

Visualization Technique of Spatial Statistical Data and System Implementation

Baek Ryong[†] · Gwang-Soo Hong^{**} · Seung-Hoon Yang^{***} · Byung-Gyu Kim^{****}

ABSTRACT

In this paper, a system technology-based algorithms and visualization is proposed to show a space data. Also the proposed system provides analysis function with combination of usual map and automatic document generation function to give a useful information for making an important decision based on spatial distributed data. In the proposed method, we employ the heat map analysis to present a suitable color distribution for 2 dimensional map data. The buffering analysis method is also used to define the spatial data access. By using the proposed system, spatial information in a variety of distribution will be easy to identify. Also, if we make a use of automatic document generation function in the proposed algorithm, a lot of time and cost savings are expected to make electronic document which representation of spatial information is required.

Keywords : Data Visualization, Statistical Analysis Data, Map Mapping

공간 통계 데이터의 시각화 기술 및 시스템 개발

백 룡[†] · 홍 광 수^{**} · 양 승 훈^{***} · 김 병 규^{****}

요 약

본 논문에서는 공간 데이터를 기반으로 한 도시의 다양한 통계 정보를 제안된 알고리즘을 통해서 시각화하고 시각화된 데이터를 지도와 사상(Mapping)하여 분석할 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. 제안된 기법은 2차원 지도 데이터에서 특징데이터 값에 대해 색을 이용하여 표현하는 기법인 히트 맵 분석(Heat Map Analysis) 기법과 공간적 근접성을 정의할 때 이용되는 버퍼링 분석(Buffering Analysis) 기법을 활용한다. 본 시스템을 통해서 공간적 정보를 시각화를 한다면 지역의 분포된 다양한 공간적 정보를 쉽게 파악할 수 있을 것이다. 또한 분석된 정보를 기반으로 제공되고 있는 문서 자동생성기능을 활용한다면 표현된 공간적 정보의 문서화에 필요한 많은 시간과 비용을 절감할 수 있을 것으로 기대된다.

키워드 : 데이터 시각화, 통계 분석 데이터, 지도 사상

1. 서 론

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:GIS)을 구축하고 활용하는 사례가 점점 증가하고 있다[1]. 지리 정보 시스템에 사용되는 데이터는 그 자체만으로는 큰 의미

를 가지고 있지 않기 때문에 용도에 맞는 정보로 만들기 위하여 가공 단계를 거쳐야 한다. 이러한 요구와 맞물려 데이터 마이닝에 관한 연구가 활발하게 진행되고 있다[2].

Shigeru Obayashi et al.은 목적함수 공간에서 Pareto 해의 trade-off를 시각화하기 위하여 자기 구성형 맵(Self-organizing maps)을 활용하였다[9]. 또한 데이터 시각화 연구의 한 분야인 텍스트 시각화에 대하여 그 시각적 표현의 재료가 되는 텍스트 데이터의 종류 및 본질, 특성에 대하여 고찰하고, 다양한 텍스트 시각화의 사례에 대해 고찰하였다[10].

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

[†] 비 회 원 : 선문대학교 컴퓨터공학과 석사과정
^{**} 비 회 원 : 선문대학교 컴퓨터공학과 박사과정
^{***} 준 회 원 : 선문대학교 컴퓨터공학과 석사과정
^{****} 정 회 원 : 선문대학교 컴퓨터공학과 부교수
논문접수: 2013년 7월 1일
수정일: 1차 2013년 7월 16일
심사완료: 2013년 7월 16일

* Corresponding Author : Byung-Gyu Kim(bg.kim@mpcl.sunmoon.ac.kr)

파악하기는 어렵다는 단점이 있다. 또한 제공하는 프로그램은 전문가가 아니면 사용하기가 매우 어렵다는 단점도 가지고 있다.

Fig. 1에 나타난 것처럼 지도에 수치만 나타나 있어 사전 지식이 없으면 지도상에서 나타나는 의미를 파악하기가 어렵다. 본 논문에서는 전문가가 아닌 비전문가도 쉽게 접근할 수 있게 직관적인 인터페이스를 제공해 주고 있다. 이로 인해 전문 지식이 없음에도 불구하고 비교할 대상만 적절히 선택하여 시각적으로 결과를 예측 및 분석할 수 있다.

본 논문에서는 기존에 제안된 기술을 개선하기 위하여 사용자 편의 기능과 자동 문서생성 기술을 추가하여 공간 데이터의 시각화 변환기술을 제안한다.



Fig. 1. Visualization of spatial data in Statistics Korea's GIS system

본 논문은 공간 데이터를 효과적으로 분석하기 위한 방법의 하나로 데이터의 시각화와 분석된 데이터 기반으로 문서화하는 시스템을 제안한다. 이를 위해서 필요한 데이터베이스를 설계하여 데이터를 가공하고 가공된 데이터를 바탕으로 밀도분석 알고리즘을 통해 공간 데이터를 시각화한다.

시각화된 데이터는 Naver 등 포털에서 제공하고 있는 지도 공개 API와 매쉬업 (Mash-up)하여 지도와 함께 볼 수 있도록 인터페이스를 제공함으로써 개발된 공간데이터 시각화 시스템 (Spatial Data Visualization System : SDVS)을 통해서 해당 기술을 사용할 수 있다. 개발된 시스템은 간편한 인터페이스를 제공하여 사전 지식이 없어도 자신이 원하는 데이터를 한눈에 볼 수 있으며 사용자 관점에서 데이터를 시각화 하여 비교 할 수 있다.

2. 공간 데이터 시각화 시스템

Fig. 2는 본 논문에서 제안하는 시스템의 전체적인 구조를 나타내고 있다. 가공이 되지 않은 공간 데이터(Raw Data)를 전처리 (Pre-processing) 과정을 통하여 데이터베이스를 구성한다. 이후 히트 맵 (Heat Map) 기법과 버퍼링 공간 (Buffering Capacity) 기법을 통하여 공간 데이터를 시각화하고 맵 뷰어 (Map Viewer)를 통하여 시각화된 공간 데이터를 볼 수 있다. 또한 분석된 데이터는 Documentation 모듈을 통하여 공간 데이터를 문서화 할 수 있다.

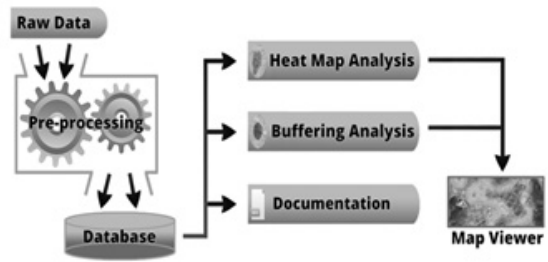


Fig. 2. Overall system structure

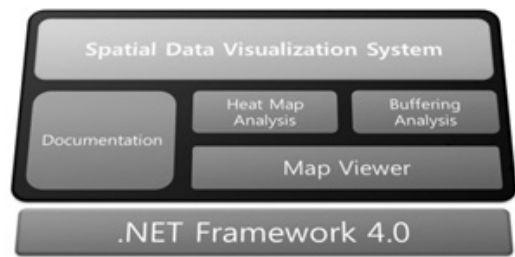


Fig. 3. Software structure of the developed system

Fig. 3은 본 논문에서 제안된 기술을 사용할 수 있도록 구현된 시스템의 소프트웨어 구조를 보여주고 있다. 공간 데이터 시각화 시스템 (SDVS)은 .NET Frame Work 4.0을 기반으로 하며, 히트 맵 분석 엔진과 버퍼링 분석 (Buffering Analysis) 엔진을 포함하고 있는 맵 뷰어 (Map Viewer) 모듈과 문서 자동화를 위한 문서화 (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 특정 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 히트 맵 기법과 버퍼링 분석 기법을 활용 한다.

3.2 히트 맵 (Heat Map Analysis) 기법

히트 맵 분석은 2차원 지도 데이터에서 특정 데이터 값에 대해 색을 이용하여 표현하는 기법이다. 히트 맵 분석은 크게 마스크 생성과 색상 화 (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 히트 맵을 완성 한다[5]-[6].

마스크 생성은 공간 데이터의 좌표를 기준으로 반경 50m로 생성한다. 그 이유는 히트 맵 기법을 활용하기 위해 수집한 공간 데이터가 중심점으로부터 반경 50M의 정보를 담고 있기 때문이다. 이후 수치화된 데이터를 농도로 표현한다.

본 논문에서는 영역의 중심부로부터 가장자리로 [0, 1] ~ [0, 255]으로 10구간에 대해 균등 분할하여 그 값을 사상(Mapping)시킨다. 마스크 영역이 생성되고 나면, 앞서 설명한 10구간 균등 분할 사상(Mapping)을 적용하여 내부로부터 외각으로 그라디언트 값(gradient value)로 채운다. 그라디언트 값 (gradient value)로 채우기 위해서 가우시안 함수를 사용하게 된다. 아래의 수식 (1)은 가우시안 함수의 수식을 나타내고 있으며 Fig. 4는 가우시안 분포 곡선의 표준 모델을 보여주고 있다. 또한 Fig. 6는 균등 분할 사상(Mapping)에 의해 설정되는 영역 내의 농도를 나타내고 있다[7]. 이를 기반으로 하여 본 논문에서는 히트 맵 분석기법을 사용하여 공간데이터를 시각화하는 기법을 제시한다.

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

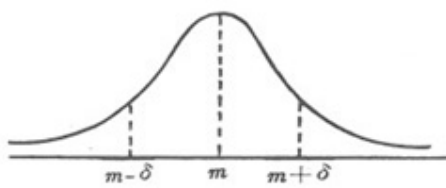


Fig. 4. A curve of Gaussian distribution

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

Fig. 5. Density control by equivalent assignment mapping

농도 값은 좌표 중심으로부터 멀어질수록 통계 데이터의 가중치 (Weight)가 낮음을 의미한다. 반면 중앙 일수록 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. Fig. 7은 마스크 생성 과정을 보여주고 있다[6].

앞서 설명한 마스크 생성이 완료되면 농도의 값을 치환하는 과정을 거쳐 히트 맵을 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 Red를 감소시켜 (0, 255, 0)으로 만들면 적색으로 시작하여 황색으로

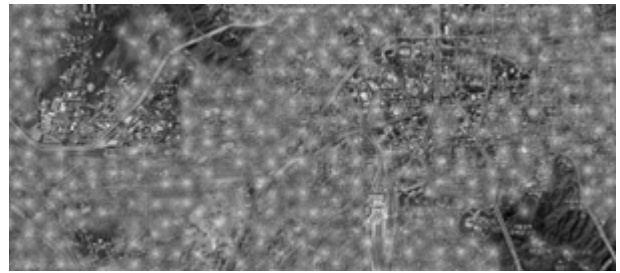


Fig. 6. A process of mask generation

변화하고 마지막으로 녹색으로 변화하는 팔레트가 완성된다. 이와 같은 과정을 통해 생성된 팔레트는 마스크의 농도 값에 의해서 치환되며 최종 히트 맵을 완성 한다[7].

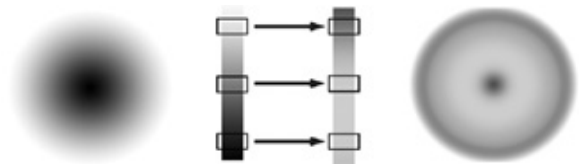


Fig. 7. Colorization process of Heat Map

3.3 버퍼링 분석 (Buffering Analysis) 기법

버퍼링 분석은 공간적 근접성(Spatial Proximity)을 정의할 때 이용되는 것으로서 점, 선, 면 또는 면 주변에 지정된 범위의 면 사상으로 구성한다. 버퍼링 분석을 위해서는 먼저 버퍼 영역(Buffer Zone)의 정의가 필요하다. 버퍼 영역은 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 영역으로 표현하는데, 아래와 있는 수식 (2)와 같은 유클리디안 거리(d)에 대한 정의에 의해 버퍼 영역을 형성 한다[8].

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

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

또한, 본 연구에서는 버퍼(Buffer)의 색상화 부분을 추가하여 색상에 따라서 거리를 판별할 수 있도록 한다. 색상은 각각 빨간색과 노란색, 그리고 회색으로 표현한다. Fig. 8은 색상화 과정을 위해서 마스크를 생성하고 농도 값이 0 ~ 0.33인 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 ~ 1까지를 (255, 0, 0)으로 설정하여 빨간색으로 표현하여 팔레트를 만들고 이를 적용하여 버퍼(Buffer)를 완성한다.

3.4 공간 데이터 가중치(Weight)의 정규화

공간 데이터의 가중치는 히트 맵 분석기법을 활용하여 공

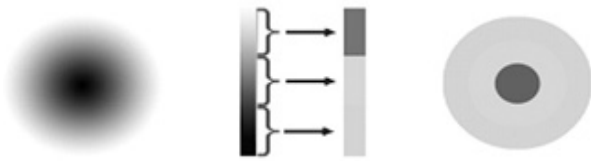


Fig. 8. Buffer colorization process

간 데이터를 시각화 할 때 중요한 근거 데이터로 사용된다. 하지만 공간 데이터 간의 가중치의 편차가 클 경우 효과적으로 시각화 할 수 없다(Fig. 9(a)). 이와 같은 문제점을 해결하기 위해서 본 연구에서는 공간 데이터의 가중치 중에서 최대값 과 최소값을 추출해서 기준점을 만들고 총 공간 데이터 수를 계산하여 10단계로 균등하게 데이터를 나누고 각각의 데이터마다 시각화 시에 사용되는 임시 가중치를 부여한다(Fig. 9(b)).

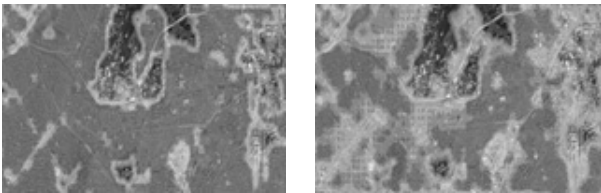


Fig. 9. (a) Before normalizing, (b) After normalizing

3.5 문서 생성기능

공간 데이터의 경우 데이터의 양이 방대하기 때문에 분석 결과를 수동으로 문서화하려면 엄청난 시간과 비용이 소요된다. 하지만 Microsoft사에서 제공하는 Office Document Automation 라이브러리를 활용하면 시간과 비용을 절약할 수 있다.

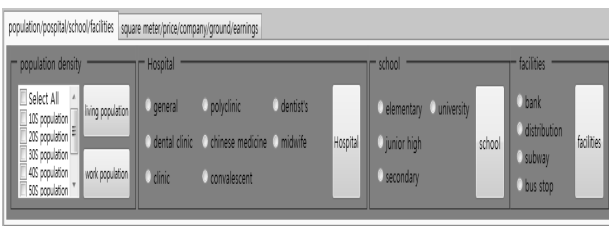


Fig. 10. Search parameters

본 연구에서 제안하는 시스템에서는 해당 라이브러리를 적용하여 Documentation 모듈을 만들어 적용하여 분석된 공간데이터를 쉽게 문서화 할 수 있도록 인터페이스를 제공한다. Fig. 11은 분석한 데이터를 차트로 보여주며 분석한 데이터를 상세하게 출력하여 데이터 편집이나 보고서로 용이하게 사용할 수 있다.

Fig. 10에서 표시된 컨트롤박스에서 사용자가 전문가가 아니더라도 제공되는 콘텐츠에서 분석이 필요한 인자들을 선택하여 비교를 통해 시각화하고 대응 하는 데이터나 통계 데이터를 문서로 처리 할 수 있도록 개발하였다.

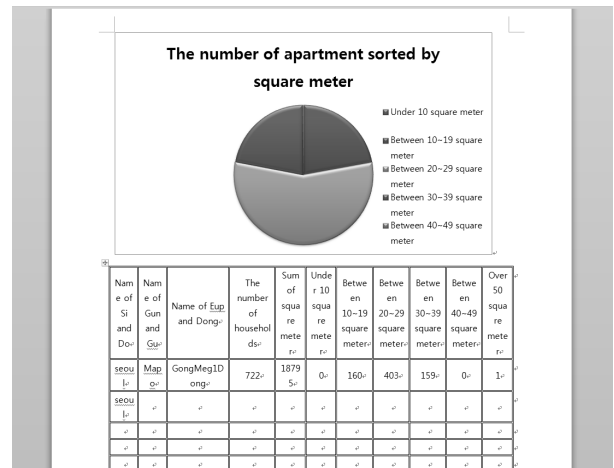


Fig. 11. Automatic documentation result

4. 실험결과 및 고찰

4.1 히트 맵 기법 기반 인구밀도 분석결과

Fig. 12는 히트 맵 분석기법을 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.



Fig. 12. A result of the proposed Heat Map Analysis

4.2 버퍼링 분석 (Buffering Analysis) 기법을 활용한 학교 시설 분포 분석 결과

Fig. 13은 버퍼링 분석기법을 적용해서 수도권 지역의 고등학교의 분포를 시각화 한 결과이다. 버퍼링 분석에서는 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 필요한 지를 판별할 수 있는 기초 자료로 사용할 수 있을 것이다.

4.3 히트 맵 분석 기법을 활용한 인구분포도 비교

Fig. 14와 Fig. 15는 각각 서울시의 주거인구와 직장 인구를 시각화 한 결과이다. 시각화 한 결과를 통해서 주거인구의 밀도가 높은 지역과 직장인구의 밀도에 관한 공간 데이

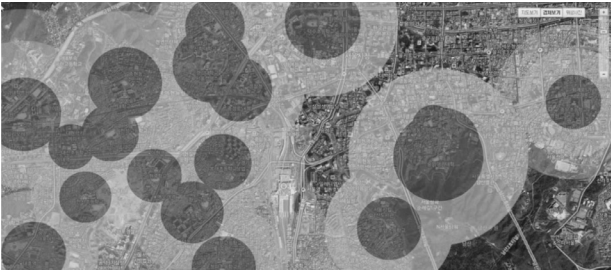


Fig. 13. A result of the proposed buffering Analysis



Fig. 14. Visualization of population distribution on Kangnam 3gu



Fig. 15. Visualization of working population distribution on Kangnam 3gu

터를 수치 데이터가 아닌 지도와 사상(Mapping)된 시각화 데이터로 확인할 수 있다.

또한 시각화 된 데이터를 기반으로 주거 인구는 주로 지하철역을 중심으로 많이 분포해 있는 모습을 볼 수 있었으며 직장 인구는 대로변을 중심으로 많이 분포되어 있었다. 공간 데이터를 시각화 한다면 이와 같이 수치 데이터일 때는 알 수 없는 부분을 유추할 수 있다.

기존의 통계청에서 제공하는 GIS 시스템에서는 비전문가 입장에서는 접근하기 힘든 인터페이스를 제공하고 있으며 단순한 수치 데이터만 보여주고 있으므로 의미 전달이 어려우며 필요에 따라서 분석 결과를 얻고자 할 경우 추가적인 시간이나 비용을 소모하기 때문에 효율적이지 못하다는 단점을 가지고 있다.

5. 결론 및 Future Work

본 연구에서는 지리 정보 시스템을 좀 더 효율적으로 활용하기 위한 방안 중 공간 데이터의 시각화와 문서 자동화

생성 기법에 대하여 제안하고 이를 기반으로 공간 데이터 시각화 시스템을 개발하였다. 제안된 공간 데이터 시각화 시스템을 활용하면 통계 데이터 기반의 지역 정보를 쉽게 시각화 할 수 있으며 이를 통해서 지역의 공간 데이터를 분석을 효과적으로 할 수 있다. 이러한 분석 데이터는 도시 개발 계획 및 복지 인프라 확충 등에 필요한 기반 분석 데이터로 활용할 수 있을 것이다. 또한 간단한 인터페이스 때문에 전문 지식이 필요 없이 원하는 데이터를 검색하여 사용자 관점에 따로 비교 및 분석을 할 수 있다.

제안된 시스템 기술은 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.

참고 문헌

- [1] Hee-Yeon Lee, Jae-Heon Shim, "GIS Geographical Information", Bubmun press, 2011.
- [2] Deuk-Ho Bae et al., "Design and Implementation of a Spatial Data Mining System," Korea Spatial Information Society Journal, Vol.11, No.2, June, 2009.
- [3] Yong-Bong Jang, "A study on risk analysis and the optimal path of dangerous materials by using GIS," Seoul National university Master Thesis, 1996.
- [4] Sung-Won Kim, "Visualization of Coverage with Heat-map on Wireless Sensor Network based on GIS," WonKwang university, Mater Thesis, 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 with 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.
- [9] Shigeru Obayashi, Daisuke Sasaki, "Visualization and Data Mining of Pareto Solutions Using Self-Organizing Map," Evolutionary multi-criterion optimization, LNCS 2632, pp796-809, 2003.
- [10] Hyo Young Kim, Jin Wan Park, "A Review on Expressive Materials and Approaches to Text Visualization," Korea Contents Society 13(1) pp.64-72, 2013.



백 동

e-mail : padigen@mpcl.sunmoon.ac.kr
2003년 선문대학교 컴퓨터공학과(학사)
2013년~현재 선문대학교 컴퓨터공학과
석사과정
관심분야: Video coding, Motion estimation



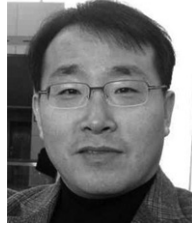
양 승훈

e-mail : smhotkdg@mpcl.sunmoon.ac.kr
2013년 선문대학교 컴퓨터공학과(학사)
2013년~현재 선문대학교 컴퓨터공학과
석사과정
관심분야: Image processing, object
segmentation



홍 광수

e-mail : hongzolv@sunmoon.ac.kr
2011년 선문대학교 컴퓨터공학과(학사)
2013년 선문대학교 컴퓨터공학과(석사)
2013년~현재 선문대학교 컴퓨터공학과
박사과정
관심분야: Feature extraction, Object
segmentation, 3D video coding



김 병규

e-mail : bg.kim@mpcl.sunmoon.ac.kr
1995년 부산대학교 전기공학과 (학사)
1998년 한국과학기술원 전기및전자공학과
(석사)
2004년 한국과학기술원 전기및전자공학과
(박사)
2004년~2008년 한국전자통신연구원 선임연구원
2009년~2012년 선문대학교 컴퓨터공학과 조교수
2013년~현재 선문대학교 컴퓨터공학과 부교수
관심분야: Video signal processing, Coding, Object detection
and tracking