Cloud and shadow removal from LANDSAT TM data

RI Pyongsop\textsuperscript{1, 2}, MA Zhangbao\textsuperscript{1}, QI Qingwen\textsuperscript{1}, LIU Gaohuan\textsuperscript{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

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...
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 14000km². 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>
<thead>
<tr>
<th>No.</th>
<th>Satellite</th>
<th>Date</th>
<th>Sensor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Landsat 5</td>
<td>2006-08-19</td>
<td>TM</td>
</tr>
<tr>
<td>2</td>
<td>Landsat 5</td>
<td>2007-08-22</td>
<td>TM</td>
</tr>
</tbody>
</table>

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.

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).

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.
\[ Y_i = a_i X_i + b_i \]  

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>
<thead>
<tr>
<th>Band</th>
<th>( a )</th>
<th>( b )</th>
<th>( R )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.86</td>
<td>7.75</td>
<td>0.90</td>
</tr>
<tr>
<td>2</td>
<td>0.86</td>
<td>3.53</td>
<td>0.87</td>
</tr>
<tr>
<td>3</td>
<td>0.82</td>
<td>3.51</td>
<td>0.87</td>
</tr>
<tr>
<td>4</td>
<td>0.92</td>
<td>0.52</td>
<td>0.96</td>
</tr>
<tr>
<td>5</td>
<td>0.94</td>
<td>0.02</td>
<td>0.96</td>
</tr>
<tr>
<td>7</td>
<td>0.86</td>
<td>1.70</td>
<td>0.93</td>
</tr>
</tbody>
</table>

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.
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.

\[
\text{CAEM} = \text{MD} \times \text{CDF} \\
\text{MD} = \frac{\sum^n_{i=1} |B_{\text{Refi}} - B_{\text{Auxi}}|}{n} \\
\text{CDF} = \text{Sign} \left( \sum^n_{i=1} \text{Sign} (B_{\text{Refi}} - B_{\text{Auxi}}) - n + 1 \right)
\]

where \( B_{\text{Refi}} \): the \( i \) band’s gray value of the main data; \( B_{\text{Auxi}} \): 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.

\( \text{CAEM: Cloud area enhancement model; MD: Mean absolute difference; CDF: Cloud discriminate function; } n: \text{ the number of bands.} \)

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 ≤ 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.

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.

\[
\text{SAEM} = \text{SDF} \times \left[ (B_{\text{Aux}5} - B_{\text{Ref}5}) + (B_{\text{Aux}7} - B_{\text{Ref}7}) \right] / 2
\]

\[
\text{SDF} = \text{Sign}(1 - \sum_{i=1}^{3} \text{Sign}(B_{\text{Ref}i} - B_{\text{Aux}i} - 2))
\]

where \(B_{\text{Ref}i}\): the \(i\) band’s gray value of the main data; \(B_{\text{Aux}i}\): 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 ≥ 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 ≥ 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
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 unsupervised 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

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.7 Cloud free TM image](imageURL)

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

Fig. 6 Diagram of automatic cloud removal
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.

5 CONCLUSIONS

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

Landsat TM

李炳燮, 马张宝, 齐清文, 刘高焕

1. Landsat TM
2. Landsat ETM+

摘要：Landsat TM

关键词：Landsat TM, ETM+, 

中图分类号：A

文献标识码：A

1

Landsat TM/ETM+ (Song, 2004)

2

Landsat TM/ETM+ (Tapasmini & Suprava, 2008)

E-mail: pyongsop@gmail.com
2.1 2.1

表 1 数据目录

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Landsat5</td>
<td>2006-08-19</td>
<td>TM</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Landsat5</td>
<td>2007-08-22</td>
<td>TM</td>
<td></td>
</tr>
</tbody>
</table>

3.1 3.1

\[ Y_i = a \cdot X_i + b_i \] (1)

2007-08-22 Landsat TM 5, 4, 3

2006-08-19 Landsat TM

2007-08-22 Landsat TM 5, 4, 3
表 2 线性回归分析结果

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.86</td>
<td>7.75</td>
<td>0.90</td>
</tr>
<tr>
<td>2</td>
<td>0.86</td>
<td>3.53</td>
<td>0.87</td>
</tr>
<tr>
<td>3</td>
<td>0.82</td>
<td>3.51</td>
<td>0.87</td>
</tr>
<tr>
<td>4</td>
<td>0.92</td>
<td>0.52</td>
<td>0.96</td>
</tr>
<tr>
<td>5</td>
<td>0.94</td>
<td>0.02</td>
<td>0.96</td>
</tr>
<tr>
<td>6</td>
<td>0.86</td>
<td>1.70</td>
<td>0.93</td>
</tr>
</tbody>
</table>

3.2

3.2.1

(1) 2007-08-22 Landsat TM
(2) 2007-08-22, 3(a)
(3) 3, 4, 5

CAEM = MD×CDF

\[ MD = \frac{\sum_{i=1}^{n} |B_{Refi} - B_{Auxi}|}{n} \] (3)

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

3.2.2

[图片]

3.2.2.1

3.2.2.2

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果

线性回归分析结果
SAEM = SDF \times [((B_{Aux1}−B_{Ref})+(B_{Aux5}−B_{Ref}))]/2 \quad (5)

SDF=\text{Sign}(1−\text{Sign}(\sum_{i=1}^{3}\text{Sign}(B_{Ref}−B_{Auxi})−2)) \quad (6)

where, \( B_{Ref} \) \( i \) \( B_{Aux} \) \( \alpha \) \( \beta \) \( \gamma \) (Shadow Area Enhancement Model); SDF \( \alpha \) \( \beta \) \( \gamma \) (Shadow Discriminate Function); Sign() \( 1, 0, -1 \) \( \alpha \) \( \beta \) \( \gamma \).

Landsat TM 

\( \alpha \) \( \beta \) \( \gamma \)
3.3 Landsat TM 数据处理流程

```
开始

Landsat TM 数据预处理

采用CAEM云地城提取

采用SAEM可能的地形提取

采用非监督分类模型的非云阴影区域

对于云及其阴影区域进行象元替换处理

无云或者弱化云影响的TM影像数据提取

结束
```

4. 2007-08-22 Landsat TM 6
REFERENCES


附中文参考文献


Landsat TM遥感影像中厚云和阴影去除

作者: 李炳燮, 马张宝, 齐清文, 刘高焕, RI Pyongsop, MA Zhangbao, QI Qingwen, LIU Gaohuan

作者单位: 李炳燮, RI Pyongsop(中国科学院地理科学与资源研究所, 北京100101; 朝鲜遥感与地理信息系统研究所, 朝鲜, 平壤), 马张宝, 齐清文, 刘高焕, MA Zhangbao, QI Qingwen, LIU Gaohuan(中国科学院地理科学与资源研究所, 北京, 100101)

刊名: 遥感学报

英文刊名: JOURNAL OF REMOTE SENSING

年, 卷(期): 2010,14(3)

参考文献 (14条)

1. 赵忠明; 朱重光. 遥感图像中薄云的去除方法. 1996(03)
2. 党安荣; 王晓栋; 陈晓峰; 张建宝. ERDAS IMAGINE遥感图像处理方法. 2007
3. Zhao Z M; Zhu C G. The cloud removal method in remote sensing images. 1996(03)
4. Wu W L. The cloud elimination in remote sensing images. 2005(01)
6. Song X N; Liu Z H; Zhao Y S. Cloud detection and analysis of MODIS image. 2004
7. Song Y Y; Lin L Y; Li C J; Wang J H and Zhao C J. Study on the cloud removal based on the single remote sensing image data. 2006(02)
9. Li X; Ye J A. The accuracy improvement of the land use change monitoring based on remote sensing using principal component analysis. 1997(04)
10. 吴为禄. 遥感图像中的云层消除处理[期刊论文]. 铁路航测. 2003(01)
11. 宋晓宇; 刘方云; 李存军; 王纪华. 赵春江. 基于单景遥感影像的去云处理研究[期刊论文]. 光学技术. 2006(02)
12. 宋小宁; 赵英时. MODIS图像的云检测及分析[期刊论文]. 中国图像与图形学报. 2003(09)
13. 勃复; 叶嘉安. 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. 1997(04)
14. Dang A R; Wang X D; Chen X F; Zhang J B. ERDAS IMAGINE remote sensing image processing methods. 2007

本文链接: http://d.g.wanfangdata.com.cn/Periodical_ygxb201003011.aspx