정수웨이블릿변환(IWT)을 이용한 지형 자료의 압축 및 정밀 지형 효과 계산을 위한 활용 방법 고찰

정호준¹ · 이희순^{1,*} · 오석훈² · 박계순³ · 임형래³

¹(주)휴먼앤어스, 138-160, 서울특별시 송파구 가락동 11 베니스타워빌딩 4층

²강원대학교 에너지자원공학과, 200-701, 강원도 춘천시 강원대학길 1

³한국지질자원연구원 광물자원연구본부, 305-350, 대전광역시 유성구 과학로 124

Compression of Terrain Data using Integer Wavelet Transform (IWT) and Application on Gravity Terrain Correction

Hojoon Chung¹, Heuisoon Lee^{1,*}, Seokhoon Oh²,
Gyesoon Park³, and Hyoungea Rim³

¹Human and Earth, Seoul 138-160, Korea

²Department of Energy and Resources Engineering, Kangwon National University, Gangwon 200-701, Korea

³Mineral Resources Research Division, Korea Institute of Geoscience and Mineral Resources,
Daejeon 305-350, Korea

Abstract: Terrain data is one of important basic data in various areas of Earth science. Recently, finer DEM data is available, which necessary to develop a method that deals with such huge data efficiently. This study was conducted on the lossless compression of DEM data and efficient partial reconstruction of terrain information from compressed data. In this study, we compressed the wavelet coefficients of DEM, obtained from integer wavelet transform (IWT) by entropy encoding. CDF (Cohen-Daubechies- Feauveau) 3.5 wavelet showed the best compression ratio of about 45.4% and the optimum decomposition level was 3. Results also showed that a small region of terrain could be restored from the inverse wavelet transform with a part of the wavelet coefficients that are related to such region instead of whole reconstruction. We discussed the potential applications of the terrain data compression for precise gravity terrain correction.

Keywords: terrain data, wavelet transform, digital elevation model, compression

요 약: 지형 자료는 지구과학 여러 분야에서 중요한 기초 자료 중 하나이다. 최근 들어, 상세한 분해능을 가지는 DEM 자료가 활용가능하며 따라서 방대한 양의 자료를 효율적으로 다루는 방법이 필요하다. 본 연구에서는 방대한 DEM 자료의 무손실 압축 및 효율적인 복원에 대해 알아보았다. 이를 위해 정수웨이블릿 변환과, 엔트로피 부호화의 개념을 이용하여, 웨이블릿 계수의 부호화 및 일부 영역의 지형복원 방법을 고안하였다. 또한, 정밀 중력 지형보정 과정에서 이러한 연구 결과의 활용성을 검토하였다. DEM의 압축률이 가장 좋은 웨이블릿은 CDF3.5이며, CDF3.1 또는 CDF3.5 웨이블릿을 사용하여 3단계 정도로 분해를 하는 것이 최적의 선택이다 (약 45.4%의 압축률). 또한 웨이블릿변환의 다중단계분석 특성을 활용하여 웨이블릿계수의 일부만을 추출하여 지형의 일부만을 복원할 수 있었다.

주요어: 지형자료, 웨이블릿 변환, 수치표고모델, 압축

*Corresponding author: yihsoon@geolux.co.kr

Tel: +82-2-554-5451

Fax: +82-2-554-5452

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

서 론

지구과학분야에서 지형자료는 매우 중요한 기초 자료이다. 지질학, 지구물리학, 해양학, 기상학 등 거의 모든 지구과학 분야에서 지형자료는 기초자료로서 매우 중요한 의미를 갖는다. 최근 위성으로부터 얻어진 지형자료는 매우 정밀해지는 추세이며, 이에 따라 지

형자료의 양도 기하급수적으로 증가하는 추세이다. 본 연구는 이러한 방대한 지형자료를 효과적으로 저장, 복원할 수 있는 방법을 개발하여 지구물리학에서의 중력자료 처리를 기준으로 제시하려 한다.

중력탐사는 지하 매질의 밀도변화에 의해 생기는 중력 변화를 측정하여 지하 광체나 지질구조를 해석하는 탐사법이다. 그러나 중력 측정값은 지하 매질의 밀도뿐만 아니라 측정간의 고도 및 위도 차이, 측정 시각 차이에 따른 조석효과 차이, 중력계 특성과 변화 등의 영향을 받아 변한다. 따라서 지하 매질의 밀도변화를 제외한 다른 요인에 의한 측정간의 중력 변동들을 제거하기 위하여 위도보정, 조석보정, 계기보정, 고도보정, 부계보정 및 지형보정 등의 자료처리 과정을 거친다.

중력자료의 지형보정은 해머차트(Hammer, 1939)에 의한 방법이 제시된 이후로 디지털화된 방법이 적용되어 사용되어 왔으며, 1970년대에 컴퓨터 및 수치해석 기술이 발달되면서 종이지도에서 읽은 고도 자료들로부터 multiquadric equation으로 생성한 지형모델을 이용하여 지형보정을 실시하는 방법이 발표되었다(Hardy, 1971; Krohn, 1976). 국내에서도 Kwon et al. (1990), Lee and Lim(2010) 등이 multiquadric equation 및 수치표고모델(DEM)을 이용하여 지형효과를 자동으로 계산하는 프로그램을 개발하였다.

근래의 중력탐사는 정밀 GPS를 활용한 측정 좌표 정확성 향상 및 중력계 정밀도 향상에 힘입어 천부 소규모 대상체(공동, 지반 침하 등)를 대상으로 하는 고정밀 중력탐사에 대한 연구가 이루어지고 있다. 고정밀 중력탐사에서 정교한 지형보정은 탐사와 성패를 좌우하는 매우 중요한 요소로 작용한다(Butler, 1984; Aiken and Cogbill, 1998). Lee and Lim(2010)는 주어진 지형자료에 측정점 주변의 지형자료를 추가하여 지형효과를 좀 더 정확하게 보정할 수 있음을 보였다.

지형효과 계산의 정확도는 수치지형자료 분해능과 정밀도, 중력효과계산 알고리즘의 정확도 등에 의존한다. 중력효과 계산 알고리즘 중 계산 대상 지역을 사각기둥 또는 원기둥의 조합으로 근사하여 계산하는 방식은 광역 중력탐사에서는 효과적으로 적용 가능하나 정밀중력탐사에서는 정확도가 다소 떨어지며, 경사면 등을 효과적으로 근사할 수 있는 삼각요소법을 이용한 계산방법이 정확도가 더 높다(Lim et al., 2010).

지형효과 계산 정확도를 좌우하는 또 다른 요인은 정밀한 수치지형자료이다. 현재 무료로 배포되는 수

치표고모델로는 SRTM 90 m 수치표고자료가 널리 쓰이고 있는데 이 자료는 NASA에 의해 제작되었고 3"x3"(약 90 m) 해상도로 USGS를 통해 무료로 제공된다. SRTM 90 m 자료는 광역 중력탐사 자료의 지형보정에 효율적으로 적용될 수 있으나 정밀중력탐사를 위해서는 해상도가 너무 떨어지는 단점이 있다. 현재 국토지리정보원에서는 10 m 간격으로 전국 수치표고모델을 완성하였고, 주요 도시의 경우 1 m 간격 수치표고모델을 완성하였는데, 이 모델은 전 국토를 대상으로 구축되고 있는 중이다. 이러한 정밀 수치표고모델을 활용한다면 정밀중력탐사 지형보정을 지금보다 훨씬 정확하게 수행할 수 있게 되어 중력탐사의 활용범위를 넓히는데 도움이 될 것이다.

정밀 수치표고모델을 활용하는데 있어서 고려해야 되는 요인 중 하나가 방대한 양의 데이터를 효율적으로 다루는 문제이다. 개인용 컴퓨터의 연산속도나 인터넷 통신 속도 등이 매우 빨라졌고 하드디스크 등의 저장매체 용량도 획기적으로 증가되어 왔으나 그에 비례하여 다루어야 되는 데이터의 양도 엄청나게 증가하였기 때문에, 방대한 양의 데이터를 효율적으로 다루는 수단은 여전히 실용적인 측면에서 매우 중요한 기술이 된다.

방대한 양의 데이터를 효율적으로 다루기 위해서는 데이터를 압축하는 것이 일차과제이다. 그리고 압축된 데이터 중에서 필요한 부분만을 추출하여 원래 데이터로 복원하여 원하는 목적으로 활용하는 과정이 매우 효율적으로 이루어져야 한다. 또한 데이터 압축과 복원 과정에서 데이터에 포함된 정보 손실이 발생하지 않아야 한다.

본 연구에서는 이와 같은 요구조건을 만족하는 정수웨이블릿 변환을 이용한 수치표고모델의 압축 기법을 개발하고 압축된 데이터의 효율적인 활용법에 대해 고찰 하고자 한다.

정수웨이블릿변환을 이용한 자료의 압축과 처리

웨이블릿은 특정한 수학적 조건을 만족하면서 시간(또는 공간)적으로 유한한 범위 내에서만 값을 가지는 파형요소이다. 푸리에 변환의 기저함수인 사인곡선은 무한히 반복되는 부드럽고 예측 가능한 함수인 반면에 웨이블릿은 불규칙적이고 비대칭인 경향을 가진다. 푸리에 변환이 어떤 신호를 다양한 주파수를

Fig. 1. Tree representations of wavelet analysis (a) and wavelet packet analysis (b).

가지는 사인곡선들로 분해하는 것과 유사하게 웨이블릿분석은, 원래의 웨이블릿을 이동(shift)하고 폭을 변화(scaling)시켜 얻어지는 다양한 웨이블릿들로 신호를 분해한다.

웨이블릿변환은 웨이블릿이라 불리는 시간적 또는 공간적으로 유한한 폭을 가지는 기저함수들의 선형결합으로 함수나 데이터를 표현하는 방법이다. 유한한 폭을 가지는 웨이블릿의 특성으로 인하여 시간적 또는 공간적으로 집중되어 있는 형태의 데이터에 대해서는 푸리에변환 보다 훨씬 적은 개수의 유의미한 계수로 변환되는 특성을 가진다.

여기에서는 정수웨이블릿의 특징 설명과 이를 지형 자료 압축에 적용하였을 경우에 어떻게 효과적으로 압축되는지, 또한 원래의 지형자료로 다시 환원했을 때 얼마나 정확하게 환원되는지 등의 연구 결과를 보여줄 것이다.

데이터압축

데이터 압축은 데이터를 원래의 형태보다 더 적은 저장 공간에 기록하는 것을 말하며, 데이터를 작은 크기로 변화시키는 부호화(encoding) 과정과 압축된 데이터를 다시 원래의 데이터 형태로 복원하는 복호화(decoding) 과정으로 구성된다. 데이터 압축 기법들은 다시 압축과 복원 과정에서 데이터에 포함된 정보의 손실 발생 유무에 따라 무손실 압축과 손실압

축으로 구분된다. 손실압축은 무손실 압축보다 높은 압축률을 얻을 수 있으며, 특정 해상도 또는 주파수 이상의 자료는 구분하지 못하거나 듣지 못하는 인간의 감각 기관의 특성을 역이용하는 방법으로 음성, 정지화상, 동영상 등 데이터의 종류에 따라 각각 다른 압축 알고리즘을 사용한다. JPEG, MPEG, MP3 등이 많이 사용되는 손실 압축 규격들이다.

반면, 원래의 데이터를 정확히 복원해야 하는 텍스트나 의료영상 등에서는 무손실 압축 기법을 사용하여야 한다. 무손실 압축기법에도 다양한 알고리즘들이 존재하며 엔트로피 부호화(entropy encoding) 범주에 속하는 허프만부호화(Huffman coding), 산술부호화(arithmetic coding) 등이 많이 사용된다. 엔트로피 부호화는 어떤 값이 나올 확률에 따라 그 값을 나타내는 코드의 길이를 다르게 부호화하는 방법이다. 일반적으로 엔트로피 인코더는 모든 값에 대해 같은 길이를 가지는 코드 값이 나올 확률값을, 음의 로그에 비례하는 길이와 코드로 바꾸어 부호화함으로써 빈도수가 높은 값일수록 짧은 코드를 사용하여 데이터를 압축한다. 허프만부호화나 산술부호화는 다양한 형태의 자료 압축에 효과적으로 이용될 수 있으나 자료압축과정에서 입력데이터의 통계적 특성에 따라 코드 할당이 동적으로 변화하고, 최적 코드할당을 위해 트리구조 또는 범위계산을 하는 등 구조가 복잡하다.

웨이블릿 변환 개요

웨이블릿변환은 크게 연속웨이블릿변환 및 이산웨이블릿변환으로 구분된다. 연속웨이블릿변환은 계산 가능한 모든 스케일 및 위치에서 웨이블릿 계수를 구하게 되며 매우 많은 시간이 소요된다. 반면, 이산웨이블릿변환은 2의 멱수에 기반한 스케일 및 위치에서만 웨이블릿 계수를 계산하는 방법으로 정확도를 유지하면서 매우 효율적으로 변환을 수행하는 방법이다. 이산웨이블릿 변환은 Mallat(1989)에 의해 개발된 필터링 기법으로 구현된다. 어떤 신호는 고주파수통과필터와 저주파수통과필터를 거쳐 근사(approximations) 및 상세(details) 성분으로 분해되며 이러한 분해 과정을 반복 수행하여 다중분해능분석(multiresolution analysis) 또는 다중단계분석(multiple-level analysis)을 수행하게 된다.

이와 같이 여러 단계에 걸쳐 분석을 할 때 근사 성분만 필터링을 하여 다음 단계 근사와 상세 성분을 구하는 것이 통상적인 웨이블릿분석(wavelet analysis)이며, 근사 및 상세성분 각각을 필터링하여 각 각의 새로운 근사 및 상세 성분들을 구하는 방법을 웨이블릿패킷분석(wavelet packet analysis)이라 한다. Fig. 1은 웨이블릿분석 및 웨이블릿패킷분석의 트리 구조를 보여준다.

정수웨이블릿변환(Integer Wavelet Transform)과 데이터 압축

웨이블릿변환(wavelet transform)을 이용한 데이터 압축은 정지영상 및 동영상 등을 압축하는데 적절한 방법 중의 하나이며 손실 또는 무손실 압축 모두가 가능하다. 웨이블릿 압축은 부드럽고 주기적으로 변하는 데이터 보다는 급격히 변하는 데이터에 적용할 때 높은 효율을 보이게 된다.

웨이블릿 압축은 다음과 같은 과정으로 이루어진다. 먼저 압축할 데이터는 웨이블릿 변환을 하는데 웨이블릿 변환을 거치게 되면 원 데이터에 포함된 정보는 더 적은 개수의 중요계수(significant coefficient)들에 집중되게 되며 이러한 과정을 변환부호화(transform coding)라 한다. 이러한 웨이블릿변환의 고유한 특성에 의해 자료압축이 쉬워진다. 변환된 웨이블릿 계수들은 양자화(quantization)되고 양자화 된 값들에 대해 엔트로피 부호화 같은 압축알고리즘을 적용하게 된다. 일반적인 웨이블릿 변환은 부동 소수점 형태의

Fig. 2. Diagram illustrating one lifting step.

웨이블릿 계수를 계산하게 되는데 이러한 계수들을 양자화 하는 과정에서 정보의 손실이 발생한다. 입력 자료 자체가 부동소수점 형태인 경우는 양자화 과정의 손실은 피할 수 없는 부분이다.

입력 자료가 정수인 경우는 정수 웨이블릿 변환을 이용하여 웨이블릿 계수의 양자화 과정을 거치지 않음으로써 정보 손실이 없는 압축 및 복원이 가능해진다. 정지 영상, 수치표고모델(DEM) 등은 대표적인 정수로 이루어진 데이터이며 이러한 데이터에 정수 웨이블릿 변환을 적용하면 무손실 압축 및 복원이 가능해진다.

본 연구에서는 엔트로피 부호화의 기본 개념 위에 정수 웨이블릿 변환 및 수치표고모델의 데이터 특성에 맞는 간단한 부호 할당 및 저장, 그리고 압축 데이터 전체를 복원하지 않고 일부만을 복원하기 위한 데이터 저장 방식 등을 고안하였고 정밀 중력 지형 보정에서 그 활용성을 검토하였다.

정수웨이블릿 변환 웨이블릿은 기존의 웨이블릿으로부터 Sweldens(1998)에 의해 고안된 리프팅(lifting) 기법을 통해 구해진다. Fig. 2는 리프팅 기법의 개념을 보여주는 그림이다. 리프팅은 크게 스플릿(split), 예측(prediction) 및 갱신(update) 단계로 구성된다. 스플릿은 주어진 신호 X를 짝수 위치 데이터 X_e 와 홀수 위치 성분 X_o 로 나누는 과정이다. 예측 단계에서는 짝수위치 데이터 X_e 로부터 홀수 위치 성분을 예측하여 예측된 값과 원래 X_o 와의 차이(difference)를 구한다. 갱신 단계에서는 예측 단계에서 구한 홀수위치에서의 차이(difference)로부터 짝수위치 값을 예측하고 그 값들을 이용하여 X_e 를 갱신하여 최종적으로 근사 및 상세 성분을 구하게 된다.

정수웨이블릿 변환 전후 데이터의 통계적 특성 변화

Fig. 3은 SRTM 90 m 수치표고모델 자료 중 한반

Fig. 3. Image map of SRTM 90m DEM data. White square represents data range used in this study.

도 대부분을 포함하는 데이터를 이용한 것이며, 흰색 사각형 내의 고도자료를 이용하여 각 종 특성들을 분석하였다. 분석영역을 사각형영역으로 제한한 것은 SRTM 데이터에는 바다 부분의 수심 자료가 포함되어 있지 않아 바다영역을 포함하는 경우 분석 결과에 오류가 발생하기 때문에 제외하였다. Fig. 3의 사각형 내 자료의 크기는 $2,048 \times 2,048$ 이며 2바이트 정수형으로 기록할 경우 파일 크기는 약 8.39 MB이다.

Fig. 4는 Fig. 3의 사각형 영역 지형자료들의 통계적 특성을 보여준다. Fig. 4a는 고도 히스토그램으로 고도 1200 m 이하에 대부분의 자료가 분포하고, 고도 700 m 이하에 90% 정도의 자료가 분포한다.

Fig. 5는 Fig. 3의 사각형 영역의 지형 자료들로부터 변환된 웨이블릿 계수들의 통계적 특성을 보여준다. 변환에 사용된 웨이블릿은 CDF3.5이며 분해 단계는 5로 하였다. CDF (Cohen-Daubechies-Feauveau) 웨이블릿은 이중 직교(biorthogonal) 웨이블릿의 일종으로 CDF m,n 형태로 표기하는데 m 과 n 은 각 각 분해(decomposition) 및 복원(reconstruction)의 배니싱모멘트(vanishing moment)를 의미한다. 웨이블릿 변환된 웨이블릿 계수의 최소값은 -16,895, 최대값은 16,884로 나와서 매우 넓은 범위에 분포하는 것처럼 보이나, 출현 빈도수를 살펴보면 최소 1번 이상 나타나는 웨이블릿 계수는 1,649개이다. 이는 최대 11비트의 가변길이를 가지는 코드 1,649개를 사용하여 모든 계수를 부호화 하는 것이 가능하다는 것을 의미한다.

(a)

(b)

Fig. 4. Statistical characteristics of DEM data inside the square on Fig. 1, (a) histogram of elevations, (b) cumulative probability distribution of elevation.

Fig. 5에서 볼 수 있는 또 다른 중요한 특징은 앞에서 설명한 변환부호화(transform encoding) 현상에 따른 계수의 집중현상이다. 웨이블릿 계수들은 -10-10 범위에 83%의 값들이 포함되고, -40-40까지 범위를 확장하면 99%의 값들이 포함되는데, 이와 같이 원래 지형자료의 분포범위보다 훨씬 좁은 범위로 계수들이 압축되어 나타나는 특성을 이용하여 지형자료를 매우 효율적으로 압축하는 것이 가능하게 된다. 정수값을 가지는 웨이블릿 계수들 값이 -40-40 구간에 99%가 분포한다는 것은 길이 7비트 이하의 코드들을 이용하여 99%의 자료를 표현할 수 있다는 점을 의미한다.

정수웨이블릿 변환 데이터의 압축 기법

본 연구에서는 DEM 자료를 정수 웨이블릿 변환하

Fig. 5. Statistical characteristics of wavelet coefficients of DEM data inside the square on Fig. 1, (a) histogram of wavelet coefficients, (b) cumulative probability distribution of wavelet coefficients.

고 웨이블릿 계수들의 통계적 특성을 고려하는 엔트로피 부호화 개념에 기반하여 압축 기법을 고안하였다. 압축을 위하여 DEM 자료는 먼저 정수웨이블릿 변환을 통해 웨이블릿 계수들로 변환된다. 웨이블릿 계수들은 빈도수에 따라 정렬되며 부호는 0부터 시작하는 정수를 사용하였다. 즉 빈도수가 가장 높은 계수값은 부호 '0'으로 저장되고 두 번째로 빈도수가 많은 계수는 부호 '1'로 저장된다.

Fig. 3의 사각형 영역에 포함된 자료를 정수웨이블릿 변환한 결과 1번 이상 나타나는 웨이블릿 계수는 2,000개 미만이었다. 11비트 이진수가 표현할 수 있는 정수는 0-2,047까지 이므로 Fig. 3의 예에서는 최

대 11비트 길이로 데이터를 저장할 수 있게 된다.

본 연구에서 사용한 데이터 부호화를 좀 더 구체적으로 설명하면 다음과 같다. 한 개의 웨이블릿 계수는 부호의 길이를 나타내는 부분과 부호를 나타내는 부분으로 구성되어 저장된다. 어떤 계수가 부호 '10'으로 부호화 된다고 가정하면, 이 부호는 4비트로 표현되고 이 부호 비트 앞에 부호 비트의 길이를 저장하게 된다. 만약 부호의 길이를 나타내는 부분을 4비트로 한다면 0-15 비트까지를 표현하게 된다. 15비트 길이의 이진수로 표현 가능한 정수는 0-32,767이므로 최대 32,767개의 코드 길이까지 4비트로 표현할 수 있다. 부호의 길이를 나타내는 비트를 3비트로 할 경우는 0-7 비트까지 표현하게 되고 7비트 길이의 이진수는 최대 128개의 부호를 표현할 수 있다.

본 연구에서 사용한 압축파일의 전체 구조는 Table 1과 같다. 데이터 셋의 크기, 변환되는 영역의 좌상 및 우하 모서리의 좌표, 변환에 사용된 웨이블릿의 종류, 변환 레벨이 순서대로 기록된다. 그 다음에 빈도수가 1 이상인 계수들의 개수(NoCf)가 처음 16비트(2바이트)에 기록되며 그 다음에 NoCf 개의 계수들의 실제 값들이 2바이트 정수로 기록된다. 그 다음부터는 각 계수들의 부호가 기록되는데 앞에서 설명한 대로 부호를 표현하는 비트의 길이를 나타내는 4비트가 먼저 나오고 그 다음에 부호비트가 나오는 형태가 계수의 개수만큼 반복되며 기록되게 된다. 처음 4비트가 나타내는 십진수가 'a'라면 부호는 다음 'a'비트에 기록되고 이 'a'비트의 십진수가 'n'이라면 최종적으로 이 데이터는 (n+1)번째 계수에 매핑된다. 그리고 길이 4비트가 나타내는 십진수가 '0', '1'인 경우는 부호를 나타내는 비트를 따로 기록하지 않고 각 첫 번째, 두 번째 부호를 나타내는 것으로 해석하면 된다. 이와 같은 방식으로 기록할 때 최종적으로 각 각의 부호들을 기록하는데 필요한 데이터 길이는 Table 2와 같다.

정수웨이블릿변환을 이용한 지형자료 저장과 이를 이용한 처리

위에서는 정수웨이블릿을 이용하여 어떻게 자료를 압축하고 다시 복원시키는지 알아보았다. 여기에서는 좀 더 구체적으로 지구물리에서 매우 중요한 기초 자료인 지형자료를 어떻게 이에 적용하여 압축하고 처리하는데 사용하는지를 보여줄 것이다.

Table 1. Structure of compressed file

Location	length(bits)	Description
1-32	32	(NoX, NoY): No. of Data points, 2 bytes integer
33-96	64	(X, Y) of left-top point, 4 byte floating point number
97-160	64	(X, Y) of right-bottom point, 4 byte floating point number
161-176	16	WN: 2byte integer indicating kind of wavelet
177-192	16	Level: Decomposition level of wavelet transform
193-208	16	NoCf=No. of Coefficients
209-209+16*NoCf-1	16*NoCf	Coefficients
D0-D0+3	4	Length of code of 1 st coefficient, D0=17+16*NoCf. Decimal of (D0-D0+3) is length of code part, 'a'.
D0+4-D0+4+a-1	a	Binary representation of (n+1)-th code, where n is decimal of these bits
D1-D1+3	4	Length of code of 1 st coefficient, D1=D0+4+a. Decimal of (D1-D1+3) is length of code part, 'b'.
D1+4-D1+4+b-1	b	binary representation of (n+1)-th code, where n is decimal of these bits
...	...	...

Table 2. Record lengths and examples of binary representation of codes

Code	Length (bits)			Example
	Length	Code	Total	
0	4	0	4	0=0000
1	4	0	4	1=0001
2-3	4	2	6	2=001010 3=001011
4-7	4	3	7	4=0011100 5=0011101
8-15	4	4	8	8=01001000 15=01001111
16-31	4	5	9	...
32-63	4	6	10	...
64-127	4	7	11	...
128-255	4	8	12	...
256-511	4	9	13	...
512-1023	4	10	14	1023=1010111111111
1024-2047	4	11	15	1024=101110000000000

웨이블릿 및 분해 단계에 따른 데이터 압축률

DEM 자료를 웨이블릿 변환한 후 앞에서 설명한 압축방식으로 압축을 할 때 압축률은 웨이블릿 변환에 사용한 웨이블릿 및 분해 단계에 따라 달라진다. Fig. 6 및 Table 3은 다양한 웨이블릿 및 분해 단계

에 따른 압축률의 변화를 그래프와 표로 나타낸 것이다. 압축률은 DEM을 웨이블릿 계수로 압축할 때의 데이터량을 2바이트 정수 이진화일로 기록할 때의 데이터량으로 나눈 값이다. 웨이블릿 변환에 사용한 자료의 크기는 2,048×2,048이며 이를 아스키(ascii) 형식으로 저장하면 약 19 MB, 2바이트 정수형으로 이진 화일로 저장하면 8.39 MB의 용량을 차지한다. 그리고 이 자료를 웨이블릿 변환을 하지 않고 본 연구에서 제안한 방법으로 압축을 하게 되면 약 6.4 MB의 저장공간이 필요하고 이때 압축률은 약 75.9%이다.

Fig. 6 및 Table 3에서 보면 사용한 웨이블릿에 따라 압축률은 최대 8%까지 차이가 난다. 압축률이 제일 나쁜 웨이블릿은 Haar 웨이블릿이며 압축률이 가장 좋은 웨이블릿은 CDF3.5이다. CDF3.1도 CDF3.5와 비슷한 압축률을 보여준다. 분해 단계에 따른 압축률의 변화를 살펴보면, 분해단계에 따라 5-6%의 압축률의 차이가 나며 초기 단계에서 압축률이 급격하게 좋아지며 3단계 이후는 대부분의 웨이블릿에서 압축률의 변화가 매우 작다. 이러한 결과만으로 보면, CDF3.1 또는 CDF3.5 웨이블릿을 사용하여 3단계 정도로 분해를 하는 것이 최적의 선택이 되며 약 45.4%의 압축률을 얻게 된다. 웨이블릿 변환을 적용하지 않았을 때의 압축률과 비교하여 보면, 웨이블릿 변환을 통해 압축률이 30%정도 증가함을 알 수 있다.

Fig. 6. Variations of compression ratios of DEM data inside the square on Fig. 1 for various wavelets and decomposition levels of wavelet transform.

Table 3. Compression ratios (%) of DEM data inside the square on Fig. 6 for various wavelets and decomposition levels of wavelet transform

Wavelets													
Level	haar	db2	db3	cdf2.2	cdf2.6	cdf3.1	cdf3.5	cdf4.2	cdf4.6	cdf5.1	cdf5.5	cdf6.2	cdf6.6
1	58.7	55.3	54.9	52.4	52.5	50.9	50.9	50.4	50.2	51.0	50.6	51.4	48.9
2	54.2	50.3	49.7	47.5	47.7	46.3	46.3	46.5	46.2	48.0	47.1	49.8	48.4
3	53.3	49.2	48.5	46.5	46.8	45.4	45.4	45.9	45.5	48.2	46.9	50.4	48.8
4	53.1	48.9	48.3	46.3	46.6	45.3	45.3	45.8	45.4	48.3	47.0	50.5	48.9
5	53.0	48.9	48.2	46.2	46.5	45.3	45.2	45.9	45.4	48.3	47.0	50.6	48.9
6	53.0	48.9	48.2	46.2	46.5	45.3	45.2	45.9	45.4	48.3	47.0	50.6	49.0
7	53.0	48.9	48.2	46.2	46.5	45.3	45.2	45.9	45.4	48.3	47.0	50.6	49.0

효율적인 지형 복원 방법의 고안

앞에서 설명한 방법으로 압축되어 저장되어 있는 데이터로부터 효율적으로 지형을 복원하는 과정 또한 매우 중요한 과정이다. 방대한 양의 데이터를 모두 컴퓨터 메모리에 로드한 후 전체를 다 복원하는 것은 실제 작업에 필요하지 않는 부분까지 복원을 하게 되어 매우 비효율적이므로 특정 작업에 필요한 부분만을 선택적으로 복원하는 방법이 필요하다.

Fig. 7은 다중 단계 웨이블릿분석의 근사 및 상세 계수 데이터 행렬의 구조를 보여준다. 분해 단계가 증가하는 것은 스케일의 변화를 의미하며 각 계수들의 행렬 상에서 위치는 실제 데이터의 공간상 위치와 관계된다. 따라서 최종 분해단계 근사행렬의 한 개 계수는 이전 분해 단계들의 상세 계수들과 공간적으로 연관되어 있으며, 연관된 계수의 수는 단계가 한 단계씩 작아질수록 2배씩 늘어난다. 즉, 최종 n 단계의 근사 계수 하나는 $(n-1)$ 단계의 2×2 크기의 계수들과 연관되며, $(n-2)$ 단계의 4×4 계수들과 연관된

다. 그리고 최종 n 단계의 근사 계수들간의 공간적인 거리는 원래 데이터 간격의 배가 된다. 이와 같은 웨이블릿 변환 계수들의 특성을 이용하면 필요한 부분만을 추출하여 효율적으로 지형을 복원할 수 있게 된다. Fig. 7은 3단계까지 분해된 결과를 보여주며 A3의 회색으로 표시된 영역의 계수와 연관된 상세 계수들도 회색영역으로 표시되어 있다. A3의 1×1 의 계수는 실제 데이터의 8×8 영역에 해당하며 Fig. 7의 회색으로 표시되는 영역들만 추출하여 새로운 8×8 행렬을 구성하여 역변환을 실시하면 지형데이터가 복원된다.

이산웨이블릿 변환 및 역변환 과정은 유한한 길이를 가지는 고주파 및 저주파 통과 필터링에 해당하는데 앞에서 설명한 내용처럼 지형을 복원하게 되면 가장자리 부분에서 오차가 발생하게 된다. 이는 필터링 과정에서 복원하고자 하는 영역의 경계 부분 정보 일부가 복원에 사용되는 웨이블릿 계수 영역의 바깥에 위치하는 계수들에 포함되어 있기 때문이다.

Fig. 7. Structure of coefficients of multi-level wavelet analysis. Gray squares represent coefficients at each level those are related to certain spatial points.

Fig. 8은 부분적인 지형 복원의 정확도를 확인하기 위한 데이터 부분이다. 즉, Fig. 3의 흰색 사각형 영역의 웨이블릿 변환 계수들 중 일부를 이용하여 Fig. 8에 해당하는 영역을 복원하여 부정확하게 복원되는 영역을 살펴보았다.

Fig. 9는 CDF3.1, CDF3.3, 그리고 CDF3.5 웨이블릿을 사용하여 3단계까지 분해한 후, 부분 복원한 결과 및 복원된 지형과 원래의 지형의 차이를 나타낸 것이다. 앞에서 설명한 것처럼 복원되는 영역의 내부는 정확하게 복원되며, 가장자리 부분에서 오차가 발생하는 것을 볼 수 있다. 오차가 발생하는 영역의 크기는 웨이블릿의 종류에 따라 달라지는데, CDF3.1은 8, CDF3.3은 16, 그리고 CDF3.5는 24 정도의 폭에서 부정확하게 복원된다. A3 단계의 근사계수 1개는 실제 데이터 영역의 데이터 8개에 해당하므로 각 각 근사계수 1, 2, 3개에 해당하는 폭만큼 가장자리에서 부정확하게 복원되었다. 따라서 부분복원 시 실제 필요한 영역보다 조금 넓게 복원영역을 설정하여 복원을 하여야 하며, 복원영역을 얼마만큼 확장할 지는 사용되는 웨이블릿의 종류 및 분해단계에 따라 결정하면 된다.

Fig. 8. Example image for the test of the partial reconstruction from compressed DEM data.

정밀 중력 지형효과 계산을 위한 활용방법 고찰

이전의 중력탐사는 광역의 개략적인 지하구조 분석이 주목적이었으며 따라서 낮은 해상도의 지형자료를 이용한 지형보정 등으로도 충분히 그 목적을 달성할 수 있었다. 그러나 최근에는 석회암지대에 건설되는 각종 사회기반 시설물 하부에 석회공동에 의한 침하, 폐광산 지역 폐갱도에 의한 지반함몰 문제 등이 보고되고 있으며, 지하공동을 탐지를 위한 천부 고분해능 중력탐사의 필요성이 대두되고 있다.

천부의 소규모 이상체를 탐지하기 위해서는 중력탐사의 정밀도를 높여야 하는데, 이는 중력계의 정밀도 향상, 정확한 위치측정 기술 및 정밀한 지형효과 계산을 통해 이루어질 수 있다. 최신 중력계의 분해능은 1 수준으로 정밀해졌으며 정밀 DGPS를 이용해서 cm 단위의 오차로 현재 위치를 결정할 수 있다. 중력 측정점 주변 지형의 정확한 지형효과를 계산하기 위해서는 정밀한 DEM이 필요하다. 현재 전 국토에 대해 1 m 해상도의 DEM이 구축되고 있으며 향후 이를 이용한 정밀한 지형효과 계산이 가능해질 것으로 예상된다.

정밀 DEM의 자료량은 매우 크기 때문에 본 연구에서 개발된 효율적인 압축 및 복원기법을 통해 다

Fig. 9. Images reconstructed from part of wavelet coefficients those are spatially related to image in Fig. 8 according to the concept in Fig. 7.

루면 정보 저장과 처리 시간 등에 많은 도움을 줄 것이다. 또한 정밀 DEM 자료는 여러 가지 이유로 일반에 완전하게 공개되기는 어려울 것이기 때문에,

기존 연구들에서 개발된 정밀한 지형효과 계산 알고리즘 및 인터넷을 이용한 서버-클라이언트 통신모델을 결합한 시스템이 구축된다면 일반 사용자는 중력

측정점의 좌표만을 서버에 송신한 후 서버에서 계산된 정확한 지형의 중력효과를 전송받을 수 있을 것이다. 이러한 시스템의 구축을 통해 자원탐사, 광해방지, 토목구조물 지반조사 등 각 분야에서 정밀 중력탐사법의 활용성이 대폭 향상될 것으로 예상되며, 본 연구의 결과들은 이러한 시스템을 구축하는데 주요 기반기술의 일부가 될 것이다.

결 론

지형자료는 지구과학의 많은 분야에서 중요한 기초 자료 중 하나로 사용된다. 특히, 지구물리와 물리탐사 분야에서는 중력탐사자료 처리 시 반드시 필요하다. 최근 위성으로부터 얻어진 지형자료는 매우 정밀해서 여러 가지 세밀한 자료처리에 도움을 주고 있으나, 처리해야할 지형자료의 양이 방대해져서 이의 처리 시간과 저장용량의 문제를 야기하기도 한다. 본 연구에서는 이러한 지형자료를 쉽게 압축, 복원할 수 있는 효율적인 방법을 제시하였다.

정밀 수치표고모델을 활용하는데 어려운 문제 중 하나는 방대한 양의 데이터를 효율적으로 다루는 문제이다. 개인용 컴퓨터의 연산속도나 저장매체의 용량도 획기적으로 증가되어 왔으나 그에 비례하여 다루어야 되는 데이터의 양도 매우 빠르게 증가하였다. 따라서 방대한 양의 데이터를 효율적으로 다루는 수단은 여전히 실용적인 측면에서 매우 중요한 기술이 된다.

본 연구에서는 엔트로피 부호화의 기본 개념 위에 정수 웨이블릿 변환 및 수치표고모델의 데이터 특성에 맞는 간단한 부호 할당 및 저장, 그리고 압축 데이터 전체를 복원하지 않고 일부만을 복원하기 위한 데이터 저장 방식 등을 고안하였다.

DEM 자료를 웨이블릿 변환한 후 압축을 할 때 압축률은 웨이블릿 변환에 사용한 웨이블릿 및 분해 단계에 따라 달라지며, 압축율은 DEM을 웨이블릿 계수로 압축할 때의 데이터량을 2바이트 정수 이진 화일로 기록할 때의 데이터량으로 나눈 값이다. 웨이블릿 변환에 사용한 자료의 크기는 $2,048 \times 2,048$ 이며 이를 아스키(ascii) 형식으로 저장하면 약 19 MB, 2 바이트 정수형으로 이진 화일로 저장하면 8.39 MB의 용량을 차지하였다. 그리고 이 자료를 웨이블릿 변환을 하지 않고 본 연구에서 제안한 방법으로 압축을 하게 되면 약 6.4 MB의 저장공간이 필요하고 이때

압축률은 약 75.9%이었다.

압축률이 가장 좋은 웨이블릿은 CDF3.5이며, CDF3.1도 CDF 3.5와 비슷한 압축률을 보여준다. 이 결과는 CDF3.1 또는 CDF3.5 웨이블릿을 사용하여 3단계 정도로 분해를 하는 것이 최적의 선택이 되며 약 45.4%의 압축률을 얻게 됨을 보여준다. 웨이블릿 변환을 적용하지 않았을 때의 압축률과 비교하여 보면, 웨이블릿 변환을 통해 압축률이 30%정도 증가함을 알 수 있었다.

웨이블릿 변환의 다중레벨 구조의 특성을 활용하여 압축된 자료의 일부분만을 추출하여 지형의 일부 영역을 복원하는 것이 가능하며, 복원영역의 가장자리 부분은 부정확하게 복원되므로 필요한 부분보다 조금 넓게 복원영역을 설정하여야 한다.

사 사

이 논문은 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 2010-0007031). 본 논문을 심사해주신 심사자분들께 감사의 뜻을 전합니다.

참고문헌

- Aiken, C.L. and Cogbill, A.H., 1998, Recent developments in digital gravity data acquisition on land. The Leading Edge, 17, 93-97.
- Butler, D.K., 1984, Microgravimetric and gravity gradient techniques for detection of subsurface cavities. Geophysics, 49, 1094-1096.
- Hammer, S., 1939, Terrain Correction for Gravimeter Stations. Geophysics, 4, 181-194.
- Hardy, R.L., 1971, Multiquadric Equations of Topography and Other irregular Surfaces. Journal of Geophysical Research, 76, 1905-1915.
- Krohn, D.H., 1976, Gravity Terrain Corrections using Multiquadric Equations. Geophysics, 41, 266-275.
- Kwon, B.D., Kwon, J.W., and Lee, H., 1990, Terrain correction of gravity data by using a multiquadric equation. Journal of the Korean Earth Science Society, 11, 156-164. (in Korean)
- Lee, H. and Lim. H., 2010, Precise Gravity Terrain Correction of Gravity Exploration for small anomalous bodies, Journal of the Korean Earth Science Society, 31, 1-7. (in Korean)
- Lim, H., Lee, H., Park, Y.S., Lim, M., and Jung, H.K., 2010, Gravimetric terrain correction using triangular

element method. Jigu-Mulli-wa-Mulli-Tamsa, 13, 169-174. (in Korean)

Mallat, S., 1989, A theory for multiresolution signal decomposition: The wavelet representation. Pattern

Analysis and Machine Intelligence, 11, 674-693.

Sweldens, W., 1998, The lifting scheme: A construction of second generation of wavelets. SIAM Journal on Mathematical Analysis, 29, 511-546.

2013년 1월 17일 접수
2013년 1월 25일 수정원고 접수
2013년 2월 1일 채택