Landsat-8을 이용한 자동화된 구름 제거 영상 생성

Fully Automated Generation of Cloud-free Imagery Using Landsat-8

김병희¹⁾ · 김용현²⁾ · 한유경³⁾ · 최원석⁴⁾ · 김용일⁵⁾

Kim, Byeong Hee · Kim Yong Hyun · Han, You Kyung · Choi, Won Seok · Kim Yong Il

Abstract

Landsat is one of the popular satellites for observing land surface that is used in various areas including monitoring, detecting and classifying changes in land surface. However, shades, which cloud itself and its shadow, interrupted often clear observation and analysis of ground surface. For this reason, the process of removing shades and restoring original ground surfaces are critical for geospatial users. This study is planned to recommend a methodology for more accurate and clear images of Landsat-8 sensor, which provided two additional bands of costal/aerosol and cirrus. In fact, those bands are known as functioned effectively in detecting and restoring shades. Otsu's thresholding technique to detect clouds, we replaced those detective shades by using experimental and reference images. In accurate assessment, the overall accuracy and kappa coefficients were about 85% and 0.7128, respectively. This indicates that the proposed technique is effective for recovering the original land surface.

Keywords : Cloud-free image, Landsat-8, Otsu thresholding, Land-use and land-cover map

초 록

Landsat은 대표적인 지구관측 위성 중 하나로 지표면 모니터링, 변화탐지, 분류 등 다양한 분야에서 사용되고 있다. 하지만 구름과 구름의 그림자는 지표의 관측과 분석을 제한하는 장애물 중 하나로, Landsat을 사용하기 전 구름을 제거하고 원래의 지표 피복으로 복원하는 과정은 필수적이다. 최근에 발사된 Landsat-8은 기존위성에 비해 2개의 추가적인 costal/aerosol, cirrus 밴드를 제공하며, 이는 구름을 탐지하고 복원하는데 효율적으로 사용될 수 있다. 따라서 본 연구에서는 Landsat-8의 영상에서 구름을 효과적으로 탐지하고, 복원하는 기법을 단계적으로 제안하였다. Otsu 임계화 기법을 통하여 구름과 구름의 그림자 지역을 탐지하였고, 탐지된 구름 및 그림자 지역은 실험 영상과 참조영상을 이용하여 원래의 지표 피복으로 복원 하였다. 복원영상의 정확도 평가에서는 전체정확도가 약 85%, 카파계수가 0.7128로 본 연구에서 제안한 알고리즘이 효율적임을 확인하였다.

핵심어 : 구름 제거 영상, Landsat-8, Otsu 임계화, 토지이용 및 토지피복도

Received 2014. 03. 12, Revised 2014. 04. 07, Accepted 2014. 04. 29

1) Department of Civil and Environmental Engineering, Seoul National University (E-mail: kbh1no@snu.ac.kr)

2) Member, Department of Civil and Environmental Engineering, Seoul National University (E-mail: kimyh@snu.ac.kr)

3) Member, Department of Civil and Environmental Engineering, Seoul National University (E-mail: han602@snu.ac.kr)

4) Department of Civil and Environmental Engineering, Seoul National University (E-mail: youn0603@kyonggi.ac.kr)

5) Corresponding Author, Member, Department of Civil and Environmental Engineering, Seoul National University (E-mail: yik@snu.ac.kr)

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.

1. 서론

Landsat은 1978년 발사 이래 지표면 모니터링 분야에서 다양하게 활용되고 있는 지구관측위성으로, 현재 많이 발사되고 있는 고해상도 위성영상에 비해 공간 해상도는 떨어지지만 넓은 촬영 폭(swath), 높은 분광해상도(spectral resolution), 16일의 재방문(revisit) 주기를 가지고 있다. 이러한 특성을 바탕으로 Landsat은 지표면의 변화탐지, 토지피복 분류, 농업 생산량 예측, 식생 모니터링 등의 다양한 지표면 관측 수단으로 사용되고 있다(Vogelmann *et al.*, 2001; Park, 2010). 하지만 Landsat 영상은 지형보정만 된 Level-1의 상태로 무료로 제공하기 때문에 영상에서 구름과 구름의 그림자는 지표를 관측하고 분석하는데 어려움을 주는 장애물 중 하나이다. 특히 한반도의 경우 청천일이 일년에 100일 내외로, 맑은 날의 영상을 취득하는데 제한적이다. 따라서 Landsat 자료를 다양한 원격탐사 분야에서 활용하기 위해서는 구름을 탐지하고 제거하여, 원래의 지표 피복으로 복원하는 전처리 과정이 필수적이다(Zhu *et al.*, 2012).

Landsat 데이터에서 구름 탐지 및 복원 연구와 2007년 Landsat ETM+의 SLC(Scan Line Corrector)-off 로 인해 영상이 촬영되지 않는 지역에 대한 복원 연구가 많이 진행되어 왔지만 각각의 연구들은 장단점을 가지고 있다. Li *et al.*(2002)은 구름, 그림자, 식생, 시가화지역에 대한 임계값을 계산하여 이 임계값과 가까이 있는 화소(pixel)는 해당 클래스로 분류하는 화소 랭킹(pixel-ranking)을 통해 구름 및 그림자를 탐지하였다. Ma *et al.*(2010)은 Landsat 영상의 각 밴드별 임계값과 의사결정나무(decision tree)를 통해서 구름 및 구름에 의한 그림자를 탐지하였으며, 탐지된 구름 및 그림자 지역은 여러 장의 Landsat영상을 통해 최신의 정보를 업데이트 하는 방식으로 복원하였다. 위에 언급한 두 연구들은 결과적으로는 우수한 복원영상을 생성하였으나, 임계값이 영상마다 수동적이고 주관적으로 결정되기 때문에 영상에 적절한 임계값을 찾는 데 시간이 걸린다는 단점이 있다. 이러한 단점을 해결하기 위해 Roy *et al.*(2010)과 Shiu *et al.*(2011)은 자동으로 임계값을 정해 구름을 추출 할 수 있는 알고리즘을 개발하였으나 구름의 그림자 지역은 연구범위에 포함하지 않았다는 한계점이 있다. 구름 및 그림자 지역을 원래의 지표면으로 복원하기 위한 연구 역시 다양한 방식으로 진행이 되었다. 다수의 연구가 구름이 존재하는 영상과 청천 영상의 방사보정, 히스토그램 매칭(histogram matching), 회귀분석(linear regression) 등을 통해 구름을 마스크(masking)한 부분에 청천 영상을 모자이크(mosaic)하여 영상을 복원하였

다(Helmer and Ruefenacht, 2005; Li *et al.*, 2002; Ma *et al.*, 2010). Zhu *et al.*(2012)은 Landsat ETM+에서 SLC-off 영상을 복원하는데 사용하였던 영상 복원 방식을 변형시켜 구름이 있는 영상을 복원하였으며, Jin *et al.*(2013)은 청천 영상에서 영상 좌표와 비슷한 밝기값을 갖는 밝기값들을 그룹화 하여 영상을 복원하였다.

한편, 2013년 2월에 발사된 Landsat-8은 기존 Landsat 센서들의 지구관측 임무를 이어가기 위해 기존 Landsat과 동일한 파장대의 정보를 수집할 뿐만 아니라 새롭게 추가된 2개의 밴드로 인해 추가적인 정보를 제공하고 있다. 새로 추가된 밴드 중 딥 블루 채널(deep blue visible channel, band 1)은 해안연구 및 에어로졸 연구에 적합하며, 다른 하나는 새로운 파장대의 적외선 밴드(band 9)로 구름탐지에 적합하도록 정보를 수집한다. 이 밴드는 1.360~1.390 μ m의 파장대로 수분과 수증기에 의해 강력하게 흡수되는 파장대를 포함하고 있다. 대부분의 수분 및 수증기는 낮은 고도의 대기에서 존재하기 때문에(Hutchison and Choe, 2007), 구름의 상층부에서는 새로운 적외선 밴드에서의 파장대의 태양 에너지가 구름 입자에 의해 반사되거나 산란되어 위성에서 센서에 입력이 되고, 이로 인해 상대적으로 높은 방사도 값을 갖는다. 반면에, 구름이 없는 지역은 태양 에너지가 지표면으로 이동하면서, 그리고 지표면에서 반사되어 센서에 입력되면서 대부분의 태양에너지가 수분에 의해 흡수되어 낮은 방사도를 나타내게 된다.

이에 본 연구에서는 새롭게 추가된 파장대의 정보를 활용하여 구름이 존재하는 지역의 영상과 동일한 지역의 맑은 날 촬영된 영상만을 이용하여 구름이 없는 맑은 날의 영상으로 자동으로 복구하는 알고리즘을 제안하였다. 구름이 존재하는 지역의 영상에 대해 최초 자동 임계화 기법을 통해 구름 및 그림자 지역을 탐지하여 마스크를 하였다. 이후 마스크된 지역을 복구하기 위해 구름이 존재하는 영상과 맑은 날의 영상에서의 좌표와 밝기값 정보만을 이용하여 영상을 복원하였다. 또한 이에 대한 정확도 평가를 통해 본 연구에서 제안하는 알고리즘이 인접한 시기에 촬영된 구름이 있는 영상과 맑은 날의 영상만을 가지고 구름과 구름의 그림자를 제거하고 원래의 지표 피복으로 영상을 복원하는데 효율적임을 확인하였다.

2. 연구 지역 및 사용 데이터

Landsat-8은 기존의 Landsat 위성들에 비해 2개의 추가된 밴드를 제공하는 동시에 QA(Quality Assessment) 밴드

를 추가적으로 제공하고 있다(Table 1). QA 밴드는 영상이 취득될 당시의 지표, 대기, 센서의 상태 등의 정보를 포함하고 있다. QA 밴드는 구름과 관련된 정보를 16비트의 4단계 ‘Yes, No, Maybe, Not determined’로 제공하지만, 동시에 구름 외의 눈, 식생, 수계 등의 정보도 제공하기 때문에 정확한 구름정보를 추출하는데 제한적이다. 또한 구름은 구름이외에 구름에서 비롯한 그림자로 인한 문제를 야기하는데, QA 밴드에서는 이를 누락하고 있다. 따라서 본 연구에서는 이러한 QA 밴드의 단점을 극복하기 위해 자동으로 구름 및 그림자 지역을 추출하는 동시에 이 지역을 복원하기 위한 연구를 수행하였다.

Table 1. Band properties of Landsat-8

Bands	Wavelength (micrometers)	Resolution (meters)
Band 1 – Coastal aerosol	0.43 – 0.45	30
Band 2 - Blue	0.45 – 0.51	30
Band 3 - Green	0.53 – 0.59	30
Band 4 – Red	0.64 – 0.67	30
Band 5 – Near Infrared	0.85 – 0.88	30
Band 6 – Short Wave Infrared 1	1.57 – 1.65	30
Band 7 – Short Wave Infrared 2	2.11 – 2.29	30
Band 8 - Panchromatic	0.50 – 0.68	15
Band 9 - Cirrus	1.36 – 1.38	30
Band 10 – Thermal Infrared 1	10.60 – 11.19	100
Band 11 – Thermal Infrared 2	11.50 – 12.51	100

본 연구에서는 30m의 공간해상도를 가진 Landsat-8 다중영상을 사용하였고, 영상은 한반도 지역 내에서 구름이 많은 여름과 상대적으로 구름이 적은 가을시기인 2013년 8,9,10월에 취득된 영상으로 영상 내 구름이 존재하는 실험영상(experimental image)과 영상에 구름이 존재하지 않는 맑은 날에 촬영된 참조영상(reference image)으로 3개의 데이터셋을 구성하였다(Table 2, Fig. 1). 실험영상과 참조영상 간 촬영 간격은 모두 16일로 최소화하여 영상 간 계절 및 지형의 변화를 최소화 하였다. Landsat-8 영상은 12 비트(bit)로 영상을 촬영하고 16비트로 사용자에게 제공을 하고 있어 각 영상의 화소 범위(range)의 차이가 매우 큰 문제를 발생하게 되므로 본 연구에서는 영상을 8비트로 선형 변화하여 실험을 수행하였다.

Table 2. Landsat-8 data sets for generation of cloud free image

Path/Row	Area	Experiment Image	Cloud Cover	Reference Image
116/ 34	Seoul, Gyeonggi-do	2013. 08. 31	33%	2013. 9. 16.
114/ 35	Gyeongsang-do	2013. 09. 02	16%	2013. 9. 18.
115/ 35	Chungcheong-do, Jeolla-do	2013. 10. 11.	9%	2013. 10. 27.

Fig. 1. Experimental (left) and reference (right) images; displayed as bands 5, 4, and 3

3. 연구 방법 및 적용결과

기존의 Landsat을 통한 구름 탐지 연구는 대부분 구름을 탐지하여 복원하기 위해 블루밴드, 근적외선 밴드, 열적외선 밴드 등 다양한 밴드의 속성 정보를 통해서 구름 추출을 위한 임계값을 결정하기 때문에 알고리즘이 복잡하다는 단점이 있다. 이를 보완하기 위해 상대적으로 적은 수의 밴드를 통해 밴드별 밝기값의 임계값을 통해 구름 추출 및 복원 연구가 진행되어 왔으나 이는 영상마다 사전 관찰을 통하여 영상에 맞는 임계값을 설정해야 해서 자동화된 알고리즘이라

고 말하기에는 한계가 있다.

반면 Landsat-8은 기존의 Landsat 위성들에서 제공하던 동일한 밴드들 이외에 앞서 언급한 1번 밴드, 9번 밴드를 추가적으로 이용할 경우 구름 탐지에 용이하기 때문에 이를 활용한 구름탐지 연구는 기존 구름탐지 연구들에 사용된 알고리즘보다 간편한 알고리즘을 구성할 수 있는 장점이 있다. 본 연구에서 사용된 구름탐지 및 복원 알고리즘은 크게 세 단계로 구성된다(Fig. 2). 첫 단계에서는 구름 지역과 그림자에 의한 그림자 지역을 Otsu 임계값 결정방식을 통해 구름과 그림자 지역으로 판단되는 곳을 마스크하였다(Otsu, 1975). 이 후 마스크 된 지역을 제외한 지형에 대해 히스토그램 매칭을 수행한 후 Jin *et al.*(2013)이 제안한 SSG(Spectral Similarity Group) 복원 방식을 변형시켜 구름 지역을 복원 하였다. 마지막으로 복원된 영상과 참조영상 간 정확도 평가를 수행하여 구름이 포함된 영상의 복원정도를 판단하였다.

Fig. 2. Flow chart of generation cloud-free image

3.1 구름 및 그림자 지역 마스크

구름 지역의 마스크 과정은 두꺼운 구름의 마스크 과정과 얇은 구름의 마스크 과정으로 나누어진다. 마스크 과정에서 새롭게 추가된 1번 밴드, 9번 밴드, 그리고 열적외선의 10번 밴드를 사용하였으며, 자동 임계화 방법인 Otsu 임계값 기법을 사용하였다. 여기서 두꺼운 구름이란 구름층으로 인해 구름아래의 지형이 보이지 않는 경우로, 9번 밴드의 파장대의 경우 태양에너지의 대부분이 반사, 산란되어 위성에 기록되기 때문에 상대적으로 9번 밴드에서 높은 값을 갖게 된다. 반면 10번 밴드(열적외선)에서 구름 지역은 상대적으로 지표면 보다 온도가 낮기 때문에 구름 지역에서의 밝기값은 지표면에서의 밝기값보다 상대적으로 작은 값을 갖게 된다.

얇은 구름의 경우 두꺼운 구름과 달리 구름이 존재해도 지표면이 어느 정도 보이는 구름으로 연무(haze)형태의 구름을 의미한다. 이 경우, 구름이 존재하더라도 태양 에너지가 구름을 투과하기 때문에 빛이 지표면에서 다시 반사되어 위성 센서로 도달하면서 대기의 수분에 의한 흡수로 인해 9번 밴드에서는 낮은 값을 갖게 된다. 또한 10번 밴드의 경우 구름에 의해 반사된 값으로 인해 지표면보다 상대적으로 낮은 밝기값을 갖게 되며, 블루 계열이 좀 더 강조된 1번 밴드의 경우 구름이 존재하기 때문에 지표에 비해 상대적으로 높은 밝기값을 갖게 된다. 두꺼운 구름과 얇은 구름을 추출하기 위한 각 밴드에 대한 임계값 조건은 Table 3과 같이 나타내었고 여기서 'threshold B1, threshold B9, threshold B10'은 각 밴드별 밝기값에 대한 임계값을 의미한다.

그림자의 경우 위치적으로 구름의 근처 화소에 위치하며 구름과 그림자의 거리가 100 픽셀 이내에 존재하는 기존 연구(Jin *et al.*(2013))와 이를 실험적으로 확인하여 100 픽셀 버퍼 존을 생성하였다. 또한 그림자만의 특성을 찾기 위해 각 밴드별로 실험영상의 밝기값을 조사한 결과 그림자 지역이 적외선 파장대에서 매우 낮은 값을 갖는 것을 확인하였다. 따라서 그림자를 추출하여 마스크하기 위해서 기존의 Otsu 임계화 기법과 동일한 원리면서 임계값을 여러개 정할 수 있는 Otsu N 임계화 기법을 통해 적외선 영역에 해당하는 5번, 6

Table 3. Thick and thin cloud mask threshold conditions

Type of cloud mask	Threshold condition
Thick cloud	Band 9 > threshold B9 & Band 10 < threshold B10
Thin cloud	Band 1 > threshold B1 & Band 9 < threshold B9 & Band 10 < threshold B10

번 밴드에 대해서 각각 9개의 임계값을 선정하였고 10개의 클래스 중 가장 낮은 클래스에 해당하는 값을 그림자 지역으로 마스킹하였다(Deepa and Bharathi, 2013).

Fig. 3은 구름(빨간색)과 구름에 의한 그림자(노란색)를 마스킹한 결과로 대부분의 구름지역이 추출되어 마스킹 된 것을 확인할 수 있다. 반면 그림자의 경우 구름과 달리 그림자 지역 이외의 지역들이 마스킹 된 것을 확인할 수 있다. 그림자를 제외한 대부분은 수계지역으로 그림자와 수계지역은 모든 밴드에서 스펙트럼 정보가 유사하기 때문에 그림자와 함께 추출이 되었다고 판단 할 수 있다. 본 연구의 최종 목적은 구름이 없는 영상과 동일한 상태로 영상을 복원하는 것으로 수행오차(commission error)를 최소화하는 것이 좋지만, 추가적인 정보를 활용하여 수행오차를 최소화 할 시 그림자 지역에서 누락오차(omission error)가 증가하는 경향이 있다. 따라서 수계지역 등에서 다소 과대 추출된 부분이 있더라도 모든 그림자 지역을 포함하여 영상을 구름이 없는 영상으로 생성하는 것이 타당하다고 판단하였다.

P116 R34

P114 R35

P115 R35

Fig. 3. Experimental images (left) and cloud and shadow masks (right); red: cloud, yellow: shadow

3.2 구름 및 그림자 지역 복원

영상은 영상마다 방사적 특성이 다르기 때문에 밝기값 차이가 발생한다. 구름이 존재하는 영상의 경우 구름이 매우 높은 복사휘도(radiance)를 갖기 때문에 상대적으로 구름 외 지역은 구름지역에 비해 낮은 복사휘도를 가지게 된다. 따라서 구름이 포함된 영상은 구름을 포함하지 않는 맑은 날의 영상에 비해서 동일지역에서 더 낮은 밝기값을 갖게 된다. 이러한 밝기값 차이를 무시하고 맑은 날의 참조영상을 이용해 실험 영상을 복원할 경우 영상에서 색상 왜곡(color distortion)이 발생하기 쉽기 때문에 두 영상 간 밝기값의 차이를 최소화 하기 위한 기술은 필수적이라 할 수 있다. 본 연구에서는 실험영상과 참조영상을 통해 3.2절에서 마스킹한 구름 및 그림자 지역을 제외하고 나머지 지형에 대해서 참조영상을 기준으로 실험영상을 선형 히스토그램 매칭을 하여 실험영상의 밝기값을 참조영상에 맞도록 변화시켰다.

추출한 구름 및 그림자 마스킹 지역의 복원을 위하여 Jin et al.(2013)이 제안한 SSG 기법을 변형하여 적용시켰다. 기존에 진행된 대부분의 구름복원 알고리즘은 마스킹된 지역을 복원 시 단순히 참조영상을 모자이크(mosaic)하였다. 하지만 촬영주기가 16일인 Landsat은 계절변화 시점에서 실험영상과 참조영상에서 차이가 발생하기 때문에 이 시기에 대해서는 자연스러운 복원영상을 구현하는데 한계가 있다. 반면 SSG 기법은 참조영상을 그대로 이용하는 것이 아니라 참조영상의 좌표만을 이용해 실험영상 내 밝기값을 통해 마스킹 지역을 복원하는 데 장점이 있다. SSG 기법은 Fig. 4와 같이 크게 세 가지 단계로 구성 된다. 첫 째로 구름 및 그림자 지역에서 마스킹 된 좌표와 동일한 위치의 밝기값을 참조영상으로부터 구한다(Fig. 4 (a)의 붉은색). 이후 참조영상에서 구해진 좌표의 위치를 실험영상으로 이동시킨 후(Fig. 4 (b)의 노란색), 그 좌표의 밝기값과 비슷한 밝기값을 갖는 좌표는 동일한 클래스라고 가정하여 실험영상 내에서 비슷한 밝기

Fig. 4. SSG cloud filling method; (a) experimental image, (b) reference image

값을 갖는 좌표값을 구해준다(Fig. 4 (b)의 파란색). 마지막으로 실험영상의 좌표들 중 마스킹 된 지역과 겹치는 좌표를 제외하고 나머지 좌표를 다시 참조영상으로 이동 시킨 후(Fig. 4 (a)의 녹색) 참조영상의 밝기값의 평균값을 계산해 최초 마스킹된 좌표에 입력하여 최종 밝기값을 구한다.

하지만 SSG 기법은 마스킹된 좌표를 제외한 모든 좌표를 대상으로 하여 평균값을 계산하기 때문에 실험영상과 참조

영상 사이에 미세한 변화에 의한 오차를 포함하는 단점이 있다. 따라서 본 연구에서는 SSG 기법으로 마스킹된 좌표들의 평균 밝기값을 계산하되 마지막에 모든 좌표들의 밝기값의 평균값을 대입하는 것 대신 평균값을 기준으로 양쪽으로 표준편차 내에 있는 좌표들의 밝기값만 대상으로 하여 평균값을 다시 계산하여 마스킹된 영상을 복원시켰다.

Fig. 5와 Fig. 6의 가장 우측의 영상은 좌측의 실험영상으

Fig. 5. Experimental (left), reference (middle), and cloud and shadow filled (right) images; displayed as bands 5, 4, and 3

Fig. 6. Experimental (left), reference (middle), and cloud and shadow filled (right) images; displayed as bands 4, 3, and 2

로부터 가운데 참조영상을 이용해 영상을 복원한 것을 나타낸다. 전체적으로 일부 구름이 남아있지만, 대부분의 구름과 구름의 그림자가 효과적으로 제거되고 복원된 것을 확인할 수 있었다. 시각적으로도 결과에서 수계, 도심지, 초지로 대략적으로 구분이 가능할 정도로 복원이 되었으며, 지형적으로도 많은 오차가 발생하지 않은 것을 확인하였다. Fig. 5와 Fig. 6의 Path 116, R 34(서울, 경기지역)의 경우 복원영상에서 오른쪽에 미세하게 구름이 남아 있는 것을 확인할 수 있었는데, 이는 대부분이 연무와 같은 얇은 구름이었으며 이러한 얇은 구름 지역은 구름 아래로 영상에서 지표 피복이 많이 관측되고, 구름에 의한 효과가 적기 때문에 자동적으

로 임계값을 추출할 때 구름지역에 해당되지 않은 것으로 판단된다. 또한 구름 마스크 영역 복원 시 참조영상의 좌표만을 이용하여 평균과 거리가 먼 영상의 밝기값은 영상 복원에서 제거하였기 때문에 Fig. 6의 Path 115, Row 35(충청도, 전라도 지역)와 같이 영상이 계절적 변화에 영향을 받지 않고 복원된 것을 확인하였다. Fig. 6의 Path 115, Row 35의 경우 해당지역이 한국의 중남부 지방으로 참조영상의 경우 10월 27일, 실험영상의 경우 10월 11일에 촬영된 것으로 실험영상과 참조영상이 촬영된 시기동안 산에 단풍이 많이 들었음에도 불구하고 복원된 영상에서는 단풍을 들지 않게 영상이 복원이 되었다. 이는 참조영상에서는 좌표만 이용하고 이를

실험영상의 밝기값을 통해 복원을 했기 때문이라고 할 수 있다. 또한 공통적으로 도심지가 밀집되어 있고 그 주변을 둘러 초지나 산림으로 둘러싸인 경우 산림에서 색상 왜곡이 발생하여 주변 산림보다 파란색 밴드에서 좀 더 밝은 값을 갖는 것을 볼 수 있다. 이는 참조영상에서 왜곡이 발생한 산림 지역에서의 밝기값이 다른 산림 지역의 밝기값보다 크기 때문에 일부 시가지 지역의 밝기값이 같이 계산된 원인이라고 판단된다.

4. 정확도 평가

구름은 시간에 따라 지속적으로 위치가 변하기 때문에 정확도 평가를 위한 명확한 참조자료의 구성이 제한된다. 통상적으로 정확도 평가를 위해서 대략적으로라도 구름지역을 디지털타이징하여 참조지도를 구성해 이를 통한 정확도 평가 방식을 통해 구름을 얼마나 제거했는지 수치적으로는 파악할 수 있지만 참조영상과 복원영상이 어느 정도의 일치도를 갖는지에 대한 판단은 힘들다는 단점이 있다. 이러한 단점을 극복하기 위해 참조영상과 복원영상의 감독분류를 통한 정확도 평가를 수행할 수 있다. Shiu *et al.*(2011)은 효과적으로 복원된 영상은 분류 후에도 참조영상과 같은 클래스를 가져야 한다는 가정 하에 감독분류를 통하여 참조영상과 구름의 복원영상을 시각적으로 비교하여 복원한 영상의 복원 정도를 판단하였다. 따라서 본 연구에서는 시각적인 비교 외에도 정확한 정확도판단을 위해 참조영상을 참조데이터로 하여 영상의 분류를 수행하여 정확도를 수치적으로 분석하였다. 우선 수계(Water), 시가지화지역(Built-up area), 초지(Forest), 나대지(Bare Soil)의 4개에 클래스에 대해서 훈련자료를 수집하였다. 훈련자료의 경우 보통 위성영상의 분류를 주된 목적으로 한 경우의 연구는 위의 4개의 클래스로만 훈련지역을 할당하기에는 한계가 있다. 본 연구는 영상을 세세하게 분류하여 다양한 토지피복 클래스를 분류하는 것이 아닌 구름이 어느 정도 복원되었는지 확인하는 것이기 때문에 훈련자료를 세분화 하지 않고 기본적인 지표특성을 대표할 수 있는 것으로 한정을 지었다. 이렇게 기본적인 클래스에 대해서만 훈련자료를 수집할 시 다양한 클래스의 훈련자료를 수집하면서 발생할 수 있는 분류 오차를 감소시킬 수 있다는 장점이 있다. 수집한 훈련자료를 바탕으로 MLC(Maximum Likelihood Classification)기법을 통해 참조영상을 감독분류 수행하였고, 동일한 훈련자료를 사용하여 복원영상 역시 MLC기법으로 감독분류를 수행하여 참조영상의 분류결과를 참조자료로 하여 복원영상의 정확도를 픽셀 단위로 평가

하였다. 정확도 평가에 사용한 영상은 실험영상들 중에서 도심지와 수계, 산림 등 자연피복을 모두 포함하고 있는 Path 116, Raw 34(서울, 경기지역) 영상을 사용하였다. Fig. 7은 참조영상과 복원영상을 분류한 분류결과를 나타내는 것으로 시각적 평가에서 대부분의 복원된 지역의 화소들이 동일한 클래스로 할당된 것을 확인하였고, 참조영상의 분류결과를 참조자료로 하여 복원영상의 분류결과와 오차행렬을 Table 4와 같이 구성하였으며 각 클래스별 분류 정확도 및 전체 정확도, 카파계수(kappa coefficient)를 Table 5와 같이 나타냈다. 나대지를 제외한 지역에서 분류결과는 70%이상의 높은 정확도를 보였으며 전체정확도가 85.25%를 나타내고 있다. 특히 사용자 정확도의 경우 참조영상을 기준으로 복원한 영상의 분류 정확도를 나타내는 것으로 구름의 복원영상이 어느 정도 정확도를 가지는지 나타낼 직접적인 지표가 될 수 있다. 또한 전체정확도는 전체 픽셀 수에 대해 바르게 분류된 픽셀 수를 의미하는 것으로 85%이상의 높은 정확도로 보아 제안한 복원 기법이 효과적임을 나타내고 있다.

카파계수 역시 분류정확도를 나타내는 지표 중 하나로 일반적으로 0.8이상일 경우 일치도가 강함을 의미하며, 0.4이하일 경우 일치도가 낮다고 판단하는 기준이 된다(Jensen, 2005). 본 연구에서는 카파계수가 0.7128의 값을 나타내 제안한 복원 기법이 강한 일치도는 아니지만 효과적임을 나타내고 있다.

초지의 경우 시가지화지역으로 대부분 누락 오차가 발생하였는데 이는 산림지역에서 일부 구름이 제거되지 못한 지역이 시가지로 분류되었기 때문에 발생하는 오차라고 판단이 된다. 또한 시가지지역과 초지의 경우 서로 누락오차와 수행 오차가 많이 발생하였는데 이는 영상에서 시가지지역과 초지지역이 많은 부분 경계를 이루고 있다는 점에서 오차의 원인을 파악할 수 있다. 시가지지역과 초지 지역이 많은 부분 경계를 이루고 있기 때문에 경계 지점에서는 Landsat의 낮은 해상도로 인한 초지와 시가지지역이 서로 혼합 화소(mixed pixel)로 존재할 가능성이 높고 이점이 시가지지역과 초지가 서로 오차가 발생한 가장 큰 원인이라고 할 수 있다. 나대지의 경우도 대부분이 시가지지역에 속해있기 때문에 지역의 경계 부분에서 혼합 화소로 존재할 가능성이 높고 이러한 원인으로 인해 오차가 많이 발생했을 것으로 판단된다. 이러한 이유로 인해 나대지의 정확도가 낮게 나왔지만 전체 화소에서 매우 적은 비율을 차지하고 있기 때문에 전체정확도에는 크게 영향을 주지 못한다.

Fig. 7. Classification results of path 116, row 34;
left: reference image, right: cloud and shadow filled image

Table 4. Error matrix of path 116, row 34

Unit(pixel)		Reference Image				
		Forest	Water	Built-up	Bare Soil	Total
Cloud-free Image	Forest	1,665,069	290	268,929	36	1,934,324
	Water	423	69,170	27,286	647	97,526
	Built-up area	61,430	5,000	737,093	60,363	913,832
	Bare Soil	29	2,677	15,532	26,080	44,318
	Total	1,726,951	77,137	1,108,786	87,126	3,000,000

Table 5. Results of accuracy test

	Prod. Acc.(%)	User Acc.(%)	Overall Acc.(%)	Kappa Coefficient
Forest	96.42	86.08	85.25	0.7128
Water	89.94	70.92		
Built-up area	71.88	86.28		
Bare Soil	29.93	58.85		

5. 결론

대표적인 지구관측 위성 Landsat은 변화탐지, 식생관측, 토지피복 분류와 같은 분야에서 다양하게 활용되었고, Landsat-8 또한 기존의 Landsat 위성 군(constellation) 임무 연속에 선상에 있다. 하지만 구름은 이러한 Landsat을 활용을 제한하는 가장 큰 요소 중 하나로 본 연구에서는 구름 및 구름의 그림자를 제거하기 위해서 Landsat-8의 1번, 9번, 10번 밴드의 Otus 임계값을 통하여 자동으로 구름을 탐지하였고, 구름의 그림자 지역은 5번, 6번 밴드와 Otsu N 임계화 기법을 통해 탐지하였다. 이후 탐지된 구름 지역을 제거하고 변형된 SSG 기법을 이용하여 영상을 복원한 결과 약 85%의 정

확도로 본 연구에서 제안한 알고리즘이 구름탐지 및 복원에 효과적임을 확인하였다. 본 연구는 추가자료 및 추가적인 주관적 판단 없이 인접한 시기에 촬영된 청천영상만 존재한다면 구름이 있는 지역에 대해서 자동적으로 구름 및 그림자지역을 복원할 수 있다는데 그 의미가 있다. 또한 기존에 구름을 복원하기 위해 사용되던 모자이크 방식과는 달리 본 연구에서 제안한 알고리즘은 계절적 변화에 구애받지 않고 영상을 복원할 수 있다는 장점이 있다.

본 연구는 구름탐지 및 복원에 적합한 알고리즘이 효율적이고 적합하다는 결론을 도출할 수 있었지만, 일부 지역에서 색상 왜곡이 발생하며 얇은 구름이 일부 복원되지 못한 문제점이 발생하였고 이를 향후 연구를 통해 해결하

고자 한다. 이와 더불어 그림자의 경우 수계와 밝기값 특성이 비슷해 그림자 지역 추출 시 수계지역도 추출이 되는 문제점을 확인하였다. 이는 향후 태양의 고도각과 위성의 촬영 고도각을 이용하여 그림자의 위치를 예측하는 방법을 통해 해결하고자 한다. 또한 본 연구에서 제시한 알고리즘이 효율적임을 확인하였기 때문에 향후 연구에서는 각기 다른 위치에 구름이 존재하는 두 영상에 대해서도 구름을 탐지하고 복원하는 연구를 진행할 예정이다.

감사의 글

이 논문은 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행되었습니다(No. 2012R1A2A2A01045157).

References

- Deepa, S. and Bharathi, V. S. (2013), Efficient ROI segmentation of digital mammogram images using Otsu's N thresholding method, *Indian Journal of Automation and Artificial Intelligence*, Vol. 2, No. 1, pp. 51-56.
- Helmer, E.H. and Ruefenacht, B. (2005), Cloud-free satellite image mosaics with regression trees and histogram matching, *Photogrammetric Engineering & Remote Sensing*, Vol. 71, No. 9, pp. 1079-1089.
- Hutchison, K. D. and Choe, N. J. (1996), Application of 1-38 μ m imagery for thin cirrus detection in daytime imagery collected over land surfaces, *International Journal of Remote Sensing*, Vol. 17, No. 17, pp. 3325-3342.
- Jensen, J. R. (2005), *Introductory Digital Image Processing: A Remote Sensing Perspective 3rd Edition*, Pearson Prentice Hall, Upper Saddle River, NJ.
- Jin, S., Homer, C., Yang, L., Xian, G., Fry, J., Danielson, P., and Townsed, P. A. (2013), Cloud-free satellite image mosaics with regression trees and histogram matching, *International Journal of Remote Sensing*, Vol. 34, No. 5, pp. 1540-1560.
- Li, M., Chin, S., and Kwoh, L. K. (2002), Generating "Cloud Free" and "Cloud-Shadow Free" mosaic for SPOT panchromatic images, *2002 IEEE International Geoscience and Remote Sensing Symposium*, IEEE GRSS, 24-28 June, Toronto, Canada, pp. 2480-2482.
- Ma, Y., Jiao, W., Wang, D., Long, T., and Wang, W. (2010), An integrated method to generate a cloud-free image automatically based on Landsat5 data, *2010 International Conference on Intelligent Computation Technology and Automation*, IEEE Computational Intelligence Society, 11-12 May, Changsha, China, pp. 453-456.
- Otsu, N. (1975), A threshold selection method from gray-level histograms, *IEEE Transactions on System, Man and Cybernetics*, Vol. 9, No. 1, pp. 62-66.
- Park, J. K. (2010), Change Detection of Vegetation Using Landsat Image, *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, Vol. 28, No. 2, pp. 239-246. (in Korean with English abstract)
- Roy, D. P., Ju, J., Kline, K., Scaramuzza, P. L., Kovalsky, V., Hansen, M., Loveland, T. R., Vermote, E., and Zhang, C. (2010), Web-Enabled Landsat Data (WELD): Landsat ETM+ composited mosaics of the conterminous United States, *Remote Sensing of Environment*, Vol. 114, No.1, pp. 35-49.
- Shiu, Y., Lin, M., and Chu, T. (2011), Mapping and recovering cloud-contaminated area in multispectral satellite imagery with visible and near-infrared bands, *2011 IEEE International Geoscience and Remote Sensing Symposium*, IEEE GRSS, 24-29 July, Vancouver, Canada, pp. 543-546.
- Vogelmann, J. E., Howard, S. M., Yang, L., Larson, C. R., Wylie, B. K., and Van Driel, N. (2001), Completion of the 1990s national land cover data set for the conterminous United States from Landsat TM data and ancillary data sources, *Photogrammetric Engineering & Remote Sensing*, Vol. 67, No. 6, pp. 650-62.
- Zhu, X., Gao, F., Liu, D., and Chen, J. (2012), A modified neighborhood similar pixel interpolator approach for removing thick clouds in Landsat images, *IEEE Geoscience and Remote sensing*, Vol. 9, No. 3, pp. 521-525.