법을 동시에 활용하는 압축 기법을 실험을 통해 확인한 결과, 비교적 적은 양의 오브젝트와 꼭지점으로 이루어진 벡터 맵 데이터에 대해서도 기존의 압축 기법보다 우수한 압축 성능을 가지는 것을 확인하였다.

본 논문의 구성은 2장에서 벡터 맵 데이터의 압축 방법과 관련된 선행 연구들을 서술하고, 3장에서 제안하는 벡터 맵 압축 방법에 대해 설명한다. 여기에서는 DP 단순화 알고리즘, 영역분할 및 좌표 점 재정의, 카테고리 빈, 방향성 빈, 정밀도 빈의 순서로 서술한다. 그리고 4장에서는 본 논문에서 제안한 압축 방법의 실험과 선행 연구의 실험 결과를 비교, 성능을 평가하며, 5장에서 최종 결론을 서술한다.

2. 관련연구

GIS는 실세계의 다양한 지리정보를 디지털 수치 맵 형태로 제작을 하는데 이 맵을 표현하는 방법은 크게 벡터 데이터와 래스터 데이터로 나눌 수 있다. 벡터 데이터의 경우 점, 선, 면 등의 도형 요소들로 표현되며, 행정 구역의 경계 및 도로, 하천과 같은 사물의 표현에 적합하다. 그리고 래스터 데이터의 경우는 셀 형태로 표현되고, 주로 고도, 강우량, 온도, 식생, 토양 유형, 항공사진 등의 연속 데이터에 사용된다[9-11]. GIS 맵 데이터는 여러 개의 레이어 들로 구성되어 있으며, 하나의 레이어는 산림, 고속도로, 하천과 같은 실세계의 지형들을 하나의 주제별로 표현 한다[12,13]. DP 벡터 단순화 알고리즘은 복잡한 형태의 선 구조를 단순화 시키는 선형 단순화 기법으로 가장 대표적인 벡터 단순화 기법이다. 이 알고리즘은 여러 꼭짓점으로 구성된 벡터 오브젝트를 최대 정밀도에 따라 설정된 임계값을 통해 비 중요 꼭짓점을 판별 후 제거함으로써 단순화를 수행한다.

DP 단순화 알고리즘과 같이 좌표 점을 제거하여 데이터를 압축하는 기법과 달리 사전 기법을 활용하여 벡터 데이터를 압축하는 방법 또한 활발히 연구되어져 왔다. FHM(Fibonacci, Huffman, Markov)압축 방법[14]은 2단계의 과정을 통해 데이터를 압축하게 되는데, 첫 번째 단계는 벡터 맵 데이터에서 상대적 좌표를 이용하여 점의 위치를 표현하는 디퍼런셜 벡터를 추출하는 과정으로, 한 좌표의 x와 y 좌표를 오브젝트의 시작점 또는 그 바로 앞의 좌표를 이용하여 차를 구해서 그 차를 이용하여 디퍼런셜 벡터를 추출해낸다. 두 번째 단계는 앞의 단계에서 구해진 디퍼런셜 벡터를 미리 정의된 사전에 근사화 하는데 이 근사화 단계에서 사용되는 사전을 제작하는 방법이 FHM이다. FHM 방법의 사전 제작 시 Fig. 1에서와 같이 원점에서 수평 수직 방향으로의 거리가 피보나치수열의 형태로 증가하는 사각형을 이용하는데, 사각형 S1의 경우 원점에서 수직 수평 거리가 1이고, 사각형 S2의 경우는 2이다.

이상과 같이 원점에서 수직 수평 거리를 피보나치수만큼 증가시켜 나가면서 사각형을 만들고 이렇게 만들어진 사각형 위에 정수 형태로 각 좌표들이 존재하게끔 엔트리를 설정하여 사전을 제작한다. 이렇게 제작된 사전을 이용하여 각 좌표들을 근사화하게 되면 압축 과정이 완료가 되는데, 빠른 압축 성능을 가지는 장점이 있지만, 사전을 제작할 시 피보나치 수열을 이용하여 사각형을 만들기 때문에 사각형의 크

기가 빠르게 커지게 된다. 이로 인해 바로 한 단계 앞의 정사각형과의 크기 차이 역시 커지게 되어 근사화 시 두 사각형 사이에 존재 하는 좌표 점의 경우 두 사각형 중 어떤 사각형으로 근사화 되더라도 정밀도에 오차가 발생하게 된다. 이런 식으로 정밀도가 누적이 되면 원본 벡터 데이터와 압축된 벡터 데이터 간의 정밀도 차이가 크게 발생하게 된다는 단점이 있다. 이런 단점을 보완하기 위해 데이터의 특성을 반영하여 사전을 제작하는 방법 또한 연구 되었는데, CBC(Clustering Based Compression) 압축 방법[15, 16]은 레이어 내에서 오브젝트를 이루는 좌표점들 간의 디퍼런셜 벡터들을 K 평균 군집화 기법을 이용하여 군집화 시킨 결과로 사전을 제작함으로써, FHM 방법에 비해 조금 더 나은 정밀도를 기대할 수 있게 하였다. CBC 압축 방법에서 사전을 제작하는 과정은 K 평균 군집화 기법을 이용하는 것으로, 각 미분 벡터들의 중심 값을 계산하여 이 중심 값에 가장 유사한 군집에 각 좌표들을 군집화 하게 된다. 이 군집화 과정에서 Euclidean 거리를 이용하여 유사도 측정을 하게 되는데, n 차원 공간에서 적용 가능한 Euclidean 거리를 계산할 수 있도록

$$d = \sqrt{\sum_{i=1}^n (b_i - a_i)^2} \tag{1}$$

where $a = (a_1, a_2, a_3, \dots, a_i)$ $b = (b_1, b_2, b_3, \dots, b_i)$
 과 같이 중심 값과 각 미분 벡터들의 유사도를 측정하고, 각 미분 벡터들이 유사도가 가장 높은 군집으로 모두 할당될 때까지 새로운 군집 중심을 선정하여, 일련의 과정을 반복 수행한다. 최종적으로 할당이 일어나지 않으면 가장 마지막에 이루어진 군집의 군집 중심들을 엔트리로 하는 사전을 제작하게 된다. 이렇게 제작된 사전을 이용하여 FHM 방법과 같이 미분 벡터를 근사화하는 방법으로 압축을 하게 되는 CBC 압축 방법은 압축에 이용되는 사건의 제작 과정 시 압축하고자 하는 벡터 데이터로부터 구해진 디퍼런셜 벡터를 이용하기 때문에 미리 제작되어진 사전을 이용하는 FHM에 비하여 향상된 정밀도를 가지지만 이 방법으로 정확한 정밀도를 얻기 위해서는 군집화 과정에서 사건의 엔트리가 되는 군집 중심의 수를 많이 늘려주어야 하고, 압축을 수행하기 전 사전을 제작하는 과정을 거치기 때문에 압축 수행 시간이 늘어나게 된다는 단점이 있다. 하지만 이마저도 근사값을 이용하여 압축을 수행하는 근사화 압축 기

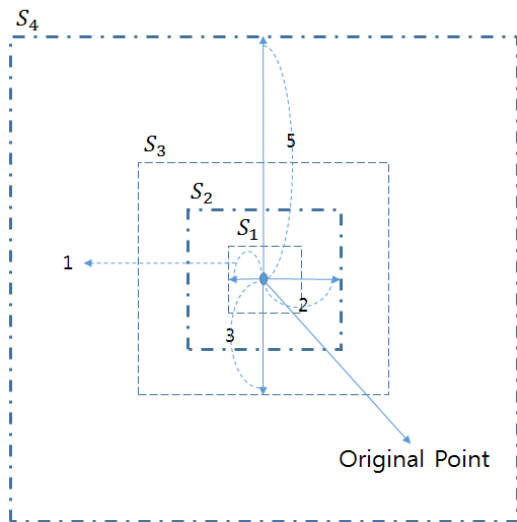


Fig. 1. Encoding by the FHM Algorithm.

법의 특성상 완벽한 정밀도를 가지기는 어렵다. Jang[7]이 제안한 빈 분류 기반 벡터 맵 데이터 압축 기법은 오브젝트를 이루는 좌표점의 개수를 줄여 데이터의 용량을 줄이는 단순화 기법이나 원래의 좌표 값을 근사값으로 근사화 하여 압축하는 근사화 기법과 달리, 레이어 내 오브젝트들 간의 인접성을 고려하여 빈(Bin)을 정의, 이 빈을 이용하여 압축을 시도한다. 빈 분류 기반 벡터 맵 데이터 압축 기법은 실수 값에서 정수 값으로 재 정의된 하나의 레이어를 분할하여 분할된 레이어 내 좌표 값 중 최소 값을 가지는 좌표 값을 기준으로 하여 분할된 레이어 내 속하는 모든 좌표 점을 재 표현하는 전처리 단계를 거친다. 이후 레이어 내의 오브젝트 분포 특성을 이용하여 Category Bin(CB)을 정의, 이에 따라 레이어 내에서 나뉘어진 CB의 연결성 검사를 통해 CB의 압축을 수행하고, 각 CB 내에 포함된 각 오브젝트가 갖는 좌표 점들의 대략적인 방향성을 압축하는 Direction Bin(DB)을 정의, 오브젝트를 이루는 좌표 점들의 연결 방향성을 이용하여 DB의 압축을 시도한다. 마지막으로 각 좌표들의 정밀한 위치 정보를 저장하는 Accuracy Bin(AB)을 정의하여 전송 과정에서 실시간성을 확보하기 위하여 좌표값을 MSB(Most Significant Bit)부터 LSB(Least Significant Bit)의 순서로 전송하여 점진적인 벡터 맵 데이터 압축을 수행한다.

3. 제안한 방법

제안한 DP 단순화 알고리즘과 빈 분류 기반 벡터 맵 데이터 압축 과정은 Fig. 2에서와 같다. 빈 분류 기반의 벡터 맵 데이터 압축 기법을 수행하기에 앞서 DP 단순화 알고리즘을 이용하여 임의의 벡터 맵 데이터 레이어 내에 속해 있는 오브젝트들에 대하여 오브젝트를 이루고 있는 좌표점 개수를 줄이는 단순화를 실시한다. DP 단순화 단계를 거친 후의 레이어에는 원본 레이어에 비해 좌표점이 줄어든 오브젝트들이 담기게 된다. 이렇게 DP 단순화 과정 후 단순화된 레이어에 대하여 전처리 단계로 일정하게 구간을 나누고(Partitioning) 실수를 가지고 있는 좌표값을 정수로 재정의(Re-Definition) 하여 빈 분류를 위한 각 좌표값들 상에서의 비트할당(Bit Per Bin: BPB)을 한다. 전처리 과정을 거치고 난 후 CB와 DB 그리고 AB까지 순서대로 압축 과정이 모두 수행되고 나

면, 본 제안 압축 기법은 종료하게 된다.

3.1 DP 단순화 알고리즘

DP 단순화 알고리즘은 복잡한 형태의 선 구조를 단순화 시키는 선형 단순화 기법으로, 이 알고리즘은 여러 꼭짓점으로 구성된 폴리곤, 폴리라인 등의 벡터 데이터에 대해 비 중요 꼭짓점을 제거함으로써 단순화를 수행하는데, 대상이 되는 레이어에 대해 DP 단순화를 실시하여 단순화 된 오브젝트로 구성된 레이어를 생성한다. DP 단순화 알고리즘의 수행 단계는 다음과 같다. 단순화될 해당 벡터 오브젝트의 시작점과 끝점을 연결한 선분과 각 꼭짓점들이 수직으로 연결되는 수선의 길이 중 임계치 ϵ 를 벗어나는 최대 수선의 길이 d_{\max} 를 찾는다. 이 최대 수선의 길이 d_{\max} 를 가진 꼭짓점을 새로운 중간점으로 설정하여 시작점과 끝점을 새로 설정된 중간점과 연결, 오브젝트를 분할하고, 임계치 ϵ 를 벗어나는 수선의 길이를 가지는 꼭짓점이 존재하지 않게 되면, 분할된 오브젝트 내의 꼭짓점을 제거하는 과정을 재귀적으로 반복 수행한다. 최종적으로 시작점, 끝점,

$$v_n = \begin{cases} \text{Preservation, if } d_{\max} \geq \epsilon \\ \text{Elimination, if } d_{\max} \leq \epsilon \end{cases} \quad (2)$$

ϵ 에 따라 선택된 분할점들만으로 구성된 단순화된 벡터 1오브젝트를 생성하는 것으로 벡터 단순화는 종료된다. 이렇게 생성된 레이어 내의 오브젝트들은 원본과 비교하였을 때 시작점과 끝점만 동일할 뿐, 오브젝트를 구성하는 꼭짓점들의 개수와 꼭짓점들의 좌표 데이터는 다르다[4].

3.2 빈 분류 기반 벡터 맵 데이터 압축

3.2.1 영역분할 및 좌표 점 재정의

단순화 된 전체 벡터 데이터로부터 N_A 개의 영역들 $\mathbf{A} = \{A_1, A_2, A_3, \dots, A_{N_A}\}$ 로 분리한 후, 각 분리된 영역에 대해서 레이어 들을 분리해낸다. 이 분리된 레이어 내 좌표 값 들은 레이어에 대한 가상의 좌표 값을 기준으로 정수 형태를 갖는 상대적 좌표 값으로 재정의 된다.

$$v_{p_k, n} = \lfloor (v_{p_k, n} - v_{\min}) \times 10^c + 0.5 \rfloor, \quad 0 \leq n < N_{V, p_k} \quad (3)$$

$v_{p_k, n}$ 은 오브젝트 p_k 에 포함된 n 번째 좌표점을 나타내며, v_{\min} 은 레이어 $L_{i,j}$ 가 갖는 가상의 최소 좌표

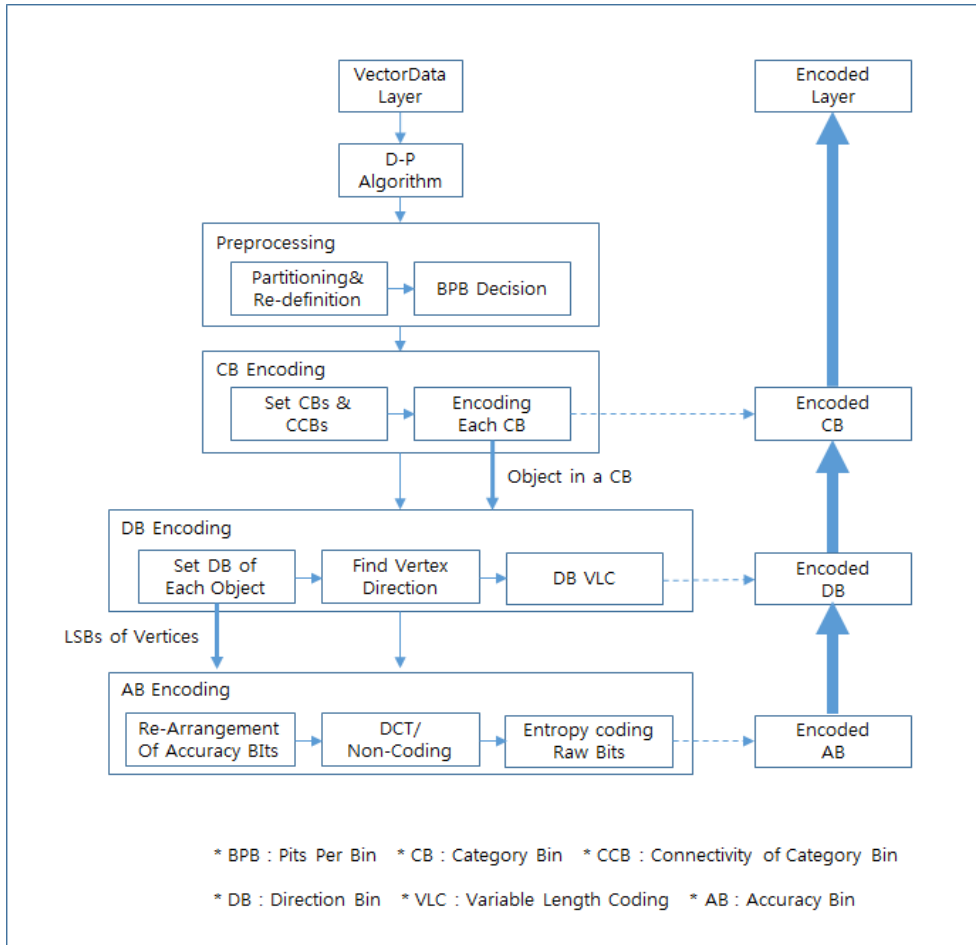


Fig. 2. The proposed vector map data compression method.

점을 나타낸다. 여기서 가상의 좌표 값은 해당 레이어 내에서 최소의 x와 y 값들로 정의된다[7].

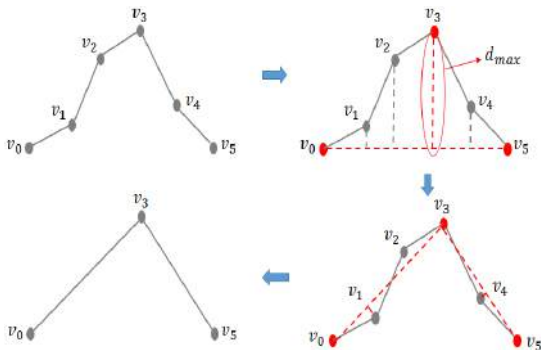


Fig. 3. DP(Douglas-Peucker) algorithm process.

3.2.2 카테고리 빈 인코딩 단계

제안 기법의 CB 압축은 레이어 내에서 Δ_{CB} 범위로 나눈 모든 CB들의 검색을 통해 의미 있는 중요 CB와 그렇지 않은 비 중요 CB들로 분리한다. 중요 CB의 검색이 끝나고 나면 하나 이상의 오브젝트가 존재하는 중요 CB들에 대해서 수직 및 수평 방향 연결성을 결정한다. [7]. 중요 CB탐색은 현재 CB를 기준으로 레이어 $L_{i,j}$ 의 공간 좌표계에서 ‘좌(left)->우(right)’ 및 ‘하(down)->상(up)’ 방향의 순서로 이루어지며, CB 압축은 분리된 CB들 중 중요 CB에 대해서만 검색하여 확인된 연결성으로 압축을 실시한다. 압축 과정에서 연결성 탐색과정에 의한 각 CB들의 분류를 아래의 표로 나타내었다. 인접 CB들 중 중요 CB가 하나 이상이고 가장 먼저 탐색되는 CB를

CB_{head} 로 정의하고, 자신의 오른쪽과 위쪽을 나타내는 방향 지시자에 할당된 2비트가 더해지는 것을 아래의 표를 통해 알 수 있다. 또, 다른 중요 CB들과 연결성이 없는 CB_{iso} 의 경우 CB_{iso} 에 속한 모든 오브젝트들의 각 좌표 값들을 표현하는 비트 중 CB에 할당된 비트들은 CB_{iso} 의 지시자로서 x, y 축에 대해 총 $2 \cdot nB_{CB}$ 의 비트를 갖는 단 하나의 대표 값 $idc_{CB_{iso}}$ 으로 표현하여 CB_{iso} 내 모든 오브젝트들의 좌표 점들을 압축한다. CB_{iso} 에서 각 좌표 점들의 압축되는(제거되는) 비트 수의 합 $eB_{CB_{iso}}$ 은 아래와 같이 계산된다. Table 1은 분류된 CB의 타입들에 대한 정보를 보여준다.

$$eB_{CB_{iso}} = 2 \cdot \left[nB_V \cdot \sum_{k=0}^{N_{P, CB_{iso}} - 1} N_{V, P_k} - \left(nB_{CB} + (nB_V - nB_{CB}) \sum_{k=0}^{N_{P, CB_{iso}} - 1} N_{V, P_k} \right) \right],$$

$$= 2 \cdot nB_{CB} \cdot \left(\sum_{k=0}^{N_{P, CB_{iso}} - 1} N_{V, P_k} - 1 \right), \text{ for } P_k \in CB_{iso} \quad (4)$$

3.2.3 방향성 빈 인코딩 단계

DB 압축은 해당 CB에 속한 각 오브젝트 단위로

수행되며, 각 오브젝트에 포함된 좌표들의 연결성으로 압축을 실시한다[7]. DB 압축은 DB들의 변위 DDB(Differential Direction Bin)를 추정한다

$$DDB_{v_q, P_k} = DB_{v_q, P_k} - DB_{v_{q-1}, P_k}, \quad 0 < q < N_{V, P_k} - 1 \quad (5)$$

이 변위의 빈도를 계산하여 사전 정의된 VLC 테이블에 의해 압축이 수행된다. 방향성 빈에서 DDB_{v_q, P_k} 는 DB의 범위가 증가할수록 점차 0에 가까운 값들의 빈도수가 증가하게 되며, DDB_{v_q, P_k} 에 대한 VLC 코드 할당량이 줄어들게 되어 DB 압축 단계의 압축효율을 향상시키는 효과가 있다. 하지만 벡터 데이터의 경우 긴 직선으로 표현되는 오브젝트 또한 포함하고 있는데, 이런 경우 좌표 점의 DDB_{v_q, P_k} 값이 커짐으로 인해 할당되는 VLC 코드 역시 길어지므로, 압축 후 좌표 점의 비트 수가 원래 x, y 좌표 점이 갖는 비트 수 $2 \cdot nB_{DB}$ 보다 증가하게 된다. 이렇게 압축된 VLC 코드가 원래의 좌표 값들이 갖는 비트 수보다 증가하는 경우를 방지하기 위하여, DB 압축 과정에서 좌표 점 간 거리가 큰 경우에도 각 좌표 점의 DB에 대한 VLC 코드의 최대길이는 th 를 설정하여 제한함으로

Table 1. CB classification by searching CB connection.

CB types	# of object	Dependency	Compression efficiency	Bits for CB indicator idc
CB_{null}	0	No	0	0
CB_{iso}	$1 \leq$	No	$\eta_{CB_{iso}}$	$2 \cdot nB_{CB}$
CB_{head}	$1 \leq$	No	$\eta_{CB_{head}}$	$2 \cdot nB_{CB} + 2$
CB_{follow}	$1 \leq$	Yes	$\eta_{CB_{follow}}$	3

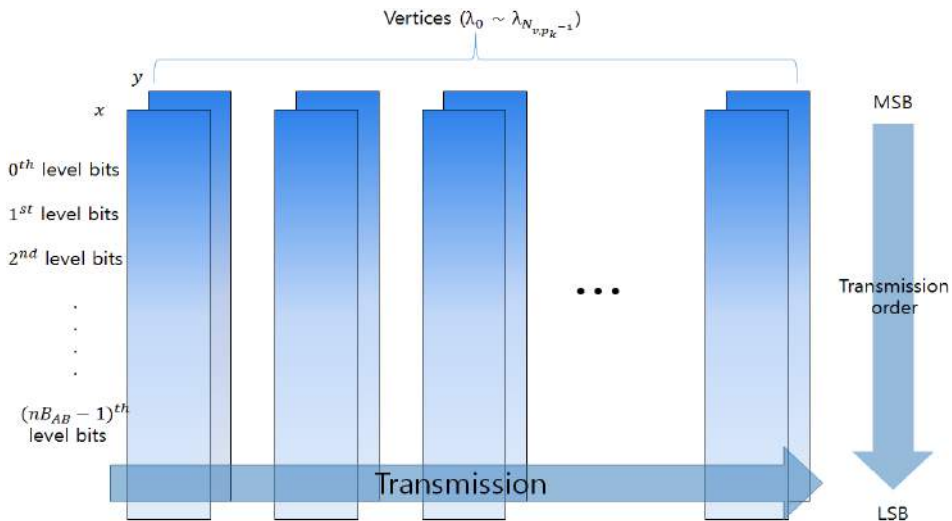


Fig 4. Progressive compression AB on AB bits in any object.

써 th 보다 클 경우 해당 좌표 점에 대해 AB(Accuracy Bin)에 할당 되는 nB_{AB} 만큼의 LSB를 제외한 원래의 값을 할당한다.

$$th = 2 \cdot (nB_{CB} + nB_{DB}) \tag{6}$$

3.2.4 정밀도 빈 인코딩 단계

모든 좌표 점들에 대해 CB와 DB에 할당되고 남은 LSB들에 대한 압축으로 AB 압축을 실시한다. AB는 요구정밀도에 대해 오차 없이 실시간으로 전송 가능한 점진적 AB 압축을 수행한다. Fig. 4는 AB 압축 과정의 예를 보여준다.

AB는 모든 좌표 점들에 대해 개의 LSB들에 대한 압축으로 표현되지만, AB에 할당된 비트 값들은 벡터 좌표계에서 정확한 정밀도를 나타내기 위해 사용되기 때문에, 실제로는 데이터 압축이 이루어지는 것이 아니라, LSB부터 점진적으로 전송되는 방식으로 저해상도에서 최종적으로 원하는 정밀도까지의 지도 탐색이 가능하게 하여 실시간성을 확보 한다는데 의의가 있다.

4. 실험 결과 및 고찰

본 논문에서 제안한 벡터 맵 데이터 압축 기법은 DP 단순화 알고리즘으로 오브젝트를 단순화하고 이 단순화 된 벡터 맵 데이터를 다시 빈 분류 기반 압축을 실시함으로써, 고용량의 GIS 벡터 맵 데이터에 대한 효과적인 압축을 하고자 하였다. 이에 아래와 같이 폴리라인 및 폴리곤 레이어를 대상으로 하여 DP 단순화 알고리즘과 빈 분류 기반 벡터 데이터 압축 기법 그리고 제안 기법의 성능 평가를 실시하였다. 이 실험은 비교적 낮은 용량의 벡터 맵 데이터의 압축과 대용량의 벡터 맵 데이터의 압축 실험을 실시하여 도출해낸 압축률을 통해 본 논문에서 제안한 압축 기법의 장점을 밝히고자 하는데 목적이 있다. 빈 분류 기반 벡터 데이터 압축 기법의 경우 각 빈에

할당되는 비트에 따라 달라지는 압축률 중 최적의 압축률을 알아내기 위해 최대 정밀도를 1m로 하여 각 빈에 할당되는 비트를 CB의 압축 정밀도에 따라 할당되는 비트를 1에서 최대 비트로 가변시켜 가며 할당하고 이후 방향성 빈과 정밀도 빈에 할당되는 비트 또한 CB에 할당한 비트를 제외한 나머지 비트들을 각각 가변 하여 실험 하였다. 빈 분류 기반 벡터 데이터 압축 기법 실험 후 본 논문에서 제안한 압축 기법의 실험은 DP 단순화 알고리즘의 경우 임계치를 1m로 고정을 하여 실행 시켰으며, 단순화를 수행하여 좌표점이 제거 된 오브젝트를 포함하고 있는 레이어에 대하여 빈 분류 기반 벡터 데이터 압축 기법을 앞서 실험한 바와 같이 각 빈에 할당되는 비트를 가변하며 실험을 실시하였다. Table 2 및 Fig. 5는 본 논문에서 사용된 대구시와 창원시의 GIS 데이터에 대한 폴리라인과 폴리곤의 정보를 나타낸 것이다.

Fig. 6에서 보이는 바와 같이, 레이어 내 포함된 오브젝트의 좌표 점 개수가 적을수록 벡터 형상에 있어서 변화가 없는 것을 확인할 수 있었는데, 이는 DP 단순화 알고리즘으로 단순화를 할 수 있는 좌표점이 많지 않았기 때문이다. 또 임계치를 높게 설정할수록 임계치의 변화에 따라 DP 단순화 알고리즘을 실시한 데이터의 벡터 형상의 변형에 더 많은 변화가 발생함을 확인하였다. 또 폴리곤의 경우, DP 단순화 알고리즘으로는 데이터의 압축 효율 면에서 제거된 꼭지점의 개수가 많지 않아 압축률 면에서도 성능이 떨어지면서 정밀도 면에서도 또한 원본 벡터 데이터와 차이가 나는 것을 확인할 수 있었다. 이처럼 DP 단순화 알고리즘과 같이 좌표 점을 제거하여 압축을 시도하는 기법의 경우 벡터 맵 데이터 자체의 상태에 따라 압축 성능을 제대로 보장 받기가 어렵다는 것을 알 수 있었다. 그리고 단순화 임계치 설정에 따라 벡터 형상에 큰 변화가 생기는 것 또한 확인이 가능하였다. 이와 달리 형상에 변화를 주지 않으면서도, 벡터 데이터의 압축에 있어서 우수한 성능을 지

Table 2. Information for tested layers.

Map	Attribute	Total Object	Total Vertices	Data Volume (Bytes)	
Daegu	Fig. 5(a)	Polyline	12,605	90,525	1,448,400
	Fig. 5(b)	Polygon	10,346	70,472	1,127,552
Chang won	Fig. 5(c)	Polyline	25,522	383,200	6,131,200
	Fig. 5(d)	Polygon	18,587	158,623	2,537,968

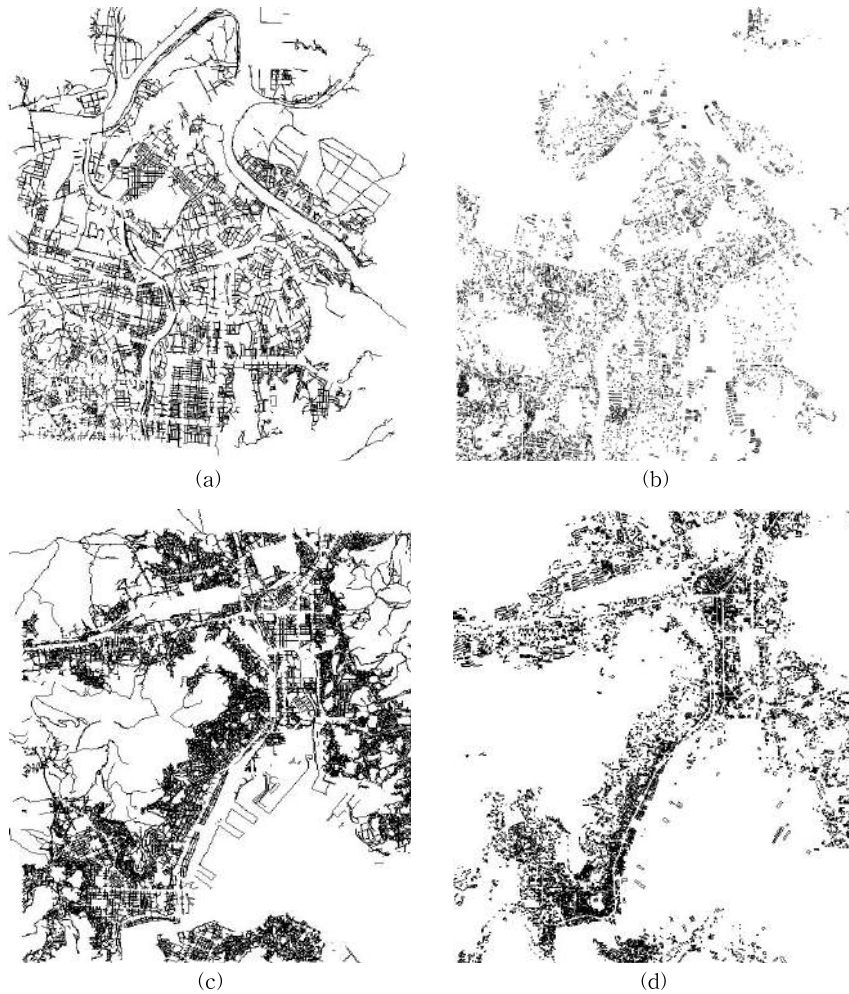


Fig. 5. (a) Polyline layer, (b) Polygon layer for Daegu and (c) Polyline layer, (d) Polygon layer for Changwon.

Table 3. Number of vertices in simplified map according to threshold using DP algorithm.

threshold (ϵ)	Vertices number after simplification [compression efficiency %]			
	Daegu		Changwon	
	Polyline layer (90525) (Fig. 5(a))	Polygon layer (70431) (Fig. 5(b))	Polyline layer (383,200) (Fig. 5(c))	Polygon layer (158,623) (Fig. 5(d))
1	49625 [51]	61643 [87]	153254[39]	115953[73]
2	43368 [47]	56938 [80]	125097[32]	102322[64]
3	40573 [44]	53506 [75]	112822[29]	96517[60]
4	38857 [42]	50399 [71]	104350[27]	91640[57]
5	37758 [41]	47356 [67]	98425[25]	85745[54]
6	36978 [40]	43897 [62]	93769[24]	79319[50]
7	36310 [40]	40307 [57]	90012[23]	73039[46]
8	35794 [39]	37334 [52]	86658[22]	67977[42]
9	35356 [39]	34984 [49]	83664[22]	64092[40]
10	34990 [38]	33081 [46]	80970[21]	60722[38]



Fig. 6. Shape changes with $\epsilon=5$ using DP algorithm: (a) Polyline layer for Changwon (b) Polyline layer for Daegu, (c) $\epsilon=9$, (c) Polygon layer for Changwon, (d) Polygon layer for Daegu.

Table 4. Compression performance in according to the bits assigned to each bin.

Layer	Daegu				Changwon			
	Bits for CB	Bits for DB	Bits for AB	Compressed Data	Bits for CB	Bits for DB	Bits for AB	Compressed Data
Polyline	8	6	0	178402 [12]	8	6	0	541771[8]
	8	5	1	162496 [11]	8	5	1	501782[8]
	8	4	2	159488 [10]	8	4	2	501002[8]
	8	3	3	154052 [10]	8	3	3	533275[8]
	8	2	4	158677 [10]	8	2	4	588243[9]
	8	1	5	168877 [11]	8	1	5	658325[10]
	8	0	6	183008 [12]	8	0	6	737574[12]
Polygon	7	6	0	119882 [10]	8	6	0	247563[9]
	7	5	1	109367 [9]	8	5	1	227386[8]
	7	4	2	104510 [9]	8	4	2	221564[8]
	7	3	3	107428 [9]	8	3	3	231626[9]
	7	2	4	114662 [10]	8	2	4	250888[9]
	7	1	5	125061 [11]	8	1	5	276889[10]
	7	0	6	137709 [12]	8	0	6	307211[12]

닌 빈 분류 기반 벡터 데이터 압축 기법에 관한 실험 결과를 아래 Table 4와 Fig. 7에 나타내었다.

Table 4와 Fig. 7에서 보여 지듯이, 빈 분류 기반 벡터 데이터 압축 기법은 원본 벡터 데이터의 형상 변화 없이 압축률 면에서도 DP 단순화 알고리즘과

최대 압축률을 비교하였을 때 폴리라인의 경우 28~39%, 폴리곤의 경우 37~75%의 큰 차이를 보였다. 마지막으로 본 논문에서 제안한 압축 기법의 실험은 앞서 실시한 DP 단순화 알고리즘의 실험을 통해 얻 어낸 결과들 중 어느 정도의 꼭지 점을 제거하면서도



Fig. 7. Compressed map by vector data compression based on bin classification; (a) Polyline layer, (b) Polygon layer for Changwon and (c) Polyline layer, (d) Polygon layer for Daegu.

형상에 시각적으로 크게 변화를 주지 않는 정밀도를 얻기 위해 임계치를 2로 고정하고 빈 분류 기반 벡터 데이터 압축 기법은 각 빈에 할당되는 비트를 가변하면서 실험 하였다. 아래 Table 5와 Fig. 8에 최적의 압축 효율을 나타낸 결과를 비교 설명하였다.

최종적으로 대구시와 창원시의 각 레이어들에 대해 DP 단순화 기법 및 빈 분류 기반 벡터 데이터 압축 기법과 제안한 기법의 실험 결과를 비교하여 Table 6으로 나타내었다. 본 실험의 결과로 대구시의 경우, 폴리라인의 압축률은 DP만을 실행 하였을 때

Table 5. Compression performance of the proposed method.

Layer	Daegu					Changwon				
	DP	Bits for CB	Bits for DB	Bits for AB	Compressed Data	DP	Bits for CB	Bits for DB	Bits for AB	Compressed Data
Polyline	2	8	6	0	116222 [8]	2	8	6	0	334965[5]
	2	8	5	1	106546 [9]	2	8	5	1	298494[4]
	2	8	4	2	99317 [6]	2	8	4	2	273826[4]
	2	8	3	3	94407 [6]	2	8	3	3	264378[4]
	2	8	2	4	91852 [6]	2	8	2	4	269255[4]
	2	8	1	5	92280 [6]	2	8	1	5	284584[4]
	2	8	0	6	95405 [6]	2	8	0	6	307058[5]
Polygon	2	7	6	0	110033 [9]	2	8	6	0	216767[8]
	2	7	5	1	98885 [8]	2	8	5	1	193639[7]
	2	7	4	2	91636 [8]	2	8	4	2	180352[7]
	2	7	3	3	91845 [8]	2	8	3	3	181614[7]
	2	7	2	4	96022 [8]	2	8	2	4	191048[7]
	2	7	1	5	103129 [9]	2	8	1	5	206639[8]
	2	7	0	6	112432 [9]	2	8	0	6	226400[8]

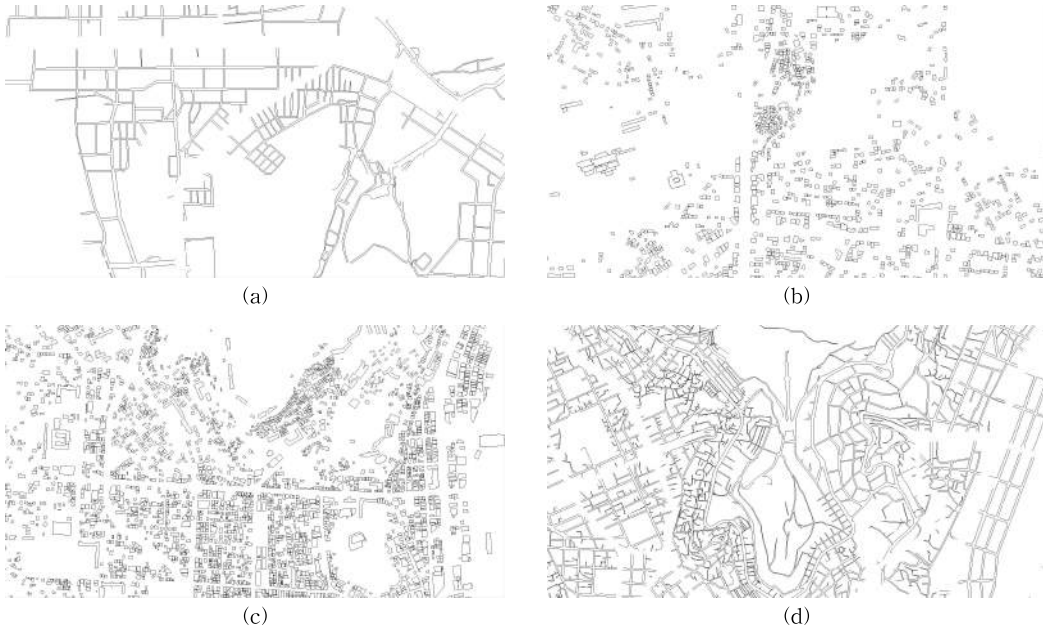


Fig. 8. Compressed vector map data by vector data compression based on bin classification; (a) polyline layer, (b) polygon layer for Changwon, and (c) polyline layer, (d) polygon layer for Daegu.

Table 6. Compression performance of proposed method and conventional methods.

Map (Daegu)	Compressed Data [Compression Ratio]			Map (Changwon)	Compressed Data [Compression Ratio]				
	DP	BCVC	Proposed		DP	BCVC	Proposed		
Layer	Polyline	794000[51]	178402 [12]	116222 [8]	Layer	Polyline	2452064[39]	541771[8]	334965[5]
		693888[47]	162496 [11]	106546 [9]			2001552[32]	501782[8]	298494[4]
		649168[44]	159488 [10]	99317 [6]			1805152[29]	501002[8]	273826[4]
		621712[42]	154052 [10]	94407 [6]			1669600[27]	533275[8]	264378[4]
		604128[41]	158677 [10]	91852 [6]			1574800[25]	588243[9]	269255[4]
		591648[40]	168877 [11]	92280 [6]			1500304[24]	658325[10]	284584[4]
		580960[40]	183008 [12]	95405 [6]			1440192[23]	737574[12]	307058[5]
		986288[87]	119882 [10]	110033 [9]			1855248[73]	247563[9]	216767[8]
	Polygon	911008[80]	109367 [9]	98885 [8]		Polygon	1637152[64]	227386[8]	193639[7]
		856096[75]	104510 [9]	91636 [8]			1544272[60]	221564[8]	180352[7]
		806384[71]	107428 [9]	91845 [8]			1466240[57]	231626[9]	181614[7]
		757696[67]	114662 [10]	96022 [8]			1371920[54]	250888[9]	191048[7]
		702352[62]	125061 [11]	103129 [9]			1269104[50]	276889[10]	206639[8]
		644912[57]	137709 [12]	112432 [9]			1168624[46]	307211[12]	226400[8]

에는 40~51%의 압축률을 가졌으며, 폴리곤의 압축률은 57~87%이며, BCVC 기법만을 사용하였을 때는 폴리라인인 경우 압축률은 10~12%, 폴리곤인 경우는 9~12%였다. 또한 제한한 기법을 사용하였을 경우 폴리라인 압축률은 6~9%이고, 폴리곤인 경우

8~9%의 높은 압축률을 가지는 것으로 나타났다. 그리고 창원시의 경우, 폴리라인의 압축률은 DP만을 실행 하였을 때에는 23~39%의 압축률을 가졌으며, 폴리곤의 압축률은 46~73%이며, BCVC 기법만을 사용하였을 때는 폴리라인인 경우 압축률은 8~12%,

폴리곤인 경우는 8~12%였다. 또한 제안한 기법을 사용하였을 경우 폴리라인 압축률은 4~5%이고, 폴리곤인 경우 7~8%의 높은 압축률을 가지는 것으로 나타났다. 따라서 제안한 기법이 기존의 기법보다 월등히 우수한 압축률을 가지는 것으로 확인하였다. 대용량의 벡터 맵 데이터인 창원시의 경우, 폴리라인의 압축률은 16~34%, 폴리곤의 압축률은 30~65% 정도의 차이를 보였다. BCVC 기법만을 이용하였을 때와 제안 기법을 이용하였을 때는 원본 벡터 데이터와 육안으로 확인 하였을 때 정밀도 면에서 크게 차이가 없으면서 압축률 면에서도 대구시의 벡터 맵 데이터를 사용한 실험에서 폴리라인 레이어의 경우 3%에서 4% 정도의 압축률 차이를 보였고, 폴리곤 레이어의 경우 1%에서 3% 정도의 압축률 차이를 보였다. 또한 창원시의 벡터 맵 데이터를 사용하여 실험한 결과를 비교 분석하였을 때 폴리라인의 경우 2%에서 4%, 폴리곤의 경우 1% 정도의 압축률 차이를 보였다.

이를 통해 본 논문에서 제안한 압축 기법의 경우 레이어 내 포함된 오브젝트의 좌표 점 개수에 압축률이 크게 달라지고 폴리곤 레이어의 경우 DP 근사화 알고리즘으로 제거한 꼭지 점의 개수가 폴리라인 레이어에 비해 현저히 줄어들기 때문에 이 같은 경우에도 역시 제대로 된 압축률을 보장 받지 못하는 DP 근사화 기법과는 달리 벡터 맵 데이터의 용량과 폴리라인, 폴리곤 레이어와 같은 레이어의 상태에 영향 받지 않고 일정한 압축률을 가지는 것을 확인하였다.

5. 결 론

본 논문에서는 벡터 데이터에 대해 DP 단순화 기법과 빈 분류 기반의 벡터 데이터 압축 기법을 제안하여 벡터 맵 데이터의 활용에 있어 용량의 문제로 발생하는 문제들을 해결하였다. 본 제안 기법은 요구되는 정밀도에 따라 임계값을 적절히 조절하여 DP 벡터 데이터 단순화 기법을 실시한 후 빈 분류 기반 벡터 데이터 압축 기법을 실시함으로써, 정밀도 면에서나 시각적 면에서도 벡터 맵 형상을 크게 손상시키지 않으면서 최적 압축률을 갖는 우수한 성능을 가짐을 확인하였다. 그리고 고용량의 벡터 맵 데이터와 본 논문의 비교적 저용량의 벡터 맵 데이터에서의 실험을 통해 DP 단순화 알고리즘 및 빈 분류 기반 벡터 데이터 압축 기법과 본 논문에서 제시한 압축 기법간의 비교를 해본 결과 DP 단순화 알고리즘의

경우는 고용량의 벡터 맵 데이터의 경우 준수한 압축률을 보여주었지만, 저용량의 벡터 맵 데이터의 경우에는 50%대 정도 압축률을 보임을 확인하였다. 이는 레이어가 포함하고 있는 오브젝트의 좌표 점의 개수 차이에 의한 것으로, DP 단순화 알고리즘으로 제거할 수 있는 좌표점이 많지 않을 경우 DP 단순화 알고리즘으로는 데이터 압축의 큰 효과를 기대하기는 어렵다는 것이 나타났다. 반면에 빈 분류 기반 벡터 데이터 압축기법의 경우 고용량과 비교적 저용량의 벡터 맵 데이터에서도 10%대 미만의 뛰어난 압축률을 보장 받을 수 있었다. 이를 통해 GIS 분야에 있어 본 논문에서 제시한 압축 기법으로 고정밀을 요구하지 않는 모바일 플랫폼 또는 웹 기반의 일반 사용자들을 대상으로 한 분야의 서비스를 제공하는데 있어 큰 이점이 있으리라 판단된다.

REFERENCE

- [1] GIS Introduction-1 Period : GIS Introduction, http://gis.seoul.go.kr/GisWebDataStore/Gis_Edu/html/S0101/SGIS-HTML.jsp?sgis=0101 (accessed Aug., 24, 2014).
- [2] A. Alesheikh, H. Helali, and H. Behroz, "WebGIS : Technology and Its Applications," *Proceeding of Symposium on Geospatial Theory, Processing and Application*, pp. 1-9, 2002.
- [3] S. Park, "3D GIS and Map Data Technology Trends," *National IT Industry Promotion Agency*, Vol. 1497, 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 Carcature," *The International Journal for Geographic Information and Geovisualization*, Vol. 10, No. 2, pp. 112-122. 1973.
- [5] S. Shekhar, Y. Huang, J. Djughash, and C. Zhou, "Vector Map Compression: A Clustering Approach," *Proceeding of the 10th ACM International Symposium on Advances in Geographic Information Systems*, pp. 74-80, 2002.
- [6] B. Yang and Q. Li, "Efficient Compression of

- Vector data Map based on a Clustering Model,” *Geo-Spatial Information Science*, Vol. 12, Issue 1, pp. 13-17, 2009.
- [7] B. Jang, *Compression and Cryptomarking Techniques for Secure Transmission and Storage of Vector Map Data*, Doctor’s Thesis of Pukyong National University of Interdisciplinary Program of Information Security, 2013.
- [8] S. Shekhar, Y. Huang, and J. Djugash, “Dictionary Design Algorithms for Vector Map Compression,” *Proceeding of IEEE Data Compression Conference*, pp. 471, 2002.
- [9] D. Im, B. Jang, S. Lee, S. Kwon, and K. Kwon, “Hybrid Polyline Simplification for GIS Vector Map Data Compression,” *Journal of Korea Multimedia Society*, Vol. 16, No. 4, pp. 418-429, 2013.
- [10] J. Park, S. Lee, E. Lee, and K. Kwon, “Efficiency Comparison of GIS Vector Map Data Compression by Vector Map Components,” *The Institute of Electronics and Information Engineers*, Vol. 37. No. 1, pp. 153-154, 2014.
- [11] P.N. Giao, G. Kwon, S. Lee, and K. Kwon, “Selective Encryption Algorithm based on DCT for GIS Vector Map,” *Journal of Korea Multimedia Society*, Vol. 17. No. 7, pp. 769-777, 2014.
- [12] GIS File Format, Raster Data, http://en.wikipedia.org/wiki/GIS_file_formats, (accessed Aug., 24, 2014).
- [13] S. Han, *GIS Theory and Practice-Arc GIS Desktop ver.10*, Gumi Library, 2013.
- [14] D. Salomon, G. Motta, and D. Bryant, *Data Compression: the Complete Reference*, Springer Science & Business Media, 4th edition, 2007.
- [15] S. Shekhar, Y. Huang, J. Djugash, and Z. Changqing, “Vector Map Compression: A Clustering Approach,” *Proceeding of the 10th ACM International Symposium on Advance in GIS*, pp. 74-80, 2002.
- [16] D. Lee, *Vector Data Compression using a Clustering Method*, Master’s Thesis of Inha University, 2005.



박진혁

2013년 2월 동의대학교 컴퓨터 소프트웨어공학과 학사 졸업(공학사)
 2013년 3월~현재 부경대학교 IT 융합응용공학과 석사과정
 관심분야: 멀티미디어 신호처리, 데이터 압축



장봉주

2002년 2월 부산외국어대학교 전자공학과 학사 졸업(공학사)
 2004년 2월 부산외국어대학교 전자컴퓨터공학과 석사졸업(공학석사)

2004년~2006년 ㈜아크로엠 연구원
 2007년 9월~2013년 8월 부경대학교 정보보호협동과정 박사과정
 2013년 7월~현재 연구원, 한국건설기술연구원
 관심분야: 멀티미디어 신호처리/압축/정보보호, 레이더 신호처리 및 응용



권오준

1986년 2월 경북대학교 전자공학과 학사 졸업(공학사)
 1992년 2월 충남대학교 전산학과 석사졸업(이학석사)
 1998년 2월 포항공과대학교 전자계산학과 박사졸업(공학박사)

1986년~2002년 한국전자통신연구원 선임연구원
 2000년~현재 동의대학교 컴퓨터소프트웨어공학과 부교수
 관심분야: 컴퓨터네트워크, 정보보호, 무선 인터넷, 패턴 인식, 인공지능망



정재진

1990년 2월 성균관대학교 독어독문학 학사
 1996년 2월 연세대학교 행정학(행정학석사)
 2004년 8월 성균관대학교 경영학(경영학박사)

2005년 9월~2009년 8월 동의대학교 디지털콘텐츠 공학과 교수
 2009년 9월~현재 단국대학교 공과대학 응용컴퓨터공학과 부교수
 관심분야: IoT/M2M, Future Internet Technology, Web Technology, Smart Application Development



이석환

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

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



권기룡

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

2000년~2002년 University of Minnesota, Post-Doc.
 1996년~2006년 부산외국어대학교 디지털정보공학부 부교수
 2006년~현재 부경대학교 IT융합응용공학과 교수
 2011년~2012년 Colorado State University 연구교수
 2015년~현재 한국멀티미디어학회 회장
 관심분야: 멀티미디어정보보호, 영상처리, 멀티미디어 통신 및 신호처리