

원격탐사 데이터베이스 구축 기술

이용주
yongju@sangju.ac.kr
상주대학교

정진완
chungcw@islab.kaist.ac.kr
한국과학기술원

Techniques for Building Remote Sensing Databases

Yong-Ju Lee
Sangju National University

Chin-Wan Chung
Korea Advanced Institute of Science and Technology

요약

본 연구에서는 원격탐사 데이터베이스를 구축하기 위한 요구사항들을 분석하고, 원격탐사 자료의 메타데이터 스키마를 설계하고 이를 구현한다. 특히, 현행 지형 공간 표준화 스펙을 따르는 자료구조가 연구되고, OMEGA(Object Management system for Geo-Spatial Applications) 객체지향 데이터베이스 관리시스템을 사용하여 SAIF(Spatial Archive and Interchange Format)를 기반으로 한 스키마 구현에 대해 기술한다. 이 연구는 차세대 데이터베이스 기술(예를 들면, 복잡한 데이터 타입 지원, 병렬처리, 사용자 편의성 인터페이스, 그리고 웹(Web) 데이터베이스 기술)에 대한 사전 기초 연구로써 실험적인 대용량 데이터베이스를 구축하기 위하여 수행되어 졌다.

1. 서론

복잡하고 대용량적인 특성을 가지고 있는 지형 공간(geo-spatial) 데이터의 처리는 최근 데이터베이스 분야에서 혁신적인 기술 변화를 요구하고 있다. 대용량 지형 공간 데이터베이스의 대표적인 예는 인공위성으로부터 수집되어지는 수십 테라 바이트(multi-terabyte) 분량의 원격탐사(remote sensing) 데이터베이스이다[1, 2, 3]. 원격탐사란 항공기나 인공위성 등의 플랫폼에 탑재된 센서(sensor)를 이용하여 지구상의 대상으로서 반사되는 전자파를 수집하여 대상으로에 관한 정보를 얻는 기술이다. 이 정보는 실시간(real-time)으로 자동 수신되기 때문에 그 데이터 량은 엄청나다. 더군다나 하나의 수신된 영상에 대하여 시스템 내에서는 여러 가지 영상 분석 작업이 수행되어 저장된다.

원격탐사 데이터베이스 시스템에서는 대용량 데이터의 저장 및 관리 뿐만 아니라 다양한 데이터 타입에 대한 질의 처리를 요구한다. 즉, 단순히 센서 이름, 날짜와 같은 텍스트 형태의 속성정보를 사용하여 질의할 수도 있고, 주어진 도면정보(예, 행정구역)에 의해 래스트(raster) 이미지의 잘라내기(cropping)를 요구할 수도 있다. 또는 내용 기반 영상 검색(content-based image retrieval)을 요구할 수 있다. 예를 들면, "1999년도에 입수한 서울지역에 관한 위성 영상들 중 구름분포가 30% 이하인 Landsat 영상들을 검색하라"와 같은 질의를 요구할 수 있다. Sequoia 2000 벤치마크[4]에서는 이러한 형태의 질의들을 볼 수 있다.

지금까지 데이터베이스 분야에서 지형 공간 데이터를 처리하기 위하여 많은 연구가 수행되어 졌다. 그들 대부분은 점, 선, 면과 같은 벡터(vector) 데이터에 대해서만 주로 연구되어 졌고, 위성 영상과 같은 래스트 데이터에 대한 연구는 거의 되어지지 않았다[2]. 벡터 데이터의 수집 및 정리, 그리고 지도제작 등에는 많은 비용과 노력이 요구된다. 어떤 지도가 수신된 혹은 그 이상의 기간동안 무리 없이 사용될 수 있다면 지도 제작비가 많이 든다는 것은 큰 문제가 되지 않으나, 비교적 짧은 기간 내에 넓은 지역에 걸쳐 발생하는 환경변화나 기후변화와 같은 그때그때 지도와해야 하는 경우 벡터 데이터베이스의 구축은 적당하지 못하다. 최근에는 항공사진이나 위성영상을 사용하여 시간에 따른 환경의 변화를 조사하는 것이 가능해짐에 따라 원격탐사 소 인은 래스트 데이터와 지상에 서 관찰한 벡터 데이터를 결합한 벡터/래스트 데이터베이스의 구축이 일반화되고 있다.

최근에 선진 각국에서는 날로 고갈되어 가는 국토자원의 새로운 탐사 수단으로써 자위탐사 위성, 기상 위성, 해양관측 위성 등 각종의 위성 상을 발사하여 환경변화, 국토개발, 자원개발, 그리고 기상예보 등 다양한 분야에 효율적으로 활용하고 있다. 그러나 우주 과학기술 개발이 뒤늦어 최근에 무궁화 인공위성을 발사한 국내 여건에서는 위성 자료들 다양한 응용분야에 실용화하기 위해서는 우선 극복해야 할 기술적 장애가 많다. 그 중에서도 위성에서 실시간으로 수신되는 막대한 분량의 원격탐사 자료를 데이터베이스화하여 일반 사용자들이 손쉽게 신속히 검색하여 사용할 수 있는 환경 구축이 가장 시급히 요구된다. 이는 앞으로 발생 가능한 긴급한 자료의 수요에 대비하며, 많은 노력과 시간이 요구되는 원격탐사 분석 자료를 공동 이용하여 일반 수요자가 분석 자

료를 손쉽게 활용할 수 있도록 해주는 사용자 측면에서 매우 중요하다. 본 연구에서는 원격탐사 데이터베이스를 구축하기 위한 요구분석 및 메타데이터 스키마 설계에 관한 연구를 수행한다. 현행 지형 공간 표준화 스펙을 따르는 자료구조가 연구되고, OMEGA(Object Management system for Geo-Spatial Applications) OODBMS (Object-Oriented Database Management System)를 사용하여 SAIF (Spatial Archive and Interchange Format)를 기반으로 한 스키마 구현에 대해 기술한다.

본 논문의 구성은 다음과 같다. 2장에서는 원격탐사 데이터베이스 구축을 위한 요구사항들을 분석 조사하였다. 이러한 요구사항들은 시키마 설계 시 지침서로 활용될 수 있다. 3장에서 원격탐사 데이터베이스의 메타데이터를 설계하고 4장에서 이를 구현하였다. OMEGA OODBMS를 이용하여 지형공간 표준화 스펙인 SAIF를 기반으로 한 스키마를 구현하였다. 마지막으로 5장에서 결론을 내렸다.

2. 원격탐사 데이터베이스 구축을 위한 요구분석

2.1 원격탐사 자료 개요

국내에서 현재 활용 가능한 대표적인 원격탐사 위성으로는 Landsat, SPOT, NOAA, GMS, 그리고 우리별 등이 있다. Landsat은 1972년 1호가 발사되어 지구자원의 위성탐사 시대를 열었고, 1984년 5호가 발사되어 자위위성의 실용화를 구현하게 되었다. Landsat에는 두 개의 센서 MSS와 TM이 있는데, MSS는 지상해상력 80m*80m, TM은 30m*30m이고 한 장면의 크기는 MSS, TM 모두 185km*185km이다. 불란서는 1986년 지상해상력 20m*20m (multispectral mode), 10m*10m (panchromatic mode)인 SPOT 위성을 성공리에 발사하여 운영 중에 있다. 기상위성은 근극궤도(near-polar orbit)인 미국 NOAA 위성과 정지위성인 일본의 GMS 3가 국내에 자료를 공급하여 주고 있다. 한편 국내에서는 금년 한국과학기술원 인공위성 연구센터에서 우리별 3호를 성공적으로 발사하여 시험 운영 중에 있다.

2.2 요구분석

일반적으로 원격탐사 자료 처리 시스템은 자료의 획득으로부터 변환, 전처리, 그리고 정보분석까지 전반적인 다양한 기능을 요구한다. 그러나 원격탐사 정보에 대한 분석 기능은 영상처리시스템(Image Processing System)과 지리정보시스템(Geographic Information System) 등을 통해 이용 가능하므로 본 연구에서는 데이터베이스 구축에 관한 요구사항 분석에 중점을 두었다.

원격탐사 데이터베이스를 구축하기 위한 상세한 요구사항은 아래와 같다.

- 위성영상 데이터베이스는 근본적인 디지털 영상자료(grid 정보) 이외에 영상에 관련된 속성정보도 포함하고 있다. 원격탐사 응용에 있어서 영상과 관련된 다양한 속성정보를 모으고는 디지털 영상을 이해하기 어렵다. 예를 들어 위성이름, 센서, 획득 영상의 계절 및

- 날짜, 기타 관련정보를 모르고는 입수된 영상을 이용하기는 어렵다.
- 하나의 위성영상을 다른 영상들과 구별하기 위하여 각 영상에 대한 유일한 식별자(image ID)가 요구된다. 식별자의 한 예는 센서, 영상 획득 날짜 및 시간을 결합한 형태이다. 즉 'TM.2Mar1999.101322'는 센서 TM으로 1999년 3월 2일 10시 13분 22초에 획득된 영상으로 참조된다.
- 위성영상을 이용할 때 동일지역을 같은 시간에 서로 다른 센서로 수집한 자료(multi-sensor image)간, 또는 동일지역을 같은 센서로 서로 다른 시간에 수집한 자료(multi-temporal image)간 서로 비교 분석해야 될 경우가 많다. 이를 위해 각 영상은 표준 지도투영법을 사용하여 2차원 좌표계에 공간 일치(spatial registration)시켜야 한다.
- 하나의 위성영상에 대해서 시스템 내에서는 여러 가지 영상 분석 작업이 수행될 수 있다. 따라서 원본으로부터 여러 버전(version)들이 생성될 수 있으며, 각 버전을 생성하기 위해서 수행된 처리과정(연행)을 조사하는 것은 중요한 일이다.
- 시스템간 상호연동(interoperability)을 향상시키고, 데이터베이스 개발을 촉진시키기 위하여 현행 지형 공간 표준화 스펙을 따르는 자료 구조가 제공되어야 한다.
- 막대한 분량의 위성영상 자료에 대한 효율적인 저장 및 관리를 위하여 영상압축 기법이 적용되어야 한다.

3. 위성영상의 메타데이터(metadata)

일반적으로 원격탐사 데이터베이스에 관한 질문들은 다음과 같은 것들이다.
 "1999년도에 입수한 위성영상들 중 구분본도가 30% 이하인 Landsat 영상들을 검색하라."
 "파일을름 TM2Mar1999101322.img 위성영상을 생성하기 위하여 수행된 처리과정은 무엇인가?"
 "위도 북위 33' ~ 38' 경도 동경 120' ~ 129' 사이에 위치하며 UTM 지도 투영법을 적용한 모든 영상 식별자를 검색하라." 즉 이러한 질의를 지원하기 위한 메타데이터에는 영상 그 자체뿐만 아니라 영상과 관련된 속성정보, 시간정보, 위치정보 등이 포함되어야 한다. <표 1>은 2장에서 수행한 요구분석을 기반으로 하나의 위성영상에 대한 메타데이터를 나타내고 있다.

속성정보	
satellite name	Landsat-5
sensor	TM
elevation	700 (km)
no. bands	7
bands (μ m)	0.45-0.52 0.52-0.60 0.63-0.69 0.76-0.90 1.55-1.75 10.40-12.50 2.08-2.35
image ID	TM.2Mar1999.101322
Grid 정보	
no. row	2983
no. columns	4320
unit of measure	meters
pixel size	30 (meters)
Spatial registration	
map projection	UTM
lat/lon of center	129 East 23 North
nw corner	1090000 752000
bounding box:	
lower left lat/lon	128.5722899 22.5837576
upper left lat/lon	128.7300591 23.5030555
upper right lat/lon	130.9946489 23.7048989
lower right lat/lon	130.7163527 22.8193919
처리과정	
geometric correction	least square method
image classification	density slicing
image enhancement	contrast enhancement

<표 1> Landsat Metadata

4. 스키마 구현

본 연구에서는 OMEGA(Object Management system for Geospatial Applications)[5] OODBMS를 이용하여 SAIF(Spatial Archive and Interchange Format)[6]을 기반으로 한 스키마를 구현하였다. OMEGA는 ODMG 표준[7]에서 정의하고 있는 객체모델을 따른다

ODL(Object Definition Language), OQL(Object Query Language), C++OML(Object Manipulation Language)를 지원한다. 또한 공간객체를 위한 확장된 데이터 모델 및 연산 등을 지원한다[8]. SAIF¹⁾는 현재 국제 표준화로 구체화되고 있는 OpenGIS(Open Geodata Interoperability Specification) 및 SQL/MM(Multimedia) Spatial 개발과 밀접하게 연관되어 있다.
 <표 2>는 <표 1>의 Landsat 메타데이터를 SAIF 타입으로 변환한 것이다. SAIF 클래스의 애트리뷰트(attribute)들은 왼쪽 열에 나타내고, Landsat 메타데이터는 오른쪽 열에 나타내고 있다.

ImageInfo	
agency	NASA
imagingSystem	Lansat-5
bands	1 0.45-0.52 2 0.52-0.60 3 0.63-0.69 4 0.76-0.90 5 1.55-1.75 6 10.40-12.50 7 2.08-2.35
imageID	TM.2Mar1999.101322
beginningDate	2Mar1999
Grid	
positionOfOrigin	1090000 752000
OrderAndOrigin	rowOrder
ordering	nw
originCenter	false
cellPositionAtCorner	0,0
originCoordinates	0,0
spacing	gridDimensions 2983, 4320 gridSpacing 30, 30 units meters, meters
OffsetStructure	none
StructuredData	
contentValues	actual image
indexingScheme	sequentialArray
GeneralLocation	
BoundingBox	UTM
coordSystem	
minPoint	128.5722899 22.5837576
maxPoint	130.9946489 23.7048989
Lineage	
processHistory	
ActionHistory	geometric correction least square method image classification density slicing image enhancement contrast enhancement

<표 2> SAIF type

<표 2>는 5개의 클래스(ImageInfo, Grid, StructuredData, GeneralLocation, Lineage)로 구성되어 있다. ImageInfo는 자료 수신기관, 이미지 시스템, 밴드, 영상 식별자 및 취득 날짜를 포함하고 있다. 래스트의 지형정보를 묘사하고 있는 Grid는 다음과 같은 애트리뷰트들을 포함하고 있다.
 - positionOfOrigin: 영상 기준점의 좌표값
 - OrderAndOrigin: ordering은 데이터 기록 순서, originCenter는 positionOfOrigin의 위치, cellPositionAtCorner는 셀의 기원(origin)이 모퉁이에 위치하면 'true', 중심에 위치하면 'false'를 나타낸다. originCoordinates는 positionOfOrigin의 grid cell 좌표값을 나타낸다.
 - spacing: gridDimensions는 각 좌표의 크기, gridSpacing은 픽셀의 크기(단위는 meter), units는 gridSpacing을 위한 측정 단위, 그리고 OffsetStructure는 triangular나 hexagonal grid를 위해 사용된다.
 StructuredData는 다음을 포함하고 있다.
 - contentValues: 실제 영상 자료를 저장한다.
 - indexingScheme: run-length encoded와 같은 압축된 형태로 저장되었는지 명시한다. sequentialArray는 압축 안된 상태를 나타낸다.
 GeneralLocation은 사용된 지도투영법 및 영상에 대한 좌소경계사각형을 포함하고 있으며, Lineage는 사용된 처리과정, 알고리즘 등을 포함하고 있다.
 본 연구에서는 OMEGA 내에서 SAIF 타입을 구현하였다. <그림 1>은 SAIF 클래스를 ODMG 2.0 ODL로 변환한 것이다. 이는 OMEGA에서 정의될 수 있다. 여기서 Agency, SceneID, DateTime.

1) SAIF 스키마는 300여 개의 클래스를 정의하고 있다. 이를 광범위한 스키마를 모두 다루려는 것은 비효율적인 일이다. 따라서 SAIF에서는 필요한 일부만 표현할 수 있는 프로파일(profile)을 제공하고 있다.

GridOrder, OriginReference, UnitMeaning, OffsetStructure, 그리고 indexStructure은 원시(primitive) 타입으로부터 유도된 것이며, Point, BoundingBox, Image는 OMEGA 공간 데이터 타입으로부터 유도되었으므로 공간 검색을 수행할 수 있다. ImageBand와 같은 구조체 타입은 이태릭체로 표현하였다.

```

struct ImageBand {
    string    bandID;
    float    wavelengthMin;
    float    wavelengthMax;
};
class ImageInfo {
    attribute Agency    agency;
    attribute string    imagingSystem;
    attribute ImageBand bands;
    attribute list<SceneID> imageID;
    attribute DateTime  beginningDate;
};
struct GridReference {
    GridOrder    ordering;
    OriginReference    originCorner;
    boolean      cellPositionAtCorner;
    array<short>  originCoordinates;
};
struct GridFramework {
    array<short>  gridDimensions;
    array<long>   gridSpacing;
    array<UnitMeaning>  unit;
    OffsetStructure    offsetStructure;
};
class Grid {
    attribute Point    positionOfOrigin;
    attribute GridReference    orderAndOrigin;
    attribute GridFramework    spacing;
};
class StructuredData {
    attribute indexStructure    indexingSchema;
    attribute Image             contentValues;
};
class GeneralLocation {
    attribute BoundingBox    MBR;
};
struct ActionHistory {
    string    operationPerformed;
    string    methodUsed;
};
class Lineage {
    attribute ImageInfo    dataSource;
    attribute list<ActionHistory>    processHistory;
};
class Raster {
    attribute string    satellite;
    attribute string    sensor;
    attribute unsigned short    noBonds;
    attribute ImageInfo    imageInfo;
    attribute Grid         grid;
    attribute StructuredData    structuredData;
    attribute GeneralLocation    generalLocation;
    attribute Lineage        lineage;
};
    
```

<그림 1> ODMG 2.0 ODL

다음은 3장에서 예시한 3개의 질문에 대해 OQL 질의문으로 표현한 것이다. 이 질의에서는 확장된 공간 연산자 CloudCover와 inside를 포함하고 있다. OMEGA에서 CloudCover와 같은 래스트 데이터 처리 연산자는 아직 구현되어 있지 않다.

```

select r.contentValues
from r in Raster
where r.satellite = "Landsat" and r.beginningDate between 1Jan1999
and 31Dec1999 and CloudCover(contentValues) < 30;

select r.lineage
from r in Raster
where r.imageID = "TM2Mar1999101322";

select r.imageID
from r in Raster
    
```

where r.MBR.coordSystem = "UTM"
and r.MBR inside rectangle(120, 33, 129, 38);

5. 결론

본 연구에서는 원격탐사 데이터베이스를 구축하기 위한 관련 기술들을 서술하였다. 설계된 스키마는 현행 지형 공간 표준화 스펙을 따른다. 따라서 시스템간 상호연동을 향상시키고 관련 데이터베이스 개발을 촉진시킬 수 있는 기반이 될 수 있다.

원격탐사와 같은 복잡하고 대용량적인 자료의 효율적인 처리를 위해서는 많은 차세대 데이터베이스 기술들이 요구된다. 예를 들면, 이미지, 비디오, 오디오, 다차원 array 데이터 타입이 지원되어야 하고, 사용하기 쉽고 시각적인 새로운 차원의 사용자 편의성 인터페이스가 설계되어야 한다. 또한 시스템 성능 향상을 위한 병렬 처리 기법 및 영상 압축 기법, 그리고 정보 공동 이용을 위한 웹(Web) 데이터베이스 기술이 연구되어야 한다. 본 연구는 이와 같은 차세대 데이터베이스 기술에 대한 사전 기초 연구로서 실험적인 대용량 데이터베이스를 구축하기 위하여 수행되어 졌다.

참고문헌

- [1] T. Barclay, R. Eberl, J. Gray, J. Nordlinger, G. Raghavendra, D. Slutz, G. Smith, and P. Smoot, "Microsoft TerraServer", June 1998, white paper, <http://www.terraserver.microsoft.com>.
- [2] J. Patel et al., "Building a Scalable Geo-Spatial DBMS: Technology, Implementation, and Evaluation," Proceedings of the ACM SIGMOD International Conference on Management of Data, pp. 336-347, 1997.
- [3] P. Brown and M. Stonebraker, "BigSur: A System for the Management of Earth Science Data," Proceedings of the 21st International Conference on Very Large Data Bases, pp. 720-728, 1995.
- [4] M. Stonebraker, J. Frew, K. Gardels and J. Meredith, "The SEQUOIA 2000 storage benchmark," Proceedings of the ACM SIGMOD International Conference on Management of Data, pp. 2-11, 1993.
- [5] Y. J. Lee and C. W. Chung, "Implementing an OMEGA Query Processor," The 4th International Conference on Computer Science and Informatics, Research Triangle Park, NC, U.S.A., October 23-28, pp. 221-223, 1998.
- [6] <http://www.env.gov.bc.ca/gdbc/fmebc>
- [7] R. G. G. Cattell, The Object Database Standard: ODMG 2.0, Morgan Kaufmann Publishers, Inc., 1977.
- [8] 박창원, 박호현, 홍남희, 정진완, "OMEGA 객체모델에 기반한 OMEGA 공간 데이터 모델," 한국정보과학회 데이터베이스 연구회지, 제13권, 제4호, 1998, pp.39-50.