Cloud and shadow removal from Landsat TM data

RI Pyongsop[1, 2], MA Zhangbao[1], QI Qingwen[1], LIU Gaohuan[1]

1. *Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing* 100101, *China*;
2. *Institute of Remote Sensing and Geo-informatics, Pyongyang, Democratic People's Republic of Korea*

Abstract: Cloud removal is an important step in remote sensing image process. In this paper, the author proposed a new algorithm for cloud removal using multi-temporal Landsat TM image data based on spectral characteristics analysis. Through the spectral characteristics analysis of the thick cloud region and its shadow region, the thick cloud and its shadow identification models were designed. Using image regression, unsupervised classification and pixel replacing techniques as well as these models, the influence of thick clouds and its shadows can be eliminated or reduced in the Landsat TM images. The result shows that the algorithm can eliminate or significantly reduce the cloud influence from Landsat TM image data.

Key words: Landsat TM, image data, cloud and shadow, spectral analysis, cloud removal

CLC number: TP751.1 **Document code:** A

Citation format: RI Pyongsop, Ma Z B, Qi Q W and Liu G H. 2010. Cloud and shadow removal from LANDSAT TM data. *Journal of Remote Sensing*. **14**(3): 534—545

1 INTRODUCTION

The earth observing satellite Landsat TM/ETM+ remote sensing image data, have been widely used as the main data source for the study of spatial/ temporal land use/cover change due to its enhanced spectral characteristics, short data acquisition cycle, wide survey field, data usability and other properties (Li *et al.*, 1997). It has been used as an ideal remote sensing image data source for the research of regional-scale natural resource and environment.

However, due to climate reasons, it is difficult to obtain completely cloud free remote sensing image data. Most of the remote sensing image data include, more or less, clouds and their shadows projected on the ground. These give some trouble to a number of users of remote sensing image data. It becomes often the most important issue how to remove the influence of clouds from the remote sensing image data (Song *et al*., 2006). So, cloud removal is an essential step in the image pre-processing process (Song *et al*., 2003).

A great number of work has been carried out on the cloud detection and removal research such as dynamic filtering method (Zhao, 1996; Wu, 2003), multi-spectral synthesis method, light temperature value difference method, the index method (Song *et al*., 2003), cloud processing algorithms based on remote sensing image classification results and the cloud detection results (Song *et al*., 2006), image fusion method based on neural network and wavelet transform (Tapasmini *et al*., 2008), cloud detection method based on the texture analysis and neural network (Song *et al*., 2004).

The conventional cloud processing algorithms differ depending on the cloud status.

Dynamic filtering method is suitable for the case that there exists relatively wide range of cloud in the image. Dynamic filtering method, combining the frequency filtering and gray value change, separates cloud and background features, and finally removes the cloud influence from the remote sensing image. Because this approach relates to the frequency filters and needs the choice of cutoff frequency, the useful information sometimes is lost in the filtering process. Moreover this approach can not be used for the thick cloud.

For the local cloud distribution regions, the time average method is generally used. This algorithm can only be used in the region in which the surface feature change along the time is very small. For the dense vegetation cover region, due to vegetation growth is closely related with time, vegetation indices of different time are significantly different from each other. Therefore, such a simple substitution algorithm can not be used in this case.

To solve the above problem, in this paper, the author suggests a new cloud removal method that uses the Landsat TM image data of the same region at different time. It uses the TM remote sensing image data of the same period, or nearly the same season in different years. Based on each band's relative change of spectral characteristics, an enhancement model of thick cloud and its shadow is designed. In conjunction with these models and conventional unsupervised automatic classification method, image matching technique using linear regression analysis and pixel replacing operation, the cloud influence

Received: 2009-06-12; **Accepted**: 2009-09-25

Foundation: Study on Methodologies in Geography (No. 2007FY140800).

First author biography: RI Pyongsop (1970—), male, doctor student. He graduated in 1993 from Kim Chaek University of Technology, Democratic People's Republic of Korea. His research interest is remotely sensed imagery. E-mail: pyongsop@gmail.com

can be eliminated or reduced from the Landsat TM remote sensing image data. The result shows that the algorithm can eliminate or reduce the cloud influence from the Landsat TM image data

2 STUDY AREA AND DATA SOURCE

2.1 Brief description of the study area

The study area is located in the central and western region of Korean Peninsula ranging from 125° 00 'E to 126° 10'E and from 38° 15 'N to 39° 30'N. The distance of a straight line between eastern and western end is about 105.84km, the distance of a straight line between north and south end is about 137.46km. The study area is about 14000km^2. The region contains a variety of terrains including mountains, plains, sea, and so on. It contains a variety of land use/cover types including woodland, grassland, paddy fields, dry fields, salt, urban and industrial land, bare soil, reservoirs, lakes, canals, and tideland and so on. The spatial distribution structure of the land use/cover is very complex and the block is relatively small in the region. This kind of region is helpful to the researchers who study the effect of the cloud influence removal from Landsat TM image data under the different terrain conditions and different land use/cover types.

2.2 Data source

Two Landsat TM images of the study area were taken in August of 2006 and 2007 in Landsat 117-33 orbit. These Landsat TM image data were obtained through the network. The remote sensing image data used in this paper were shown in Table 1.

Table 1 List of used data

No.	Satellite	Date	Sensor
1	Landsat 5	2006-08-19	TM
2	Landsat 5	2007-08-22	TM

Among the two remote sensing image data, the one in August 2007 includes relatively large area of locally distributed thick cloud and its shadow regions. Another one in August 2006 includes small area of cloud and its shadow regions. The main purpose of this study is to remove the thick cloud and its shadow. So we use the remote sensing image data in August 2007 as the main data, the remote sensing image data in august 2006 as the auxiliary data. The color composite image of the Landsat TM remote sensing image data in the study area is shown in Fig. 1.

Fig. 1 The color image of Landsat TM (red-band 5, green-band 4, blue-band 3) on August 22, 2007

3 METHOD

3.1 Data pre-processing

The data pre-processing procedure is shown in Fig 2. First, using AutoSync module in ERDAS IMAGINE 9.2, image registration was carried out between the two Landsat TM image data (Dang *et al.*, 2007).

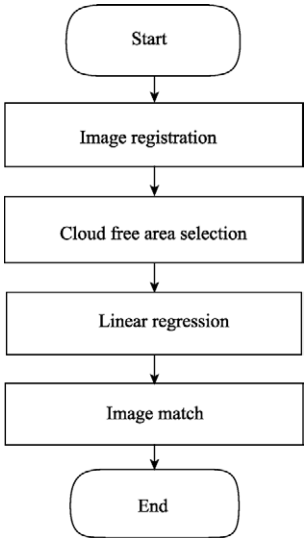


Fig. 2 Data preprocessing diagram

Then for the two Landsat TM image data, image match was performed between each corresponding bands. First, the naked eye visual interpretation method is used to select the cloud free region ROI in the two Landsat TM image data. In these regions, for each corresponding bands of the two Landsat TM image data, linear regression analysis was performed. Using linear regression model, each band of the auxiliary data was matched to the corresponding band of the main data.

The linear regression model is described as follow.

$$Y_i = a_i X_i + b_i \quad (1)$$

where X_i: the i band's gray value of the auxiliary data; Y_i: the i band's transformed gray value; i: the band number.

Table 2 shows that the correlation between each corresponding bands of the two Landsat TM image data is very high. In particular, the pairs of band 4 and band 5 have the maximum value respectively.

Table 2 Linear regression results

Band	*a*	*b*	*R*
1	0.86	7.75	0.90
2	0.86	3.53	0.87
3	0.82	3.51	0.87
4	0.92	0.52	0.96
5	0.94	0.02	0.96
7	0.86	1.70	0.93

3.2 Cloud and shadow enhancement model

3.2.1 Spectral characteristics of the thick cloud region

To extract the thick cloud and its shadow region, the comparison analysis of the spectral characteristics between cloud free region and cloud region was performed in this paper. Fig. 3 is the 5, 4, 3 bands color composite image of the Landsat TM image data on August 22, 2007 (a) and spatial profile curve (b). In Fig. 3(a), the straight line lies across the region which includes vegetation, water, thick cloud and its shadows.

In the color composite image, the white region is the thick cloud region and the black part beside it is its shadow region.

From Fig. 3, the spectral characteristics of cloud and its shadow region are as follows:

(1) In all bands, the spectral reflectance value in thick cloud region is significantly higher than cloud free region (The difference is over 150).

(2) In water bodies and cloud shadow regions, the spectral reflectance value of band 4, 5 and 7 is significantly reduced.

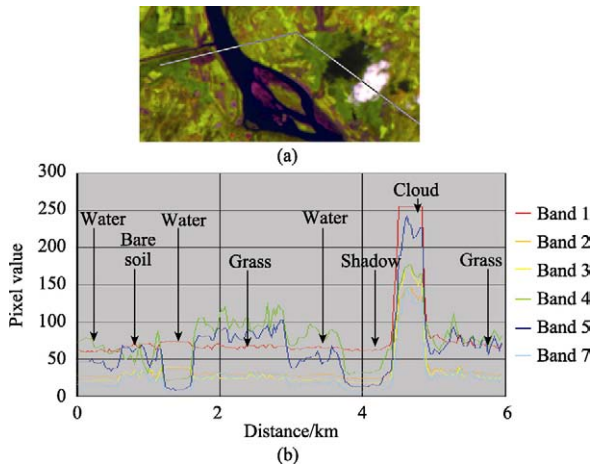


Fig. 3 Spectral characteristics of cloud region

(a) Color image of band 5, 4, 3; (b) Profile curves of the interest region

(3) In cloud shadow regions, the spectral reflectance value of band 1, 2 and 3 is reduced a little. But in water bodies, the spectral reflectance value of band 1, 2 and 3 is increased.

Although the variation amount of the spectral characteristics is not the same with the difference of cloud thickness, but such variation trend caused by thick cloud and its shadow is the same.

3.2.2 Cloud enhancement model

Based on the analysis result above, we proposed the cloud and its shadow region enhancement model.

The change content of the spectral characteristics between two TM image data includes the change part caused by cloud and the change part caused by land use/cover change. So, in order to extract cloud and its shadow region, first of all, it is necessary to distinguish between the change part of the spectral characteristics caused by cloud and the one caused by land use/cover change.

Firstly, the thick cloud enhancement model is designed in this paper.

From Fig. 3, the differences between the change of the spectral characteristics caused by thick cloud and the one caused by the land use/cover change are as follows:

First: the change of the spectral characteristics caused by thick cloud is very great. The variation of the spectral reflectance values of each band is more than 150.

Second: the change trend of the spectral characteristics of each band is the same, which is increased.

Based on the above two features, thick cloud enhancement model is designed as follow.

$$\mathrm{CAEM} = \mathrm{MD} \times \mathrm{CDF} \quad (2)$$

$$\mathrm{MD} = \frac{\sum_{i=1}^{n} \left| B_{\mathrm{Refi}} - B_{\mathrm{Auxi}} \right|}{n} \quad (3)$$

$$\mathrm{CDF} = \mathrm{Sign}\left[\sum_{i=1}^{n} \mathrm{Sign}(B_{\mathrm{Refi}} - B_{\mathrm{Auxi}}) - n + 1 \right] \quad (4)$$

where B_{Refi}: the i band's gray value of the main data; B_{Auxi}: the i band's gray value of the auxiliary data

In this paper, the Landsat TM image data on August 22, 2007 was used as main data and the one at August 19, 2006 which was transformed by linear regression model was used as auxiliary data.

CAEM: Cloud area enhancement model; MD: Mean absolute difference; CDF: Cloud discriminate function; n: the number of bands.

If the gray value of all bands of the main TM data is greater than the one of the auxiliary data, CDF = 1; otherwise, CDF$\leqslant$0.

In the region in which there is a great change of the spectral characteristics, MD value is high, but in the region in which there is no change of the spectral characteristics, MD value is low.

Among the region in which MD value is high, if CDF = 1, this region is just the cloud region.

In order to reduce the influence of the variation of sun's position, different atmospheric conditions and a number of other

reasons, the auxiliary data was matched to the main TM image data. For each corresponding band of the two TM image data, the linear regression analysis was performed in the cloud-free region. Using this linear regression model, the auxiliary TM image data was transformed.

Fig. 4 is the cloud region enhancement result calculated by the above model. Fig. 4 shows that the difference between thick cloud region and other region is very clear. Using a simple threshold, we can distinguish thick cloud regions and other regions; moreover the range of the usable thresholds is relatively big.

Fig. 4 Enhancement result of cloud regions

In this paper, two threshold values of 30 and 50 were used to compare the two result images and analyze the influence of selection of different threshold value. The analysis result shows that the change between two result images is less than 0.05% of the entire image area.

Cloud shadow extraction method is more complex than the thick cloud extraction method. Because the change of the spectral characteristics caused by cloud shadow is very little, sometimes it is confused with the change caused by land use/cover change. So this paper suggests the combined method of cloud shadow enhancement model and conventional unsupervised classification method to extract cloud shadows.

From Fig. 3, the differences between the changes of the spectral characteristics caused by cloud shadows and the other changes in the cloud free region are as follows:

First: the reflectance values of the band 4, 5 and 7 are significantly reduced.

Second: In cloud shadow regions, the reflectance value of the band 1 is reduced, but in water regions, the reflectance value of the band 1, 2 and 3 is increased.

The spectral characteristics of water regions and cloud shadow regions in Landsat TM image data are similar. So, in order to extract cloud shadow regions, firstly water region must be extracted. Based on the above change features of the spectral characteristics, the cloud shadow enhancement model is designed.

$$\mathrm{SAEM} = \mathrm{SDF}\times[(B_{\mathrm{Aux5}}-B_{\mathrm{Ref5}})+(B_{\mathrm{Aux7}}-B_{\mathrm{Ref7}})]/2 \quad (5)$$

$$\mathrm{SDF}=\mathrm{Sign}\left(1-\mathrm{Sign}\left(\sum_{i=1}^{3}\mathrm{Sign}(B_{\mathrm{Refi}}-B_{\mathrm{Auxi}})-2\right)\right) \quad (6)$$

where $B_{\mathrm{Ref}\,i}$: the i band's gray value of the main data; B_{Auxi}: the i band's gray value of the auxiliary data; SAEM: Shadow area enhancement model; SDF: Shadow discriminate function; Sign(): the sign function which extracts the sign of a real number.

If the values of the band 1, 2 and 3 of the main TM data are higher than the one of the auxiliary data, SDF = 0 (it is water region); otherwise, SDF$\geqslant$1 (it is not water region).

Among the region in which the reflectance values of the band 5 and band 7 of the main TM data are reduced, if it is not water region (SDF$\geqslant$1), SAEM value would be high.

The cloud shadow enhancement image calculated using the above model also contains the spectral characteristics change region caused by land use/cover change. In Fig. 5, the pink regions are thick cloud regions and the green regions are likely cloud shadow regions.

In order to distinguish between cloud shadow regions and the spectral characteristics change regions caused by land use/cover change, this paper used unsupervised automatic classification method.

For the result image calculated using the SAEM model, the threshold operation was performed and as a result, the likely

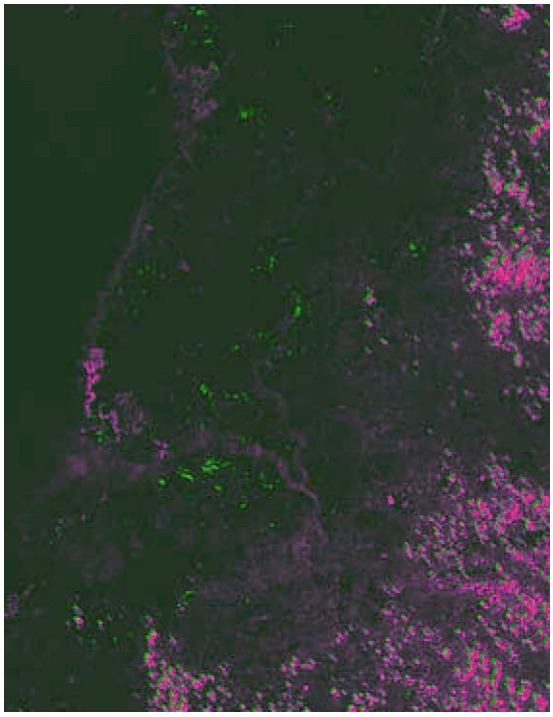


Fig. 5 Enhancement result of cloud and shadow region in August 22, 2007

cloud shadow regions were extracted. For these result regions, unsupervised classification was performed using Landsat TM original image data. Because each band's spectral characteristics in the cloud shadow regions are different from the one of vegetation, bare soil and other types of land use/cover category, the cloud shadow regions can be separated from the other land use/cover change regions by using conventional unsupervised classification method.

For different cloud thickness, different land use/cover types and different atmospheric conditions, although the extent of such change in spectral characteristics may not be the same, but such change tendency always exists and in the images obtained under better atmospheric conditions, such tendency is more pronounced. Based on this principle, it is possible to obtain cloud free or cloud influence minimized Landsat TM image data using two or more Landsat TM image data of the same period or the same season in different years.

3.3 Automatic cloud removal process

The overall process of cloud influence removal or reduction is shown in Fig. 6.

For two Landsat TM image data of the same period or the same season in different years, the image registration and image matching operation is performed. In the cloud free regions, for each pair of the corresponding bands of two Landsat TM image data (one is a main data and another one is an auxiliary data), linear regression analysis is performed and as a result, the new auxiliary TM image data matched with main data is obtained. It can reduce the influence caused by different sun position, different atmospheric conditions and some other conditions.

Using the above cloud enhancement model (CAEM) and cloud shadow enhancement model (SAEM), using unsuper-vised classification method and using the Modeler module, the programming Model Maker, in ERDAS IMAGINE 9.2, the cloud and its shadow region enhancement image can be calculated from the original TM image data

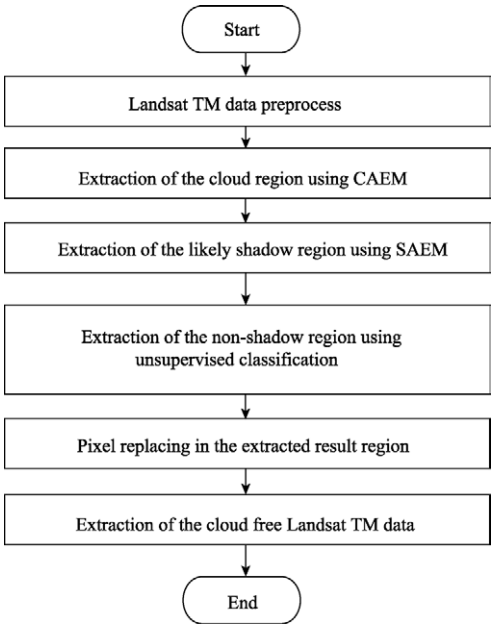
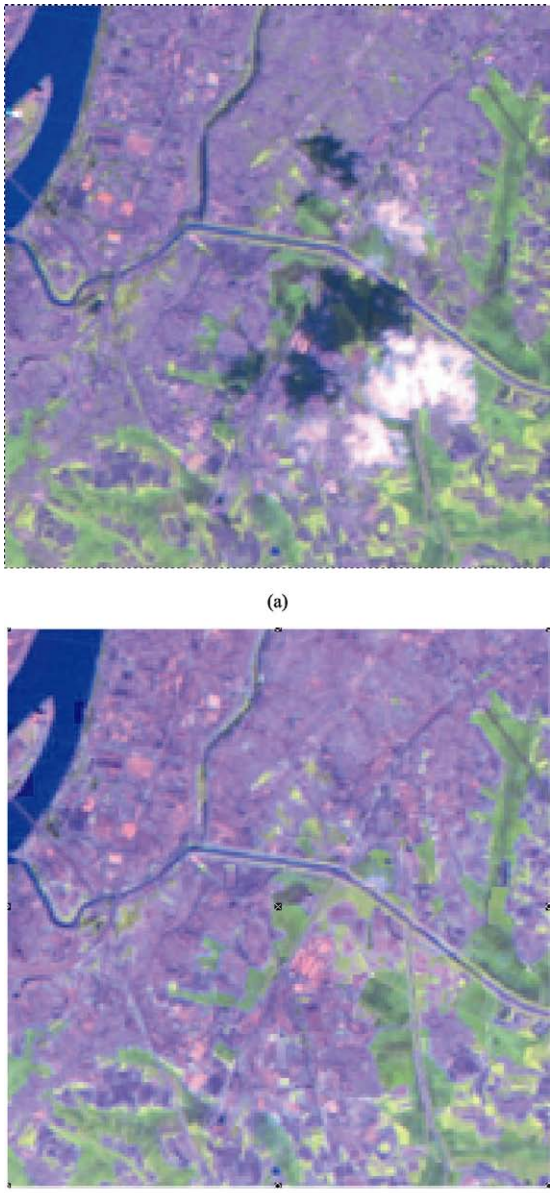


Fig. 6 Diagram of automatic cloud removal

And then using the newly calculated auxiliary TM image data which is matched with the main TM image data and cloud and its shadow region enhancement image, pixel replacing process is performed in the cloud influence regions. As a result, the new TM remote sensing image data in which the cloud influence is removed or reduced is obtained. This process is shown in Fig. 6.

4 RESULTS AND ANALYSIS

Using the cloud influence enhancement model CAEM and SAEM described in part 3.2, through the cloud removal process described in part 3.3, the cloud free Landsat TM image data was obtained from the August 22, 2007 Landsat TM image data. Fig. 7 shows the band 5, 4 and 3 color composite image of the result image data. The result shows that cloud and its shadow was removed. Fig.8 is the entire reswlt of the study area.



(a)

(b)

Fig. 7 Cloud free TM image

(a) Cloud and shadow region; (b) Cloud free image

The content of the non-shadow region among the shadow enhancement image calculated using SAEM model is as follows: (1) paddy regions, specifically, in the 2006 year TM data (auxiliary data), rice was already planted, but in the 2007 year TM data (main data) rice was not planted yet in paddy fields. It is similar to a paddy in water region. Such effect is rather good for the automatic classification of paddy fields because its nature is not water, just paddy; (2) tideland region and water bodies; depending on the water elevation change, some regions may be enhanced as shadow. Through unsupervised classification, such regions can be distinguished with shadow regions.

Fig. 8 Entire cloud free TM image

5 CONCLUTIONS

Using two Landsat TM image data on August 22, 2007 and August 19, 2006, the study on the removal method of the cloud influence was performed in the central and western region of the Korean Peninsula. The result shows that: The above method can effectively remove or weaken the cloud influence and the process is relatively simple. Moreover, in the process, it does not need any strict threshold value adjustment. It needs only the spectral characteristics of the Landsat TM image data. It does not need any other information. So, for another regions, another times Landsat TM image data, the above-mentioned process can be automatically applied.

Although this study used only two Landsat TM image data to remove cloud influence, but in order to obtain better result, it can be used more than two Landsat TM image data.

Above-mentioned operation is performed in each pixel as a unit, the requirement of an image registration is very strict and in order to obtain better result, it needs high-precision topographic correction, atmospheric correction and image matching operation.

REFERENCES

Dang A R, Wang X D, Chen X F and Zhang J B. 2007. ERDAS IMAGINE remote sensing image processing methods. Beijing: Qinghua University Press

Li X and Ye J A. 1997. The accuracy improvement of the land use change monitoring based on remote sensing using principal component analysis. *Journal of Remote Sensing*, **1**(4): 282—288

Song X N and Zhao Y S. 2003. Cloud detection and analysis in MODIS images. *Chinese Journal of Image and Graphics*, **8**(9): 1079—1083

Song X Y, Liu L Y, Li C J, Wang J H and Zhao C J. 2006. Study on the cloud removal based on the single remote sensing image data. *Optical Technology*, **32**(2): 299—303

Song X N, Liu Z H and Zhao Y S. 2004. Cloud detection and analysis of MODIS image. IEEE Geoscience and Remote Sensing Symposium

Tapasmini Sahoo and Suprava Patnaik. 2008. Cloud removal from satellite images using auto associative neural network and stationary wavelet transform. *IEEE Computer Society*, **6**: 100—105

Wu W L. 2003. The cloud elimination in remote sensing images. *Railway Aerial Survey*, (1): 6—7

Zhao Z M and Chu Z G.1996. The cloud removal method in remote-sensing images. *Environmental Remote Sensing*, **11**(3): 195—199

Landsat TM 遥感影像中厚云和阴影去除

李炳燮[1,2], 马张宝[1], 齐清文[1], 刘高焕[1]

1. 中国科学院地理科学与资源研究所, 北京 100101;
2. 朝鲜遥感与地理信息系统研究所 朝鲜 平壤

摘 要: 提出了一种新的利用多时相 Landsat TM 影像数据进行的厚云及其阴影去除的方法。该方法通过分析厚云及其阴影的光谱特征, 设计了厚云和云阴影识别模型。该算法的实现是采用图像配准技术、非监督分类、像元替换等运算, 计算出厚云和云阴影区域的 TM 影像替换数据, 进而得到消除或者减少云影响的 TM 遥感影像。试验结果表明本文提出的厚云及其阴影去除方法效果很好, 能消除或者弱化云对 TM 影像数据的影响。

关键词: Landsat TM, 厚云, 云阴影, 光谱特征分析, 去云

中图分类号: TP751.1 文献标识码: A

引用格式: 李炳燮, 马张宝, 齐清文, 刘高焕. 2010. Landsat TM 遥感影像中厚云和阴影去除. 遥感学报, **14**(3): 534—545
RI Pyongsop, Ma Z B, Qi Q W and Liu G H. 2010. Cloud and shadow removal from LANDSAT TM data. *Journal of Remote Sensing*. **14**(3): 534—545

1 引 言

由于陆地观测卫星 Landsat TM/ETM+遥感影像数据具有光谱特征强、数据获取周期短、覆盖面积广、数据利用性强等特点, 所以被广泛地作为土地利用/土地覆被时空变化研究的基本数据来源(黎夏等, 1997)。它是当前进行区域尺度资源环境遥感监测较为理想的数据源。不过, 由于气候的原因, 很难获取完全无云的遥感影像, 大部分遥感影像在获取时会或多或少地受到云以及云在地面投射的阴影的影响。这给许多遥感影像的应用者带来了麻烦, 如何从遥感影像数据中去除云的影响, 往往是许多应用者所面临的首要问题(宋晓宇等, 2006)。所以, 去云处理也是图像预处理中的一个必要环节(宋小宁等, 2003)。

目前有许多对影像云检测和处理的研究, 如同态滤波去云法(赵忠明等, 1996; 吴为禄, 2003)、多光谱综合法、亮温差值法、指数法(宋小宁等, 2003)、基于遥感影像分类结果及云检测结果的去云处理算法(宋晓宇等, 2006)、基于神经网络和小波变换的图像融合法(Tapasmini & Suprava 等, 2008)、基于空间纹理分析和神经网络的检测法(Song 等, 2004)。

常规的云处理算法随云下垫面的不同而不同。对大范围存在薄云的影像, 采用同态滤波法较好。同态滤波法把频率过滤与灰度变化结合起来, 分离云与背景地物, 最终从影像中去除云的影响, 这种方法由于涉及滤波器以及截至频率的选择, 在滤波的过程中有时会丢失一些有用信息。而且对于有厚云的影像, 不能用这种方法。

对于局部有云的影像, 一般使用时间平均法, 这种算法适用于地物特征随时间变化较小的地域。对于植被覆盖茂密的地域, 由于植被的长势与时间有密切的关系, 不同时相的植被长势在影像中有明显的区别, 所以不能用这种简单的替代算法。

为解决上述的问题, 在本文中, 作者利用同一地域不同时相的 Landsat TM 影像, 采用同周期近时相或不同年份同一季节的 TM 遥感影像数据,根据每个波段的光谱特征的相对变化, 设计了厚云及其阴影地域增强模型, 结合该模型和常规的非监督自动分类模型, 采用线性回归分析的图像匹配法和像元替换运算, 提出了消除或者减小 Landsat TM 遥感影像数据中的云影响的步骤。研究结果表明, 该算法能

收稿日期: 2009-06-12; 修订日期: 2009-09-25
基金项目: 科技基础性工作专项项目(编号: 2007FY140800)。

第一作者简介: 李炳燮(1970—), 男, 朝鲜平壤, 博士研究生, 1993 年毕业于朝鲜 Kim Cheak 大学自动化系。目前主要从事遥感影像信息处理方面研究。E-mail: pyongsop@gmail.com。

消除或者弱化云对 Landsat TM 影像数据的影响。

2 研究区概况及数据来源

2.1 研究区概况

研究区位于朝鲜半岛中西部。地理坐标处在东经125°00′—126°10′，北纬38°15′—39°30′之间，东西端直线距离约 105.84km，南北端直线距离约137.46km。地上面积约 1.4 万 km^2。该地区覆盖山地、平原、海洋等各种不同的地形特征，覆盖多种土地利用/土地覆被类型，包含林地、草地、水田、旱地、盐田、城镇以及工业用地、裸地、水库、湖泊、河渠、海涂等。这些用地的空间分布复杂，地块比较小。这样的区域有助于研究在不同的地形条件下、不同的土地利用类型条件下，对去除云的影响效果。

2.2 数据来源

本文利用 2006 年和 2007 年的 8 月的 Landsat 117-33 轨道 覆盖的 TM 遥感影像数据。这些 Landsat 遥感影像数据是通过网络获取的。研究区遥感影像数据如表 1。

表 1 数据目录

编号	卫星	数据获取时间	传感器
1	Landsat5	2006-08-19	TM
2	Landsat5	2007-08-22	TM

两个时期遥感影像数据中， 2007-08-22 的遥感影像数据包含比较多的局部分布的厚云及其阴影地域，2006-08-19 的遥感影像中，云及其阴影地域很小。本研究的主要目的就是去除厚云及其阴影，因此以 2007-08-22 的遥感影像数据为基本目标数据，2006-08-19 的遥感影像数据利用为辅助数据。研究区 Landsat TM 遥感影像数据的彩色合成图如图 1。

3 研究方法

3.1 数据预处理

数据预处理的流程如图 2。首先利用 ERDAS IMAGINE 9.2 中的 AutoSync 模块进行两个时期 Landsat TM 遥感影像数据之间的图像配准处理(党安荣等，2007)，然后对于它们中每个对应波段进行匹配处理。为了匹配处理，选择数据均无云地域。本文采用人工目视判读的方法选出两个时期均无云地域 ROI。在这个地域，对于两个时期 TM 数据中

图 1 2007-08-22 Landsat TM 5, 4, 3 波段的彩色合成图像

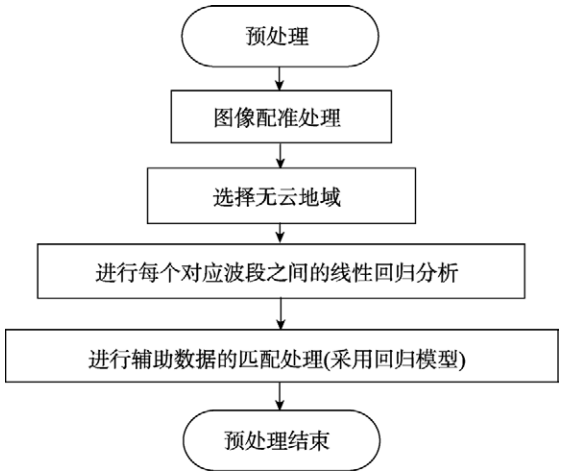


图 2 数据预处理过程

每个对应波段进行线性回归分析。采用线性回归分析结果，以 2007-08-22 的 TM 数据为基准，进行 2006-08-19 TM 数据每个波段的匹配处理。

两个 TM 数据中每个对应波段之间的线性回归方程如式(1)。

$$Y_i = a_i X_i + b_i \quad (1)$$

式中 X_i 为 2006-08-19 TM 数据 i 波段的光谱反射值；Y_i 为以 2007-08-22 TM 数据为基准变换的 2006-08-19 TM 数据 i 波段的光谱反射值；i 为波段号码。

从表 2 看出，两个时期 TM 遥感影像数据中每个对应波段之间的相关性较高。特别是，波段 4、5

的两个时期对应波段之间的相关性最高。

表 2 线性回归分析结果

波段	*a*	*b*	相关系数
1	0.86	7.75	0.90
2	0.86	3.53	0.87
3	0.82	3.51	0.87
4	0.92	0.52	0.96
5	0.94	0.02	0.96
7	0.86	1.70	0.93

3.2 厚云及其阴影地域增强模型

3.2.1 云地域光谱特征分析

为提取厚云及其阴影地域，首先进行了云地域跟无云地域间的光谱特征对比分析。

图 3 是 2007-08-22 的 Landsat TM 遥感影像数据中云地域的 5, 4, 3 波段彩色合成图像和空间剖面曲线图。图 3(a)中直线经过的区域有植被、水体、云的阴影、厚云。

彩色合成图像中白色区域就是厚云地域，它旁边的黑色部分是它的阴影。

分析图 3，得出云地域及其阴影地域的光谱特征如下:

(1) 每个波段的光谱反射值在厚云地域比无云地域明显增高(变化值 150 以上);

(2) 波段 4，5，7 的光谱反射值在水体和云的阴影地域明显减小。

(3) 在云的阴影地域中，波段 1，2，3 的光谱反射值略有减小，在水体地域中波段反射值增高。

虽然随云层厚度的变化，其相应的光谱特征变化量不一样，但是上述的厚云及其阴影导致的光谱特征变化的趋势不变。

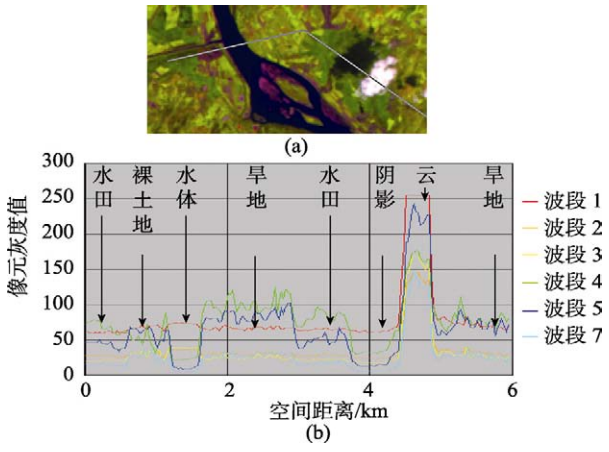


图 3 云地域光谱特征

(a) 波段 5, 4, 3 彩色合成图像 (b) 在(a)中切线处的空间剖面曲线

3.2.2 云地域增强模型

根据上述研究内容，提出了云地域和云阴影地域增强模型。

两个时期 TM 遥感影像数据之间的光谱特征变化包括云导致的变化和土地利用/土地覆被变化导致的光谱特征变化。为提取云及其阴影地域，首先应该区分云导致的光谱特征变化与土地利用/土地覆被变化导致的光谱特征变化。

本文首先设计了厚云地域增强模型。

从图 3，厚云导致的光谱特征变化跟土地利用/土地覆被变化导致的光谱特征变化的差异如下:

第一，厚云导致的光谱特征的变化量很大，每个波段的光谱反射值变化量在 150 以上;

第二，每个波段的光谱特征变化趋势一样，都增高。

根据上述的两个特征，提出了厚云地域增强模型。

$$\mathrm{CAEM} = \mathrm{MD}\times\mathrm{CDF} \tag{2}$$

$$\mathrm{MD} = \frac{\sum_{i=1}^{n}\left|B_{\mathrm{Refi}} - B_{\mathrm{Auxi}}\right|}{n} \tag{3}$$

$$\mathrm{CDF}=\mathrm{Sign}\left[\sum_{i=1}^{n}\mathrm{Sign}(B_{\mathrm{Refi}} - \mathrm{B}_{\mathrm{Auxi}}) - n + 1\right] \tag{4}$$

式中, B_{Refi} 为基准时期 TM 数据波段 i 的反射值; B_{Auxi} 为辅助时期 TM 数据波段 i 的反射值; 本文利用 2007-08-22 图像为基准数据，以 2006-08-19 图像经过线性变换后的图像为辅助数据; CAEM 为云地域增强模型(cloud area enhancement model); MD 为两个时期 TM 数据每个波段之间差的平均绝对值图像(mean difference); CDF 为云地域判别函数(cloud discriminate function); n 为波段数。

如果基准时期 TM 数据的每个波段值都增高则 CDF=1，否则 CDF≤0

在光谱特征变化大的地域中 MD 值增高，在没变化的地域中 MD 值减低。

在 MD 值高的地域中 CDF=1 的就是云地域。

为减少不同太阳位置、不同大气条件等一些原因的影响，本文利用与基准 TM 影像数据匹配的辅助 TM 影像数据。对于无云地域，对两个 TM 影像数据各个对应波段进行线性回归分析。以 2007-08-22 的 TM 影像数据为基准，进行辅助 TM 影像数据的线性变换。

图 4 是采用上述的模型计算得到的云地域增强结果图像。从图 4 看出，厚云地域跟其他地域差别明显。用一个单纯的阈值能区别厚云地域和其他地域，并且选择阈值的范围比较宽。本文采用 30 和

图 4　2007-08-22 云地域增强图像

50 两个阈值对两个结果图像进行对比分析。分析结果说明两个图像之间的变化部分是整个图像面积的0.05%以下。

云的阴影地域提取方法比厚云地域提取方法复杂，因为云阴影导致的光谱特征变化程度不大，有时跟土地利用/土地覆被变化导致的光谱特征变化混淆。因此，本文提出了一个云阴影地域增强模型与常规的非监督分类模型相结合的方法提取云阴影。

从图 3 看出，云阴影导致的光谱特征变化跟其他无云地域的光谱特征变化的差异如下:

第一，波段 4, 5, 7 的反射值明显减低；第二，在云阴影地域波段 1 的反射值减低，但在水体地域波段 1, 2, 3 的反射值增高。

Landsat TM 遥感影像数据中水体跟云阴影地域的光谱特征相似。因此，为提取云阴影地域首先提取水体地域。根据上述的光谱特征变化特点，提出了云阴影增强模型。

$$SAEM = SDF\times[(B_{Aux5}-B_{Ref5})+(B_{Aux7}-B_{Ref7})]/2 \quad (5)$$

$$SDF=Sign(1-Sign(\sum_{i=1}^{3} Sign(B_{Refi}-B_{Auxi})-2)) \quad (6)$$

式中，$B_{Ref\ i}$ 为基准时期 TM 数据波段 i 的反射值; B_{Auxi} 为辅助时期 TM 数据波段 i 的反射值; SAEM 为云影地域增强模型(Shadow Area Enhancement Model); SDF 为云影地域判别函数(Shadow Discriminate Function); Sign()为符号函数，选择 1, 0, –1 中的一个值。

如果基准时期 TM 数据的 1, 2, 3 波段值增高则 SDF=0(水体地域)，否则 SDF≥0(非水体地域)。

基准时期 TM 数据的波段 5 和波段 7 的反射值降低的地域中非水体地域的(SDF≥0 的)SAEM 的值增高。

通过上述模型计算得到的云阴影地域增强图像中也包含一些因土地利用/土地覆被变化导致的光谱特征变化的地域。在图 5 中紫色的地域是厚云地域，绿色地域可能是云的阴影的地域。

为区别云导致的阴影地域跟土地利用/土地覆被变化导致的光谱特征变化地域，本文采用非监督自动分类模型。对于采用 SAEM 计算得到的图像，进行阈值处理，提取可能是云阴影的地域。对于结果部分地域，利用 Landsat TM 源数据，进行非监督分类。因为云阴影地域的每个波段的光谱特征跟植被、裸土地等其他土地类型的光谱特征不一样，采用常规的非监督自动分类模型能区别云的阴影地域跟其他土地利用/土地覆被变化地域。

对于不同厚度的云、不同土地类型的地域、不同大气条件下的影像，虽然上述的光谱特征变化程度可能不一样，但是这种变化特征一直存在，在大气条件较好的图像中更明显。依据这一原理，利用一定周期内的或者不同年份同一季节的几幅 Landsat TM 影像数据能得到没有云影响或者影响最小的 Landsat TM 影像数据。

图 5　2007-08-22 云及其阴影地域增强图像

3.3 自动消除或者减少云影响过程

消除或者减少云影响的总体流程如图 6。

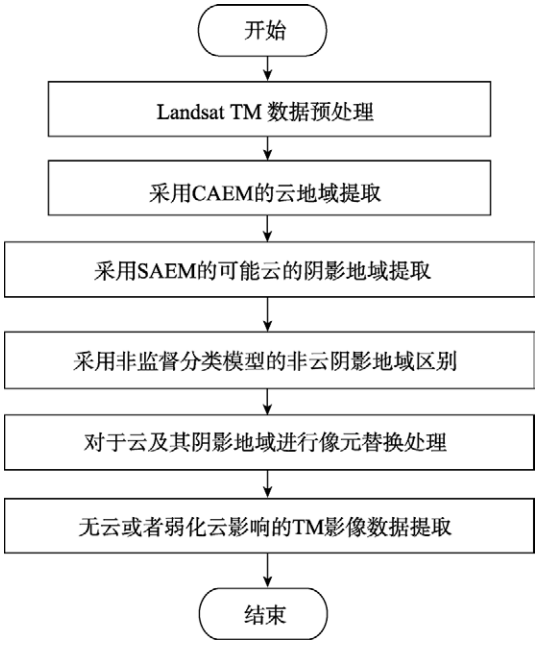


图 6 自动消除或者减少云影响流程图

对一定周期内或者不同年份同一季节的两个时期 Landsat TM 遥感影像数据，进行配准和匹配处理。 对于无云地域，进行两个 Landsat TM 源数据(一个基准 TM 影像数据和另外一个辅助 TM 影像数据)中每个相应波段之间的线性回归分析，得到互相匹配的新的辅助 TM 影像数据，减少不同太阳位置、不同大气条件等影响。

采用上述的云和云阴影增强模型 CAEM、SAEM 和非监督自动分类模型，利用 ERDAS IMAGINE 9.2 中的 Modeler 模块 Model Maker 编程实现，可以算出 TM 遥感影像数据中云和云阴影地域增强图像。

然后利用与基准 TM 影像数据匹配的新辅助 TM 影像数据和云及其阴影地域增强图像，对于有云影响的地域进行像元替换处理，得到消除或者减少了云影响的 TM 遥感影像数据(图 6)。

4 结果与分析

采用 3.2 节所述的云影响增强模型 CAEM、SAEM，通过 3.3 节所述的消除或者减少云影响处理流程，得到了从 2007-08-22 Landsat TM 遥感影像数据的消除云影响的结果影像数据。图 7 是结果影像数据的波段 5，4，3 的彩色合成图像，从结果可以看出，云和云阴影已经消除。图 8 是消除或者减小影响后的研究区的整个结果图像。

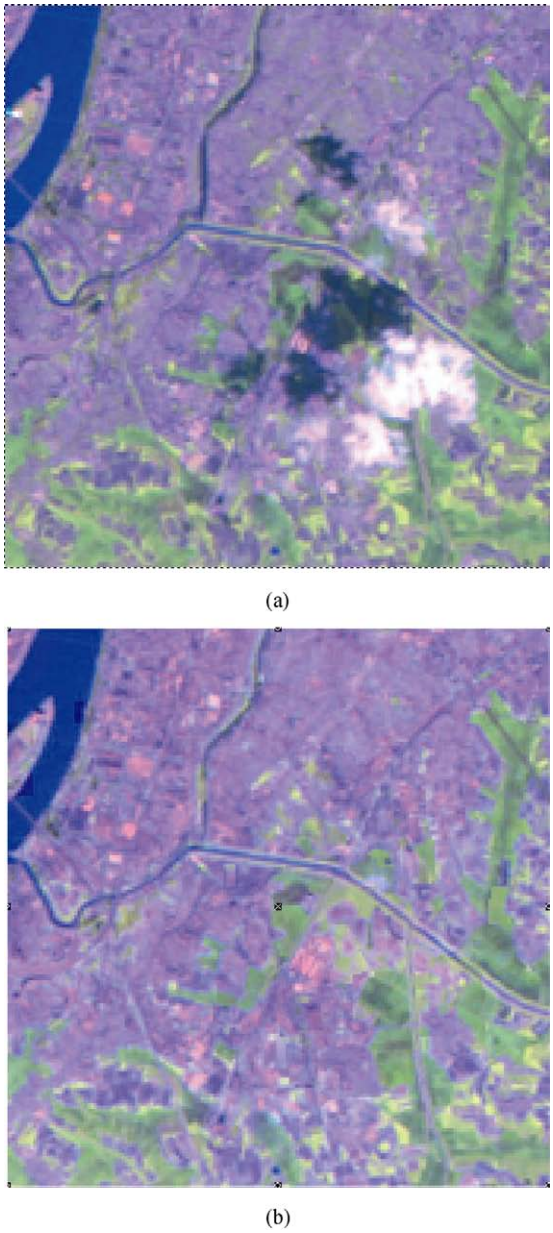


图 7 消除或者减小云影响前后的图像
(a) 云及其阴影地域; (b) 结果图像

采用 SAEM 模型算出的云阴影地域增强图像中非阴影地域增强的地域大概如下: (1) 水田地域，具体来说, 2006 年的 TM 数据(辅助数据)中，种植水稻，但是 2007 年的 TM 数据(基准数据)中还没种植水稻、就是类似于水体的水田地域。这样的效果反而对水田自动分类有好处。因为它的本质不是水体，而是水田; (2) 海涂地域和水域; 按照水平面的高程变化，有的地域增强为阴影地域。这种地域通过非监督自动分类模型能与阴影地域区分开来。

图 8　消除或者减小云影响后的整个结果图像

5　结　论

利用本文提出的方法，使用 2007-08-22 和 2006-08-19 的两个时期 Landsat TM 遥感影像数据，对于朝鲜半岛中西部地区进行 Landsat TM 遥感影像的去除或弱化云影响的研究。结果表明：上述方法可以有效地去除或者弱化云的影响，处理流程简单，并且在这个过程中，不需要任何严格的阈值调整，除了需要 Landsat TM 遥感影像数据中每个波段的光谱反射信息以外，不需要其他信息。所以对于其他地域、其他时期的 Landsat TM 遥感影像数据可以自动实行上述流程。

虽然本文仅利用了两个时期的 Landsat TM 遥感影像源数据来计算云影响，但是为获取更好的结果，可以利用两个以上时期的 Landsat TM 遥感影像源数据。

本文提出的运算以每个像元为单位进行，对于影像数据之间的配准要求很严格，同时为了获取更好的结果，应该进行高精度的地形校正、大气纠正和匹配处理。

REFERENCES

Dang A R, Wang X D, Chen X F and Zhang J B. 2007. ERDAS IMAGINE remote sensing image processing methods. Beijing: Qinghua University Press

Li X and Ye J A. 1997. The accuracy improvement of the land use change monitoring based on remote sensing using principal component analysis. *Journal of Remote Sensing*, **1**(4): 282—288

Song X N and Zhao Y S. 2003. Cloud detection and analysis in MODIS images. *Chinese Journal of Image and Graphics*, **8**(9): 1079—1083

Song X Y, Liu L Y, Li C J, Wang J H and Zhao C J. 2006. Study on the cloud removal based on the single remote sensing image data. *Optical Technology*, **32**(2): 299—303

Song X N, Liu Z H and Zhao Y S. 2004. Cloud detection and analysis of MODIS image. IEEE Geoscience and Remote Sensing Symposium

Tapasmini Sahoo and Suprava Patnaik. 2008. Cloud removal from satellite images using auto associative neural network and stationary wavelet transform. *IEEE Computer Society*, **6**: 100—105

Wu W L. 2003. The cloud elimination in remote sensing images. *Railway Aerial Survey*, (1): 6—7

Zhao Z M and Zhu C G.1996. The cloud removal method in remote sensing images. *Environmental Remote Sensing*, **11**(3): 195—199

附中文参考文献

党安荣，王晓栋，陈晓峰，张建宝． 2007. ERDAS IMAGINE 遥感图像处理方法. 北京：清华大学出版社

黎夏，叶嘉安. 1997. 利用主成分分析改善土地利用变化的遥感监测精度 ——以珠江三角洲城市用地扩张为例. 遥感学报，**1**(4): 282—288

宋小宁，赵英时. 2003. MODIS 图像的云检测及分析.中国图像与图形学报， **8**(9): 1079—1083

宋晓宇，刘良云，李存军，王纪华，赵春江. 2006. 基于单景遥感影像的去云处理研究. 光学技术, **32**(2): 299—303

吴为禄. 2003. 遥感图像中的云层消除处理. 铁路航测, (1**)**: 6—7

赵忠明，朱重光. 1996. 遥感图像中薄云的去除方法. 环境遥感，**11**(3):195—199