

(19)대한민국특허청(KR) (12) 등록특허공보(B1)

(51) 。 Int. Cl.

G06T 1/00 (2006.01)

H03M 7/30 (2006.01)

H04N 7/24 (2006.01)

(45) 공고일자

2006년07월10일

(11) 등록번호

10-0598134

(24) 등록일자

2006년06월30일

(21) 출원번호

10-2005-0024399

(65) 공개번호

(22) 출원일자

2005년03월24일

(43) 공개일자

(73) 특허권자

인하대학교 산학협력단
인천 남구 용현동 253 인하대학교

(72) 발명자

박수홍
서울 양천구 신정동 목동아파트 1403동 106호

이동현
인천 남구 용현4동 83-1 4층

전우제
서울 서초구 서초3동 1546-5 현대 하이즈 나동 302호

(74) 대리인

이원희

(56) 선행기술조사문헌

JP2002228467 A

JP2004093671 A

KR1020040004611 A

논문

* 심사관에 의하여 인용된 문헌

심사관 : 이상헌

(54) 케이평균 군집화를 이용한 벡터 데이터 압축 방법 및시스템

요약

본 발명은 K평균 군집화 및 사전기반 압축 기법을 통해 공간 데이터가 차지하는 저장 공간의 크기를 효율적으로 줄일 수 있도록 하는 K평균 군집화를 이용한 벡터 데이터 압축 방법 및 시스템을 제공한다.

이러한 본 발명은 각 기하 객체를 시작점 좌표와 디퍼런셜 벡터의 나열로 분리한 후, 상기 디퍼런셜 벡터를 길이와 각도로 각각 분리하고, 분리된 길이와 각도에 대하여 각각 K평균 군집화를 수행하여 군집 중심을 엔트리로 갖는 사전을 제작하며, 각각의 길이와 각도에 대하여 근사하여 사전의 엔트리를 가리키는 포인터로 변환하는 과정을 거쳐 모바일 환경에 적합한 벡터 데이터 압축 방법을 제공함을 특징으로 한다.

대표도

도 4

색인어

모바일, 디퍼런셜 벡터, K평균 군집화, 사전기반압축, 벡터 데이터

명세서

도면의 간단한 설명

도 1은 사전기반 압축의 설명을 위한 라인 스트링의 예를 나타낸 도.

도 2는 도 1의 사전 구성도.

도 3은 일반적인 군집화 과정을 나타낸 흐름도.

도 4는 본 발명에 따른 K평균 군집화를 이용한 벡터 데이터 압축 방법의 전체적인 흐름도.

도 5는 본 발명에서 인접 시작점이 같은 군집이 된 경우에 대한 설명도.

도 6은 본 발명을 구현하기 위한 하드웨어 구성도.

도 7은 본 발명에서의 실험 데이터를 나타낸 도.

도 8은 도 7의 실험 데이터에 대한 디퍼런셜 벡터를 나타낸 도.

도 9는 시작점을 제외한 디퍼런셜 벡터의 확대도.

도 10은 본 발명의 실험에서 조합 가능한 벡터 사전을 나타낸 도.

도 11은 종래 압축 방법에 따른 벡터 사전을 나타낸 도.

도 12는 본 발명에 대한 결과 분석을 위한 표.

도 13 내지 도 15는 본 발명에서 각각 길이와 각도의 사전 크기를 늘려가면서 실험한 결과를 나타낸 도.

도 16는 종래 압축 방법에 대한 연구 결과 분석표.

도 17은 압축률과 위치오차의 관계를 나타낸 그래프.

<도면의 주요부분에 대한 부호의 설명>

100 : 입력부 200 : 디퍼런셜 벡터 추출부

300 : 압축부 400 : 제어부

500 : 저장부 510 : 시작점 좌표 DB

520 : 길이사전 DB 530 : 각도사전 DB

540 : 길이사전 포인터 DB 550 : 각도사전 포인터 DB

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

본 발명은 모바일 환경에서 사용되는 벡터 형태의 데이터를 압축하는 방법 및 시스템에 관한 것으로, 특히 K평균 군집화 및 사전기반 압축 기법을 통해 공간 데이터가 차지하는 저장 공간의 크기를 효율적으로 줄일 수 있도록 하는 K평균 군집화를 이용한 벡터 데이터 압축 방법 및 시스템에 관한 것이다.

최근 이동전화, PDA, 카 네비게이션 단말기 등 모바일 기기의 사용이 빠른 속도로 증가하고 있다. 이들 모바일 기기에서는 경로탐색, 지도 서비스 등을 위해서 공간 데이터의 사용이 필수적으로 요구된다.

그러나 이러한 모바일 기기는 여전히 데스크톱 환경에 비하여 제한적인 연산 수행능력과 저장 공간의 한계가 존재한다. 따라서 레스터 데이터에 비해서 상대적으로 적은 저장 공간을 차지하는 벡터 데이터조차도 여전히 큰 부담이 된다.

모바일 환경에서 사용되는 벡터 형태의 데이터에는 실제로 거리 측정 또는 경로탐색을 위한 데이터와 배경으로 사용되는 맵 데이터가 있다. 맵 데이터의 경우에는 약간의 위치 오차가 포함되더라도 그 오차정도가 눈으로 구분할 수 없을 정도의 수준이라면 받아들여 질 수 있다.

현재까지 데이터 압축 분야에서 벡터 데이터에 대한 압축 기법은 레스터 데이터에 비하여 연구가 다양하지 못하였다. 대표적인 사전기반의 벡터 데이터 압축에 관한 연구로는 2000년에 발표한 "Design Algorithms for Vector Map Compression(D. Salomon. 'Data Compression: the Complete Reference.'Springer-Verlag, 2nd edition, 2000.)"가 있는데, 이 연구에서는 벡터 데이터를 각각의 디퍼런셜 벡터로 쪼개어서 FHM(Fibonacci, Huffman, and Markov)방법을 이용하여 미리 제작된 사전에 근사화하는 방법을 통하여 데이터량을 줄이는 방법을 제시한다. 이 연구에서는 미리 제작된 사전을 사용하는 방식으로 사전을 제작하는 과정을 생략할 수 있지만 사전이 데이터의 특성을 반영할 수 없으므로 복원시에 많은 위치오차를 갖는다.

또 다른 연구로는 데이터 특성을 찾아낼 수 있는 데이터 마이닝 기법 중 K평균 군집화 기법을 적용하여 사전을 제작하는 방법에 대하여 연구한 "Vector Map Compression: A Clustering Approach(Shashi Shekhar, Yan Huang, Judy Djughash, Changqing Zhou, 'Vector Map Compression: A Clustering Approach', Proceeding of the tenth ACM international symposium on Advances in geographic information systema, 2002, 74-80)"가 있다. 이 연구의 결과는 FHM방식을 이용하는 것보다 복원시 위치 정확도 측면에서 더 좋은 성능을 보이지만 실제 데이터에 적용했을 때 여전히 만족할 만한 위치 정확도를 얻을 수 없는 단점이 있었다.

발명이 이루고자 하는 기술적 과제

본 발명은 이러한 점을 감안한 것으로, 본 발명의 목적은 각 기하 객체의 상대적인 위치를 표현하는 디퍼런셜 벡터(Differential Vector)를 길이와 각도로 나누어 각각을 K평균 군집화 기법을 적용한 군집화로 사전을 제작함으로써 벡터 데이터에 대하여 최소의 연산을 통해 구분하기 어려운 정도의 위치 정확도 손실 차이를 가지면서 향상된 압축률을 얻을 수 있도록 한 K평균 군집화를 이용한 벡터 데이터 압축 방법 및 시스템을 제공함에 있다.

발명의 구성 및 작용

상기 목적을 달성하기 위한 본 발명에 따른 K평균 군집화를 이용한 벡터 데이터 압축 방법은, 각 기하 객체를 시작점 좌표와 디퍼런셜 벡터의 나열로 분리하는 단계; 상기 디퍼런셜 벡터를 길이와 각도로 분리하는 단계; 상기 분리된 길이와 각도에 대하여 각각 K평균 군집화를 수행하여 군집의 평균값을 엔트리로 갖는 사전을 제작하는 단계; 각각의 길이와 각도에 대하여 근사화하여 사전의 엔트리를 가리키는 포인터로 변환하는 단계;를 포함하는 것을 특징으로 한다.

또한, 본 발명의 K평균 군집화를 이용한 벡터 데이터 압축 시스템은, 입력부를 통해 입력되는 각 기하 객체에 대한 데이터로부터 각 기하 객체를 시작점 좌표와 디퍼런셜 벡터의 나열로 분리하는 디퍼런셜 벡터 추출부; 상기 디퍼런셜 벡터 추출부에서 추출된 디퍼런셜 벡터를 길이와 각도로 각각 분리하여 K평균 군집화를 적용하여 길이와 각도 각각에 대한 사전을 제작하고, 제작된 사전을 이용하여 각 개체를 개체가 속하는 군집의 대표 값을 가리키는 포인터로 변환하여 압축하는 압축부; 상기 디퍼런셜 벡터 추출부로부터의 시작점 좌표와 압축부로부터의 길이사전, 각도사전, 길이사전 포인터, 각도사전 포인터 등이 저장되는 각각의 데이터 베이스를 갖는 저장부; 및 상기 각부를 제어하는 제어부;로 구성되는 것을 특징으로 한다.

이하, 본 발명을 첨부된 도면을 참조로 하여 보다 상세하게 설명한다. 단, 하기 실시예는 본 발명을 예시하는 것일 뿐 본 발명의 내용이 하기 실시예에 한정되는 것은 아니다.

우선, 본 발명에 적용되는 사전기반 압축 기법 및 K평균 군집화(K-Means Clustering) 기법에 대하여 살펴본다.

본 발명에서 사용될 사전 기반(Keith C. Clarke, 1990)의 접근 방법은 공간 데이터만을 위하여 개발된 접근 방법은 아니다. 여러 다른 분야에 적용되어 사용되고 있는 기술로, 본 발명에서는 벡터 데이터 압축 방법을 사전 기반의 접근 방법을 사용하여 설계한다.

사전 기반의 접근 방법은 표현 하고자 하는 데이터를 엔트리로 하는 사전을 제작한 후, 실제 데이터에는 값(value)을 저장하는 것이 아닌 사전의 엔트리를 가리키는 포인터의 나열로서 데이터를 표현하는 방법이다.

이를 벡터 데이터에 대하여 적용하면 다음과 같다. 하나의 라인 스트링(LineString)이 도 1과 같이 2차원 공간상에 존재할 때, 이를 저장할 수 있는 방법은 여러 가지가 있다.

OGC의 표준(OGIS, 1999)에서 제시하는 방법으로는 타입 정보, 전체 포인트의 수 등과 함께 8바이트 크기를 갖는 x와 y의 배열을 이용하여 실제 좌표를 저장하고, 표현 방법으로는 'LINESTRING (60 120, 100 180, 200 100, 260 140, 160 160, 220 240)'과 같은 방법을 이용하여 표현한다. 이것을 사전 기반의 방법으로 저장하는 방법은 다음과 같다. 우선 LineString이 갖는 개별적인 점 좌표들을 엔트리로 갖는 도 2의 표와 같이 사전을 제작한다.

도 2의 표에서는 LineString이 가지고 있는 모든 점의 좌표의 중복을 제거하여 구성된 것이다. 그리고 LineString을 표현하는데에는 실제 좌표를 넣는 것이 아닌, 사전에 구성된 엔트리를 가리키고 있는 포인터의 배열로 나타낸다. 따라서 다음과 같이 '(1, 2, 3, 4, 5, 6)'으로 저장되어 있는 구조이다. 실제 이렇게 저장된 공간 객체를 사용하고자 할 때는 객체 좌표를 대신하고 있는 배열이 가리키는 사전의 엔트리를 검색하여 좌표로 표현하게 된다.

이러한 데이터 구조를 사전 기반의 접근이라고 한다. 위의 예와 같이 단일 객체에 대하여 사전 기반 접근법을 이용하여 데이터 구조를 만든다 할지라도 전체 데이터에 대한 크기는 줄어들지 않는다. 사전을 구성하는 엔트리에 엔트리를 가리키는 포인터의 크기가 추가되어 전체 데이터량은 단순히 좌표를 기록하는 방식보다 포인터의 크기만큼 데이터량이 늘어난다. 본 발명에서는 다수의 공간 객체에 대하여 유사한 형태의 중복이 많이 일어나는 점을 엔트리로 제작하여 사전을 제작하여 데이터의 양을 줄이는 방법에 대하여 제시하게 된다.

다음은 K평균 군집화에 대하여 살펴본다.

군집분석은 대규모의 데이터베이스에서 의미 있는 정보를 추출해 분석하는 데이터 마이닝 기법 중 한 가지이다. 여기에서 군집이라는 것은 유사한 데이터들의 집합이다(Jiawei Jan, Micheline Kanber, 2000). 즉, 하나의 군집에는 유사한 데이터들이 포함되고 다른 군집에는 유사도가 떨어지는 데이터들의 집합이 되도록 만들어 주는 것이 군집화이다. 따라서 군집분석은 서로 섞여있는 이질적인 대상을 유사도(Similarity)에 의해 몇 개의 균질적인 군집(Cluster)으로 분류하는 분석방법이다.

군집화 과정은 도 3과 같이 4단계의 과정으로 이루어진다.

첫 번째 변수 측정의 단계에서는 개체들을 군집화 하는데 이용될 수 있는 각 개체의 특성을 측정하는 변수들을 구한다. 즉, n개의 개체에 대하여 m개의 변수를 측정한다. m은 개체가 가지고 있는 변수들 중 실제 군집화에 사용할 변수들의 수가 된다. 두 번째 단계에서는 유사성을 측정하는 과정인데, 측정한 m개의 변수를 이용하여 전체 개체들 사이의 거리(Distance) 또는 비유사성(Dissimilarity)을 계산하는 과정이다. 개체들 사이의 거리를 측정하는 방법은 군집화 방법에 따라서 다르게 계산될 수 있다. 이렇게 모든 개체들 사이의 거리를 계산하여 비유사성을 나타내는 거리 행렬을 만든다. 비유사성은 값이 작을수록 두 개체사이의 거리가 가깝다. 개체 사이가 가깝다는 것은 개체들에서 군집화를 위해 측정한 m개의 변수들이 유사한 특성을 갖는다는 것이다. 세 번째 단계에서는 선정된 군집화 방법으로 계산된 비유사성 행렬을 이용하여 거리가 가까운 개체들을 하나의 군집으로 묶어주는 과정이다. 마지막 분석의 단계는 각 군집의 성격, 상호 관계를 판단하는 과정이다. 이 과정은 군집화 과정이 완료된 결과를 통하여 개별 군집이 어떠한 의미를 갖는지, 군집들 사이의 관계가 어떠한지를 판단하는 분석과정이다.

군집분석에는 크게 계층적 군집화 기법과 비 계층적 군집화 기법으로 나눌 수 있다.

비 계층적 군집화 기법은 주어진 n 개의 개체를 선정된 분할 방법을 이용하여 k 개의 군집으로 분할하는 방법이다. 비 계층적 군집화 방법에는 몇 가지 조건이 필요한데 첫 번째는 항상 k 가 n 보다 작아야 한다는 것이다. 최종 군집의 수가 개체의 수보다 작아야 하고 각 군집에는 최소한 한 개 이상의 개체가 포함되어야 한다. 마지막으로 각 개체는 하나의 군집에만 속해야 한다는 것이다. 비 계층적 군집화 기법 중 대표적인 것으로는 K평균 군집화 기법이 있다.

계층적 군집화 기법은 주어진 개체 집합을 계층적으로 분해하는 군집화 기법이다. 계층적 군집화 기법에는 세부적으로 상향식(Bottom-Up)방법과 하향식(Top-Down)의 두 가지 방법이 있다. 상향식 방법은 각각의 개체를 다른 군집으로 할당하고 유사성을 측정하여 유사성이 높게 나타나는 군집을 하나로 병합하는 방법으로 최종 군집이 한 개로 병합되거나 주어진 조건을 만족할 때 까지 반복하는 방법이다. 하향식 방법은 상향식과 반대로 모든 개체를 하나의 군집으로 시작하여 하나의 군집이 하나의 개체로만 이루어지거나 주어진 조건을 만족할 때 까지 분할하는 과정을 반복하여 군집화 하는 방법이다. 대표적인 계층적 군집화 기법에는 단일 결합법, 완전 결합법, 평균 결합법, Ward 법 등이 있다. 계층적 방법은 초기의 부적절한 병합이나 분할이 일어났을 경우 회복할 수 없다는 단점이 있다.

본 발명에 적용될 K평균 군집화 기법은 MacQueen(MacQueen, 1967)에 의해 개발된 비 계층적 군집법이다. 즉 n 개의 개체를 k 개의 군집으로 유사성이 높은 것을 하나의 군집으로 묶어주는 분할하는 방법이다.

첫 번째 단계는 입력 인자로 군집 중심 K 개를 입력받고 입력받은 K 만큼의 군집 중심을 선정하는 과정이다. 초기 군집중심은 임의로 선정된다. 다음 단계는 각각의 개체를 선정된 군집에 할당하는 과정이다. 이 단계에서는 유사성을 계산하여 유사성이 가장 높은 군집으로 할당한다. 유사성을 계산하는 방법은 여러 가지가 있는데, K평균 군집화 기법에서는 각 군집에 할당된 개체들의 평균을 이용하여 새로운 군집의 중심값을 계산한다. 2차원 이상의 변수에 대한 군집화에 대해서는 평균값 대신 무게중심(Centroid)값을 이용하여 군집 중심을 재계산 한다. 이러한 과정을 주어진 조건을 만족하거나 군집 중심의 이동이 없어질 때 까지 반복한다.

군집화 과정 중 두 번째 단계의 개체를 가장 유사한 군집에 할당 하는 방법은 여러 가지가 있다. Minkowski 거리, Euclidean 거리, 표준화 거리, Mahalanobis 거리 등이 개체 사이의 유사성을 측정하는 도구로 사용이 가능하다. K평균 군집화 기법을 적용하는 데에 가장 많이 사용되는 유사도 측정 도구로는 Euclidean 거리이다. 본 발명에서는 2차원 공간상의 기하학적인 좌표를 이용하여 연산을 해야 하기 때문에 Euclidean 거리를 유사도 측정의 도구로 이용하였다. Euclidean 거리를 계산하는 식을 n 차원에 공간에서 적용가능 하도록 확장한 식은 다음과 같다.

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

$$a = (a_1, a_2, a_3, \dots, a_n)$$

$$b = (b_1, b_2, b_3, \dots, b_n)$$

이 거리 계산을 이용하면 개체와 군집의 유사성을 측정한 후 개체와 유사성이 가장 높은 군집으로 개체를 할당하여 군집을 재계산 한다. 이러한 과정을 재 할당이 일어나지 않거나 주어진 조건을 만족할 때 까지 반복하여 최종 결과를 얻는다.

K평균 군집화 기법은 전체 데이터의 내부적인 구조에 대한 사전 지식이 없어도 의미있는 정보를 찾아낼 수 있다는 장점이 있다(Anil K. Jain, M. Narasimha Murty, 1999). 또한, 관찰 값과 군집중심 사이의 거리 관계를 데이터의 형태에 맞게 정의한다면 대부분의 형태의 데이터에 적용이 가능하다. 초기의 잘못된 군집에 개체가 속하더라도 반복을 통하여 타당한 군집으로 재 할당이 이루어진다. K평균 군집화 기법은 초기값 K 이외에 다른 사전정보를 요구하지 않으므로 적용하기에 용이하다. 하지만 초기값 K 를 적합하게 선정하지 못한다면, 만족할만한 군집화 결과 얻을 수 없다. 또 비유사성 거리를 정의하는 과정에서 여러 가지 자료유형의 측정 척도(Measurement scale)(Jay lee, David w. s. Wong, 2000)가 상이한 경우 하나의 거리로서 정의하는 것이 어렵다. 또한 사전에 주어진 군집 목적이 없으므로 결과 해석에 어려움이 따르기도 한다.

본 발명은 모바일 환경에서 배경으로 사용되는 맵 데이터에 대하여 상기와 같은 K평균 군집화 기법을 이용한 손실 압축 기법을 이용함과 더불어 시전기반의 접근방법을 통해 전체 데이터에 대한 압축률 보다는 실제 사용가능하도록 데이터의 손실률을 최소화하는 방향으로 압축하며, 압축률과 데이터 손실에 따르는 위치 정확도 관계에서 위치 정확도를 높일 수 있도록 하며, 이러한 개념을 바탕으로 하는 본 발명을 도 4의 동작 흐름도를 참조로 설명한다.

먼저, 벡터 데이터를 압축하는 데에 K평균 군집화 기법을 적용하기 위해서는 기하 객체를 시작점 좌표와 디퍼런셜 벡터의 나열로 분리하는 과정이 필요하다.

디퍼런셜 벡터는 현재 점의 위치를 표현하기 위하여 이전 점의 좌표 혹은 해당 객체가 시작하는 점의 좌표와의 차이를 이용하여 상대적인 위치를 표현하는 벡터이다. 본 발명에서는 개별 좌표의 해당 좌표와 이전 좌표의 차이를 이용하여 디퍼런셜 벡터를 추출한다(S100).

그리고 추출된 디퍼런셜 벡터를 길이와 각도로 분리하여 각각에 대하여 K평균 군집화를 적용하여 사전을 제작한다(S200,S300).

이때, 디퍼런셜 벡터의 길이와 각도에 대하여 K평균 군집화 기법을 이용하여 군집화를 수행하며, 군집내의 할당된 개체들을 이용하여 중심을 계산한다. 본 발명에서는 2차원의 디퍼런셜 벡터를 1차원의 길이와 각도로 분리하였으므로 각각의 평균값을 이용하여 중심을 계산한다. 이렇게 계산된 군집내의 대표 값을 엔트리로 하는 사전을 제작한다. 그리고 사전 기반 압축 기법으로 상기 제작된 사전을 이용하여 각 개체를 개체가 속하는 군집의 대표 값을 가리키는 포인터로 바꾸어 준다(S400).

이러한 일련의 과정을 통하여 압축된 데이터로는 공간 객체의 절대적인 위치를 표현하는 개체의 시작점 집합, 시작점을 기준으로 상대적인 위치를 표현하는 디퍼런셜 벡터의 길이와 각도의 사전, 또 이를 가리키고 있는 두 개의 포인터 배열 집합 등이다.

상기에서 공간 객체의 시작점의 경우 아무런 처리를 하지 않고 그대로를 저장한다. 디퍼런셜 벡터가 개별 점 데이터의 상대적인 위치를 표현한다면, 객체의 시작점은 좌표상의 절대적인 위치를 가지고 있는 값이다. 절대적인 위치를 표현하는 기준과 각각의 상대좌표를 나타내는 디퍼런셜 벡터가 함께 존재 해야만 공간 객체를 재구성 할 수 있다.

이러한 시작점까지도 함께 군집화를 적용한다고 할 때, 객체의 개수 즉 시작점의 개수보다 군집의 개수가 단 한 개라도 적게 된다면 도 5와 같은 두 개 이상의 폴리곤이 하나로 합쳐져 나타나게 된다.

객체의 절대적인 위치를 표현할 수 있는 좌표까지도 다른 상대적인 위치를 나타내는 디퍼런셜 벡터와 함께 군집화를 적용할 때 전체 객체의 수 이하로 군집의 수를 결정한다면 전체 데이터에 대한 위치 오차는 줄어들 수도 있지만 도 5와 같은 현상으로 인하여 데이터에 대한 왜곡이 심하게 되어 사용이 불가능한 데이터가 된다. 절대적인 위치를 표현하는 점까지 군집화를 적용하면서 군집의 수를 객체의 수 이상으로 늘려준다면 참조해야 하는 사전의 엔트리 수가 많아지게 된다. 사전의 엔트리 수가 늘어나게 되면 그에 따라서 사전을 가리키고 있는 포인터의 크기가 커져야 한다. 예를 들면 사전의 엔트리 수가 256개 일 경우에는 포인터의 크기가 최소 8비트로서 사전의 엔트리를 참조할 수 있는 반면 엔트리의 수가 512개로 늘어날 경우 9비트의 포인터가 필요하다.

또한, 본 발명에서 군집화를 적용하기 전에 디퍼런셜 벡터를 길이와 각도로 분리하는 이유는 다음과 같다.

첫 번째로는 디퍼런셜 벡터 추출 과정에서 각각의 벡터가 도로와 유사한 각도로 많이 분포하기 때문이다. 이러한 경우 유사한 값을 하나의 값으로 묶어주는 군집화 기법을 적용하였을 때 더 위치 정확도 측면에서 이익을 볼 수 있기 때문이다. 두 번째로는 사전을 형성하였을 때, 가질 수 있는 값의 경우의 수가 많아진다는 것이다. 2차원 공간에서 군집화를 하여 10개의 엔트리를 갖는 사전을 만들었다면 실제 가질 수 있는 값 역시도 10개 이다. 하지만 두 개의 인자로 나누어서 각각 5개의 엔트리를 갖는 사전 두 개를 제작하였다면, 실제 표현할 수 있는 값은 5×5 개가 된다. 따라서 더 적은 사전을 가지고도 더 많은 표현이 가능하다는 점에서 분리하였다. 마지막으로 높은 위치 정확도가 요구되는 맵 데이터를 손실 압축할 경우 군집의 수 K가 커지게 된다. 이러한 경우 압축 과정에서 요구되는 비용이 높아지는 이유에서 길이와 각도를 분리하여 군집화 과정을 거치게 된다.

도 6은 이러한 본 발명을 구현하기 위한 하드웨어 구성도를 도시한 것으로, 입력부(100)를 통해 기하 객체에 대한 데이터가 입력되면 디퍼런셜 벡터 추출부(200)에서 해당 기하 객체를 시작점 좌표와 디퍼런셜 벡터의 나열로 분리하게 되며, 압축부(300)에서 디퍼런셜 벡터를 길이와 각도로 각각 분리하여 K평균 군집화를 적용하여 길이와 각도 각각에 대한 사전을 제작하고, 제작된 사전을 이용하여 각 개체를 개체가 속하는 군집의 대표 값을 가리키는 포인터로 변환하여 압축을 행하게 되며, 이러한 과정에서 생성되는 시작점 좌표, 길이사전, 각도사전, 길이사전 포인터, 각도사전 포인터 등이 제어부(400)의 제어에 따라 저장부(500)의 시작점 좌표 DB(510), 길이사전 DB(520), 각도사전 DB(530), 길이사전 포인터 DB(540), 각도사전 포인터 DB(550) 등의 데이터 베이스에 각각 저장되게 된다.

한편, 상기와 같은 과정을 거쳐 압축이 완료된 데이터를 압축 과정의 역 과정을 통하여 원 데이터로 재구성할 수 있다. 이를 위해 모바일 기기에는 상기 압축 과정의 역과정을 수행하는 복원부를 구비하게 되며, 상기와 같이 압축된 데이터를 수신받아 이를 저장 공간에 저장하여 복원부를 통해 복원하게 된다.

즉, 모바일 기기의 복원부에서는 모바일 기기의 저장 공간으로부터 각각의 길이와 각도 사전을 가리키고 있는 포인터를 실제 길이와 각도를 갖도록 엔트리를 가져온다. 이 과정을 거치면 포인터의 배열 집합이 아닌 실제 길이와 각도의 배열 집합을 얻을 수 있게 된다 이렇게 얻어진 길이와 각도를 이용하여 계산하면 점의 상대적인 위치 좌표를 가지고 있는 디퍼런셜 벡터의 집합으로 재구성이 된다. 이렇게 만들어진 디퍼런셜 벡터와 공간 객체의 절대적인 위치 좌표를 갖는 객체의 시작점을 이용하여 원래의 공간 객체로의 재구성이 가능하다.

상기와 같은 본 발명의 실험예를 살펴본다.

연구지역은 서울시 양천구와 강서구 일대(8.8km x 10.1km)를 선정하였고, 실험에 사용된 데이터는 연구 지역의 1/1,000 수치지도에서 건물 코드를 갖는 지형지물을 추출한 데이터이다. 총 72016개의 폴리곤 데이터이며, 전체 점의 개수는 566607개로, 실험 데이터는 도 7에 도시한 바와 같다.

먼저, 실험 데이터에 대하여 군집기법을 적용하기 위한 디퍼런셜 벡터를 추출하였다. 연구에서 사용한 디퍼런셜 벡터는 현재 점 좌표에서 이전 점 좌표와의 차이를 이용하여 추출하고, 각 객체의 시작점의 경우에는 원점과의 차이를 이용하였다.

다음 과정은 이렇게 계산된 디퍼런셜 벡터에서 각 폴리곤 데이터의 시작점 벡터를 따로 저장하는 것이다. 도 8의 디퍼런셜 벡터 중심부분에 시작점을 제외한 나머지 디퍼런셜 벡터가 전체 점의 수에서 폴리곤수를 뺀 만큼 밀집되어 있는 것을 확인할 수 있다. 데이터가 시작점을 제외한 디퍼런셜 벡터의 평균 길이보다 커지는 넓은 지역을 포함할 경우 벡터의 길이와 각도로 군집화를 수행하더라도 좋은 결과를 얻을 수 없었다.

도 9는 시작점을 제외한 디퍼런셜 벡터의 확대도로, 이렇게 얻어진 결과를 각각 길이와 각도로 분리하여 K평균 군집화 기법 및 사전기반 압축기법을 적용하여 두개의 사전과 포인터를 얻는 과정으로서 압축이 된다.

실험에서 K평균 군집화 과정은 통계 패키지인 SPSS v10을 이용하여 수행하였다. 초기값 K를 거리와 각도 모두 256, 512, 1024개로 늘려가면서 군집화하였다. 도 10은 이와 같이 하여 생성된 거리와 각도의 사전으로 조합 가능한 경우의 수를 표현한 것이다.

도 10에서 좌로부터 각도, 길이의 사전 수가 각각 256, 512, 1024개로서 가질 수 있는 모든 경우의 수는 65536, 262144, 1048576개가 된다. 이는 도 9에서 나타나는 디퍼런셜 벡터가 도 10의 사전에 가장 가까운 값으로 근사화될 때 사전이 클수록 적은 오차를 포함할 수 있는 것이다.

도 11은 기존의 연구 방법으로 실험 데이터에 적용하였을 때 가질 수 있는 사전의 분포로, 도 11은 각각 좌로부터 256, 512, 1024개의 벡터를 갖는 사전을 표현한 것이다. 이는 동일한 데이터 량을 갖는 사전을 두가지 다른 방식으로 제작하였을 때 사전이 표현할 수 있는 벡터수의 차이이다.

각 디퍼런셜 벡터를 사전에 근사화하는 방법으로 압축된 데이터를 복원과정을 거쳐 원본과의 비교를 통해 데이터 손실에 따르는 위치 오차 정도, 최종 데이터 크기, 압축률을 계산해 보았다. 각각 사전 크기를 65536, 262144, 1048576개로 늘려가며 실험하였다. 압축된 데이터의 크기는 폴리곤의 시작점을 저장하는 부분, 디퍼런셜 벡터의 거리와 각도를 대표하는 두개의 사전, 사전을 가리키는 두개의 포인터의 다섯 부분을 합한 것이다. 또 압축률은

압축률(%) = ((원본 데이터 크기 - 압축 데이터 크기) / 원본 데이터 크기) X 100의 식으로 계산하였다. 원본 데이터의 크기는 9065712바이트이다.

도 12는 결과 분석을 위한 표로, 세 번의 실험에서 각각 위치 오차 수준이 상당히 작은 수준으로 나타나는 것을 볼 수 있다.

도 13, 도 14, 도 15는 각각 길이와 각도의 사전 크기를 늘려가면서 실험한 결과로, 도 13은 사전크기 256X256이고, 도 14는 사전크기 512X512이며, 도 15는 사전크기 1024X1024이다. 도 15의 경우 데이터 크기를 약 25%정도로 줄이면서 정확도를 유지하는 결과를 보였다.

본 발명에서 제시한 방법의 결과와 유사한 결과를 보이도록 기존 연구의 사전 크기를 조절하였다. 또한, 인접한 두 객체가 합쳐지지 않도록 시작점을 사전에 포함하도록 제작하여 실험하였다. 실험결과는 도 16의 표와 같다. 3번의 실험에서 사건의 크기를 72272, 82528, 73040개로 늘려가면서 실험을 하였다. 사건은 시작점 72016개를 포함하고 각각 256, 512, 1024개의 초기값을 이용하여 K평균 군집화를 통한 결과이다. 도 12과 도 16의 표에서 알 수 있듯이 본 발명의 압축률이 높음을 알 수 있다.

도 17은 압축률과 위치오차의 관계를 나타낸 것으로, 기존 연구의 결과와의 비교에서 비슷한 수준의 압축률을 보이는 경우 더 향상된 위치 정확도를 보이고 있다.

상기에서 살펴본 바와 같이, 본 발명은 하나의 객체 내에서 상대적 위치를 나타내는 디퍼런셜 벡터의 다양한 경우를 포괄하기 위하여 길이와 각도 두 인자로 분리하여 군집화를 수행하고, 좌표계 내에서 절대적인 위치를 정의하기 위하여 각각의 공간 객체에서 하나의 점, 본 발명에서는 시작점의 변형을 가하지 않고 그대로 저장함으로써 압축률을 떨어트리지 않고 위치 정확도를 향상시킬 수 있으며, 벡터 데이터에 대하여 손실을 최소화하면서 데이터의 크기는 약 25% 수준으로 낮출 수 있게 된다.

상술한 바와 같이, 본 발명의 바람직한 실시예를 참조하여 설명하였지만, 해당 기술 분야의 숙련된 당업자는 하기의 특허 청구범위에 기재된 본 발명의 사상 및 영역으로부터 벗어나지 않는 범위내에서 본 발명을 다양하게 수정 또는 변경하여 실시할 수 있다.

발명의 효과

이상에서 살펴본 바와 같이, 본 발명에 따른 K평균 군집화를 이용한 벡터 데이터 압축 방법 및 시스템은, 각 기하 객체의 상대적인 위치를 표현하는 디퍼런셜 벡터를 길이와 각도로 나누어 각각을 K평균 군집화 기법을 적용하여 군집화하여 각각 길이 및 각도 사건을 제작한 후, 사건기반 압축 기법을 적용하여 군집화에 의해 제작된 사건을 이용하여 각 개체를 개체가 속하는 군집의 대표 값을 가리키는 포인터로 바꾸어 줌으로써 벡터 데이터에 대하여 최소의 연산을 통해 구분하기 어려운 정도의 위치 정확도 손실 차이를 가지면서 향상된 압축률을 얻을 수 있게 되며, 이에 따라 모바일 기기에서 차지하는 저장 공간을 줄일 수 있게 된다.

(57) 청구의 범위

청구항 1.

모바일 환경의 벡터 데이터를 압축하는 방법에 있어서,

각 기하 객체를 시작점 좌표와 디퍼런셜 벡터의 나열로 분리하는 단계;

상기 디퍼런셜 벡터를 길이와 각도로 분리하는 단계;

상기 분리된 길이와 각도에 대하여 각각 K평균 군집화를 수행하여 군집의 평균값을 엔트리로 갖는 사건을 제작하는 단계; 및

각각의 길이와 각도에 대하여 근사화하여 사건의 엔트리를 가리키는 포인터로 변환하는 단계;

를 포함하여 이루어지는 것을 특징으로 하는 K평균 군집화를 이용한 벡터 데이터 압축 방법.

청구항 2.

제 1 항에 있어서, 상기 디퍼런셜 벡터는 각 기하 객체의 개별 좌표의 해당 좌표와 이전 좌표와의 차이를 이용하여 구하는 것을 특징으로 하는 K평균 군집화를 이용한 벡터 데이터 압축 방법.

청구항 3.

제 1 항에 있어서, 상기 해당 기하 객체의 시작점을 나타내는 시작점 좌표는 변형을 가하지 않는 것을 특징으로 하는 K평균 군집화를 이용한 벡터 데이터 압축 방법.

청구항 4.

모바일 환경의 벡터 데이터를 압축하는 시스템에 있어서,

입력부를 통해 입력되는 각 기하 객체에 대한 데이터로부터 각 기하 객체를 시작점 좌표와 디퍼런셜 벡터의 나열로 분리하는 디퍼런셜 벡터 추출부;

상기 디퍼런셜 벡터 추출부에서 추출된 디퍼런셜 벡터를 길이와 각도로 각각 분리하여 K평균 군집화를 적용하여 길이와 각도 각각에 대한 사전을 제작하고, 제작된 사전을 이용하여 각 개체를 개체가 속하는 군집의 대표 값을 가리키는 포인터로 변환하여 압축하는 압축부;

상기 디퍼런셜 벡터 추출부로부터의 시작점 좌표와 압축부로부터의 길이사전, 각도사전, 길이사전 포인터, 각도사전 포인터 등이 저장되는 각각의 데이터 베이스를 갖는 저장부; 및

상기 각부를 제어하는 제어부;

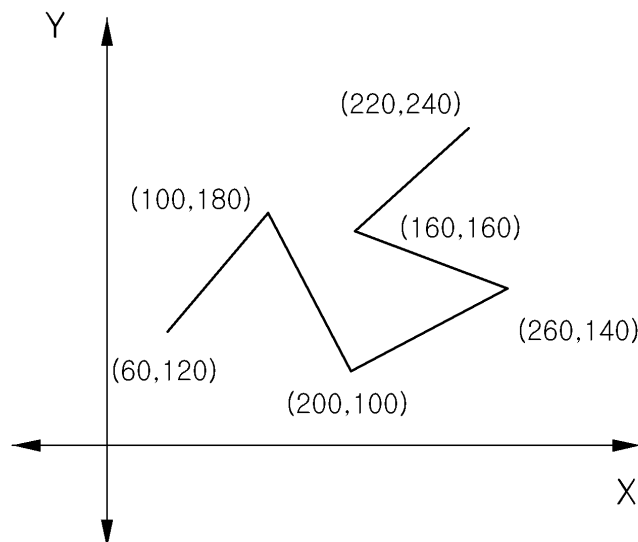
를 포함하여 구성되는 것을 특징으로 하는 K평균 군집화를 이용한 벡터 데이터 압축 시스템.

청구항 5.

제 4 항에 있어서, 상기 디퍼런셜 벡터는 각 기하 객체의 개별 좌표의 해당 좌표와 이전 좌표와의 차이를 이용하여 구하는 것을 특징으로 하는 K평균 군집화를 이용한 벡터 데이터 압축 시스템.

도면

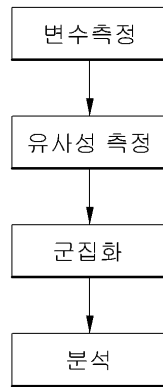
도면1



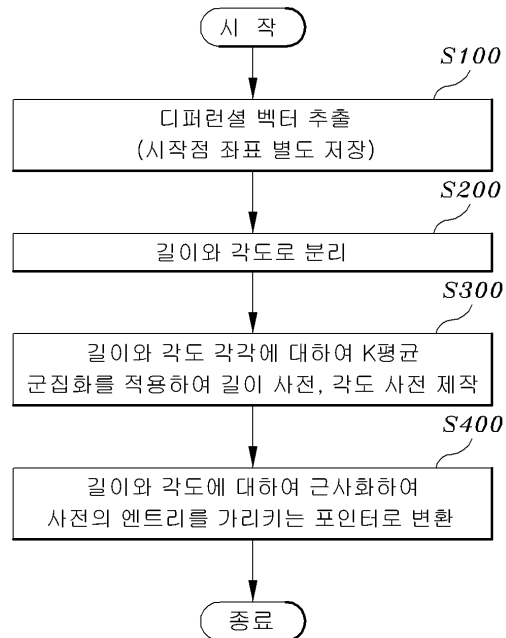
도면2

엔트리	
1	(60, 120)
2	(100, 180)
3	(200, 100)
4	(260, 140)
5	(160, 160)
6	(220, 240)

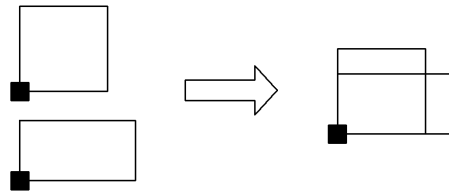
도면3



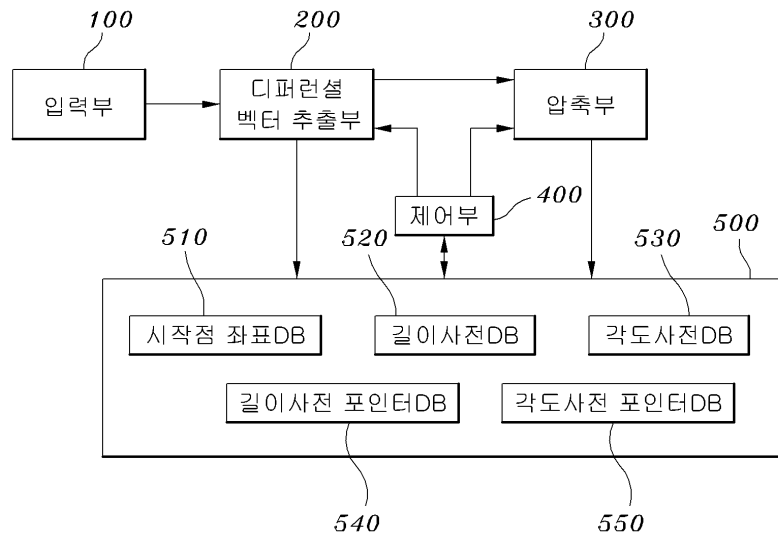
도면4



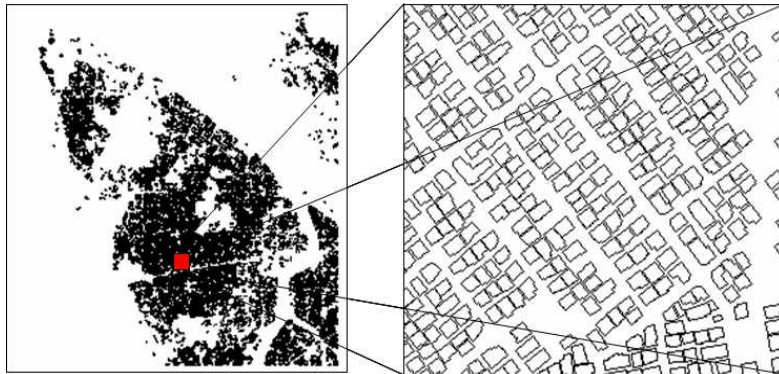
도면5



도면6



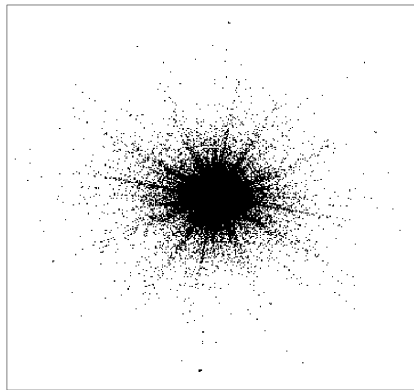
도면7



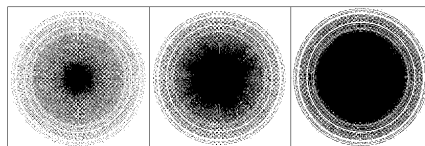
도면8



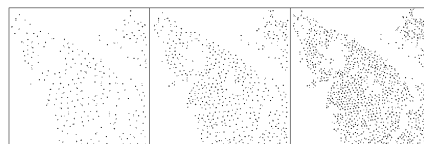
도면9



도면10



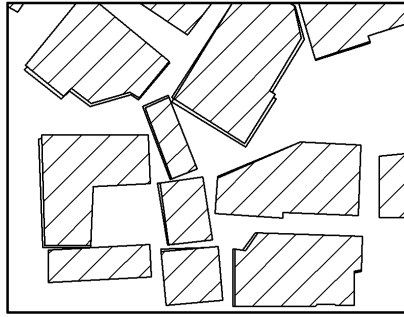
도면11



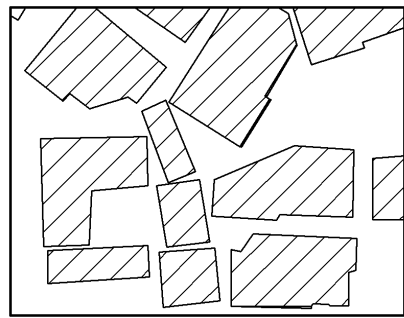
도면12

	실험 1	실험 2	실험 3
RMSE(m)	0.048840	0.010844	0.002756
Size(byte)	2001502	2111241.75	2225077.5
압축률(%)	77.922286	76.711793	75.45612

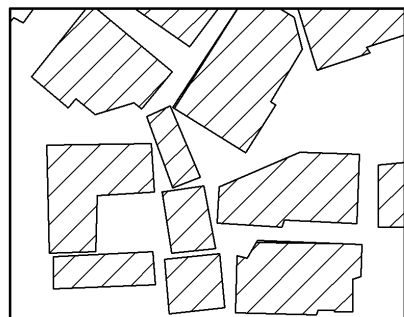
도면13



도면14



도면15



도면16

	실험 1	실험 2	실험 3
RMSE(m)	5.086660	1.144022	0.860976
Size(byte)	2207357.875	2211453.875	2219645.875
압축률(%)	75.65155773	75.60639611	75.51603366

도면17

