

통계 데이터 기반 공간 데이터 시각화 시스템 기술 개발에 관한 연구

장경순*, 노호진*, 백용*, 이창식*, 김병규*
*선문대학교 컴퓨터공학과

e-mail : {piscses, oldbird, padigen, lcs4320}@sunmoon.ac.kr, bg.kim@ieee.org

A Study on Development of the Spatial Data Visualization System Based on Statistical Data

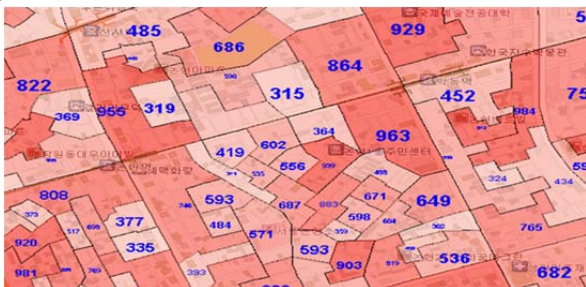
Kyung-Soon Jang*, Ho-Jin Noh*, Back*, Chang-Sik Lee*, Byung-Gyu Kim*
*Dept. of Computer Engineering, SunMoon University

요 약

본 논문에서는 공간 데이터를 기반으로 한 도시 정보를 알고리즘을 통해서 시각화하고 시각화된 데이터를 지도와 사상(Mapping)하여 볼 수 있게 할 뿐만 아니라 공간적 정보를 기반으로 의사 결정을 하는 경우 활용할 수 있는 문서를 사용자의 간단한 조작으로 프로그래밍 방식에 의해 작성해주는 시스템을 제안한다. 본 시스템을 통해서 공간적 정보를 시각화를 한다면 지역의 공간적 정보를 쉽게 파악할 수 있을 것이다. 또한 분석된 정보를 기반으로 제공되고 있는 문서 자동화를 활용한다면 공간적 정보의 문서화에 필요한 많은 시간과 비용을 절감할 수 있을 것으로 기대된다.

1. 서론

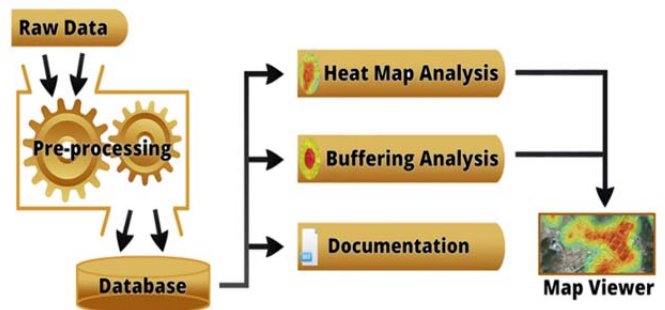
최근 컴퓨터 기술의 발달로 과거 인쇄물 형태로 이용하던 지도 및 지리정보를 컴퓨터를 이용해 작성 및 관리하고 통계를 통해 수집한 공간 데이터를 분석하고 가공하여 활용하는 지리 정보 시스템(Geographic Information System)을 구축하고 활용하는 사례가 점점 증가하고 있다 [1]. 지리 정보 시스템에 사용되는 데이터는 그 자체만으로는 큰 의미를 가지고 있지 않기 때문에 용도에 맞는 정보로 만들기 위하여 가공 단계를 거쳐야 한다. 이러한 요구와 맞물려 데이터 마이닝에 관한 연구가 활발하게 진행되고 있다 [2]. 하지만 가공된 공간 데이터를 효과적으로 표현하는 방법에 대해서는 아직 연구가 미흡하다. 한 예로 통계청에서 제공하고 있는 GIS 시스템에서 제공하고 있는 시각화는 텍스트와 함께 나타내주지만 전체적으로 한 지역의 통계 값을 파악하기는 어렵다는 단점이 있다.



(그림 1) 통계청 GIS의 공간 데이터 시각화
본 논문은 공간 데이터를 효과적으로 분석하기 위

한 방법의 하나로 데이터의 시각화와 분석된 데이터를 문서화 하는 시스템을 제안한다. 이를 위해서 필요한 데이터베이스를 설계하여 데이터를 가공하고 가공된 데이터를 바탕으로 밀도분석 알고리즘을 통해 공간 데이터를 시각화한다. 시각화된 데이터는 Nave에서 제공하고 있는 지도 Open API와 매시업(Mash-up)하여 지도와 함께 볼 수 있도록 인터페이스를 제공하며 Spatial Data Visualization System(SDVS)을 통해서 해당 기술을 사용할 수 있다.

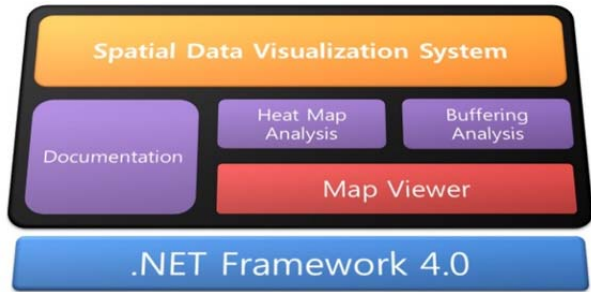
2. 공간 데이터 시각화 시스템



(그림 2) 전체 시스템 구조도

(그림 2)는 본 논문에서 제안하는 시스템의 전체적인 구조를 나타내고 있다. 가공이 되지 않은 공간 데이터(Raw Data)를 전처리(Pre-processing) 과정을 통해서 데이터베이스화 한다. 이후 Heat Map 기법과 Buffering Capacity 기법을 통해서 공간 데이터를 시

각화하고 Map Viewer 를 통해서 시각화된 공간 데이터를 볼 수 있다. 또한 분석된 데이터는 Documentation 모듈을 통해서 공간 데이터를 문서화할 수 있다.



(그림 3) 소프트웨어 구조도

(그림 3)은 본 논문에서 제안된 기술을 사용할 수 있도록 구현된 시스템의 소프트웨어 구조를 보여주고 있다. Spatial Data Visualization System 은 .NET Framework 4.0 을 기반으로 하며, Heat Map Analysis 엔진과 Buffering Analysis 엔진을 포함하고 있는 Map Viewer 모듈과 문서 자동화를 위한 Documentation 기능을 위한 모듈로 구성되어 있다.

3. 제안된 공간 데이터 시각화 기술

3.1 공간분석기법

공간분석은 지리적 특징들에 대해 새로운 정보를 추출하거나 작성하는 과정이고, 특징의 분포, 네트워크 또는 영역 및 이들 특징 사이의 관계를 결정하는 기술이라고 말할 수 있다. 구체적으로 보면 위치나 질의 (Query)를 이용하여 특징과 그와 관련된 특징을 찾을 수 있을 뿐만 아니라 교차 (Intersection)기능 등을 통해 다른 특징 레이어 범위에 포함되는 레이어의 정확한 영역 등을 분석하는 작업과 같은 고급 공간 분석기능을 필요로 한다 [3]-[4]. 본 논문에서는 공간 데이터의 시각화를 위해서 공간분석기법 중 Heat Map 기법과 Buffering Capacity 기법을 활용한다.

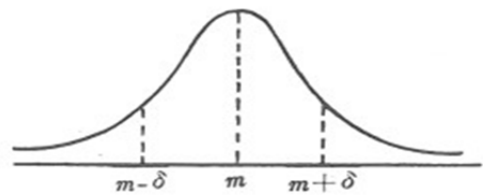
3.1.1 Heat Map Analysis 기법

Heat Map Analysis 는 2 차원 지도 데이터에서 특정 데이터 값에 대해 색을 이용하여 표현하는 기법이다. Heat Map Analysis 는 크게 마스크 생성과 색상화 (Colorize)의 두 단계를 거쳐 완성된다. 본 논문에서는 마스크 생성 이전에 지도 좌표와 공간 데이터의 좌표 간의 사상 (Mapping)을 위한 변환 과정을 거친다. 이후 사상 (Mapping)된 위치에 마스크를 생성하고 영역의 명도에 따라 미리 정해진 색상으로 치환하여 영역에 Heat Map 을 완성한다 [5]-[6].

마스크 생성은 공간 데이터의 좌표를 기준으로 반경 50m 로 생성한다. 그 이유는 Heat Map 기법을 활용하기 위해 수집한 공간 데이터가 중심점으로부터 반경 50M 의 정보를 담고 있기 때문이다. 이후 수치화된 데이터를 농도로 표현한다. 본 논문에서는

영역의 중심부로부터 가장자리로 [0, 1] ~ [0, 255]으로 10 구간에 대해 균등 분할하여 그 값을 사상 (Mapping)시킨다. 마스크 영역이 생성되고 나면, 앞서 설명한 10 구간 균등 분할 사상 (Mapping)을 적용하여 내부로부터 외각으로 그라디언트로 채운다. 그라디언트로 채우기 위해서 가우시안 함수를 사용하게 된다. 아래의 (1)은 가우시안 함수의 수식을 나타내고 있으며 (그림 4)는 가우시안 분포 곡선의 표준 모델을 보여주고 있다. 또한 (그림 6)는 균등 분할 사상(Mapping)에 의해 설정되는 영역 내의 농도를 나타내고 있다 [7].

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} \dots \dots (1)$$



(그림 4) 가우시안 (Gaussian) 분포 곡선

0~0.1	0.1~0.2	0.2~0.3	0.3~0.4	0.4~0.5	0.5~0.6	0.6~0.7	0.7~0.8	0.8~0.9	0.9~1
0~25	25~51	51~76	76~102	102~127	127~153	153~178	178~204	204~229	229~255

(그림 5) 균등 분할 사상(Mapping)에 의한 농도 설정

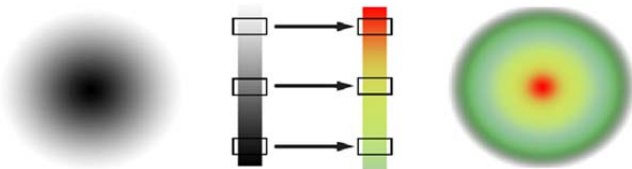
농도 값은 좌표 중심으로부터 멀어질수록 통계 데이터의 가중치 (Weight)가 낮음을 의미한다. 반면 중앙일 수록 밀도 데이터의 가중치 (Weight)가 높음을 의미한다. 이와 같은 방법으로 구해진 농도는 앞서 설명한 마스크 영역 생성 후, 영역 내부의 중심점으로부터 외각으로 그라디언트로 채워 마스크를 생성한다. (그림 7)은 마스크 생성 과정을 보여주고 있다 [6].



(그림 6) 마스크 생성 과정

앞서 설명한 마스크 생성이 완료되면 농도의 값을 치환하는 과정을 거쳐 Heat Map 을 완성한다. 가장 먼저 치환할 팔레트를 만들어야 하는데, 이는 다음과 같은 순서로 진행된다. 먼저 RGB 값을 각각 (255, 0, 0)으로 설정하고 Green 값을 점차 증가시켜 (255, 255, 0)을 만든다. 이후

Red 를 감소시켜 (0, 255, 0)으로 만들면 적색으로 시작하여 황색으로 변화하고 마지막으로 녹색으로 변화하는 팔레트가 완성된다. 이와 같은 과정을 통해 생성된 팔레트는 마스크의 농도 값에 의해서 치환되며 최종 Heat Map 을 완성한다 [7].



(그림 8) Heat Map 의 색상화(Colorize) 과정

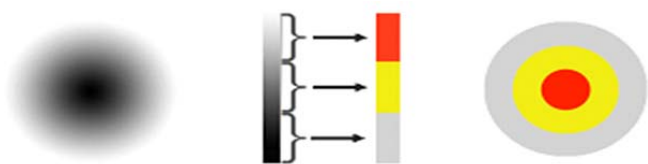
3.1.2 Buffering Analysis 기법

Buffering Analysis 는 공간적 근접성 (Spatial Proximity)을 정의할 때 이용되는 것으로서 점, 선, 면 또는 면 주변에 지정된 범위의 면사상으로 구성한다. Buffering Analysis 을 위해서는 먼저 버퍼 존 (Buffer Zone)의 정의가 필요하다. 버퍼 존은 입력사상과 버퍼를 위한 거리 (Buffer Distance)를 지정한 이후 생성한다. 일반적으로 거리는 단순한 직선거리인 유클리디안 거리 (Euclidean Distance)를 이용한다. 즉, 입력된 자료의 점으로부터 직선거리를 계산하여 이를 버퍼 존으로 표현하는데, 아래와 있는 (수식 2)와 같은 유클리디안 거리 계산공식에 의해 버퍼 존을 형성한다 [8].

$$\sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2}$$

$$= \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \dots \dots (2)$$

또한, 본 연구에서는 Buffer 의 색상화 부분을 추가하여 색상에 따라서 거리를 판별할 수 있도록 한다. 색상은 각각 빨간색과 노란색, 그리고 회색으로 표현한다. 색상화 과정을 위해서 마스크를 생성하고 농도 값이 0 ~ 0.33 인 영역은 RGB 값을 각각 (255, 255, 255)으로 설정하여 회색으로 표현하고 0.34 ~ 0.66 까지는 (255, 255, 0)으로 설정하여 노란색으로 표현한다. 마지막으로 0.67 ~ 1 까지를 (255, 0, 0)으로 설정하여 빨간색으로 표현하여 팔레트를 만들고 이를 적용하여 Buffer 를 완성한다.

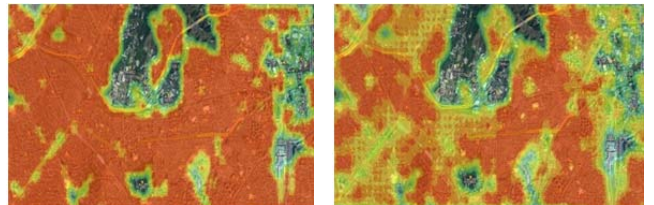


(그림 9) Buffer 의 색상화(Colorize) 과정

3.2 공간 데이터 가중치(Weight)의 정규화

공간 데이터의 가중치는 Heat Map Analysis 기법을 활용하여 공간 데이터를 시각화 할 때 중요한

근거 데이터로 사용된다. 하지만 공간 데이터 간의 가중치의 편차가 클 경우 효과적으로 시각화 할 수 없다. 이와 같은 문제점을 해결하기 위해서 본 논문에서는 공간 데이터의 가중치 중에서 최대값과 최소값을 추출해서 기준점을 만들고 총 공간 데이터 수를 카운팅하여 10 단계로 균등하게 데이터를 나누고 각각의 데이터마다 시각화 시에 사용되는 임시 가중치를 부여한다.



(그림 10) 정규화(Normalize)의 적용:전(좌), 후(우)

3.3 문서 자동화

공간 데이터의 경우 데이터의 양이 방대하기 때문에 분석 결과를 수동으로 문서화 하려면 엄청난 시간과 비용이 소요되게 된다. 하지만 Microsoft사에서 제공하는 Office Document Automation 라이브러리를 활용하면 시간과 비용을 절약할 수 있다. 본 논문에서 제안하는 시스템에서는 해당 라이브러리를 적용하여 Documentation 모듈을 만들어 적용하여 분석된 공간데이터를 쉽게 문서화 할 수 있도록 인터페이스를 제공한다.

4. 실험결과 및 고찰

4.1 Heat Map 기법을 활용한 인구밀도 분석 결과

(그림 12)는 Heat Map Analysis 기법을 적용해서 공간 데이터를 시각화 한 결과이다. 붉은 색은 가중치(인구밀도)가 높음을 의미하고 있고 녹색은 가중치가 낮음을 의미하므로 분석 데이터의 종류에 따른 공간 데이터의 밀도 분포를 시각적으로 볼 수 있다. 이를 통해 넓은 범위에 분산되어 존재하는 공간 데이터가 의미하는 바를 유추할 수 있다.

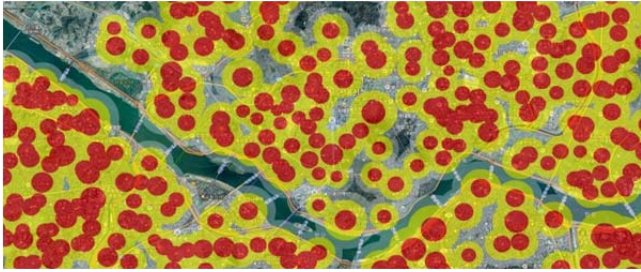


(그림 11) 제안된 Heat Map Analysis 기법의 적용

4.2 Buffering Analysis 기법을 활용한 학교 시설 분포 분석 결과

(그림 12)은 Buffering Analysis 기법을 적용해서 수도권 지역의 고등학교의 분포를 시각화 한 결과이다. Buffering Analysis 에서는 3 가지 색상을 통해서 학교시설과의 거리를 나타내고 있다. 빨강은 학교 시설과 매우 가까운 지역, 노란색은 근접한 지역,

회색은 조금 떨어져있는 지역이라고 볼 수 있다. 이와 같은 시각화 데이터를 통해서 지역 내의 학교 시설이 더 필요 한지를 판별할 수 있는 기초 자료로 사용할 수 있을 것이다.



(그림 12) 제안된 Buffering Analysis 기법의 적용 예

4.3 Heat Map Analysis 기법을 활용한 인구 분포도 비교

(그림 13)과 (그림 14)는 각각 서울시의 주거인구와 직장인구를 시각화 한 결과이다. 시각화 한 결과를 통해서 주거인구의 밀도가 높은 지역과 직장인구의 밀도에 관한 공간 데이터를 수치 데이터가 아닌 지도와 사상 (Mapping)된 시각화 데이터로 확인할 수 있다.



(그림 13) 강남 3구 주거인구 분포 시각화 결과



(그림 14) 강남 3구 직장인구 분포 시각화 결과

또한 시각화 된 데이터를 기반으로 주거인구는 주로 지하철 역을 중심으로 많이 분포해 있는 모습을 볼 수 있었으며 직장인구는 대로변을 중심으로 많이 분포되어 있었다. 공간 데이터를 시각화 한다면 이와 같이 수치 데이터일 때는 알 수 없는 부분을 유추할 수 있다.

5. 결론 및 Future Work

본 연구는 지리 정보 시스템을 좀 더 효율적으로 활용하기 위한 방안 중 공간 데이터의 시각화와 문서 자동화 방법에 대하여 제안하고 이를 기반으로 Spatial Data Visualization System (SDVS)을 개발하고 시스템에

서 사용될 공간 데이터로 통계청이나 기타 정부부처에서 제공하고 있는 분산되어 있는 통계 데이터를 수집하고 가공하여 데이터베이스를 구축하였다. SDVS를 활용하면 통계 데이터 기반의 지역 정보를 쉽게 시각화 할 수 있다. 이를 통해서 지역의 공간 데이터를 분석을 효과적으로 할 수 있다. 이러한 분석 데이터들은 도시 개발 계획에 필요한 기반 데이터로 활용할 수 있을 것이다.

본 연구에서는 공간 데이터의 시각화의 장점을 부각하기 위해서 한정된 공간 데이터만을 포함하고 있다. 하지만 공간 데이터를 추가한다면 좀 더 폭넓은 분석을 할 수 있을 것이다. 또한 시각화에 사용되는 색상의 폭을 넓혀서 좀 더 세부적인 시각화를 가능하게 한다면 좀 더 세부적인 분석이 가능할 것이다.

Acknowledgements

본 논문의 내용은 선문 비트 교육센터 고급과정 23기 프로젝트 결과물을 토대로 하고 있으며, 컴퓨터공학과 김병규 교수님과 선문 비트교육센터 관계자 여러분의 협조에 감사 드립니다.

참고문헌

- [1] 이희연, 심재현, “GIS 지리정보학”, 법문사, 2011
- [2] 배덕호, 백지행, 오현교, 송주원, 김상욱, 최명희, 조현주, “공간 데이터 마이닝 시스템의 설계 및 구현” 한국공간정보시스템학회 논문지 : 제 11 권 제 2 호, June 2009.
- [3] 장용봉, “GIS 를 이용한 위험도분석과 위험물질의 최적이동경로에 대한 연구”, 서울대학교 대학원 석사논문 1996.
- [4] 김성원, “GIS 기반 무선 센서 네트워크에서 Heat-map 을 이용한 Coverage 가시화”, 원광대학교 대학원 석사논문, Dec 2009.
- [5] D. Fisher, Hotmap: Looking at geographic attention. IEEE Transactions on Visualization and Computer Graphics, 13(6):1184-1191, 2007.
- [6] D. Fisher, “The Impact of Hotmap. Microsoft Reserch”, available from <http://research.microsoft.com/apps/pubs/default.aspx?id=81244>, 2009.
- [7] Johannes Trame, Carsten Kebler, “Exploring the Lineage of Volunteered Geographic Information width Meat Maps”, GeoViz: Linking Geovisualization with Spatial Analysis and Modeling, 10-11 March 2011.
- [8] Nagapramod Mandagere, “buffer Operation in GIS”, University of Minnesota available from http://www-users.cs.umn.edu/~npramod/enc_pdf.pdf.