

# 격자 기반의 통계정보 표현을 위한 데이터 변환 방법

## A Data Transformation Method for Visualizing the Statistical Information based on the Grid

김문수\* · 이지영\*\*

Munsu Kim · Jiyeong Lee

**요약** 본 논문에서는 다양한 형태로 존재하는 통계정보를 일정한 모양과 크기를 갖는 격자로 표현하기 위해 필요한 데이터 변환 방법론에 대하여 제시한다. 격자는 기존 통계지도 서비스에서 활용하고 있는 통계공간단위인 행정구역과 집계구와 비교하였을 때 모양과 크기가 일정하여 통계정보를 객관적으로 파악할 수 있게 하며, 지도 축척 변화에 유연하게 적용될 수 있는 특징이 있다. 한편, 기존 통계지도 서비스에서는 면 보간법을 활용하여 통계공간단위로 변환하고 있는데, 이것을 다양한 형태로 존재하는 통계정보에 적용시키기 위해서는 추가적인 프로세스가 필요하다. 이에 따라, 본 논문에서는 다양한 형태로 존재하는 통계정보의 격자 변환을 위해 1)지오코딩을 통한 공간데이터로의 변환, 2)공간 관계 정의를 통한 위치정보 변환, 3)데이터 척도를 고려한 속성정보 변환을 수행하는 방법론을 정리하였다. 제시한 방법론은 서울시 A지역의 인구 밀도 통계정보를 격자로 변환하기 위해 적용하였다. 특히, 동일한 통계정보를 표현하는 참조데이터가 서로 다르더라도 유사한 격자 표현이 가능해야 함을 검증하기 위해 공간 자기상관성을 통해 분석하였다. 그 결과, 집계구와 건물을 통해 표현되는 인구 밀도를 각각 격자로 변환하였을 때, 두 데이터 모두 유사한 격자 분포를 표현할 수 있었다. 이러한 결과를 통해 본 연구에서 제안하는 방법론은 일관된 결과를 표현할 수 있음을 확인하였다.

**키워드** : 데이터 변환, 통계공간단위, 격자, 참조데이터, 데이터 척도

**Abstract** The purpose of this paper is to propose a data transformation method for visualizing the statistical information based on the grid system which has regular shape and size. Grid is better solution than administrator boundary or census block to check the distribution of the statistical information and be able to use as a spatial unit on the map flexibly. On the other hand, we need the additional process to convert the various statistical information to grid if we use the current method which is areal interpolation. Therefore, this paper proposes the 3 steps to convert the various statistical information to grid. 1)Geocoding the statistical information, 2)Converting the spatial information through the defining the spatial relationship, 3)Attribute transformation considering the data scale measurement. This method applies to the population density of Seoul to convert to the grid. Especially, spatial autocorrelation is performed to check the consistency of grid display if the reference data is different for same statistic information. As a result, both distribution of grid are similar to each other when the population density data which is represented by census block and building is converted to grid. Through the result of implementation, it is demonstrated to be able to perform the consistent data conversion based on the proposed method.

**Keywords** : Data Transformation, Spatial Unit, Grid, Reference Data, Data Scale Measurement

### 1. 연구배경 및 목적

공간정보는 ICT, 환경, 의료, 방재, 건축 등 다양한 분야들과 융합되어 활용되고 있다. 최근에는 통계와 공간정보 분야 간 융합 및 활용 사례가 증가하고 있으며, 특히 UN-GGIM의 경우 통계 및 공간정보 통합 전문가 위원회(United nations expert group on the

integration of statistical and geospatial information)를 통해 세계 각국의 다양한 통계 및 공간정보 융합 사례를 소개하고 있다. 대표적으로 현재 가장 많이 활용되고 있는 사례는 통계정보를 지도위에 표현하여 시각적으로 통계정보를 파악할 수 있게 하는 ‘통계지도’이다. 통계지도에 사용되는 통계데이터는 수집 방법에 따라 다양한 형태로 존재 할 수 있는데, 지도상에 표현하

† This research was supported by a grant from Development Plan of National Territorial Statistic Monitoring and Construction of Platform Project in 2014 funded by National Geographic Information Institute of Korean government.

\* Munsu Kim, Master's Student, Dept. of Geoinformatics, University of Seoul. rlaanstn01@gmail.com

\*\* Jiyeong Lee, Professor, Dept. of Geoinformatics, University of Seoul. jlee@uos.ac.kr (Corresponding Author)

기 위해서는 좌표를 기반으로 직·간접적으로 표현되어야 하며 형태는 점, 선, 면으로 통계정보를 표현할 수 있다. 그러나 통계정보의 종류에 따라서는 원천데이터를 그대로 맵핑하여 통계지도 서비스를 할 경우에는 개인정보나 보안이 요구되는 정보들을 그대로 노출할 수도 있기 때문에 이를 보완하기 위해 일정한 기준으로 통계정보를 집계하여 표현하며, 집계를 위해 활용되는 단위를 통계공간단위(Spatial Unit)라 한다. 통계지도 서비스를 위해 사용되는 대표적인 통계공간단위는 행정구역, 집계구, 격자 등이 있으며, 현재 행정구역이나 집계구가 가장 많이 활용되고 있다. 그러나 행정구역이나 집계구의 경우 모양과 크기가 불규칙하며, 시간의 변화에 따라 경계가 조정되어 통계공간단위의 모양과 크기가 바뀔 수 있고, 집계구의 경우 지도의 축척과는 무관하게 고정된 모양과 크기를 갖기 때문에 축척에 따른 유기적인 통계공간단위를 제공하지 못하는 한계점이 있다. 따라서 이러한 한계점을 극복할 수 있는 적절한 통계공간단위에 대해 많은 의견들이 제시되고 있으며, 그 중 일정한 모양과 크기를 갖는 ‘격자’를 통해 통계정보를 표현하는 방안이 가장 주목받고 있다.

격자는 행정구역이나 집계구와는 다르게 데이터의 수집단위로 활용되고 있지 않기 때문에 격자를 기반으로 통계정보를 표현하기 위해서는 원천데이터의 격자 변환이 필요하다. 이에 따라, 본 연구에서는 다양한 형태로 존재하는 통계정보를 지도상에 통계공간단위를 기반으로 표현하기 위한 체계적인 데이터 변환 방법을 제시한다. 이와 같은 연구를 위한 연구 수행 과정은 다음과 같다.

2장에서는 기존 통계지도 서비스에서 활용되고 있는 통계공간단위에 대한 현황 분석을 수행하며, 또한 통계정보를 GIS에서 활용하기 위해 수행하는 대표적인 데이터 변환 방법인 Areal Interpolation에 관한 연구를 살펴본다. 3장에서는 다양한 형태로 존재할 수 있는 통계정보의 격자변환을 위한 체계적인 방법론을 제시한다. 제안하는 방법론은 1) 비공간데이터로 존재하는 통계정보의 공간데이터 변환을 위한 지오코딩 단계, 위치 정보 변환을 위한 2)공간관계(Spatial Relation ship) 정의 단계, 3)데이터 척도(Data Scale Measurement)를 고려한 속성정보 변환 단계로 세분화하여 제시된다. 4장에서는 제시된 방법론을 적용하여 서울시 A지역에 대한 인구 밀도 통계를 격자로 변환하도록 한다. 방법론을 적용하여 올바른 데이터 변환을 확인하기 위한 한 가지 방법으로 원본 데이터가 갖는 공간 자기상관성을 그대로 유지하고 있는지를 확인하여 데이터 변환

의 결과를 검증하도록 한다.

## 2. 관련 연구

현재 통계공간단위를 기반으로 다양한 종류의 통계지도 서비스들(통계지리정보서비스, 국토포털, 서울시 통계지도 서비스 등)이 제공되고 있다. 이러한 통계지도 서비스에서는 통계정보를 표현하기 위한 통계공간단위로 행정구역과 집계구를 주로 사용하고 있다. 그러나 현재 서비스되고 있는 통계공간단위는 통계정보를 표현함에 있어 다음과 같은 한계점을 갖는다.

첫 번째로 통계공간단위의 모양과 크기가 일정하지 않아 통계정보의 분포, 추세, 패턴 등을 비교하기 어렵다. 행정구역과 집계구는 다양한 크기와 모양을 갖는다. 인구가 많이 거주하는 도심지역의 경우 비도심지역과 비교하여 상대적으로 행정구역과 집계구의 크기가 작고 조밀하게 분포되어있다. 이에 따라, 행정구역과 집계구를 통해 표현되는 통계정보는 지도상에 다양한 모양과 크기로 표현되며, 이것은 통계정보의 분포, 추세, 패턴이 공간상에 동일한 기준으로 표현되지 않아 사용자가 객관적인 분석을 수행하기에 제한적이다. 이에 따라, 집계구를 격자로 변환하거나 행정구역을 격자로 변환하기 위한 기본 연구들이 진행되었다[1,2].

두 번째로 행정구역과 집계구의 모양과 크기가 시간에 따라 변할 수 있다. 행정구역의 경우 행정구역 개편에 따라, 집계구는 정의하기 위한 인구수에 따라 모양과 크기가 변할 수 있다. 이것은 첫 번째 문제점으로 지적한 다양한 모양과 크기뿐만 아니라 시간에 따라 경계가 변할 수 있기 때문에 통계정보의 시계열 분포를 비교하기에 부적합하다. 이에 따라, 시계열 분석을 수행한 다양한 연구들은 격자를 기반으로 연구 결과를 표현하고 있다[3,4].

마지막으로 현재의 통계공간단위는 지도의 다양한 축척을 기반으로 통계정보를 표현하기에 제한적이다. 행정구역의 경우 시도, 시군구, 읍면동 등으로 체계화되어 있지만 읍면동 수준보다 더 세밀하거나 행정구역 이외의 크기나 모양의 통계공간단위로 표현되어야 하는 통계정보의 표현을 위해 활용되기 힘들다. 또한, 집계구의 경우에는 축척에 따른 모양과 크기의 변화가 없기 때문에 다양한 축척에서 집계구를 활용한 통계정보 표현이 어렵다. 따라서 다양한 분야의 선행연구들에서 다양한 격자 크기를 통해 분석결과를 비교하고 있었다[5,6].

한편, 수집되는 통계정보의 가장 이상적인 형태는 각 객체에 대한 정보를 점 데이터로 수집하는 것이지

만 많은 비용과 시간이 소모되는 사항이기 때문에 자주 수행되지 않는다. 그 대신 다수의 통계정보는 우편 구역, 서비스 구역, 선거 구역과 같이 목적에 맞는 특정 영역들을 기준으로 수집된다[7]. 이렇게 공간단위 기반으로 수집된 통계정보를 행정구역 등과 같은 통계공간단위로 표현하기 위해 가장 많이 활용되는 방법론이 면 보간법(Areal Interpolation)이다[7,8,9]. 면 보간법은 공간단위로 수집된 통계정보를 다른 형태의 공간단위(예-격자 등)로 변환하는 과정을 의미한다. 이러한 면 보간법은 다양한 세부 방법론들이 발전해 왔는데, 대표적으로 ‘면적 가중 보간법(Areal Weighted Interpolation)’, ‘Pycnophylactic Smoothing 방법’, ‘중심점(Centroid)을 이용한 밀도면 생성 방법’ 등이 있다. 이 중, 가장 많이 활용되는 방법은 면적 가중 보간법으로 공간단위에 해당 통계정보가 고르게 분포되어 있다는 가정을 기반으로 다른 형태의 공간단위로 변환 할 때 면적에 비례하여 값을 할당하는 방법이다. 하지만 앞서 언급한 것처럼 통계정보는 다양한 형태로 수집됨에 따라 면 보간법을 그대로 적용하기에는 다음과 같은 제약사항이 존재한다.

우선, 면 보간법을 통해 변환하고자 하는 소스데이터의 유형은 면형 데이터로 존재해야 하지만 통계정보는 다양한 형태로 존재할 수 있기 때문에 면 보간법을 활용하기 위해서는 추가적인 프로세스가 필요하다. 또한 통계정보는 표현하고자 하는 값에 따라 명목척도, 서열척도, 등간척도, 비율척도로 구분 가능하지만 면 보간법을 적용하기 위해서는 통계정보가 비율척도로 표현될 수 있는 값으로 존재해야 한다. 이에 따라 비율척도 이외의 값을 표현하는 통계정보에 대한 변환을 수행할 수 있는 방법론도 정립되어야 한다.

기존 통계지도 서비스에서 활용하는 통계공간단위인 행정구역, 집계구 등은 통계정보를 일정하게 표현하지 못하는 한계점을 갖는 것으로 다양한 선행연구들을 통해 파악되었다. 또한, 기존 통계공간단위로 활용되고 있는 행정경계 또는 집계구의 한계점을 보완할 수 있는 통계공간단위로 격자가 많이 활용되고 있음을 확인하였으며, 이에 본 연구에서는 격자를 기반으로 다양한 통계정보의 통계공간단위로의 변환에 대하여 체계적인 방법론을 제시하도록 한다. 또한, 현재 수집된 통계정보를 통계공간단위로 표현하기 위해 가장 많이 활용되는 방법론인 면 보간법은 통계정보가 공간단위로 존재한다는 가정 하에 적용될 수 있는 방법론이며, 통계정보의 데이터 척도 측면에서 보았을 때 비율척도에 속하는 통계정보를 변환하기에 적합한 방법론이다. 이에 따라, 다양한 형태로 존재하는 통계

정보에 일반적으로 적용될 수 있는 방법론으로는 한계가 있다. 따라서 다양한 형태로 존재하는 통계정보를 통계공간단위로 표현하기 위해서는 일정한 모양과 크기를 갖는 격자로 변환할 수 있는 일반적인 방법론이 필요하며, 본 연구에서는 이러한 방법론을 기존에 존재하는 이론들을 기반으로 통계정보에 적용하기 위해 보완되어야 하는 추가 프로세스를 포함하여 제시한다.

### 3. 통계정보의 격자 데이터 변환 방법론

#### 3.1 개요

본 연구에서는 다양한 형태로 존재하는 통계정보를 격자로 변환하기 위해 Figure 1과 같은 프로세스 기반의 데이터 변환 방법론을 제안한다.

데이터 변환을 수행하기 위한 입력데이터로 통계정보와 격자가 정의되어야 한다. 통계정보는 통계공간단위를 통해 표현하고자 하는 통계 값을 포함하여야 하며, 격자는 표현하고자 하는 통계정보의 목적에 알맞은 크기가 정의되어야 한다.

격자로의 데이터변환을 위한 첫 번째 단계는 지오코딩 단계이다. 대부분 통계정보는 비공간데이터로 수집되기 때문에 격자로 변환하기 위해서는 지오코딩을 통해 공간데이터로 변환한다. 만약 수집되는 통계정보가 공간데이터 형태로 존재하면 본 단계를 수행

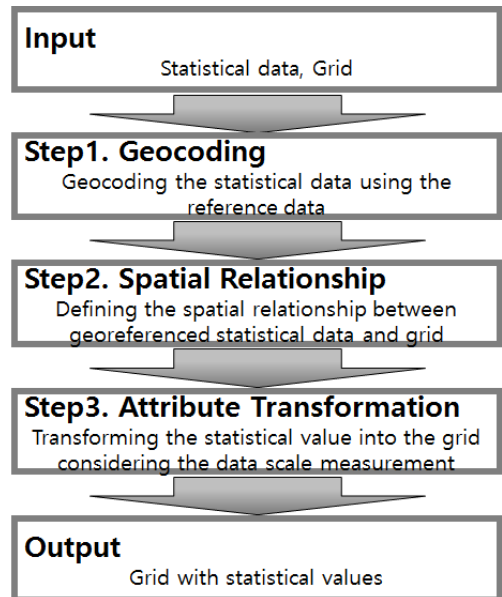


Figure 1. Overview of the Method

하지 않고 공간관계 정의 단계를 바로 수행하도록 한다.

두 번째는 공간관계 정의 단계로, 공간데이터로 존재하는 통계정보와 격자와의 공간관계를 파악하여 통계정보가 어느 격자 객체로 변환되어야 하는지를 파악하는 단계이다.

세 번째는 속성정보 변환 단계로, 공간관계 정의 단계를 통해 확인된 격자 위치에 통계 값을 변환하기 위한 단계이다. 본 과정을 수행함에 있어 통계 값이 갖는 데이터 척도를 고려한 변환이 되도록 한다. 이러한 프로세스 통해 통계정보는 입력데이터에서 정의된 격자를 통해 표현될 수 있다. 각 단계별 세부 방법론은 다음과 같다.

### 3.2 지오코딩 단계

지오코딩은 주소, 우편번호, 장소명 등 지리적 참조가 가능한 요소를 통해 지리적 좌표로 변환하는 과정을 의미한다. 따라서 지오코딩을 수행하기 위해서는 지리적 참조를 가능케 하는 참조데이터(Reference Data)가 정의되어야 하며, 참조데이터와 매칭되도록 하는 방법론(Method)이 정의되어야 한다. 이러한 두 요소를 로케이터(Locator)라 하며, 지오코딩을 수행하기 위해서 반드시 정의되어야 한다[10].

통계정보는 수집하는 방법에 따라 통계항목에 대한 통계 값만 있는 비공간데이터로 수집되거나, GPS 등을 통해 공간데이터로 수집될 수도 있다. 그러나 대부분의 통계정보는 비공간데이터로 수집되기 때문에 통계격자로 변환하기 위해서는 지오코딩을 통해 공간데이터로 변환시켜줘야 한다. 수집된 통계정보의 지오코딩을 수행하기 위해서는 로케이터가 정의되어야 하며 그 중에서도 통계정보를 실제와 가장 가깝게 표현하는 참조데이터의 정의가 중요하다.

본 논문에서는 통계정보의 지오코딩을 위해 정의되는 참조데이터를 다음과 같이 2가지로 구분한다. 첫 번째는 직접 참조데이터로서 통계정보가 표현하고자 하는 값을 직접적으로 표현할 수 있게 하는 참조데이터를 의미한다. 예를 들어, 지가에 대한 통계 값을 표현하기 위해 참조데이터로 필지데이터를 활용하는 것은 직접 참조데이터를 활용하는 것이다. 두 번째는 간접 참조데이터로서 통계정보가 표현하고자 하는 값을 가장 실제와 가깝게 표현해 줄 수 있는 참조데이터를 의미한다. 예를 들어, 인구에 대한 통계 값을 표현하기 위해 집계구를 활용하는 것이 간접 참조데이터를 활용한 지오코딩이라 할 수 있다.

수집한 통계정보를 지오코딩하여 공간데이터로 변

환할 때 가장 이상적인 상황은 직접 참조데이터를 사용하는 것이겠지만, 각 통계정보를 직접적으로 표현할 수 있는 참조데이터를 사용하는 것은 제한적이다. 따라서 대부분의 통계정보는 간접 참조데이터를 통한 지오코딩을 수행할 수 밖에 없기 때문에 통계정보가 올바르게 격자로 변환되기 위해서는 간접 참조데이터 정의 시 통계정보가 표현하고자 하는 실제 데이터의 분포를 가장 잘 반영할 수 있는 참조데이터가 정의되는 것이 중요하다.

### 3.3 공간관계 정의 단계

데이터가 공간데이터로 존재하기 위해서는 데이터를 공간상에 표현할 수 있게 하는 위치정보와 객체의 특성을 묘사할 수 있는 속성정보가 존재해야 한다. 마찬가지로, 본 연구에서 변환하고자 하는 격자는 지도상에 통계정보를 표현할 수 있게 하는 공간데이터이므로 격자로의 데이터 변환방법은 이러한 공간데이터의 특성을 고려하여 위치정보 측면과 속성정보 측면을 구분하여 변환 방법을 제시한다.

공간관계 정의 단계의 목적은 통계정보와 격자간의 공간관계를 데이터 변환 될 수 있는 관계로 정의하여 통계정보가 어느 격자 객체로 변환 될 지를 정의하는 단계이다. 데이터 변환을 위한 통계정보와 격자와의 공간 관계는 벡터로 표현되는 두 객체 사이의 공간관계를 정의한 국제표준인 DE-9IM (Dimensionally Extended Nine-Intersection Model)을 참고할 수 있다[1]. DE-9IM에서는 기하유형에 따른 다양한 공간관계를 정의하고 있는데, 통계정보가 격자로 올바르게 변환되기 위해서는 Figure 2와 같이 격자가 통계정보를 포함(Contains)하도록 공간관계가 정의 되어야한다.

그러나 통계정보가 길이나 면적을 갖는 참조데이터로 표현될 경우에는 하나의 통계정보 객체가 다수의 격자에 영향을 주는 교차(Crosses) 관계가 정의될 수 있다. 이러한 경우 올바른 변환을 위해 통계정보 객체가 분포하는 공간적인 영향력을 유지하면서 포함 관계로 정의되도록 해주어야 한다. 따라서 길이나 면적을 갖는 객체가 격자와 교차 관계가 성립할 경우에는 격자를 기준으로 통계정보 객체를 분할하여 통계정보 객체의 공간적인 영향력은 유지하되 포함 관계가 성립될 수 있도록 한다.

### 3.4 속성정보 변환 단계

통계정보와 격자사이의 공간관계 정의를 통해 위치 정보 변환을 수행할 수 있다면, 본 단계에서는 격자에

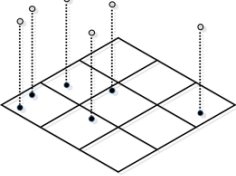
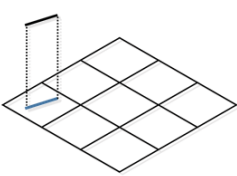
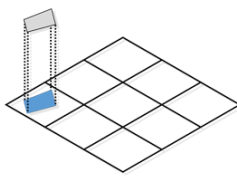
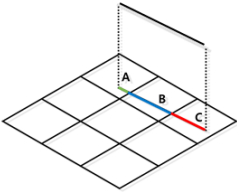
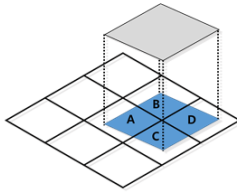
Spatial Relationship	Geometry Type of Statistical Data		
	Point	Line	Polygon
Contain			
Crosses	No crossing relationship possible		

Figure 2. Defining the Spatial Relationship in Geometry Type

Table 1. Main Method of Attribute Transformation in Data Scale Measurement

Geometry type	Point	Line	Polygon
Data scale measurement			
Nominal/ Ordinal	Raster Encoding Method (Most important, Predominant, Centroid, etc.)		
Interval/ Ratio	Aggregation	Length weighted interpolation	Areal weighted interpolation

통계정보가 표현하고자 하는 통계 값을 반영하도록 속성정보 변환을 수행한다. 통계정보를 통해 표현하고자 하는 값은 데이터 척도 측면에서 명목, 서열, 등간, 비율척도로 구분 할 수 있다. 명목척도는 데이터를 분류할 수 있는 정도의 측정수준을 가리키며, 서열척도는 데이터의 차이점뿐만 아니라 크고 작음 등의 순위화 된 측정수준을 말한다. 등간척도는 데이터의 속성간의 차이를 양적인 크기로 나타낼 수 있다. 마지막으로 비율 척도는 앞서 언급한 척도에 관한 특성을 모두 표현함과 동시에 기준을 의미하는 수치(예-0)를 갖고 있어 비율 계산이 가능한 척도를 의미한다[10]. 이에 따라, 통계 값을 격자로 변환하기 위해서는 통계 값 사이의 사칙연산이 불가능한 명목척도, 서열척도와 연산이 가능한 등간척도, 비율척도를 구분하여 통계정보의 데이터 척도를 고려한 속성정보 변환이 수행되어야 한다.

Table 1은 통계정보의 기하 유형과 데이터 척도를 고려하였을 때 속성 값 변환을 위해 적용될 수 있는 방법론을 나타낸다. 명목척도 또는 서열척도로 표현

되는 통계정보 예/아니요, 남자/여자, 고속도로/국도/지방도와 같이 사칙연산이 불가능한 값을 갖기 때문에 하나의 격자에 다수의 통계정보 객체가 포함될 경우에는 대푯값을 선정하여 할당하여야 한다. 이 때, 대푯값을 선정하기 위한 방법론은 레스터 변환 방법론 중 가장 중요한 값을 할당하는 Most Important, 격자 내에 가장 많이 존재하는 값을 할당하는 Predominant, 격자의 중심에 가장 가까운 객체의 값을 할당하는 Centroid 방법 등이 적용될 수 있다.

등간척도나 비율척도를 통해 표현되는 데이터들은 양을 표현할 수 있기 때문에 연산이 가능하다. 이러한 척도를 표현하는 데이터는 온도, 점수, 인구, 연령, 가격, 소득 등 대표적인 통계정보들이 등간척도나 비율척도에 속한다. 등간척도나 비율척도를 표현하는 통계정보가 점으로 표현되는 경우에는 격자 객체 내 존재하는 통계정보 객체들의 집계(Aggregation)을 통해 통계공간단위에 값을 할당하면 되는데, 집계를 위한 연산은 데이터가 표현하고자 하는 목적에 따라 합계, 평균, 표준편차, 최댓값, 최솟값 등을 수행할 수 있다.

통계정보가 선 또는 면으로 표현되는 경우에는 길이 나 면적을 갖기 때문에 통계정보가 표현하는 객체의 길이 또는 면적에 따른 가중치를 부여하여 보간법을 수행하는 Length/Areal Weighted Interpolation을 수행함으로써 통계 값을 격자로 변환 할 수 있다.

### 3.5 격자 데이터 변환 프로세스

본 연구에서는 다양한 형태의 통계정보를 통계 격자로 변환하기 위한 프로세스를 Figure 3과 같이 체계화할 수 있다. 입력데이터는 통계데이터와 격자로 정의된다.

통계데이터는 공간데이터 여부를 판별하여 공간데이터로 존재하지 않는 경우에는 지오코딩을 통해 공간데이터로 변환한다. 통계데이터가 공간데이터로 수집된 경우에는 지오코딩을 수행하지 않아도 된다. 공간데이터로 존재하는 통계정보는 격자와의 공간관계를 파악하여 변환 가능한 관계인 포함 관계 여부를 판별한다. 만약, 통계정보가 다수의 격자에 영향을 끼치는 교차 관계인 경우에는 격자를 기준으로 분할하여 통계정보와 격자와의 관계가 포함관계가 될 수 있게 만든다. 통계정보와 격자와의 관계가 포함 관계로 올바르게 정의되었다면 표현하고자 하는 통계 값의 데이터 척도를 고려한 속성정보 변환을 수행한다. 이

때, 데이터척도의 구분은 연산 가능성을 고려하여 통계 값 사이에 사칙연산이 불가능한 명목척도, 서열척도는 대푯값을 할당할 수 있는 방법론이 적용되며, 사칙연산이 가능한 등간척도, 비율척도는 집계이나 가중치 보간법을 통해 통계 값을 변환하도록 한다. 이러한 과정을 거쳐 통계정보의 위치정보 및 속성정보가 조건을 만족하는 격자 객체로 변환이 완료되면 통계 격자를 생성 할 수 있다.

## 4. 방법론 적용 및 검증

### 4.1 실험 환경 및 방법

본 부분에서는 서울시 A지역의 인구 밀도에 대하여 제시한 방법론을 적용하여 격자로 변환한다. 또한 인구 밀도를 표현하는 원본데이터가 상이하더라도 격자로 변환되었을 때에는 유사한 데이터를 표현할 수 있어야 하므로 이를 확인하기 위해 집계구와 건물데이터를 참조데이터로 정의하여 격자 변환 시 데이터 표현에 차이가 있는지를 분석하도록 한다.

입력데이터는 서울시 A지역에 대한 총 인구수 정보와 다양한 크기의 격자 변환 결과를 확인하기 위해 4가지 종류(50m, 100m, 200m, 400m)의 격자를 사용하였으며, 총 인구수 정보의 경우 2014년 국토지리원의 국토조사 발전계획 수립 및 플랫폼 기반구축 사업의 일환으로 수집된 주민등록인구 정보를 활용하였다. 주민등록인구 정보는 개인정보가 모두 제거된 상태로 활용되었다[13].

제시된 프로세스를 기반으로 가장 먼저 수행되는 사항은 통계정보로 정의한 총 인구 수의 공간데이터 유무를 판단하는 것이다. 총 인구 수 데이터는 비공간 데이터로 존재하기 때문에 격자 변환을 위해서는 간접 참조데이터에 총 인구 수를 지오코딩하여 공간데이터로 변환하여야 한다. 본 실험에서는 총 인구 수를 표현할 수 있는 간접 참조데이터로 집계구와 건물을 각각 정의하여 격자로 변환한 뒤에 결과를 비교하도록 한다. Figure 4와 5는 각각 집계구와 건물에 서울시 A지역의 총 인구 수를 지오코딩하여 표현한 것의 일부분이다.

지오코딩 수행 이후 집계구, 건물을 기반으로 표현되는 총 인구수와 격자와의 공간관계를 파악하여 어느 격자 객체에 총 인구 수 정보를 변환할 지를 파악한다. 이 때, 총 인구 수 정보를 표현하고 있는 집계구와 건물은 면으로 표현됨에 따라 하나의 집계구, 건물 객체가 다수의 격자 객체에 공간적인 영향을 끼칠 수

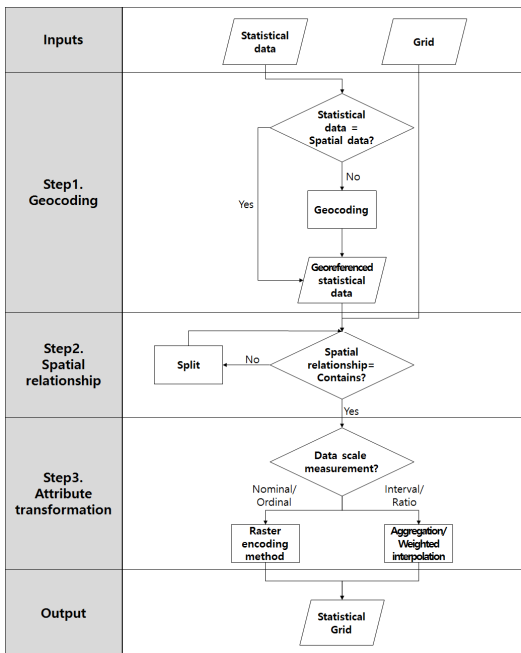


Figure 3. Process of the Method



Figure 4. Displaying the Number of Population through the Census Block

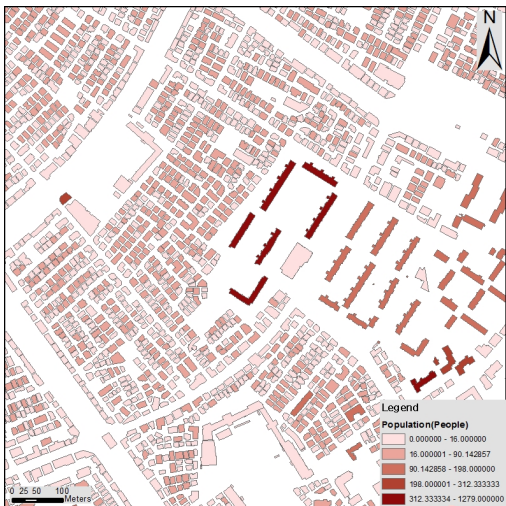


Figure 5. Displaying the Number of Population through the Building

있으며, 이러한 경우에는 격자를 기준으로 집계구, 건물 객체를 분할하여 두 객체사이의 관계가 포함 관계가 될 수 있게 한다.

공간관계 정의를 통해 위치정보 변환을 수행 한 뒤에는 격자에 통계 값을 할당하기 위한 속성정보 변환을 수행한다. 본 부분에서 다루는 총 인구 수는 비울척도로 구분되는 통계 값이며, 면으로 표현되고 있으므로 면적 가중 보간법을 통한 통계 값 변환이 가능하며 수식은 Eq. (1)과 같다.

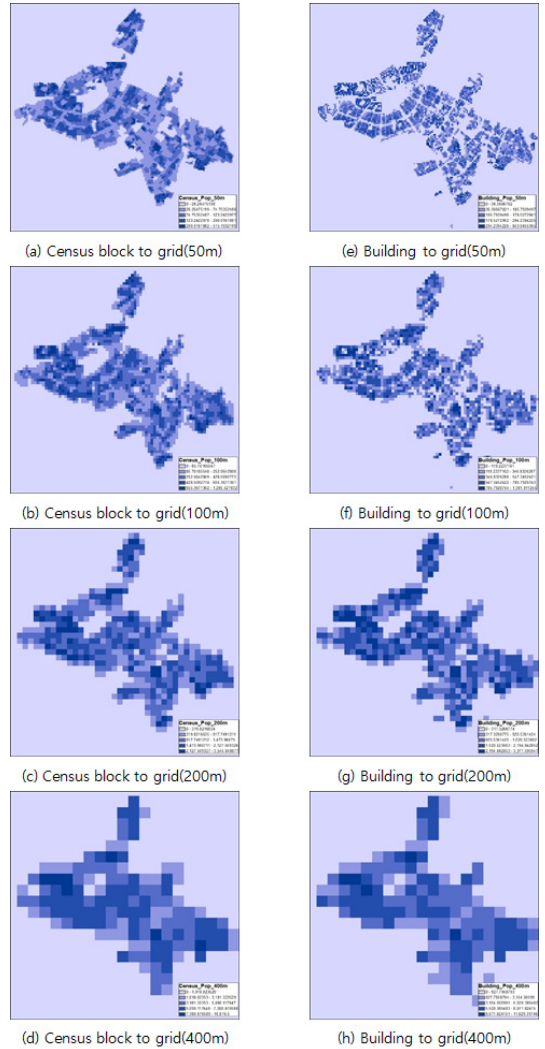


Figure 6. Results of the Visualization of Population Density on the Grid

$$\hat{P}_t = \sum_{s=1}^k \left( \frac{A_{ts}}{A_s} P_s \right) \quad (1)$$

격자로 변환되는 인구( $\hat{P}_t$ )는 해당 격자에 조금이라도 포함되는 총 인구수 객체들이 표현하는 인구 수의 합으로 표현할 수 있다. 각 총 인구수 객체들이 해당 격자 내에서 표현하는 인구 수는 원래 면적( $A_s$ )과 해당 격자에만 포함되는 면적( $A_{ts}$ )의 비를 총 인구수( $P_s$ )에 곱해 구할 수 있다.

이러한 과정을 거쳐 도출된 서울시 A지역의 총 인구수를 각 격자의 면적으로 나누어 도출한 서울시 A지역의 인구 밀도에 대한 격자 변환 결과는 Figure 6과

Table 2. Summary of the Statistical Results

Type	Size	Count	Average (person/m <sup>2</sup> )	Standard Deviation	Moran's I	Z-score	P-value
Census Block	Original	1055	0.051063	0.034247	0.109226**	18.858959	0
	50m	29575	0.00855	0.017727	0.899091**	218.0807	0
	100m	7480	0.008451	0.01679	0.8576**	104.3645	0
	200m	1892	0.008353	0.015582	0.804722**	49.05068	0
	400m	484	0.008163	0.014073	0.758281**	23.24169	0
Building	Original	25448	0.157552	1.21961	0.002673**	7.304974	0
	50m	29575	0.009048	0.023645	0.645261**	156.5308	0
	100m	7480	0.008944	0.01979	0.73295**	89.19787	0
	200m	1892	0.00884	0.017336	0.74721**	45.54458	0
	400m	484	0.008639	0.015199	0.722995**	22.16305	0

\*\* is significant at the 0.01 level

같다. Figure 6의 (a)~(d)까지는 집계구를 간접 참조데이터로 정의하여 지오코딩 된 총 인구수에서 인구 밀도를 계산하여 각각 50m, 100m, 200m, 400m의 격자로 변환하였을 때의 결과이며, (e)~(h)까지는 건물을 간접 참조데이터로 정의하여 지오코딩 된 총 인구수에서 인구 밀도를 계산하여 각각 50m, 100m, 200m, 400m의 격자로 변환하였을 때의 결과이다.

한편, 동일한 통계정보를 서로 다른 참조데이터로 정의한 뒤 격자로 변환 할 경우에도 동일한 데이터 분포 표현을 할 수 있는지 검증하기 위해 추가적인 실험을 수행하였다.

격자 변환과 같이 수집된 통계정보의 공간단위를 변경함으로써 공간단위 수정가능성의 문제(Modifiable Areal Unit Problem: MAUP)를 야기할 수 있다[3,4,14]. 이러한 문제가 정량적으로 측정되어 정의되기는 힘들지만, 데이터의 공간적 분포 정도를 확인할 수 있는 공간적 자기상관성(Spatial autocorrelation)을 통해 왜곡여부를 간접적으로 파악할 수 있다[15]. 다시 말해, 올바른 변환이 수행되었다면 격자로 변환된 뒤에도 원본데이터가 표현하는 공간적인 분포와 유사해야 한다. 따라서 본 논문에서는 서로 다른 두 간접 참조데이터의 올바른 격자 변환 수행이 이루어졌는지 여부를 검증하기 위해 집계구, 건물로 인구 밀도가 표현되었을 때와 격자로 변환하여 표현하였을 때의 인구 밀도에 대한 공간적 자기상관성을 모란지수(Moran's I)를 통해 확인하였다. 모란지수는 공간상에 존재하는 통계 값의 패턴이 1에 가까울수록 밀집(Clustered)되며, -1에 가까울수록 분산(Dispersed)되어 있으며, 0에 가까울수록 공간적 자기상관성이 없다(Random)고 본다

[16]. Table 2는 집계구 및 건물에 대한 원본과 격자 분포의 통계수치를 표현한다.

#### 4.2 실험 결과

서울시 A지역의 인구 밀도를 격자로 변환하였을 때 격자 크기에 따라 분포의 정도는 조금씩 차이가 있었지만 최소단위 격자와 원본데이터의 분포를 비교했을 때 유사한 패턴 표현하고 있는 것을 Figure 6을 통해 확인할 수 있었다. 이것은 원본의 분포와 유사하게 표현하면서 다른 공간과의 객관적인 비교를 가능하게 한다.

한편, 간접 데이터에 따른 데이터 변환 결과를 비교하기 위해 집계구와 건물의 원본을 통한 총 인구수 분포와 격자로 변환했을 때의 인구 밀도의 분포를 모란지수를 통해 정량적인 수치로 Table 2와 같이 도출해 내었다. 실험 결과, 집계구로 표현되는 데이터를 격자로 변환하여 표현한 것과 건물로 표현되는 데이터를 격자로 변환하여 표현하는 경우 모두 원본데이터의 분포와 격자 데이터의 분포가 유의수준 내에서 밀집으로 동일한 군집패턴을 보이는 것으로 나타났다. 또한 격자의 크기에 따라 표현되는 인구 밀도에 대한 평균과 표준편차도 큰 차이가 나타나지 않는 것으로 확인됨에 따라 격자의 크기가 달라짐에도 불구하고 격자가 표현하는 인구 밀도는 유사한 통계 값과 분포로 표현된다는 것을 확인하였다.

이에 따라, 참조 데이터의 종류가 다르더라도 인구 밀도를 표현하는 올바른 참조 데이터를 활용하며, 본 연구에서 제안하는 체계적인 격자변환을 수행한다면 격자 변환의 결과물은 유사하다는 것을 확인할 수 있었다.



## 5. 결론 및 향후연구

본 연구에서는 다양한 형태로 존재하는 통계정보를 격자로 표현하기 위해 필요한 데이터 변환 방법론을 체계적으로 제시를 하였다. 제시한 방법론을 검증하기 위해 서울시 A지역의 인구 밀도를 여러 가지 크기의 격자로 변환하였으며, 격자 변환 시 간접참조 데이터는 2가지 종류를 활용하여 참조데이터의 종류가 다르더라도 통계정보를 표현할 수 있는 올바른 참조 데이터를 사용한다면 격자 변환 시 최종 결과가 유사하게 변환됨을 공간 자기상관성을 기준으로 확인하였다.

본 연구에서 제시하고 있는 격자로의 데이터 변환 방법은 지속적으로 논의되는 통계공간단위로서 격자가 더 적극적으로 활용 될 수 있도록 할 것이다. 향후에는 이러한 데이터 변환 방법을 기반으로 고도화된 격자 변환을 위해 다양한 연구가 필요할 것이다. 본 연구에서는 정사각형 모양의 격자를 기반으로 한 데이터 변환 방법에 대한 연구 실험을 진행하였는데, 격자 체계라는 것은 다양한 모양이 될 수 있기 때문에 향후 연구에서는 이러한 점을 고려하여 다른 형태의 격자에도 제안된 방법론을 적용하는 연구가 필요하다. 또한 데이터 변환을 통해 생성된 격자가 유용하게 활용되기 위해서는 데이터 표현, 분석 등에 필요한 적정 격자 크기에 대한 연구도 필요하며, 이를 위해 여러 데이터를 비교 분석하여 정량적인 결과 도출 할 필요가 있다.

## References

- [1] Kim, K. 2011, Effects of the Modifiable Areal Unit Problem (MAUP) on a Spatial Interaction Model, *Journal of the Association of Korean Geographers*, 46(2):197-211.
- [2] Sudsom, N; Techato, K; Thammapalo, S; Monprapussorn, S. 2012, Grid and Census: a Geographic Sampling Strategy for Studying Dengue Vector Breeding Sites in Urban Area, Paper presented at the 33<sup>rd</sup> Asian Conference on Remote Sensing, Pattaya, November 26-30.
- [3] Shin, M. Y; Yun, J. I; Suh, A. S. 1999, Estimation of Daily Maximum/Minimum Temperature Distribution over the Korean Peninsula by Using Spatial Statistical Technique, *Journal of the Korean Society of Remote Sensing*, 15(1):9-20.
- [4] Song, M. Y; Jung, K. S; Lee, G. H; Kim, Y. S; Shin, Y. A. 2014, DAD Analysis of Yongdam Dam Watershed Using the Cell-Based Automatic Rainfall Field Tracking Method, *Journal of the Korean Association of Geographic Information Studies*, 17(3):68-81.
- [5] Kim, J. Y; Park, I. S; Park, C. Y; Park, S. K. 2010, Effects of Grid Size on Noise Prediction Results of Road Traffic Noise Map, *Korean Society for Noise and Vibration Engineering*, 20(2):199-204.
- [6] Um, D. Y; Lee, B. S. 2012, Analysis of the Effect on the Location Evaluation of Golf Course according to the Unit Grid Size applied in Slope Analysis (In flank of Environment), *Korea Society of Surveying, Geodesy, Photogrammetry, and Cartography*, 30(5):467-475.
- [7] Lee, S; Kim K. 2007, Representing the Population Density Distribution of Seoul Using Dasymetric Mapping Techniques in a GIS Environment, *Journal of the Korean Cartographic Association*, 7(2):53-67.
- [8] Reibel, M; Agrawal, A. 2007, Areal interpolation of population counts using pre-classified land cover data, *Population Research and Policy Review*, 26(5-6):619-633.
- [9] Sridharan, H; Qiu, F. 2013, A spatially disaggregated areal interpolation model using light detection and ranging-derived building volumes, *Geographical Analysis*, 45(3):238-258.
- [10] Lee, J. 2009, GIS-based geocoding methods for area-based addresses and 3D addresses in urban areas, *Environment and Planning B: Planning and Design*, 36:86-106.
- [11] Clementini, E; Sharma, J; Egenhofer, M. J. 1994, Modeling Topological Spatial Relations : Strategies for Query Processing, *Computers and Graphics*, 18(6):815-822.
- [12] Lee, H. Y; Noh, S. C. 2013, *Advanced Statistical Analysis : Theory and Practice 2nd Edition*, Moon Woo (in Korean).
- [13] National Geographic Information Institute. 2014, Development Plan of National Territorial Statistic Monitoring and Construction of Platform Project, Korea.
- [14] Lee, G. 2011, A review of Object and Field Perspective on Modifiable Areal Unit Problem,

Journal of the Korean Cartographic Association,  
11(1):25-32.

- [15] Swift, A; Liu, L; Uber, J. 2008, Reducing MAUP bias of correlation statistics between water quality and GI illness, *Computer, Environment and Urban Systems*, 32:134-148.
- [16] Zhang, N; Zhang, H. 2011, Scale Variance Analysis Coupled with Moran's I Scalogram to Identify Hierarchy and Characteristic Scale, *International Journal of Geographical Information Science*, 25(9):1525-1543.

---

Received : 2015.07.08

Revised : 2015.10.26

Accepted : 2015.10.27