GIS를 이용한 2000년 강원도 인구 분포의 시각화에 관한 연구

이 상 일*

目 次

Ⅰ. 서론
Ⅱ. 2000년 강원도 인구분포의 시각화
1. 단계구분도와 인구 분포상의 특징
2. 벡터-기반 시각화
3. 레스터-기반 시각화
Ⅲ. 결 론

Ⅰ. 서 론

1. 문제제기 및 연구목적

1990년대 이후 현대 사회를 특징짓는 용어로서 '정보사회' 라는 개념이 널리 보급되어 사용되고 있다. 정보사회를 그 이전과 구분 짓게 하는 요소들은 다양하게 열거될 수 있겠지만, 그 중 가장 중요한 것으로 두 가지를 꼽을 수 있다. 하나는 정보의 량, 정보의 종류, 그리고 정보에의 접근성에서 발생한 엄청난 진보이고, 또 다른 하나는 그러한 정보의 저장·검색·분석·소통을 가능케 하는 컴퓨터 기술의 급격한 진보이다. 통칭하여 '정보기술' 이라 일컬어지는 분야의 학문적·산업적 발달은 소위 '디지털 혁명'을 이끌고 있으며 현대 사회의 가장 중요한 추동력이 되고 있다. 이러한 관점에서 볼 때, 공간정보 역시 이러한 사회적 변화와 맞물려 과거와는 전혀 다른 형태로 생산·가공·전달되고 있다. 공간정보의 중요성에 대한 새로운 인식과 공간정보의 디지털화의 급격한 진전은 특정한 공간정보기술로서의 GIS의 발전과 함께, 사회 전반적인 '디지털 혁명'의 중요한 부분을 차지하고 있는 것이다.

* 강원대학교 사범대학 지리교육과 전임강사

지리학적인 측면에서 볼 때, GIS의 도입과 발전은 종종 '계량혁명'과 비견될 만큼의 커다란 영향력을 지리학에 행사하고 있다[1]. GIS가 지리학에서 활용되는 방식은 지리학의 하위분과나 특정 연구 과제에 따라 상이하겠지만, 가장 기본적이지만 가장 중요하기도 한 GIS의 기능 범주는 바로 '지도학적 시각화(cartographic visualization)'라 불리는 분야일 것이다[2]. 지도학적 시각화는 공간정보에 대한 단순한 시각적 표현을 의미하는 것을 넘어, 공간정보를 제시(presentation)·분석(analysis)·탐색(exploration)하는 전 과정을 포괄하는 용어이다[3]. GIS의 발전은 기본적으로 컴퓨터 지도학(computer cartography) 혹은 자동화된 지도학(automated cartography)의 발전과 밀접히 관련되어 있는데, GIS 소프트웨어는 지도학에서 발전해온 다양한 주제도 작성(thematic mapping)이나 지도디자인(map design) 기법들을 포함시켜 왔다. 따라서 GIS를 이용함으로써 연구자는 수집된 공간정보를 이용하여 연구과제에 의거하여 다양한 형태의 주제도를 손쉽게 작성할 수 있게 된 것이다.

인문지리학의 분야에서 이러한 지도학적 시각화의 문제는 주로 행정구역으로 合域되어(spatially aggregated) 제공되는 공간정보와 관련되어 있다. 이러한 域(area)-기반 공간정보는 여타의 공간객체유형 -즉, 點(point), 線(line), 그리고 面(surface)- 과는 다른 특징을 보이는 것으로, 가장 널리 쓰이고 있는 인문지리적 정보로 볼 수 있을 것이다. 또한 이러한 정보가 주로 센서스 자료와 관련되어 있기 때문에 종종 '센서스 지리학(census geography)'이라는 용어로 불리기도 한다[4]. 본 논자는 디지털적으로 전달되는 센서스 자료와 GIS의 결합은 위에서 언급한 지도학적 시각화를 매개로 하여 GIS의 유용성을 인문지리학적인 방법으로 극대화하는 하나의 방법이 될 수 있다고 본다. 즉, 행정구역에 대한 수치지도와 통계청으로부터 인터넷을 통하여 전자적으로 구득된 자료를 GIS 상에서 통합시키고, 그것을 바탕으로

---

1) GIS와 지리학의 관계에 대한 면밀한 검토는 GIS에 대한 과학철학적·지식사회학적 분석을 요구하는 것으로 이 논문의 범위를 넘어 서는 것이다. 개론적인 수준에서 다음의 글을 참고할 수 있다. Johnston, R. J., 1999, Geography and GIS, in Longley, P. A., Goodchild, M. F., Maguire, D. J., and Rhind, D. W. (eds.), *Geographical Information Systems: Principles and Technical Issues*, Volume 1, New York: John Wiley and Sons, 39-47.

2) 김감영은 지도학적 시각화라는 용어 대신 '지리적 시각화(geographic visualization)'라는 용어의 사용을 옹호하고 있지만, 여기서는 GIS와 지도학의 관계를 강조한다는 측면에서, 그리고 주제도 작성과 지도 디자인에 초점을 둔다는 측면에서, 지도학적 시각화라는 용어를 쓰고자 한다. 김감영, 2001, "GIS와 애니메이션을 이용한 시공간 자료의 지리적 시각화," 한국지도학회지, 1, 25-39.

3) Kraak, M.-J., 1999, Visualising spatial distributions, in Longley, P. A., Goodchild, M. F., Maguire, D. J. and Rhind, D. W. (eds.), *Geographical Information Systems: Principles and Technical Issues*, Volume 1, New York: John Wiley and Sons, 157-173.

4) Martin, D., 2000, Towards the geographies of the 2001 UK census of population, *Transactions of the Institute of British Geographers*, 25, 321-332.

다양한 주제도를 지도학적 시각화에 의거하여 생산하는 것은 가장 기본적이고도 가장 중요한 인문지리학적인 GIS의 활용이 될 것이다.

따라서 본 논문의 목적은 GIS의 다양한 지도학적 시각화 기능을 이용하여 2000년 인구센서스에 의해 조사된 강원도 인구의 공간분포를 재현하는 것이다.

2. 연구 자료 및 연구 과제

연구에 이용될 도형자료는 우리나라 1:25,000 지형도의 수치지도를 재구성함으로써 얻어진 우리나라 전체에 대한 시·군·구 행정구역도와 읍·면·동 행정구역도에서 강원도 부분을 추출하여 사용하고자 한다. 또한 속성자료는 통계청의 홈페이지를 통해 각 행정구역 수준 별로 2000년 총인구 자료를 다운로드하여 사용할 것이다. 사용될 GIS 소프트웨어는 ArcView이며, 데이터 조작(data manipulation)이나 특정한 주제도 작성을 위해 ArcView extensions, Avenue scripts, 혹은 여타 소프트웨어가 부가적으로 사용될 것이다.

인구의 공간적 분포를 시각화하는 방법은 우선 '벡터-기반(vector-based) 시각화' 와 '레스터-기반(raster-based) 시각화' 로 나누어 살펴 볼 것이다. 이러한 시각화의 이분법 이전에 우선적으로 단계구분도를 통해 전반적인 인구분포 양상에 대한 고찰을 하고자 한다. 단계구분도는 인구의 공간적 분포를 재현하는데 많은 제한점을 가지고 있긴 하지만, 전반적인 인구분포의 양상을 행정단위별로 기술하려는 목적을 위해서는 불가피한 시각화의 방법이라 할 수 있다. 시·군 수준과 읍·면·동 수준에서 단계구분도를 작성하고 비교함으로써 지리적 스케일에 따른 패턴 인식에 어떤 차이를 발생시키는지에 대한 문제가 논의될 것이다. 다음으로 행정구역별 인구밀도분포에 대한 단계구분도와 점묘도가 읍·면·동 수준에서 제시될 것이다. 벡터-기반 시각화를 위해 첫째 점묘도가 제작될 것이다. 임의적 점배치(random placement of dots) 방식에 의해 구현되는 점묘도의 작성 절차가 간략히 소개될 것이고, 단계구분도에 대한 점묘도의 장점이 논의될 것이다. 둘째, 왜상통계지도(cartogram)가 시·군 수준에서 그려질 것이다. 연속적 면형 왜상통계지도(contiguous area cartogram)의 작성 알고리듬에 대한 논의가 제공될 것이다. 마지막으로 시·군별 인구수에 대한 3차원 재현으로 계단형 통계면(stepped statistical surface)이 제시될 것이다.

레스터-기반 시각화에서는 우선적으로 kernel smoothing 기법에 의한 공간적 밀도면(spatial density surface)이 제시될 것이다. 비연속적 域-기반 자료를 연속적 레스터-기반 재현기법으로 전환하는 과정에 대한 설명과 관련된 논점들에 대해 소개될 것이다. 이 밀도면을 위에서 작성된 단계구분도와

비교함으로써 인구의 공간분포를 재현하는데 있어 레스터-기반 시각화 방법의 장점이 부각될 것이다. 두 번째의 레스터-기반 시각화의 방법으로 경향면(trend surface)이 제시될 것이다. 강원도 인구분포의 전역적 패턴을 가장 잘 묘사해주는 시각화로서 4차 경향면이 작성될 것이다. 마지막으로, 위에서 제작된 밀도면을 3차원으로 전환함으로써 시각화의 효과를 증대시키고자 했다.

Ⅱ. 2000년 강원도 인구 분포의 시각화

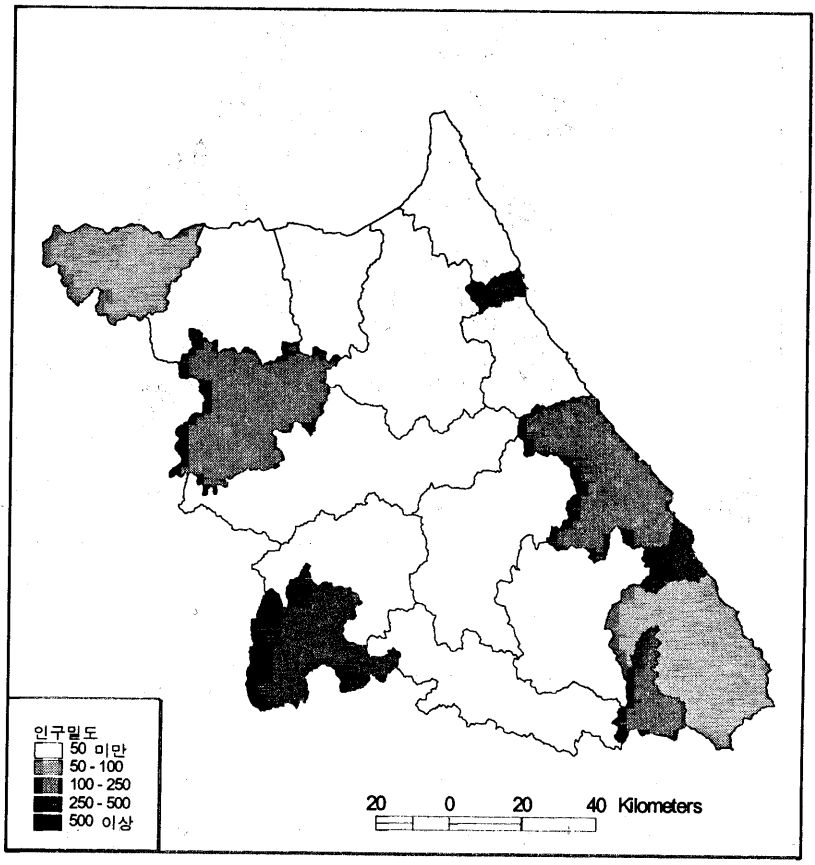
1. 단계구분도와 인구 분포 상의 특징

<그림 1>은 2000년 시군별 강원도 인구 분포에 대한 단계구분도이다. 단계구분도는 각 행정구역별로 인구밀도를 산출하여 제작되었다. 인구밀도 산정을 위한 면적은 GIS 상에서의 면적 계산 알고리듬에 의한 것으로 행정구역총람 상의 행정구역 면적[5]과는 다를 수 있다. 2000년 인구센서스 결과 우리나라 전체의 인구밀도는 462명/㎢인데, 지역별로 보면, 서울이 16,342명/㎢으로 가장 높고 강원도가 88명/㎢으로 가장 낮았다. 특히 인제군은 19명/㎢으로 전국에서 가장 낮은 인구밀도를 보였는데, 이것은 가장 높은 인구밀도인 서울시 양천구의 26,401명/㎢에 1,000분의 1도 되지 않는 수치이다[6]. <그림 1>이 보여주고 있는 인구밀도 분포의 가장 큰 특징은 지형적 영향에 의한 수직적 3지대 분포패턴일 것이다. 즉, 인구가 조밀한 서부 지대와 동부의 해안 지대, 그리고 그 둘 사이의 인구가 희박한 산간지대로 뚜렷이 구분되어 있는 것이다. 강원도내의 시·군별 인구밀도를 살펴보면, 철원군(62)을 제외한 모든 군지역이 50명/㎢ 이하의 낮은 인구밀도를 보인 반면, 삼척시를 제외한 모든 시지역은 100명/㎢이상의 인구밀도값을 보였다. 특히 속초시는 825명/㎢으로 가장 높았고, 동해시(531), 원주시(307), 춘천시(219), 강릉시(219), 태백시(178)가 그 뒤를 이었다.

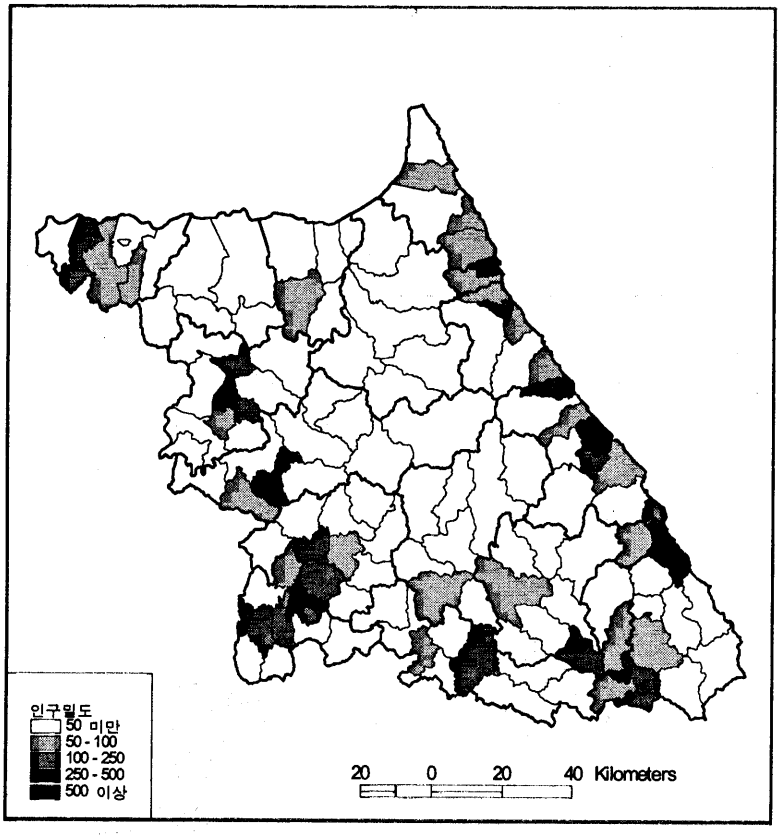
이러한 시·군 단위의 인구밀도 분포는 시군 단위 아래에서 발생하는 공간적 변이를 보여주지 못한다. 읍·면·동 수준에서의 인구밀도 분포를 나타낸 <그림 2>를 보면, 읍·면·동의 계층적 차이에 따른 인구밀도 분포 상의 차이를 명백히 보여주고 있다. 즉, 시지역의 중심부를 차지하고 있는 동지역은 높은 인구밀도를 보이는 반면, 읍면 지역은 낮은 인구밀도를 보여주고 있다. 위에서 언급한 수직적 3지대 분포패턴은 이 <그림 2>에서 보다 명확히 나타나고

5) 한국행정구역총람편찬회, 2001, 한국행정구역총람 2001-2002, 한국행정문화원.
6) 통계청, 2001, "2000 인구주택총조사 전수집계결과 (인구부문)", 보도자료, p.14.

있다. 여기에 덧붙여, 몇몇 도시 내부에서 인구분포의 특정한 패턴이 나타나고 있다는 사실에 주목할 필요가 있다. 예를 들어, 춘천시의 경우에서는 남북방향의 수직축을 중심으로 한 인구밀집 지역이 시역 내부에 한정되어 있고, 원주시의 경우에는 북북동-남남서 방향의 축이 횡성군까지 이어지고 있다.



<그림 1> 시·군별 인구밀도의 단계구분도



<그림 2> 읍·면·동별 인구밀도의 단계구분도

<표 1>은 가장 높은 인구밀도를 보이는 동지역과 가장 낮은 인구밀도를 보이는 면지역을 나열한 것이다. 여기에서 보면 춘천시는 인구밀도가 가장 높은 지역을 포함하고 있음과 동시에 인구밀도가 가장 낮은 지역 역시 포함하고 있다. 이것은 시·군별로 내적인 편차가 다양함을 함축하고 있는데, 시·군간 편차와 시·군내 편차를 비교함으로써 인구밀도 분포의 새로운 정보를 얻을 수 있다.

<표 1> 인구밀도에 의한 최상위 및 최하위 10개 동읍면

인구밀도가 가장 높은 동읍면			인구밀도가 가장 낮은 동읍면		
동읍면명	시군명	인구밀도(명/㎢)	동읍면명	시군명	인구밀도(명/㎢)
후평2동	춘천시	26,933	북산면	춘천시	4
후평3동	춘천시	22,535	가곡면	삼척시	6
원인동	원주시	20,082	노곡면	삼척시	6
교 동	속초시	19,968	왕산면	강릉시	7
명륜2동	원주시	18,599	방산면	양구군	7
명륜1동	원주시	14,582	내 면	홍천군	8
학성동	원주시	14,547	하장면	삼척시	9
일산동	원주시	14,357	수주면	영월군	9
조운동	춘천시	13,778	하동면	영월군	10
효자3동	춘천시	13,017	상남면	인제군	10

<표 2>는 시·군 내 인구밀도의 분산에 대한 통계치를 제공하고 있는데 마지막 열에 있는 변동계수(coefficient of variation)에 유의할 필요가 있다. 변동계수는 표준편차의 값을 평균으로 나눔으로써 구해지는데, 평균의 차이를 감안한 표준편차의 의미를 가지기 때문에 지역간 비교에 적절한 통계치이다. <표 2>를 보면, 전반적으로 시지역의 변동계수가 군지역에 비해서 높게 나타나 있는데, 이는 도농통합시의 개념이 도입되면서 한 시역내에 많은 비도시지역이 포함되었기 때문이다. 시·군간 변동계수는 1.41인데, 삼척시, 홍천군, 태백시, 횡성군, 양양군은 이 수치를 넘어서고 있다. 이는 이 지역들 내부에서의 인구 편포 정도가 시·군간의 인구 편포 정도에 비해 높음을 의미하는 것으로, 인구분포 상에서의 지역 내 동질성이 상대적으로 취약함을 의미한다.

<표 2> 행정구역 수준별 인구밀도 편차

시군명	동읍면 수	인구밀도(명/㎢)	평균밀도	표준편차	변동계수
춘천시	25	219	6281.37	7606.63	1.21
원주시	25	307	4921.20	6687.26	1.36
강릉시	21	219	2972.87	3596.78	1.21
동해시	10	531	1649.52	1554.43	0.94
태백시	8	178	534.94	794.19	1.48
속초시	8	825	5128.48	6396.52	1.25
삼척시	12	64	477.18	1011.94	2.12
홍천군	10	37	51.72	83.73	1.62
횡성군	9	41	49.74	72.00	1.45
영월군	9	39	37.60	41.67	1.11
평창군	8	28	27.81	12.88	0.46
정선군	9	38	51.86	49.23	0.95
철원군	7	62	59.50	41.16	0.69
화천군	5	22	22.19	9.37	0.42
양구군	5	29	29.37	23.04	0.78
인제군	6	19	17.92	6.73	0.38
고성군	5	46	53.13	24.40	0.46
양양군	6	46	101.39	145.48	1.43

인구분포의 공간적 특성을 파악하기 위해 인구밀도에 대한 단계구분도를 사용하는 것은 가장 용이하고 각 공간단위별로 인구밀도값을 비교할 수 있다는 측면에서 장점이 있지만, 다음의 두 가지 측면에서 문제점을 내포하고 있다[7]. 첫째, 행정경계가 실질적인 인구분포의 양상을 반영하지 못할 수 있다. 즉, 인구밀도가 각 공간단위 내에서 균등하게 분포할 것이라는 가정은 지지되기 어렵다. 둘째, 단계구분의 과정에 불확실성이 개입될 수 있다. 즉, 단계구분의 방식에 따라 파악되는 패턴이 달라질 수 있는 것이다.

2. 벡터-기반 시각화

1) 점묘도

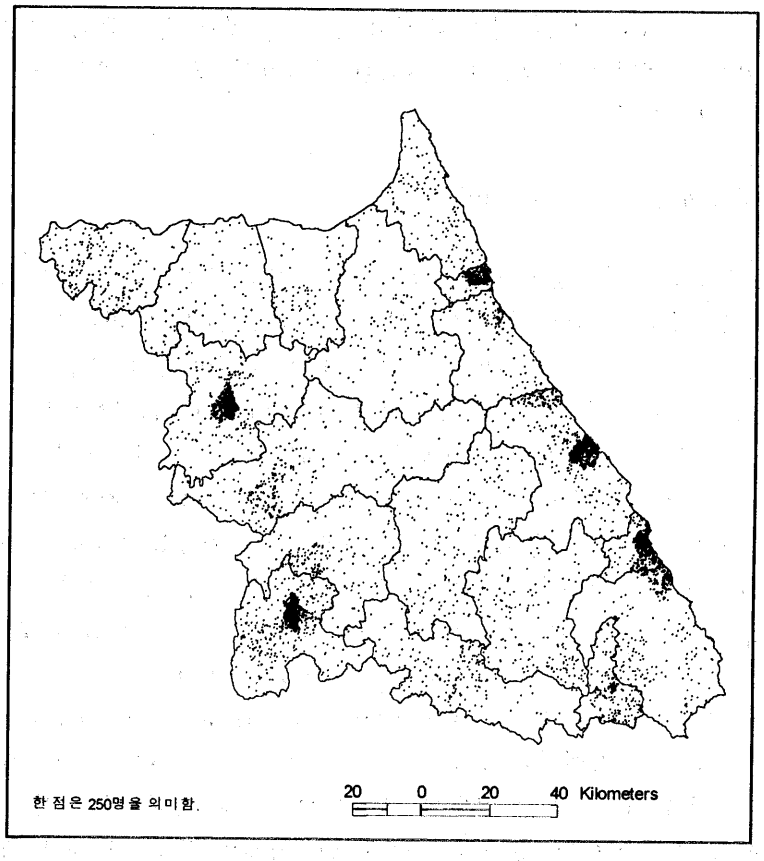
점묘도는 밀도 분포의 배열적 변이를 잘 보여준다는 측면에서 인구밀도를 표현하는 방법으로 가장 널리 쓰여 왔다. 域-단위의 인구 자료를 점묘도로 전환하기 위해서는 단위점의 수치(dot value)와 단위점의 크기(dot size)를 고려해야 한다[8]. 단위점의 수치는 하나의 점에 의해 표현되는 값인데, 인구밀도에 대한 점묘도의 경우는 몇 명의 인구가 하나의 점으로 표현될 것인가에 대한 것이다. 단위점의 수치와 크기는 서로 상호작용 하면서 다양한 점묘도로 표현되는데, 밀도분포에 대한 가독성을 극대화하는 점묘도를 획득하기 위해서는 여러 차례에 실습이 요구된다[9].

<그림 3>은 250명을 단위점의 수치로 하고 폰트 2를 단위점의 크기로 하여 작성된 강원도 인구분포에 대한 점묘도이다. GIS 프로그램은 읍·면·동의 인구를 250으로 나누어 각 행정구역에 들어갈 점의 개수를 결정한 후 임의적 점배치(random placement of dots) 방법을 통해 점묘도를 작성한다. 점의 배치가 임의적이기 때문에 동일한 자료에 대해서도 점묘도를 작성할 때마다 상이한 지도가 만들어 진다[10]. 따라서 점묘도 작성에서는 가능한 가장 낮은 수준의 공간단위에 대한 자료를 구하는 것이 매우 중요하게 된다.

<그림 3>과 <그림 2>를 비교해 보면, 점묘도의 특성을 잘 볼 수 있는데, 기본적으로 동일한 자료에서 작성된 지도이지만, 단계구분도에서는 각 공간단위의 수치 간의 차이가 설정된 계급에 의해 일반화되지만, 점묘도의 경우는 이러한 데이터의 상실이 발생하지 않는다. 따라서 밀도분포의 패턴 파악이라는 목

---

7) Martin, D., 1989, Mapping population data from zone centroid locations, *Transactions of the Institute of British Geographers*, 14, 90-97.
8) Dent, B. D.,1999, *Cartography: Thematic Map Design*, 5th Edition, New York: McGraw Hill, pp.165-166.
9) 이희연, 1995, 지도학 -주제도 제작의 원리와 기법-, 법문사, p.363.
10) Dent, *op. cit.*, pp.170-171.



<그림 3> 읍·면·동별 인구수의 점묘도

적에는 단계구분도보다 점묘도가 훨씬 유용한 시각화의 방법이다. <그림 3>을 통해 몇 가지 공간적 패턴이 확인될 수 있다. 첫째, 최고의 인구 집중 지역은 강릉시의 핵심부이다. 가장 넓은 지역에 걸쳐 높은 점 밀도를 보여 주고 있다. 둘째, 원주시가 가장 연속적인 인구밀도 분포를 보여주고 있다. 앞에서 언급된 북북동-남남서의 축을 따라 비교적 높은 인구밀도의 연속적 패턴이 확인되는 것이다. 이에 비해 춘천시와 강릉시는 핵심부에서 주변지역으로의 밀도분포가 상대적으로 단절적이다. 셋째, 몇몇 지역에서 선적인 연담화의 경향

이 나타난다. 특히, 동해시와 삼척시의 북부 지역은 뚜렷한 연담화의 패턴을 보여주고 있는데, 이러한 선적인 연담화 패턴은 강릉시의 핵심부를 지나 강릉시의 북부지역에 까지 약하게 연결되어 있다. 또한 속초를 중심으로 한 인구밀집지역이 북으로는 고성군의 남부지역, 남으로는 양양군의 북부지역과 연결되어 있다. 그리고 미약하지만 춘천의 중심부로부터 홍천군의 중심부 사이에서도 연담화의 패턴이 관찰된다. 이러한 연담화의 패턴은 레스터-기반 시각화에서 다시 다루어질 것이다.

2) 왜상통계지도(cartogram)

왜상통계지도란 지도상의 공간단위의 속성(면적 혹은 거리)이 그 공간단위가 가지고 있는 양적 크기에 따라 왜곡되어 표현되는 지도를 의미한다. 왜상통계지도에 대한 기본적인 아이디어의 탄생은 19세기 중엽까지 거슬러 올라가지만, 지도학적 맥락에서 체계적으로 정리한 사람은 미국의 지도학자 Erwin Raisz로 알려져 있다[11]. 이후 왜상통계지도는 지리학적 개념을 도식적으로 보여주는데 많이 이용되어 왔지만, 학문적인 의미에서의 진보는 거의 이루어지지 않았다. 하지만 최근 Dorling의 연구는 다시 한번 왜상통계지도에 대한 학문적 관심을 고조시키고 있다[12]. 일반적으로 왜상통계지도는 선형 왜상통계지도(linear cartograms)과 면형 왜상통계지도(area cartograms)로 나뉘어 지는데 후자는 다시 연속적 면형 왜상통계지도(contiguous area cartograms)와 비연속적 면형 왜상통계지도(noncontiguous area cartograms)로 나뉘어진다[13]. 본 논문이 작성하려고 하는 것은 바로 연속적 면형 왜상통계지도이다.

연속적 왜상통계지도는 비연속적 왜상통계지도에 비해 훨씬 복잡한 과정을 거쳐야 하기 때문에 GIS 상에서 구현가능한 형태의 알고리듬들이 개발된 것도 비교적 최근의 일이다[14]. 더 나아가 이러한 알고리듬을 특정한 GIS 상에서

---

11) *Ibid.*에서 재인용, p.209; Raisz, E., 1934, The rectangular statistical cartogram, *Geographical Review*, 24, 292-296.

12) Dorling, D., 1996, *Area Cartograms: Their Use and Creation*, Concepts and Techniques in Modern Geography, 59, Norwich: School of Environmental Sciences.

13) DeMers, M. N., 1997, *Fundamentals of Geographic Information Systems*, New York: John Wiley & Sons, pp.402-406.

14) 연속적 면형 왜상통계지도 제작을 위한 알고리듬은 개발자에 따라 조금씩 차이가 나는데, 다음의 논문들은 그 예가 될 수 있다. 특히 Kocmoud는 상이한 알고리듬에 대한 비교분석을 행하였다. Dougenik, J. A., Chrisman, N. R. and Niemeyer, D. R., 1985, An algorithm to construct continuous area cartograms, *Professional Geographer*, 37, pp.75-81; Tobler, W., 1985, Pseudo-cartograms, *The American Cartographer*, 13(1), 43-50; Guseyn-Zade, S. M. and Tikunov, V. S., 1993, A new technique for constructing continuous cartograms, *Cartography and Geographic Information Systems*, 20(3), 167-173; Kocmoud, C. J., 1997, *Constructing Continuous Cartograms: A Constraint-Based Approach*, Master's

실현하는 작업도 용이하지 않다[15]. 상업적 프로그램 상에서 연속적 왜상통계지도를 구현하려는 시도는 Jackel(1997)과 Du and Lin(1999)에 의해 시도되었는데, 둘 다 ArcView가 제공하는 스크립트 언어인 avenue로 위의 Dougenik등(1985)이 제시한 알고리듬을 구현한 것이다[16]. 이 알고리듬을 요약하면 다음과 같다.

첫째, 각 폴리곤의 중점(centroids)를 구한다.
둘째, 각 폴리곤의 속성값을 위에서 구해진 폴리곤의 중점에 부여한다.
셋째, 폴리곤의 중점들 간의 작용력(forces)을 속성값의 크기에 비례하여 계산한다.
넷째, 하나의 폴리곤을 선택한 후, 경계를 접하고 있는 폴리곤의 각 쌍에 대해, 위에서 계산된 작용력에 따라 경계 상의 버텍스(vertex)의 좌표를 변환한다. 주어진 반복(iteration) 횟수에 따라 순환과정(looping)이 계속된다.
다섯째, 특정한 준거값(tolerance)에 따라 왜상통계지도를 선택한다.

반복의 횟수가 많으면 많을수록 최상의 속성값에 따른 비례적 면적을 가진 왜상통계지도를 얻을 수 있지만, 폴리곤이 많은 버텍스로 구성되어 있을 경우 많은 시간이 소요된다는 단점이 있다.

<그림 4>는 18번의 반복을 통해 구해진 강원도 시·군별 인구수에 대한 왜상통계지도이다. 원지도와 비교해 보았을 때 각 행정구역의 면적이 인구수에 따라 비례적으로 왜곡되었음을 알 수 있다. 이 왜상통계지도의 정확성을 평가하기 위해 <표 3>를 작성하였다. 여기서 인구 비중이란 강원도 전체 인구수에서 각각의 시·군이 차지하는 비중을 나타낸 것이고, 면적 비중은 왜상통계지도의 전체 면적에서 각 시·군이 차지하는 비중을 나타낸 것이다. 그리고 마지막 열에 있는 비중 오류는 인구 비중과 면적 비중의 절대차를 의미한다. 완벽한 왜상통계지도라면 이 비중 오류는 0이 될 것이다. 비중 오류를 모두 합하여 시·군의 개수로 나누면 평균오류값을 얻을 수 있는데, 일반적으로 평균오류값이 10% 미만이면 왜상통계지도로서 받아들여진다[17]. 이렇게 보았을 때, 평균오류

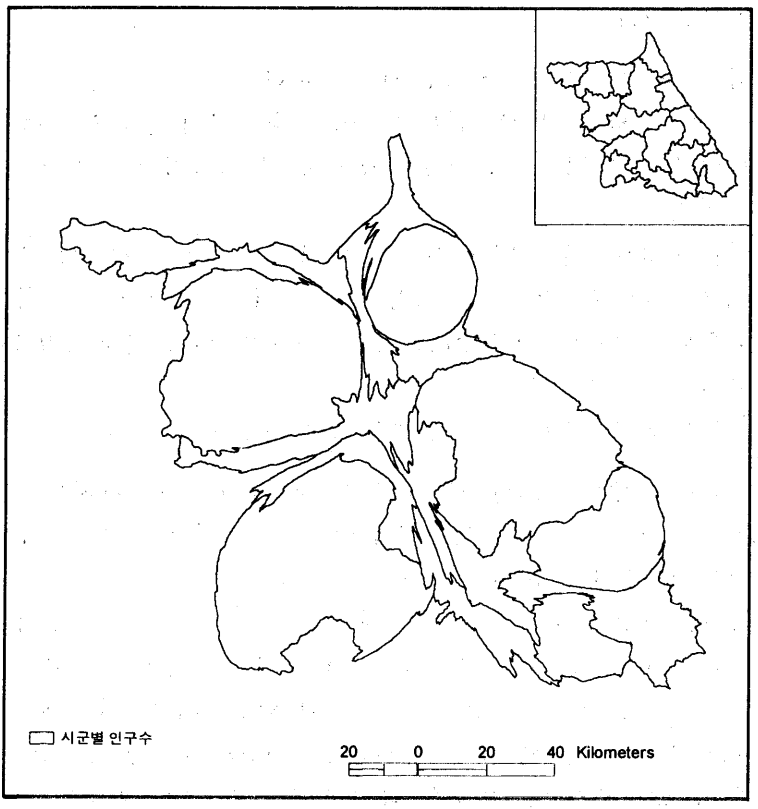
Thesis, Texas A&M Visualization Laboratory, Texas A&M University.

15) 이러한 이유로 다음의 논문에서 소개되는 프로그램은 GIS와의 연동이 없는 독립적인 소프트웨어로 개발되었다. Keim, D. A., North, S. C., and Panse, C., 2001, CartoDraw: A fast algorithm for generating contiguous cartograms, AT&T Shannon Laboratory, Florham Park, New Jersey, Working Paper.

16) Jackel, C. B., 1997, Using ArcView to create contiguous and noncontiguous area cartograms, *Cartography and Geographic Information Systems*, 24(2), 101-109; Du, C. and Lin, L., 1999, Constructing contiguous area cartogram using ArcView Avenue, *Geoinformatincs and Socioinformatics*, Proceedings of Geoinformatic '99 Conference, Ann Arbor, Michigan, 19-21 June, 1-7.

17) Jackel, *op. cit.*, p.104.

0.05%는 매우 낮은 것으로 적절하게 제작된 것으로 평가될 수 있다.



<그림 4> 시·군별 인구수의 왜상통계지도

3) 3-D 시각화

데이터를 3차원적으로 표현한다는 것은 속성값을 마치 자연적인 고도처럼 높이로 환산하여 표현한다는 것을 의미하는 것으로, 특정한 형태의 통계면(statistical surface)을 구성하게 된다. 데이터가 불연속적일 경우, 즉 공간단위별로 속상값이 주어진 경우에는 계단형 통계면(stepped statistical surface)을 작성하는 것이 바람직하다[18].

18) Kraak & Ormeling, *op. cit.*, p.142.

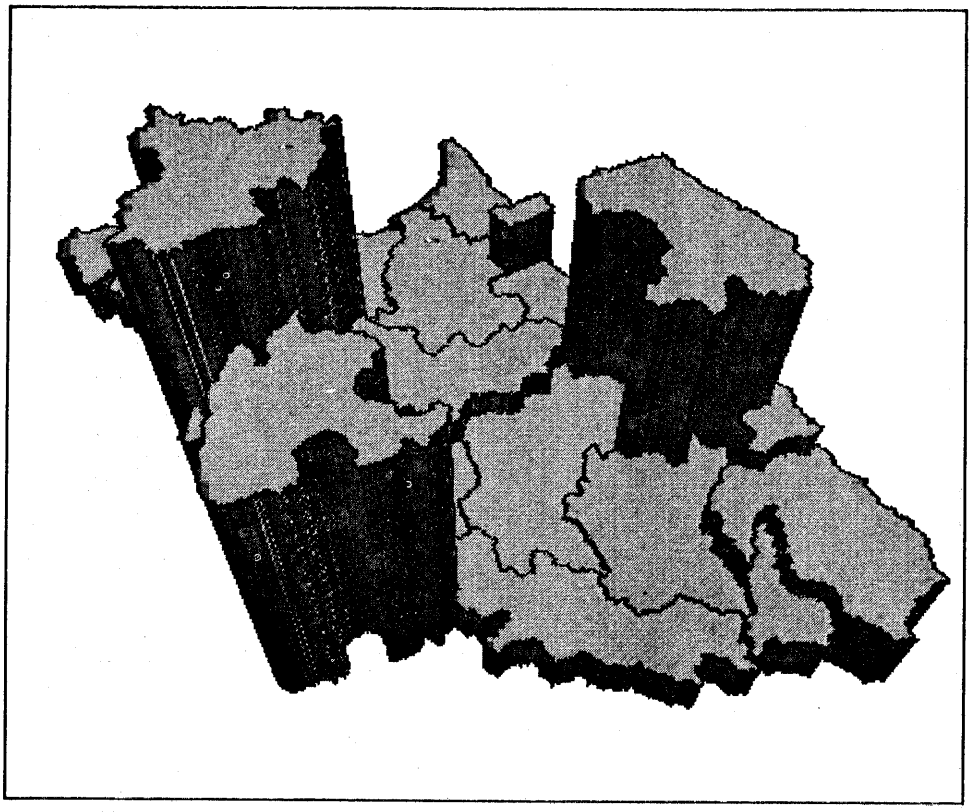
<그림 5>는 ArcView 3D Analyst를 이용하여 그린 3-D 계단형 통계면이다. 높이 값은 인구수에 비례하는데, 시각화의 효과를 높이기 위해 과장치(exaggeration value)를 조정하였다. 위의 왜상통계지도와 마찬가지로 지도로부터 정보를 인식하는데 걸리는 시간을 단축시켜 준다는 점에서 이러한 3D 지도의 강점이 있다.

<표 3> 인구수에 비례한 면적의 왜곡

시 군 명	인구 비중(%)	면적 비중(%)	비중 오류(% error)
춘 천 시	16.98	16.95	0.03
원 주 시	18.05	17.99	0.05
강 릉 시	15.35	15.40	0.06
동 해 시	6.50	6.47	0.03
태 백 시	3.64	3.51	0.13
속 초 시	5.92	5.89	0.02
삼 척 시	5.09	5.24	0.15
홍 천 군	4.52	4.54	0.02
횡 성 군	2.76	2.89	0.13
영 월 군	2.94	3.00	0.06
평 창 군	2.77	2.61	0.16
정 선 군	3.10	3.07	0.03
철 원 군	3.23	3.20	0.03
화 천 군	1.53	1.56	0.03
양 구 군	1.42	1.42	0.01
인 제 군	2.07	2.11	0.04
고 성 군	2.17	2.18	0.01
양 양 군	1.96	1.96	0.00
합 계	100.00	100.00	0.98
평균오류(mean error)	0.05		

3. 레스터-기반 시각화

레스터-기반 시각화는 벡터-기반 시각화와는 달리 연속적 데이터를 표현하는 방법으로 개발되어 온 것이다. 따라서 행정구역처럼 불연속적인 데이터를 시각화하는데 레스터-기반 시각화의 방법을 적용하는데는 기본적인 한계와 위험성이 있다. 하지만, 인구와 같은 인문사회적 데이터의 경우에도 데이터 구성 자체는 불연속적이라 하더라도 그 현상 자체는 연속성의 가정이 가능한 경우가 많다. 다시 말하면, 인구밀도의 공간분포는 강수량의 공간분포 만큼 연속적이지는 않다 하더라도 인구밀도가 높은 지역에서 주변으로 갈수록 인구밀도가 점차적으로 낮아지는 특정한 형태의 '공간적 의존성(spatial dependence)'이 존재한다는 점은 강조되어야 한다.



<그림 5> 시·군별 인구수의 계단형 통계면

불연속적인 데이터에 레스터-기반의 연속적인 시각화 기법을 적용하기 위해서는 우선적으로 폴리곤을 포인트로 전환해야만 한다. 이러한 형태의 공간객체유형 변환에서 가장 많이 쓰이는 것이 폴리곤의 중점(centroid)를 구하는 것이다. 폴리곤의 중점을 구하는 방식에는 몇가지 방법이 있는데, 여기서는 폴리곤을 다수의 삼각형으로 전환하고 각 삼각형의 무게중심의 면적가중평균법에 의해 중점의 좌표를 계산하는 방식을 택했다[19]. 생성된 중점에 폴리곤의 속성을 부여하면, 다양한 형태의 레스터-기반 시각화 도구들을 사용할 수 있게 된다.

1) Kernel 밀도면(density surface)

19) O'Rourke, J., 1998, *Computational Geometry in C*, 2nd Edition, New York: Cambridge University Press, p.43; 여타의 보다 간편한 방식에 의한 것보다 훨씬 정확한 도형의 중점을 계산해 낸다.

域-기반 자료를 이용하여 밀도면을 형성하는 시도는 1980년대 후반에 이루어졌다[20]. 공간통계학적 측면에서 보면 밀도면을 형성하는 방법은 사실상 무한하다. 본 논문에서는 그 다양한 방법 중 kernel smoothing에 의한 밀도면 형성에 대해서만 논하고자 한다. Kernel이란 이변량 확률밀도함수(bivariate probability density function)를 의미하는 것으로, 일반적으로 점분포패턴 분석에 쓰이는 것이지만, 기본 함수식에 속성을 부과하면 일종의 공간적 smoothing 기법으로 사용될 수 있다[21]. 이변량 확률밀도함수가 의미하는 바는 한 지점으로부터 2차원상에서 구성되는 모든 방향으로 거리가 멀어질 수록 확률이 감소하는 경향을 묘사하는 것이다. 이것이 공간분석에서 중요한 것은 kernel 함수가 일종의 거리조락함수를 제공하기 때문이다. 즉 거리에 따른 상대적인 영향력을 계산해 낼 수 있는 것이다. Kernel 함수는 다양한데[22], 여기서는 이변량 정규분포밀도함수(Gaussian kernel)를 이용하였다. Kernel 함수를 이용해서 밀도면을 구하는 과정을 간략히 정리하면 다음과 같다.

첫째, 읍·면·동의 중점을 구한다.

둘째, 강원도 전지역을 1㎢의 셀로 구획한다.

셋째, kernel 함수를 적용하기 위한 bandwidth를 10km로 설정한다.

넷째, GIS 프로그램은 각 셀의 중심점으로부터 10km의 반경 내에 포함되는 중점을 찾아내여, 셀 중심점으로부터 각 중점까지의 거리를 계산하여, kernel 함수에 의거하여 가중치를 부여한다. 가까운 거리의 중점에는 높은 값을 부여하고(1에 가까운 값), 멀리 떨어진 중점에는 낮은 값을 부여한다(0에 가까운 값). 가중치 벡터는 나중에 행-표준화(row-standardization) 과정을 거친다.

다섯째, 부여된 가중치와 속성값(인구수)을 곱하여 모두 더하고 셀의 면적을 고려하여 그 셀에 부여될 인구밀도값을 계산한다.

알고 있는 지점의 속성값으로부터 모르는 지점의 속성값을 유추해 낸다는 측면에서 kernel에 의한 밀도면의 생성은 공간적 내삽(spatial interpolation), 특히 kriging과 매우 유사하다. 域-기반 인구 자료의 공간 분포에 대한 시각화를 위해 밀도면을 사용하는 데는 두 가지 장점이 있다. 첫째, 위에서 언급한 벡터-기반 단계구분도의 한계를 극복할 수 있다[23]. 둘째, 공간객체유형의

---

20) Martin, D., 1989, *op. cit.*; Bracken, I. and Martin, D., 1989, The generation of spatial population distributions from census centroid data, *Environment and Planning A*, 21, 537-543.

21) Bailey, T. C. and Gatrell, A. C., 1995, *Interactive Spatial Data Analysis*, Harlow, Essex: Longman, p.85.

22) Fotheringham, A. S., Brunsdon, C. and Charlton, M., 2000, *Quantitative Geography: Perspectives on Spatial Data Analysis*, Thousand Oaks, CA: SAGE, pp.45-48. 다양한 형태의 kernel 함수에 대해 도해되어 있다.

전환에서 발생하는 위험성도 있지만, 인문사회적 자료를 연속적인 면으로 표현하는 것이 패턴인식에 보다 더 효과적인 시각화 기법일 수 있다[24]. 그러나 행정구역으로부터 중점을 계산하고 그 중점에 부여된 인구수를 가지고 밀도면을 생성하는 이러한 방법론이 공간적 분포에 대한 보다 의미 있는 재현이기 위해서는 센서스 공간단위에 대한 보다 면밀한 고려가 요구된다. 단적으로 말해 센서스 자료는 인구 규모, 공간적 형태, 그리고 사회인구학적 구성에서 표준화되어 제공될 필요가 있다[25]. 또한, 센스서 자료의 정보기밀성(confidentiality)에 위배되지 않는 범위에서 보다 작은 규모의 공간단위에 대한 정보가 제공되어야 한다. 이는 생성된 밀도면의 신뢰성과 밀접한 관련이 있을 뿐만 아니라 센서스 자료의 사용자가 공간단위를 합역하여 보다 의미 있는 '사용자 지리(user geographies)' 를 생성할 수 있게 하기 때문이다[26]. 우리나라에서도 이러한 측면에 대한 논의가 일부 이루어지고 있다[27].

<그림 6>은 작성된 밀도면을 나타낸 것이다. 이 밀도면은 앞의 벡터-기반 시각화에 비해 패턴 확인이라는 측면에서 더 우월한 시각화 기법임을 보여주고 있다. 단계구분도인 <그림 2>나 점묘도인 <그림 3>과 비교해 보면 그러한 측면을 확인할 수 있다. 우선적으로 관측되는 패턴은 앞에서도 언급되었듯이 선적인 연담화의 경향에 대한 것이다. 가장 뚜렷한 연담화의 패턴은 원주시의 남서부에서 중심부를 지나 횡성군의 남부로 이어지는 축이다. 이외에 동해안을 따라 삼척시 북부-동해시-강릉시-양양군 남부로 이어지는 축, 양양군 북부-속초시-고성군 남부로 이어지는 축, 그리고 춘천시 중심부-홍천군 중심부로 이어지는 축 등이 확인된다. 하지만 이러한 연담화 패턴에 대한 해석에는 위험성이 상존한다. 기본적으로 kernel 함수에 의한 밀도면은 원자료의 공간적 평균화를 통해 생성된 것이기 때문에 현실과의 차이를 항상 염두에 두어야 한다. 또한 결과된 패턴은 초기에 설정된 bandwidth의 값과 선택된 kernel 함수에 많은 영향을 받는다. 따라서 밀도면에는 정답이 존재하지 않으며 연구자의 연구목적에 따라 다양한 실험을 통해 도출되어야만 하고, 오로지 탐색적(exploratory) 목적에서만 사용되어야 한다.

---

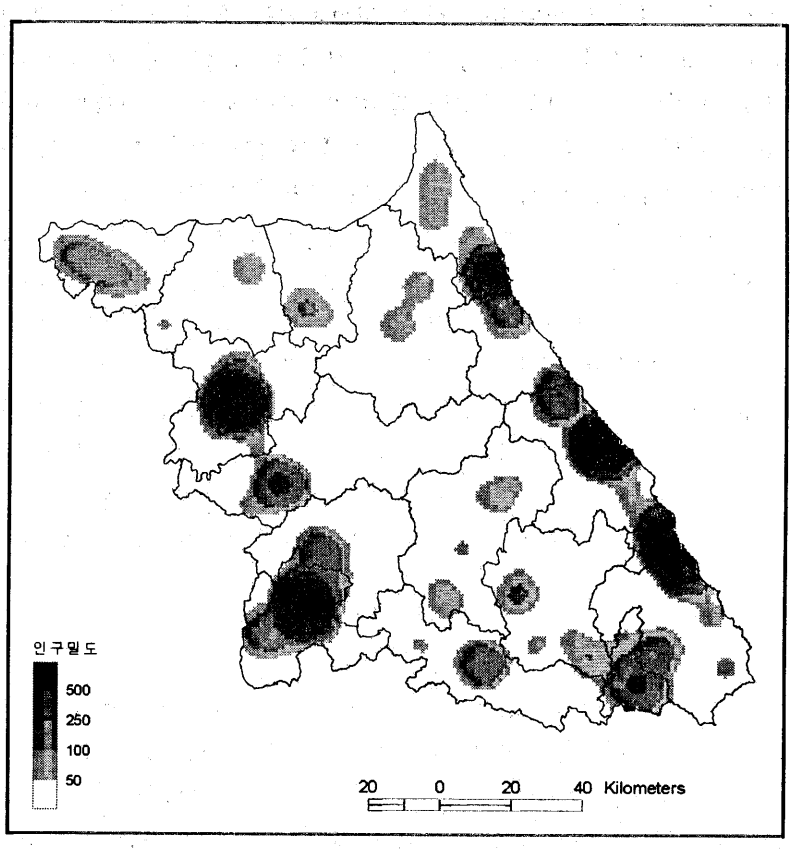
23) Martin, D. and Bracken, I., 1991, Techniques for modelling population-related raster databases, *Environment and Planning A*, 23, p.1069.

24) Langford, M. and Unwin, D. J., 1994, Generalising and mapping population density surfaces within a geogrpahical information system, *The Cartographic Journal*, 31, 21-26.

25) Martin, D., 1998, Optimizing census geogrpahy: the separation of collection and output geographies, *International Journal of Geographical Information Science*, 12(7), p.678.

26) Martin, 2000, *op. cit.*, p.329.

27) 구자문, 1999, "도시분석을 위한 인구주택센서스와 GIS의 연계활용방안 연구: 수치지도의 보완과 센서스트랙의 결정," 한국지리정보학회지, 2(2), 27-44; 김영표 · 한선희, 1999, "GIS를 이용한 국가 통계정보의 활용도 제고방안," 국토연구원.



<그림 6> 읍·면·동 수준의 kernel 밀도면

2) 경향면(trend surface)

경향면은 전반적인 패턴 확인을 위한 기법 중에서 지리학에서 가장 널리 이용되어 왔다[28]. 경향면은 일종의 공간적 내삽 기법으로 인식되고 있지만, 여타의 기법에 비해 가장 저급한 형태라고 볼 수 있다. 기본적으로 경향면은 속성값을 좌표값의 다항식(polynomial equation)으로 표현할 수 있다는 가정에

28) Unwin, D., 1975, *An Introduction to Trend Surface Analysis*, Geo Abstracts.

서 출발한다. 보다 통계적으로 말하면, 속성값을 종속변수로 좌표값을 독립변수로 한 중다회귀분석을 통해 연구지역내의 모든 지점에 대한 속성값을 추정하는 것이다. 경향면의 차수가 높아지면 보다 정확한 예측에 도달할 수 있지만, 그러한 복잡한 회귀식이 의미하는 바를 해석하기는 몹시 어렵다. 또한 차수가 높아지면서 특정 지역에서 아주 높은 음수값이 나타나는 등 현실에 대한 정확한 재현으로 보기에는 어려운 측면이 많다[29]. 따라서 비록 경향면 분석이 통계적 유의성에 대한 검정적(confirmatory) 통계치를 산출해 준다하더라도, 전반적 패턴의 파악이라는 탐색적 목적으로만 사용되는 것이 타당할 것이다.

<그림 7>은 인구밀도 분포에 대한 4차 경향면을 표현한 것이다. 4차 경향면은 강원도 읍·면·동로부터 인구에 의한 가중중심점(weighted mean center)을 각 시·군에 대해 구한 후, 각 가중중심점의 x, y자표로부터 구성되는 15개의 회귀계수에 의해 예측된 인구밀도값을 나타낸 것이다[30]. 차수가 낮을수록 단순한 패턴이 주어지고 차수가 높을수록 현실에 가까운 복잡한 패턴이 도출된다. 단순하면서도 실제와 유사한 패턴을 산출하는 차수를 고르는 작업이 경향면 분석에서는 가장 중요하다. <그림 7>의 4차 경향면은 강원도 인구분포의 가장 단순한 패턴인 수직적 3지대 형태를 잘 표현해 주고 있다.

3) 3-D 시각화

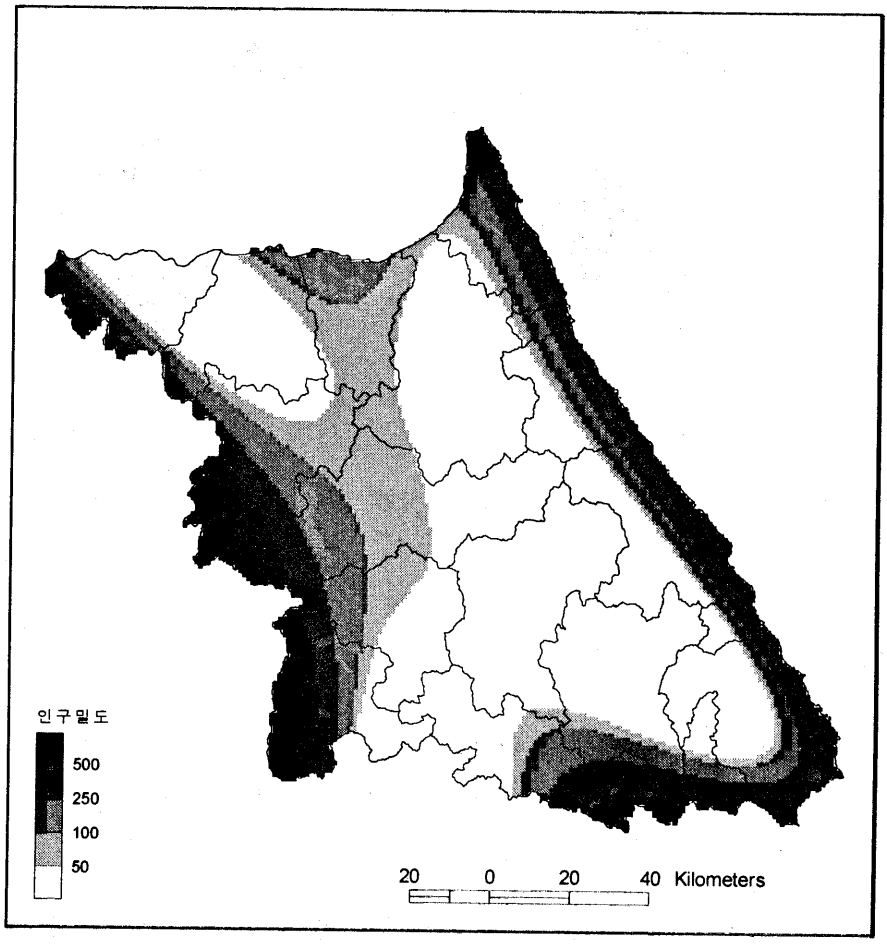
<그림 8>은 ArcView 3D Analyst를 이용하여 작성한 kernel 밀도면의 3차원 지도이다. 계단형 통계면과 마찬가지로 시각적 효과를 위해 과장치를 조정하였다. 인문사회적 자료를 공간적 내삽을 포함한 다양한 기법을 통해 레스터로 전환하고 그것을 다시 3D 시각화로 표현하는 것은 GIS의 인문지리학적 사용에 있어 중요한 위치를 차지하게 될 것이다[31].

---

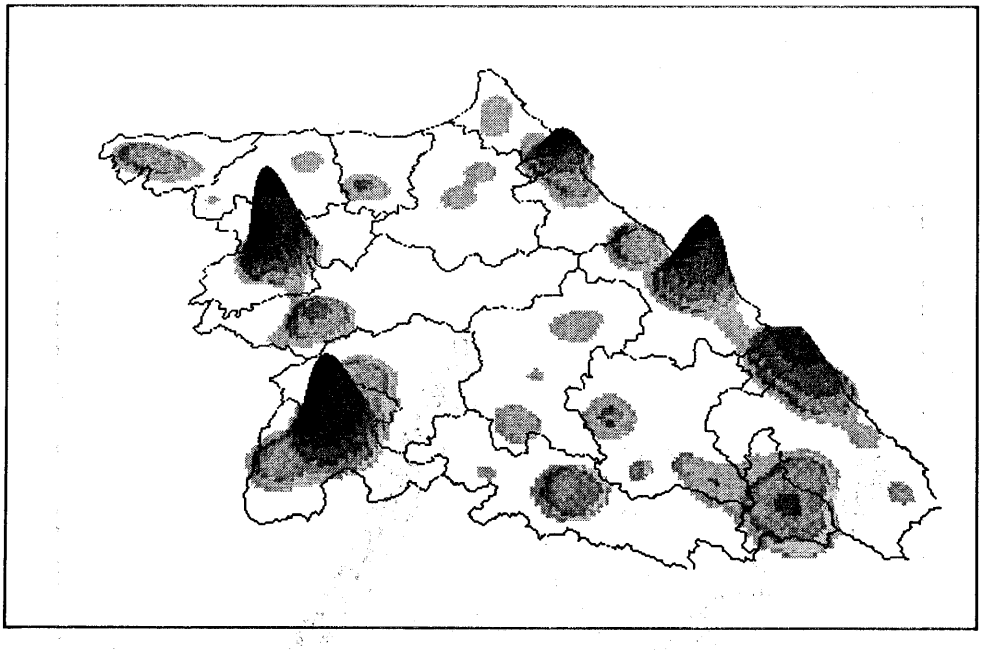
29) Burrough, P. A. and McDonnell, R. A., 1998, *Principles of Geographical Information Systems*, New York: Oxford University Press, pp.108-109.

30) p차 경향면에서 도출되는 회귀계수의 수는 $(p+1)(p+2)/2$의 공식에 의해 구해진다. *Ibid.*, p.109.

31) 이러한 측면에서 오충원은 안양시의 지가분포에 대한 연속면을 3차원으로 표현한 바 있다. 오충원, 2002, "시·공간분석을 위한 GIS기법의 시간 지도 구현에 관한 연구," 대한지리학회지, 37(2), 191-202.



<그림 7> 시·군 수준의 인구밀도 4차 경향면



<그림 8> Kernel 밀도면의 3차원 표현

Ⅲ. 결 론

본 논문은 다양한 지도학적 시각화가 GIS 상에서 어떻게 구현되는지를 2000년 강원도 인구분포를 사례로 살펴본 것이다. 벡터-기반 시각화 기법으로 점묘도, 왜상통계지도, 계단형 통계면이 소개되었고, 레스터-기반 시각화 기법으로 kernel 밀도면, 경향면, 그리고 3차원 연속 통계면이 예시되었다. 각각의 지도가 제작되는 과정에 대한 간략한 설명이 주어졌으며, 각 기법의 장단점도 논의되었다.

지도학적 시각화는 단순히 공간정보를 다양한 기법으로 표현하는 것[32]에서부터, 데이터를 분석하고 탐색하는 도구로서 그 중요성을 확장하고 있다. 즉, 특정한 지도학적 시각화는 그 기법이 사용되지 않았을 때는 발견될 수 없는 새로운 정보의 획득을 가능하게 한다. 바로 이 지점에서 지도학적 시각화는 공간분석의 새로운 조류인 '탐색적 공간자료분석(ESDA; exploratory spatial data analysis)' 과 조우하게 된다[33]. GIS에 기반한 탐색적 공간자료분석

32) 이기석·노희방 편저, 1994, 지도로 본 서울, 서울시정개발연구원, 성지문화사, 서울; Brewer, C. A. and Suchan, T. A., 2001, *Mapping Census 2000: The Geography of U.S. Diversity*, Redlands, CA: ESRI Press.

의 틀 속에서, 지도학적 시각화는 연구의 최종적인 단계로 인식되어온 지도 제작을 넘어서, 공간 패턴의 확인에서 공간 과정에 대한 가설의 설정에 이르는 공간자료분석의 모든 단계에 지속적으로 개입하는 새로운 위상을 차지하게 된다[34].

域-기반 인문사회적 자료를 레스터-기반 시각화 기법으로 재현하는 것은 매우 의미있는 작업으로 판단된다. 이러한 작업을 보다 발전시키기 위해 다음의 세 가지가 지적될 수 있다. 첫째, 여타의 공간 자료를 센서스 자료와 연결시킨다면 보다 정확한 레스터 데이터를 생성할 수 있을 것이다[35]. 예를 들어 행정단위의 중점을 계산하여 밀도면 형성의 기본 위치로 삼는 것 외에 인공위성 영상으로부터 토지피복 상태에 대한 정보를 첨가한다면, 보다 현실에 부합되는 밀도면을 생성할 수 있을 것이다. 둘째, bandwidth를 설정하는 문제에 있어 보다 심도 있는 연구가 필요하다. Bandwidth는 국지적 차원에서의 공간적 의존성에 기초하는 것이기 때문에 동일한 bandwidth를 모든 데이터 포인터에 적용하는 것은 옳지 않다[36]. 또한 bandwidth를 원으로 구성해야하는 하등의 이유가 없다. Bandwidth를 원으로 구성하고 반경을 설정하는 것은 공간적 패턴에서 방향성이 존재하지 않음(isotropy)을 가정하는 것이다. 그러나 많은 공간적 패턴은 방향의존성(anisotropy)을 보여주고 있으며, 국지적 bandwidth를 설정하는데 있어 원이 아닌 타원(ellipses)을 모든 데이터 포인터에 적용하는 것이 보다 타당하다[37]. 셋째, 제작된 밀도면을 여타의 자연환경에 대한 레스터 데이터와 결합한다면, 인구분포에 대한 다양한 해석을 할 수 있을 것이다. 예를 들어, 인구밀도 분포와 고도나 경사도와의 관계에 대한 정량적인 분석이 가능해질 것이다.

---

33) MacEachren, A. M., Kraak, M.-J., 1997, Exploratory cartographic visualization: advancing the agenda, *Computers & Geosciences*, 23:4, 335-343; Dykes, J. A., 1998, Cartographic visualization: exploratory spatial data analysis with local indicators of spatial association using Tcl/Tk and cdv., *Journal of The Royal Statistical Society Series D-The Statistician*, 47:3, 485-497; Kraak, M. J. and MacEachren, A., 1999, Visualization for exploration of spatial data, *International Journal of Geographical Information Science*, 13:4, 285-287.

34) O'Sullivan, D. and Unwin, D. J., 2003, *Geographic Information Analysis*, Hoboken, NJ: John Wiley & Sons, p. 19. Figure 1.2.

35) Martin, D., Tate, N. J., and Langford, M., 2000, Refining population surface models: experiments with Northern Ireland census data, *Transactions in GIS*, 4(4), 343-360.

36) Brunsdon, C., 1995, Estimating probability surfaces for geographical point data: an adaptive kernel algorithm, *Computers & Geosciences*, 21(7), 877-894.

37) Rosenberg, M. S., 2000, The bearing correlogram: a new method of analyzing directional spatial autocorrelation, *Geographical Analysis*, 32:3, 267-278; Paez, A. 2003, Anistropic kernel functions in geogrpahically weighted regression models, *Geographical Analysis*, on review.

GIS-Based Cartographic Visualization Methods for Exploring the Population Distribution of Kangwon in 2000

Sang-Il Lee

Summary

This paper is concerned with presenting and exploring spatial patterns of population distribution in Kangwon using GIS-based various cartographic visualization techniques. The rapid development of GIS obviously fueled by remarkable advances in computer sciences allowing for dealing with massive and digitally transferred data has increasingly reshaped not only the practice of geographical research, but the nature of geographical inquiries itself. In this sense, cartographic visualization techniques provided by GIS have attracted much attention from those who seek to effectively present, analyze, and explore spatial patterns of socioeconomic attributes including census data.

In this paper, cartographic visualization techniques are classified into two categories, vector-based and raster-based. Vector-based methods include dot maps, contiguous area cartograms, and stepped statistical surfaces along with regular choropleth maps. Raster-based techniques encompass kernel density surfaces, trend surfaces, and 3-D representation of kernel density surfaces. Basic notions and pros and cons of each technique are provided and discussed. All these methods are applied to population data of Kangwon extracted from the 2000 census to present and explore spatial patterns of population distribution. This application concludes that GIS-based cartographic visualization techniques are quite useful in detecting spatial patterns, and that they can be extended to other

socioeconomic data.

It should be noted that cartographic visualization goes beyond simply presenting data outputs usually named thematic mapping or map design to get involved in the whole process of geographical research, ranging from pattern detection to hypothesis formulation, besides presenting research results. This recognition of cartographic visualization as an active part in research has substantially contributed to the advent of a new paradigm in spatial analysis, *exploratory spatial data analysis* (ESDA). More advances should be made especially in producing raster-based surfaces from area-based socioeconomic data. First, density surfaces calibrated around area centroids can be rectified by utilizing other data sources such as satellite images. Second, kernel bandwidth could be locally adaptive and elliptical to embrace the directionality of spatial patterns. Third, surfaces of socioeconomic attributes could be overlaid by natural and environmental rasters to allow for investigating some forms of human-nature interactions.