

## GIS를 이용한 주택가 분석 모델에 관한 연구

강 영 옥\*

### An Empirical Study of Housevalue Using GIS

Young-ok Kang

#### 요 약

GIS는 지난 몇 십년간 컴퓨터 하드웨어 처리기술 발달과 가격인하로 인해 급속히 발달해 왔다. 그러나 GIS는 지도제작산업이나 시설물 관리와 같은 단순한 정보처리 면에서는 탁월한 기능을 보였으나 분석능력면에서는 아직도 미약한 상태이고 또 한편으로는 현실세계의 복잡성을 제대로 반영하지 못한다는 비판을 받고 있다. 본 연구에서는 첫째, 통계분석을 이용한 주택가 분석모델을 GIS에 연계시켰으며, 둘째, 다핵구조적 도시특성과 GIS를 이용한 도로망에 따른 거리계산등 현실세계의 복잡성을 주택가 분석모델에 반영하였으며, 셋째 서로 다른 분석단위(센서스 트랙과 블럭 그룹) 데이터 사용에 따른 주택가격모델의 다양성을 검증해 보았다.

**ABSTRACT:** GIS has been developed very rapidly during last few decades and the performance of GIS in terms of information processing such as automated mapping and facility management has been tremendous. However, its analytical capability is still very limited and it is often criticized due to lack of reality. The objective of this research is that first, linking housevalue models to GIS, second, reflecting the complexity of real world into the housevalue model using GIS in terms of incorporating polycentric urban structure and calculating distance through street network, and third, comparing the results of housevalue model at census tract to that of block group level.

\* 서울시정개발연구원(Seoul Development Institute, 171, Samsung-dong, Kangnam-ku, Seoul 135-091, Korea, Tel.(02)550-1181)

## 서론

GIS는 지리적 공간과 관련된 데이터를 입력, 저장, 분석, 검색 그리고 디스플레이하는 컴퓨터시스템이다 (Burrough, 1986; Marble et al.,1984). GIS 기술은 지난 몇 십년간 컴퓨터 하드웨어의 처리기술 발달과 가격의 인하로 급속히 발달해 왔으며 지도제작(Automated Mapping, AM) 산업이나 시설물 관리(Facility Management, FM)와 같이 단순한 작업을 위해서는 매우 유용하게 사용되었다.

그러나 현재 GIS를 이용한 분석이라는 것은 지도중첩(polygon overlay), 지형분석(terrain analysis), 그리고 기본적인 네트워크 분석(network analysis)만이 가능한 정도로서 GIS의 제한된 분석능력에 대해서는 많은 연구자들에 의해 지적되어온 사항이다.(Densham and Goodchild, 1989; Ding and Fotteringham,1992; Goodchild et al., 1992; Openshaw, 1990;1991)

GIS 연구에 있어 또 다른 지적은 GIS가 복잡한 현실세계를 반영하지 못한다는 사실이다. GIS 연구의 초점은 정보관리 측면에 치중해 왔으며 때로는 정보처리의 편의를 위해 현실세계를 단순화시킨 것도 사실이다. 예를 들어 거리계산에 있어서도 상용되어지는 GIS 소프트웨어의 거리계산은 주로 직선상의 거리 계산 능력만을 제공할 뿐이다. 그러나 실제 사람들의 교통행태를 볼 때 도심이나 기타 가고 싶은 장소로의 이동은 가로를 따라 행해지지 출발지와 목적지간 직선상의 거리개념이란

현실적이지 못한 것이다. 이와 같은 분석능력의 부족과 현실세계의 지나친 단순화는 현재 GIS가 당면하고 있는 주요문제들 중의 하나이다.

본 연구는 GIS의 이러한 문제점들이 주택가 분석 모델을 사례로 어떻게 다루어 질 수 있나를 살펴보고자 한다. 이 연구의 첫번째 목적은 통계 분석 방법을 사용한 주택가 분석 모델을 GIS에 연계시켜 보고자 하는 것이다. 도시내 주택가격에 대한 연구는 주택가격이 대지면적, 방의 수, 욕실의 수 등 주택의 물리적 특성뿐 아니라 학군이나 주변여건 등의 사회경제적 요소 그리고 접근성 요소 등 다양한 요인들에 의해 설명되어지기 때문에 주로 다변량분석에 의한 방법을 사용해 왔다. 그러나 현재 GIS는 이러한 통계분석을 GIS 내에서 하나의 모듈로서 지원하고 있지 않다. 그리고 사회과학에서 통계분석은 상당한 수준으로 발전돼 있는 반면 GIS는 분석보다는 정보처리 기능에 초점을 두어 이러한 두개의 기능들이 직접 연계되어 있지 않은 형편이다. 모델과 GIS를 연계시키는 방법은 직접연계(strongly-coupled)로부터 간접연계(loosely-coupled)까지의 연속체로 볼 수 있다(Batty and Xie, 1994). 직접연계란 소프트웨어간 데이터를 전송할 필요없이 한 시스템에 다른 하나의 기능을 첨가하는 것으로 GIS 내에 모델을 연계시키는 것(예: ESRI, location-allocation and spatial interaction model in ARC/INFO, Batty and Xie, 1994)으로부터 모델내에 GIS를 연계시키는 것(예: Hunt and Simmonds,1993 ; Putman,1992)까지를 의미

한다. 간접연계란 모델에 사용되며, 모델에 의해 만들어진 데이터를 GIS에 보내거나 받아들이는 것을 말한다(예: Wiggins and Ferreira, 1993 ; Densham, 1991). 본 연구의 첫번째 목적은 사회과학에서 널리 쓰이고 있는 통계분석 방법을 GIS에 연계시켜 보고자 하는 것이다.

두번째 연구목적은 주택가에 대한 모델정립에 있어 GIS를 이용하여 복잡한 현실을 반영하고자 하는 것이다. 즉 1980년대 초반까지만 해도 대부분의 도시는 단핵구조(monocentric structure)를 가졌으나 도시의 교외화, 기능의 다양화와 함께 최근의 도시구조는 다핵구조(polycentric structure)를 나타낸다. 또한 접근성(accessibility)의 개념은 도시구조를 나타내는 중요한 지표중의 하나인데 GIS를 이용하여 직선상의 거리가 아닌 도로를 따르는 거리를 계산하여 주택가격 분석 모델 정립에 있어 복잡한 현실을 반영하고자 하는 것이다.

세번째 연구목적은 서로 다른 분석단위 데이터 사용에 따른 주택가격모델의 다양성을 검증해 보려는 것이다. 미국에서 기존의 많은 도시연구의 기초단위는 센서스 트랙(tract)이었다. 이는 센서스 트랙이 사회·경제적 또는 인구학적 동질성을 나타내는 적정 단위여서라기 보다는 보다 세분된 단위의 데이터를 얻기

힘들었기 때문이었다. 그러나 미국의 통계청(Census of Bureau)에서는 이 보다 좀 더 세분된 블럭그룹(block group) 단위의 데이터를 제공하고 있는데 1990년의 센서스 트랙과 블럭그룹 단위 데이터를 이용해 주택가를 분석해 봄으로써 데이터 분석 단위에 따른 다양성을 살펴보고자 하는 것이다.

## 연구지역 및 연구자료

### 1) 연구지역 및 자료

본 연구의 연구지역은 미국 캘리포니아주의 북 태평양연안에 위치하고 있는 샌프란시스코시(MSA : Metropolitan Statistical Area)로서, 1990년 현재 인구는 723,969 이다. 주택가 분석을 위한 센서스 트랙과 블럭그룹의 기본도면(boundary information)은 1992년 ESRI에서 만든 Arc Census 자료<sup>1)</sup>를 이용하였으며, 각 조사 단위지역에 대한 속성 자료는 미국통계청의 STF 화일(U.S. Census Bureau's 1990 Summary Tape File 1A and 3A)<sup>2)</sup>을 이용하였다. 주택가에 대한 다핵적 도시기능의 영향력을 설명하기 위해 도심(CBD; Central Business District) 뿐만 아니라 주요 쇼핑센터, 해변, 공원 등 주요 결절점(urban node)의 위치를 확인하였다(Fig.1, Fig.2, Fig.3). 조사 단위 지역으로

1) Arc Census, California, Pre\_Release, 1990, ESRI, 1992. 이 자료는 ESRI에서 센서스 트랙과 블럭그룹의 기본도면에 미국 통계청(U.S. Census of Bureau)의 STF 화일의 일부 자료를 연결시켜 놓은 것이다.

2) 이 화일은 미국에서 10년 마다 행해지는 센서스의 인구 및 주택에 관련된 제반 자료를 포함하고 있다.

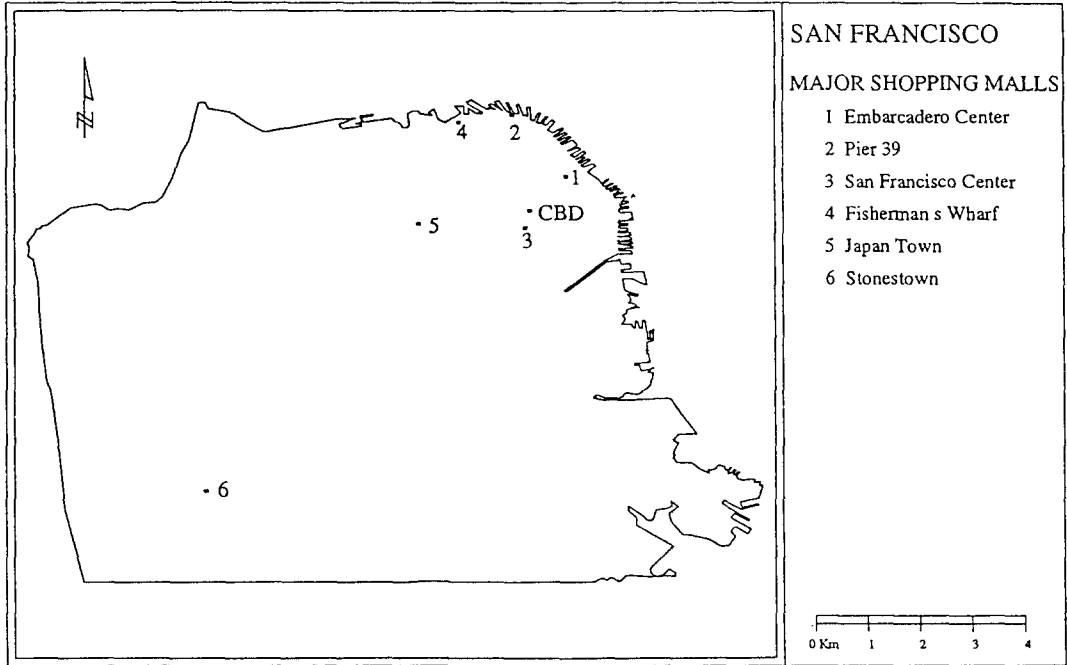


Fig. 1. The site of the CBD and shopping malls in San Francisco

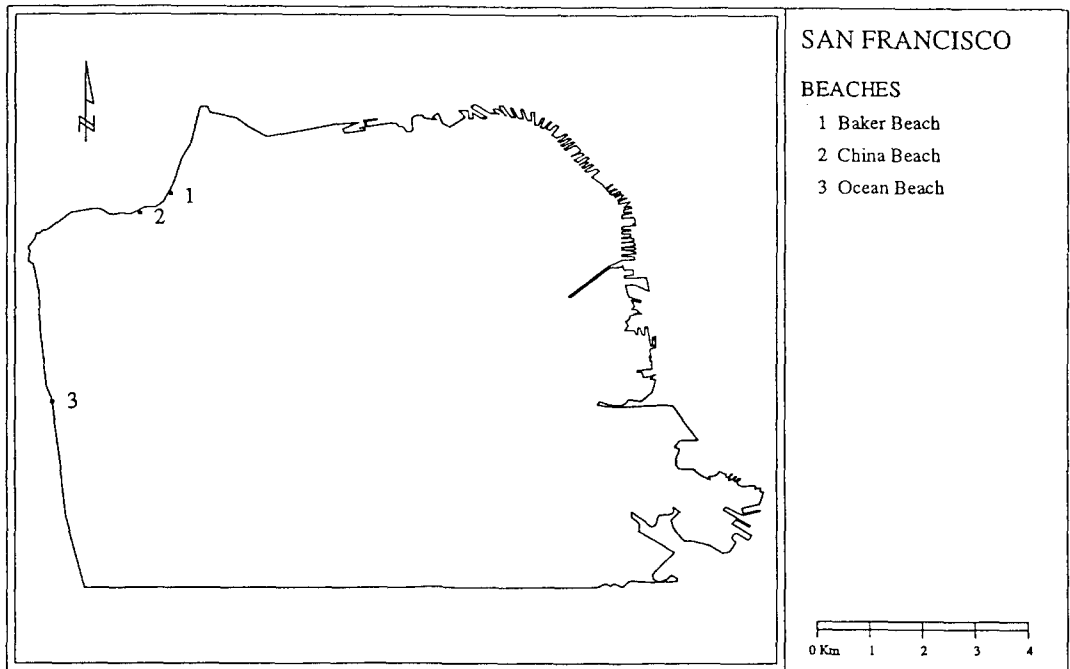


Fig. 2. The site of beaches in San Francisco

GIS를 이용한 주택가 분석 모델에 관한 연구

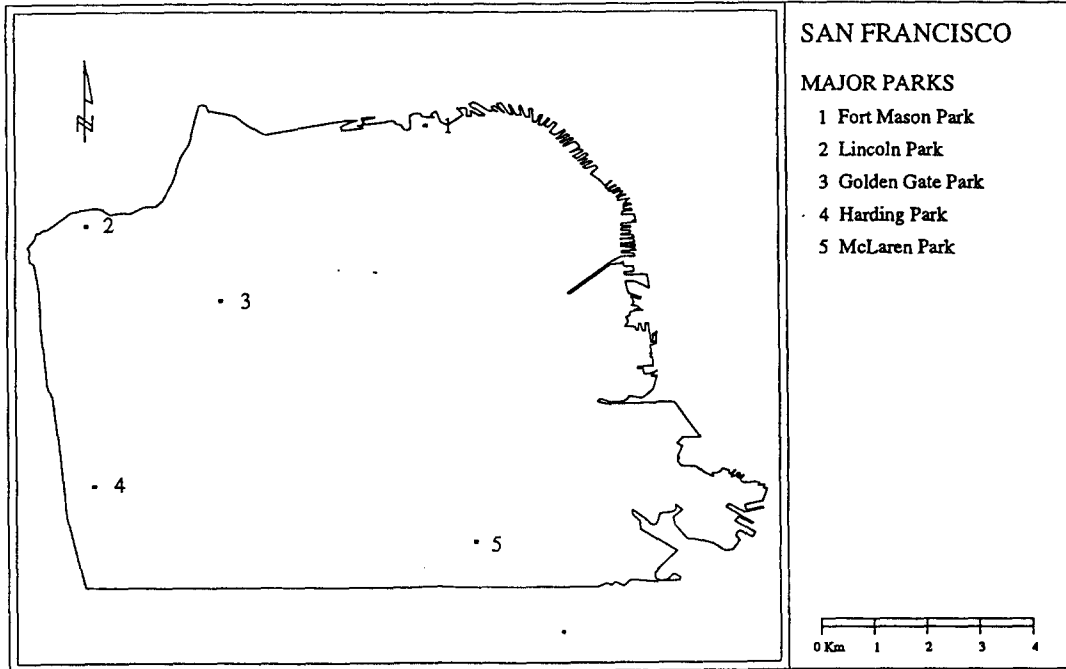


Fig. 3. The site of major parks in San Francisco

Table 1. Dependent and Independent Variables for Housevalue Model

변수유형	변수이름	설 명	
종속변수	HOUSEVALUE	자가주택의 평균 주택가격	
독립변수	사회 경제적 변수	HINCOME	1989년 가구당 평균 소득
		PINCOME	1989년 1인당 소득
		P_WHITE	백인의 비율(%)
		P_BLACK	흑인의 비율(%)
		P_NONWHITE	흑인비율 + 아메리칸 에스키모 비율 + 아시안 비율 + 기타비율
		P_POVERTY	1989년 소득이 빈곤선보다 낮은 사람의 비율
		P_HTBA	대학교 이상의 교육을 받은 사람의 비율
	주택 변수	OCCU_A	행정직, 관리직, 매니저직 및 전문직 종사자 비율
		P_PROMGT1	방당 인구밀도가 1인이상인 주택의 비율
		P_HOUSE 49	1949년 이전에 지어진 주택의 비율
		P_HOUSE 39	1939년 " "
		P_OWNEROCC	자가주택의 비율
	접근성 변수	PERS_UNIT	주택당 평균 인구수
		DIST_CBD	CBD까지의 거리
		NSDIST	가장 가까운 쇼핑센터까지의 거리
NPDIST		" 공원까지의 거리	
NBDIST	" 해변까지의 거리		

부터 도시 결절점까지의 거리를 구하기 위한 샌프란시스코의 도로망도는 TIGER/Line화일<sup>3)</sup>로부터 구하였다.

본 연구를 위하여 사용된 종속변수 및 주요 독립변수는 Table 1 과 같다.

## 2) 연구환경

본 연구에서는 ARC/INFO 6.01 workstation용 소프트웨어가 주로 사용되었으며 데이터의 전환과정에서 PC ARC/INFO 3.4 가 사용되었다. 그리고 통계분석을 위해 SPSS<sup>X</sup> /PC<sup>+</sup> 3.0이 사용되었다.

GIS와 통계분석간의 연계는 Fig.4 와 같이 나타나있다. 조사단위지역별 기본도 및 속성 데이터는 ARC CENSUS와 CENSUS DATA 를 사용하였는데 이들 자료는 workstation에 직접 사용될 수 없기 때문에 PC ARC/INFO 에서 받은후(import) ARC/INFO format으로 데이터를 전환한 후 workstation에서 다시 받아들여(import) 사용하는 형식을 취하였다. TIGER/Line 화일을 이용한 거리계산은 네트워크 모듈과 AML을 이용하여 속성데이터로 만

들어지고, 도심 및 각종 서비스를 제공하는 결절점은 ARC EDIT에서 새로운 레이어(layer)를 작성하였다. 이렇게 만들어진 속성 자료들은 PC ARC/INFO로 보내져 DBF format으로 데이터를 바꾼 후 통계분석 패키지로 보내졌다. 통계분석 후에는 모델 적합도를 공간적으로 확인할 수 있도록 결과 화일을 workstation으로 보내었다.

## 연구방법

### 1) GIS를 이용한 접근성 계산

주택가격을 결정하는 중요요소중의 하나인 접근성 변수는 다핵도시에 있어 도심 뿐 아니라 기타 다양한 서비스를 제공하는 부도심까지의 거리에 의해 측정되어질 수 있다. 실제 통행패턴은 도로를 따라 일어나는 것이기 때문에 이 연구에서는 분석단위지역(센서스트랙과 블럭그룹)의 중심점으로부터 도심 및 특정 서비스를 제공하는 가장 가까운 결절점까지의 거리를 도로를 따라 계산하였다. 샌프란시스코의 도로망은 1990년 TIGER/Line 화일로부터

3) TIGER(Topologically Intergrated Geographic Encoding and Referencing) 시스템은 미국 통계청에서 센서스와 그와 관련된 지리적 조사활동을 지원하기 위해 1970년에 처음 만든 것이다. TIGER 데이터베이스는 지형지물(도로, 철도, 강), 센서스 조사 단위지역(트랙 과 블럭), 정치적 단위지역(도시나 郡區), 지형지물의 이름, 미국 전체 1980년과 1990년 센서스 지역 코드, 그리고 대도시지역 가로에 따른 주소범위와 우편번호등을 수치 데이터로 나타낸 것이다.

TIGER 데이터베이스 자체는 일반에게 공개되지 않기 때문에 일련의 초록들을 이용할 수 있는데 이 초록중의 하나가 TIGER/Line 화일으로써 지도상에서 볼 수 있는 線과 관련된 특징들을 체계적인 네트워크로 만든 데이터 베이스이다. TIGER/Line 화일은 1:100000척척의 USGS (United State Geological Survey) 지도에 기초하여 작성한 것이기 때문에 고도의 정밀한 측정을 요구하는 부문(예: 소유권 이전, 엔지니어링 및 건축분야의 응용)에는 적합하지 못하지만 대부분의 계획 응용분야에 대해서는 적절히 사용될 수 있다.

GIS를 이용한 주택가 분석 모델에 관한 연구

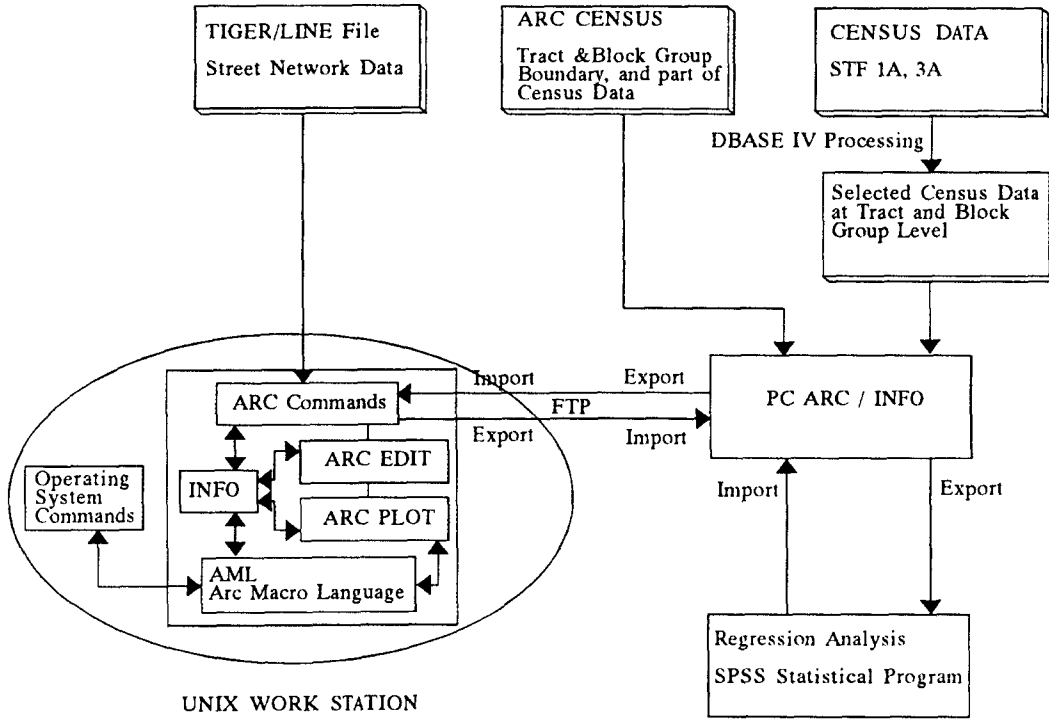


Fig. 4. Relationship among Data Source, GIS and Statistical Analysis

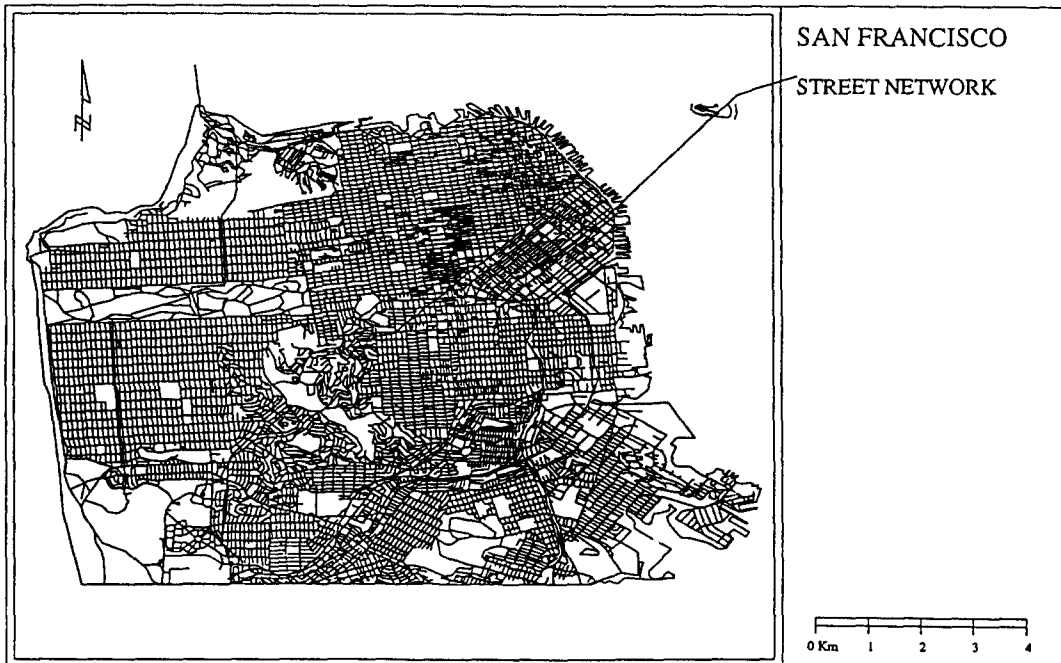


Fig. 5. Street network in San Francisco

터 조사되어졌는데 15517개의 부분(segment)으로 구성되어져 있다(Fig. 5). 실제 이 도로망은 일차도로(각주간 고속도로), 이차도로(州내 고속도로), 연계도로(郡간도로), 街路등으로 구분되어져 있으며, 도로의 등급에 따라 제한속도등이 다르기 때문에 저항(impedance)을 다르게 줄 수 있으나 본 연구에서는 총 도로연장길이만을 계산하였다.

모든 센서스 트랙과 블럭그룹의 중심점으로부터 도심 및 특정서비스를 제공하는 가장 가까운 부십지역까지의 거리는 2가지 방법에 의하여 계산될 수 있는데 첫번째 방법은 모든 조사단위지역으로부터 모든 결절점까지의 거리를 각각 계산하기 때문에 시간이 많이 소요되는 반면 두번째 방법은 시간소요는 작으나 모든 조사단위지역을 각각 계산하지 않고 가장 가까운 node에 접속시키기 때문에 여러 개의 중심점이 하나의 node에 접합되는 경우 정보를 잃을 가능성이 존재하여 이에 대한 주의가 필요하다.

(1) 첫번째 방법: ARC/INFO의 네트워크 모듈의 최단거리를 찾는 알고리즘(shortest path finding algorithm)을 이용하는 방법

이 방법은 각 분석단위 중심점으로부터 도심 및 특정 서비스를 제공하는 모든 결절점까지의 최단거리를 찾은 후 특정서비스를 제공하는 가장 가까운 결절점까지의 거리를 선택하는 방법이다. 최단거리의 결과는 INFO 화일에 수록되는데 총거리만이 계산되어지는 것

이 아니라 각 부분(segment)의 누적된 합계로 계산되어지기 때문에 총길이만을 추출하여 데이터베이스를 만드는 과정이 필요하다 즉 센서스 트랙 단위에서는 152개, 블럭그룹 단위로는 611개의 중심점으로부터 1개의 CBD, 3개의 해변 중심점, 5개의 공원 중심점, 그리고 6개의 쇼핑센터 중심점까지 모두 11,445개의 최단거리를 찾는 반복적인 과정이 필요하다. 이러한 과정은 AML(Arc Macro Language)을 사용하여 프로그래밍 되었다<sup>4)</sup>.

PATH를 이용한 거리계산 과정

1. 센서스 트랙의 모든 단위지역의 LABEL POINT를 폴리곤 내로 옮김.
2. 각 센서스 트랙 중심점의 ID와 X, Y 좌표를 아스키 (ASCII) 화일로 만들.
3. 트워 분석을 위해 사용될 도로망도의 이름을 지정해 주고 네트워크 분석의 결과 화일이 쓰여질 화일의 이름을 지정해 줌.
4. 네트워크의 저항(impedance)을 어떻게 줄 것 인지를 지정해 줌.
5. 출발지와 목적지의 입력방법을 정해준후 두 지점간 'PATH'를 찾게 함.

(2) 두번째 방법: ARC/INFO의 네트워크 모듈 중 SPATIAL INTERACTION MODEL을 이용하는 방법

기본개념은 각 분석단위 중심점을 수요지점으로 간주하고 특정 서비스별 중심점(예를 들어 6개의 쇼핑센터)을 공급지점으로 선정하여 INTERACTION이라는 ARC/INFO 기능

4) 이 과정을 위한 AML은 Kang(1994)을 보시오.



을 사용하여 가장 가까운 공급지점까지의 거리를 도로망을 따라 구하는 것을 전제로 한다. 이 과정에서 각 도로교차점(node)에는 하나의 수요 또는 공급지점만을 가질 수 있기 때문에 여러 개의 분석단위 중심점이 하나의 도로교차점에 위치하게 되는 경우 한개의 분석단위 중심점만이 수요지점으로 계산되고 나머지 분석단위 중심점은 수요지점에서 제외되기 때문에 후자의 지점에 대한 특별한 주의를 필요로 한다.

INTERACTION을 이용한 거리계산 과정(예: 센서스트랙의 각 중심점으로 부터 가장 가까운 쇼핑센터까지의 거리계산)

1. 센서스 트랙의 PAT화일에 DEMAND라는 항목을 만들어서 각 센서스 트랙 중심점의 DEMAND가 1이 되도록 PAT화일을 만들.
2. 센서스 트랙의 모든 중심점이 가로망도의 node에 위치하도록 함.
3. 쇼핑센터의 모든 중심점이 가로망도의 결절점에 위치하도록 한 후 도로망도의 쇼핑센터가 위치한 node에 대한 정보를 쇼핑센터 PAT화일로 만든 후 쇼핑센터의 PAT화일로부터 INFO화일을 만들.
4. 새로이 생성된 INFO화일에 SUPPLY라는 항목을 만들어서 값을 1로 부여함.
5. 'INTERACTION'을 이용하여 DEMAND와 SUPPLY지점간 최단거리 계산

## 2) Expansion Method를 이용한 주택가 가격 모델의 정립

기존의 주택가분석에 대한 연구는 주로 단핵도시의 가정하에 주택의 물리적 특성 및 사회경제적 환경을 강조해 왔으며 도심까지의 접근성이 접근성을 대표하는 변수로서 사용되어져 왔다. 그러나 최근의 도시는 도심 뿐 아니라 쇼핑센터, 레크리에이션 센터, 고용중심지 등 다양한 서비스를 제공하는 결절점을 갖는 다핵도시로 발전해 가고 있으며 이러한 특성이 주택가 분석에 영향을 미치고 있다.

또 하나의 중요성은 주택의 사회적 특성(사회경제학적, 인구학적, 주택구조적 특성)과 접근성이라는 요소가 상호작용할 수 있다는 사실이다. 즉 기존의 연구에서 도심까지의 접근성은 부유한 지역에서나 가난한 지역에서나 주택가격에 대해 똑같은 한계가치를 갖는 것을 내포했었다. 그러나 부유한 지역의 주택가는 도심의 공해, 소음, 혼잡 등 (-)적인 효과 때문에 외곽에서 도심으로 갈수록 하락하고, 가난한 지역의 주택가는 도심의 고용기회 제공 또는 손쉬운 교통로의 접근 때문에 외곽에서 도심으로 갈수록 비싸질 수 있다는 것이다. 그러나 이러한 상호효과적 측면은 크게 관심의 대상이 되지 못했는데 Expansion Method는 이러한 상황을 모델화시키는데 사용되어질 수 있는 방법이다(Casetti,1991). 즉 Expansion Method를 사용한 주택가 분석 모델에서는 주택가에 영향을 미치는 여러 사회적 특성들이 도시상황과 관계없이 고정된 것이 아니라 주택이 가지는 여러 지점으로의 접근성이라는 상황(context)에 따라 달라질 수 있음을 전제로 한다.

Expansion Method를 사용한 주택가격모델

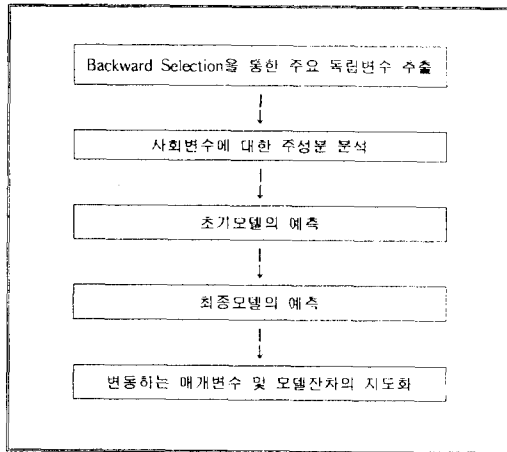


Fig. 6. Process to Develop Housevalue Model

의 정립은 Fig.6 와 같은 흐름도로 나타낼 수 있다.

첫째: 블럭그룹 단위의 데이터를 이용하여 후방선택(Backward Selection)을 통한 회귀분석을 실시한다. 이과정은 초기에 모든 독립변수를 포함한 회귀분석을 실시한 후 다음의 단계에서 가장 낮은 t값을 갖는 변수를 제거하여 모델에 포함된 모든 변수의 계수가 95% 이상의 중요도를 갖을 때 까지 계속한다.

둘째: 센서스 트랙과 블럭그룹 단위로 접근성 변수를 제외한 사회변수에 대하여 각각 Varimax 회전된 요인 분석을 실시한다(즉 후방선택(Background Selection)에 의한 회귀분석이후 남아있는 사회변수에 대해 요인분석을 실시한다). 요인 분석에 의해 eigenvalue가 1.0보다 큰 요인은 두개가 추출되었고 센서스 트랙단위와 블럭그룹 단위 데이터 분산의 85%와 78%를 각각 설명하였다. Varimax 회전에 의해 추출된 요인은 Table2, Table3 과

같다.

첫번째 요인은 대학이상의 교육을 받은 사람의 비율(P\_HTBA)과 가구주 평균소득(HINCOME)에 대해 (-)부하를 보이며, (+)부하를 보이는 것은 방당 인구가 1명 이상인 주택의 비율(PROOMGT1)과 비백인 인구의 비율(NONWHITE)이다. 이 요인은 사회경제적 지위가 낮은지역, 주택당 인구밀도가 높은 지역 그리고 비백인위주의 지역과 그렇지 않은 지역을 나누고 있다. 두번째 요인은 1939년 이전에 건축된 주택의 비율에 대하여 (+)부하량을 보여주어 이 요인은 노후주택의 존재를 나타냄을 알 수 있다.

셋째: 이 단계에서 주택가격은 앞에서 얻은 요인의 함수로 나타내어지며 초기 모델의 매개변수는 다변량회귀분석 프로그램에 의해 추정 되어진다.

넷째: 초기모델의 매개변수는 도심 및 각종 서비스를 제공하는 결절점까지의 거리의 함수로 확장되어진다. 그리고 확장된 매개변수는 초기모델에 대치되어 최종모델을 만들게 된다. 최종모델은 후방선택(Backward Selection) 과정을 통한 회귀분석방법을 통해 추정 된다. 중요하지 않은 독립변수의 제거수준은 첫번째 단계에서 나온 것과 동일하다.

다섯번째: 이 과정에서는 변동하는 매개변수를 지도화하고 해석하며 센서스 트랙과 블럭그룹 단위 주택가 모델의 적합도가 지역에 따라 얼마나 차이가 나는지를 공간적으로 나타내기 위해 잔차도(residual map)를 작성한다.

Table 2. Varimax Rotated Factor Matrix (Census Tract Data)

변수	요인1	요인 2	요인들에 의해 설명된 분산의 비율
P_HTBA	-.94647		.91169
PROOMGT1	.90984		.82872
NONWHITE	.88817		.85589
HINCOME	.79725		.67946
HOUSE39		.97901	.96058

Table 3. Varimax Rotated Factor Matrix (Block Group Data)

변수	요인1	요인 2	요인들에 의해 설명된 분산의 비율
P_HTBA	-.89863		.85413
PROOMGT1	.86688		.75391
NONWHITE	.82904		.82456
HINCOME	-.72477		.55017
HOUSE39		.96691	.96058

## 분석결과

### 1) 분석단위별 주택가 모델

#### 센서스 트랙을 이용한 주택가 모델

##### 초기모델

$$y = 348221.53 - 66153.75 F_1 + 36810.50 F_2$$

$$R^2 = 0.465 \quad (1)$$

##### 최종모델

$$y = 536310.23 - 11520.65 D_1 - 12535.13 D_2$$

$$- 11854.43 D_3 - 135124.64 F_1 + 9622.14$$

$$F_1 D_2 + 19569.88 F_2$$

$$R^2 = 0.652 \quad (2)$$

### 블록그룹을 이용한 주택가 모델

#### 초기모델

$$y = 341907.41 - 74585.80 F_1 + 32357.97 F_2$$

$$R^2 = 0.536 \quad (3)$$

#### 최종모델

$$y = 525096.68 - 11989.10 D_1 - 11838.22 D_2$$

$$- 9720.73 D_3 - 118471.32 F_1 + 6853.96 F_1 D_2$$

$$+ 61388.28 F_2 - 2577.66 F_2 D_1 - 3385.46 F_2 D_2$$

$$R^2 = 0.681 \quad (4)$$

(1)의 공식에서  $F_1$ 의 계수(-66153.75)는 요인1과 주택가와의 관계를 나타낸다. (+)적인  $F_1$  값을 갖는 센서스 트랙은 사회경제적 지위가 낮고 비백인 위주이며, 주택당 인구밀도가 높은 지역이다. 이러한 지역에서 요인1의 점

수가 한 단위 높아짐에 따라 추정된 주택가는 약 \$66153가 하락하는 것으로 나타난다. 한편 (2)의 공식에서  $F_2$  의 계수(36810.50)는 요인2와 주택가와의 관계를 나타내는데 요인2가 (+)적인 값을 나타내는 센서스트랙은 노후주택의 비율이 높은 지역이다. 즉, 샌프란시스코에서 노후주택의 비율이 한 단위( $F_2$  값) 증가함에 따라 추정된 주택가는 \$36810 증가하는 것으로 나타났다.

이러한 센서스트랙에서의 초기모델(1) 공식은 도심( $D_1$ )과 해변( $D_2$ ), 공원( $D_3$ )까지의 접근성 함수로 확장되었으며 (2)의 공식에서 도심, 해변 그리고 공원으로부터 거리가 1Km 증가함에 따라 주택가는 \$11520, \$12535, 그리고 \$11854 각각 감소하는 것으로 나타났다.  $F_1$ 의 변화에 따른 주택가의 탄력성은 초기모델보다 2배 가량 증가(계수의 절대값으로 보았을 때)했으며  $F_2$ 의 경우는 초기모델보다 감소한 것으로 나타났다. 그리고 요인1과 해변까지의 거리간 상호작용( $F_1D_2$ )<sup>5)</sup>이 주택가 추정에 주요하게 작용하는 것으로 나타났다. 그리고 초기모델에서는 사회경제적, 인구학적 주택의 특성을 나타내는 두가지 요인에 의해 주택가의 47% 정도를 설명할 수 있었던 반면, 최종모델에서는 도심, 해변, 공원 등까지의 거리를 포함한 접근성 변수와 2개의 요인 변수와 접근성 변수간 상호작용에 관한 변수를 더함으로써 주택가의 65%정도를 설명할 수 있었다.

이와 같은 현상은 블럭그룹 단위의 데이터

를 사용했을 때에도 거의 유사한 양상을 보였으며 다만 최종모델에 있어 요인1과 해변까지의 접근성간 상호작용( $F_1D_2$ )뿐 아니라 요인2와 도심까지의 접근성에 따른 상호작용( $F_2D_1$ )과 요인2와 해변까지의 접근성간 상호작용( $F_2D_2$ )도 주택가 추정에 중요한 요인인 것으로 나타났다.

## 2) 잔차도 분석

주택가를 설명함에 있어 센서스트랙과 블럭그룹 단위 데이터의 사용에 따른 다양성은 최종모델의 경우 센서스트랙 단위에서는 주택가 추정의 65%, 블럭그룹 단위에서는 68%정도로  $R^2$ 값으로는 큰 차이를 나타내지 못했으나, 잔차도(residual map) 분석을 통해 모델의 적합성이 떨어지는 지역을 공간적으로 확인하고 종속변수의 분산을 추가로 설명해 줄 수 있는 새로운 독립변수를 선정할 수 있었다.

Fig.7과 Fig.8의 지도에서 센서스트랙과 블럭그룹 지역 모두에서 South of Market 지역에서 주택가가 과대 추정된 것과 South of Market 북쪽의 샌프란시스코만과 태평양 연안으로 과소추정된 것을 볼 수 있다. South of Market 지역은 샌프란시스코에서 도심의 슬럼적 특성을 나타내는 곳이며 South of Market 북쪽으로 언덕등의 전망이 좋은 곳에는 고급주택이 위치하는데 이러한 요소들이 설명되지 못한 것이다.

블럭그룹 단위 데이터를 이용한 잔차도 분석에서는 과대추정이나 과소추정지역이 보다

5) 상호작용에 대한 설명에 대해서는 Kang(1994)을 보시오.

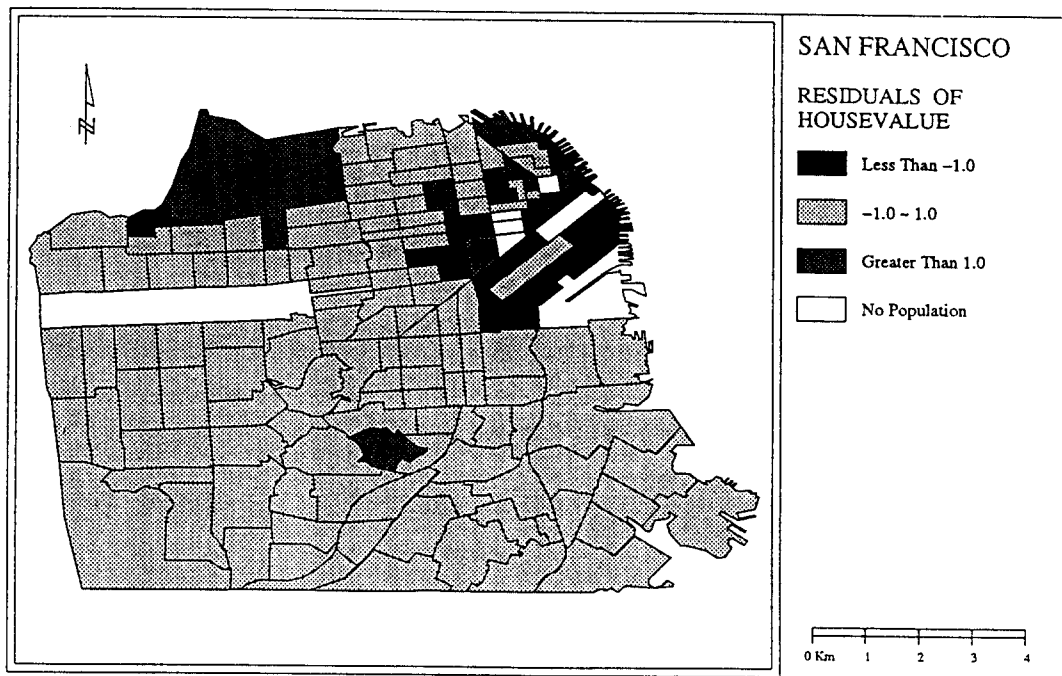


Fig. 7. Residuals of housevalue model at tract level

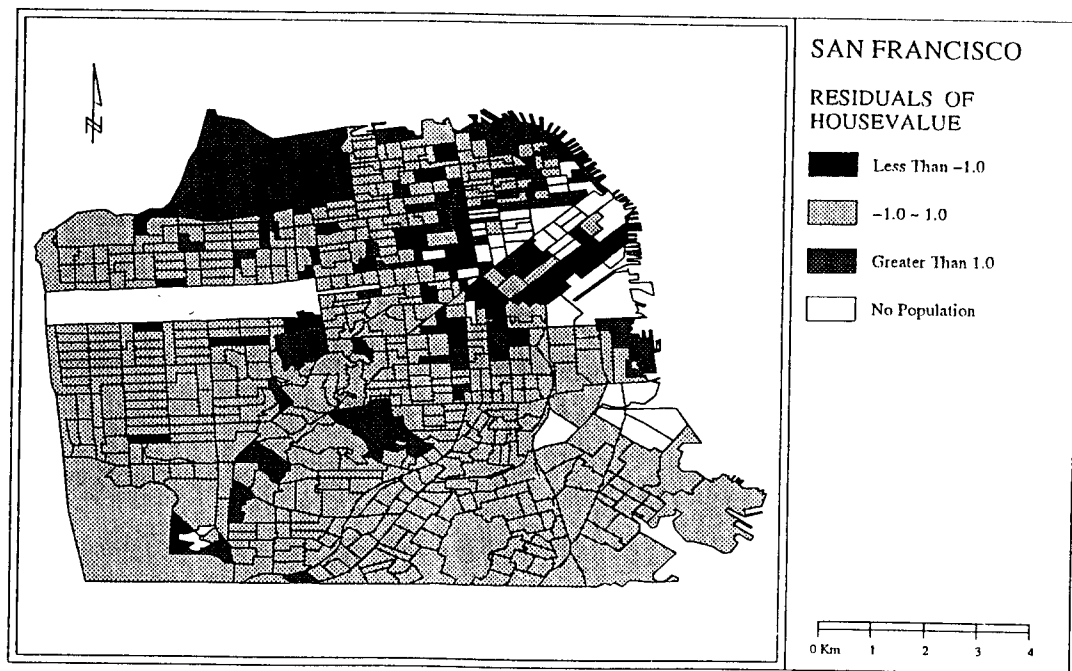


Fig. 8. Residuals of housevalue model at block group level

분산되어 있으며 특히 Golden Gate Park 주변에 몇몇 과대 추정된 지역이 존재하며 도시의 남서부 지역에 과소추정지역이 산재해 있음을 나타낸다. 샌프란시스코의 경우 결인(homeless people)등이 공원주위에 기숙하여 공원에 근접한 곳은 주택가에 (-)적 영향을 미치는 경우가 많으며 도시의 남서부에 산재한 과소추정지역은 대개 신흥주택가로서 모델에 의해 설명되어지지 못한 것으로 나타나 보다 세분된 데이터를 사용함으로써 보다 많은 공간적 정보를 입수할 수 있게 됨을 나타냈다.

## 결 론

본 연구에서는 GIS가 단순정보처리 수준을 넘어 분석능력을 향상시키기 위해 어떠한 방법들이 실제 사용될 수 있나를 주택가 분석모델을 예로 들어 보여주었다. 주택가 모델을 GIS에 간접연계하였으며, 주택가 모델을 정립함에 있어서도 GIS를 이용하여 복잡한 현실 세계를 반영하려 노력하였다. 즉 도심이외에도 해변, 공원, 쇼핑센터 등 다핵적 특성을 나타내는 도시의 결절점을 모델에 도입하였으며 이러한 결절점까지의 거리도 가로망에 따라 계산하였다. 사회경제적, 인구학적 변수이외에 이러한 접근성 변수와 변수간 상호작용 요인을 더함으로써 주택가 추정에 있어 15~20% 정도 설명력을 높일 수 있었다. 센서스트랙과 블럭그룹 단위의 데이터 사용에 따른 다양성은 모델의 잔차를 GIS를 이용하여 공간적으

로 디스플레이 함으로써 공간적 이해력을 높이고 서로 다른 분석단위 데이터 사용에 따른 차이를 가시화시킬 수 있었다.

본 연구에서는 주택가 모델과 GIS를 간접연계 시켰으나 이보다는 직접연계가 보다 다양한 분석능력 환경을 제공할 수 있으리라 생각된다. 즉 사용자 또는 연구자로 하여금 다양한 그래픽 방법으로 데이터를 디스플레이 해 볼 수 있도록 하는 환경의 제공은 GIS에 있어 분석능력 부족이라는 결점을 극복할 수 있는 한 방법으로써 가시적 표현(visualization)<sup>6)</sup>이라는 연구의 한 분야로 대두되고 있으며 앞으로 많은 연구가 필요한 분야이다.

## 참고문헌

- Batty, M. and Y. Xie, 1994, "Modeling inside GIS: Part1. Model Structures, exploratory spatial data analysis and aggregation," *Int. J. Geographical Information Systems*, Vol.8, No.3, pp.291-307.
- Burrough, P. A., 1986, *Principles of Geographical Information Systems for Land Resources Analysis*, Oxford: Clarendon Press.
- Buttenfield, B. P. and W. A. Mackaness, 1991, "Visualization", in *Geographical Information Systems:*

6) Buttenfield, B.P. and W.A. Mackaness (1991), 강영옥(1995).

- Principles and Applications*, edited by D. J. McGuire, M. F. Goodchild and D. W. Rhind, London: Longmans, pp.427-443.
- Casetti, E., 1991, "Investigation of parameter drift by expanded regressions: generalities, and a 'family-planning' example," *Environment and Planning A*, 23, PP. 1045-1061.
- Chen, L. 1993, "Modeling Housing and Demographic Diversity at Census Tract versus Block Group Level of Aggregation", *URISA 1993 Conference Proceedings*, Vol.3, 56-72.
- Densham, P. J., 1991, "Spatial decision support systems," in *Geographical Information Systems: Principles and Applications*, edited by D. J. McGuire, M. F. Goodchild and D. W. Rhind, London: Longmans, pp.403-412.
- Densham, P. J. and Goodchild, M. F., 1989, "Spatial Decision Support Systems: A Research Agenda," *GIS/LIS '89 Proceedings*, 2, pp.707-716,
- Ding, Y. and A. S. Fotheringham, 1992, "The integration of Spatial Analysis and GIS," *Computers, Environment and Urban Systems*, 16, pp.3-19.
- Environmental Systems Research Institute, 1992, *Arc Census Pre-release :California: Users Guide & Data Reference*, Redlands, CA.
- Goodchild et al., 1992, "Intergrating GIS and Spatial Data Analysis.: Problems and Possibilities," *Int. J. Geographical Information Systems*, 6(5), PP. 407-423.
- Hunt, J. D., and Simmonds, D.C., 1993, "Theory and applications of integrated land use and transport modelling frameworks," *Environment and Planning B*, 20, pp.221-244.
- Huxhold, William E., 1991, *An Introduction to Urban Geographic Information Systems*, New York : Oxford Univ. Press.
- Kang, Youngok, 1994, *An Empirical Study of Housevalue and Population Density by Expanded Regressions*, Ph.D. Dissertation, Dept. of Geography, The Ohio State University.
- Marble, D. F., H. W. Calkins, and D. J. Peuquet, eds., 1984, *Basic Readings in Geographic Information Systems*, Buffalo, NY:SPAD Systems, Ltd.
- Marx, Robert W, ed. 1990, "The Census Bureau's TIGER System," A spec-

- ial issue of *Cartography and Geographic Information Systems*, Vol.17, pp.7-113.
- Openshaw, S., 1990, "Spatial analysis and geographical information systems: A review of progress and possibilities," in *Geographical Information Systems for Urban and Regional Planning*, eds. H. Scholten and J. C. H. Stillwell, pp.153-163, Boston: Kluwer Academic Press.
- Openshaw, S., 1991, "Developing Appropriate Spatial Analysis Methods for GIS," in *Geographical Information Systems: Principles and Applications*, eds. D. J. Maguire, M. F. Goodchild and D. W. Rhind, pp.389-402, New York: John Wiley & Sons.
- Putman, S., 1992, *Intergrated Urban Models 2: New Research and Applications of Optimization and Dynamics*, London: Pion Press.
- U.S. Bureau of the Census, 1991, *TIGER/Line Census Files, 1990 Technical Documentation*, Washington, D. C.: U. S. Government Printing Office.
- Wiggins, L. L., and Ferreira, J., 1993, "Computer Technology for land use and regional planning: the case of the National Capital Planning Commission," A paper presented at *The Third International Conference of Computers in Urban Planning and Urban Management*, 23-25 July, Georgia Institute of Technology, Atlanta, GA, pp.285-299.
- 강영욱, 1995, "GIS에서 可視的 표현 (Visualization)의 중요성", 한국 여성지리학자회 회보, 제 1권 제 1호, pp. 8-10.