Development of Application Tool for Automatically Creating QuickBird Imagery Index

Pattrawuth PHUTHONG, Kativich KANTHA, and Poramet THUWAKHAM

Geo-Informatics Scientist, User Services and Business Development,
Geo-Informatics and Space Technology Development Agency (Public Organization)

120 (Building B) ChaengWattana Road, Lak Si District, Bangkok 10210, THAILAND;
Tel: +66(0)-2141-4575; Fax. +66(0)-2143-9592
E-mail: phuthong@gistda.or.th

KEYWORDS : Satellite Imagery Index, QuickBird, GDAL, and Python

ABSTRACT : The objective of this study is to develop an application tool for automatically creating a QuickBird imagery index to support works of user service. Customers can quickly search and view samples of the imageries of interest in the archive. The system is developed by raster and vector managing and processing tool set (Geo-spatial Data Abstraction Library: GDAL) and the Python programming language. It will automatically work when the server receives original imageries via the FTP. It will then execute the data file, by selecting different types of processing routines to enhance, compress, resample, extract, converse the coordinate of imageries, and obtain the final imageries in both the UTM and Geographic Coordinate Systems. The results are Shape File and a set of quick look images with a ground resolution of 2.4 meters and 25 meters. Results from the developed system were compared with the results from the manual process. It was found that the developed system is more effective in providing better object contrast, higher accuracy of attribute data with less processing time. However, the result in such application did not provide natural color and the significantly larger file size of quick look images could be seen as disadvantages. However, data duplication from server to server and image's boundary are similar for the two systems. For better operation results, the prototype will need to be adjusted in order to achieve a more accurate boundary.

INTRODUCTION

Searching satellite imageries is one of the important tasks of user service for serving the needs of customers. It is important for customers to know details of imagery in their AOI, such as sensor name, type of product, acquisition date, cloud cover percentage, processing time, cost effectiveness, etc. The completeness of details will help customers to evaluate the feasibility and then decide to receive services from GISTDA which could be a significant KPI. Creating an index of archived imageries will allow the user service team to quickly search the imageries in the archive. The index consists of Quick Look image and Vector data (Shape file) which are convenient for customer's AOI. At the moment, creating an imagery index cannot be carried out in real time because there are large amounts of satellite data from many different satellites coming to the archive each day. Therefore, it is difficult for the user service team to serve a customer's needs in time, and they may miss the opportunity to distribute archived data that has not yet been created in the index.

In this study, the prototype application for automatically creating a QuickBird imagery index has been developed to reduce the user services tasks, and to make it real time services. All types of QuickBird products in the archive are used for developing the prototype. The result of a new application will be compared to the result from the existing system, considering both quality and data duplication. It will then be adjusted to perform more accurately and effectively.

OBJECTIVE

To compare the accuracy and quality of the QuickBird imagery index from developed application, to the result from the manual operation.

SCOPE OF RESEARCH

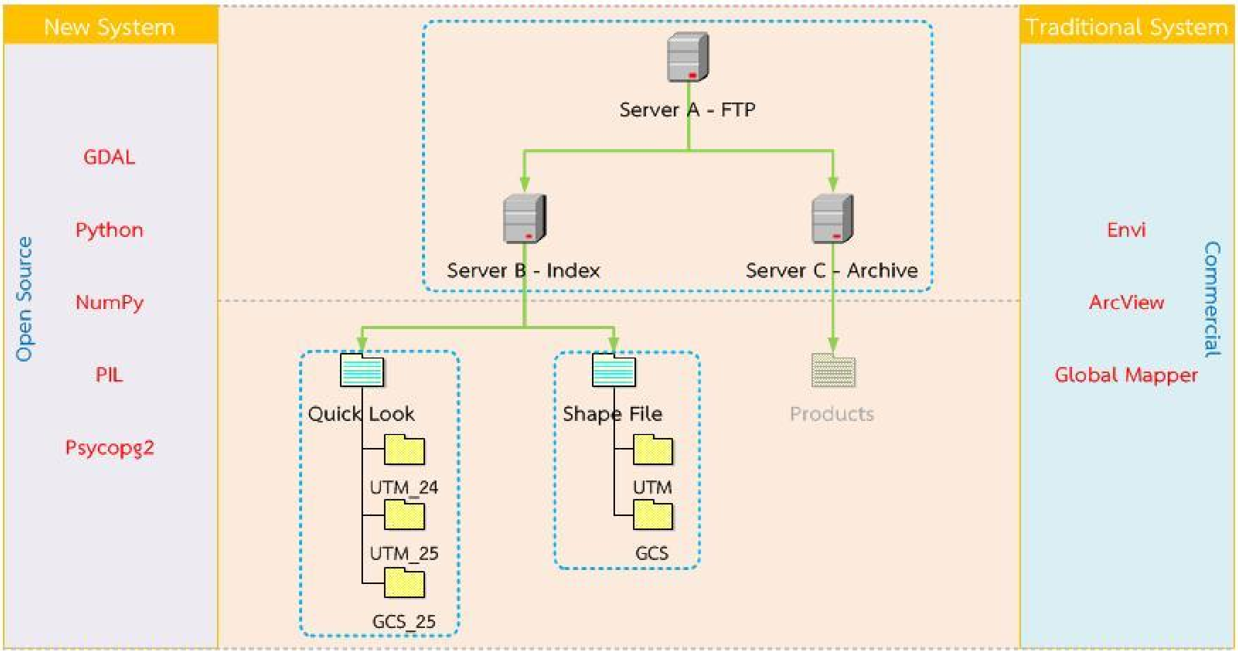


Figure 1 : Processing Work Flow

To develop the prototype application for automatically creating the QuickBird imagery index, the whole existing work flow which is processed by the user services operator was simulated. The developed application on the Server A consists of two steps as follows:

1. Duplicating the original data to server C, data in this server will be processed to serve the customer's needs such as panchromatic, multispectral, bundle, Pan-sharpened, mosaic and Ortho image.

2. Creating imagery index and transferring to server B. The index of an original imagery consists of three quick look images and two vector data (Shape file).
   - 2.1. Three Quick Look images are:
     - 2.1.1. Quick Look image at 2.4 meters GSD in UTM WGS84 Zone 47N (UTM_24)
     - 2.1.2. Quick Look image at 25 meters GSD in UTM WGS84 Zone 47N (UTM_25)
     - 2.1.3. Quick Look image at 25 meters GSD in Geo WGS84 (GCS_25)
   - 2.2. Two Vector shape file are:
     - 2.2.1. Boundary of imagery with attribute data 12 items in UTM WGS84 Zone 47N
     - 2.2.2. Boundary of imagery with attribute data 12 items in Geo WGS84

Table 1 : 12 items were included in the Field Name and Description in Attribute Data.

Field Name	Description
Sensor	Sensor Vehicle
Scene_ID	Scene Number
Product_ID	Product Number
Date	Acquisition date
Year	Acquisition Year
Scene	Number of Scenes ofSCENE_ID
Type	Type of Product
Level	Level of Product
Datum	Datum
Area	Square Kilometer
License	License
Agency	Data Distributor

Results from both systems (existing commercial software and the developed system from free and open source development tools) will be compared in four categories.

Table 2 : Comparison category.

Category	Comparison Item
Data duplication	Duplicating original data to Server C Duplicating index data to Server B
Quick Look image	Color Contrast Size on disk
Vector data	Boundary of imagery Completeness of attribute table
Processing Time	Average processing time (Quick Look image, Vector Data) per original scene

METHODS AND EQUATION

1. Products of QuickBird
Products of QuickBird are licensed by DigitalGlobe Inc. The Satellite collects panchromatic data (405-1053 nm) at 0.61 meter GSD, and multispectral data (Blue: 430 - 545 nm, Green: 466 - 620 nm, Red: 590 - 710 nm, and Near-IR: 715 - 918 nm) at 2.44 meter GSD. The total number of scenes for creating sample index is 97. The largest size is 13.4 GB, and total is 184 GB for three types of product: Basic Satellite Imagery (Bundle), Standard Satellite Imagery(4 –Band Pan Sharpened) and Standard Satellite Imagery(Bundle).

There are two levels of products, Basic Satellite Imagery and Standard Satellite Imagery. Basic Satellite Imagery comprises about 5% of all QuickBird data in the archive. In addition, this type of product is used in Ortho Rectification works for Topographic Surveying,in which the high-positioning accuracy is needed. The other type of QuickBird product, Standard Satellite Imagery, comprises the other 95%of the archive, and this product is often used for classification (where high – positioning accuracy is not necessary).

Each level of product in the archive is stored in two ways, which are 4-Band Pan Sharpened and Bundle (Panchromatic + 4-Band Multispectral). The former comprises 15% of the QuickBird in the archive, and the rest is stored as Bundle. The different types of product will be used for generating the different index.

2. Development Tools
This prototype application is developed on Windows Server 2003, using the free and open source tool, Pytho and its extension function to control the entire system, as outlined below.

2.1 GDAL (Geospatial Data Abstraction Library) is a library for reading and writing raster geospatial data formats, and is released under the permissive X/MIT style free software license by the Open Source Geospatial Foundation. As a library, it presents a single abstract data model to the calling application for all supported formats. It may also be built with a variety of useful command-line utilities for data translation and processing. The related OGR library (which is part of the GDAL source tree) provides a similar capability for simple feature vector data. Several software programs use the GDAL/OGR libraries to allow them to read and write multiple GIS formats. For example:ArcGIS, FWTools, Google Earth, MapServer, World Wind JAVA, and Quantum GIS.

2.2 Python is a high-level programming language developed by Guido van Rossum in 1990. It is a flexible and effective programing language since it has a large and comprehensive standard library. Moreover, it is free and open source simply structured piece of software that can be operated on many operating systems. Python is based on server side script which has a high-level of security. Nowadays, Python is becoming more popular especially in the Geo-Informatics field such as Pan-sharpening, Mosaicing, creating imagery indices, GIS data management, geodatabase management, etc. However, it is necessary to add extensions to the Python's standard library. There are several extensions for the Python library, but only three extensions are added for the project, namely NymPy, PIL, and psycopg2.

2.3 NumPy is a general-purpose array-processing package designed to efficiently manipulate large multi-dimensional arrays of arbitrary records.

2.4 PIL (Python Imaging Library) adds image processing capabilities which support many file formats, and provide powerful image processing and graphics capabilities.

2.5 Psycopg2 is a PostgreSQL/PostGIS database adapter for Python. It was designed for heavily multi-threaded applications.

Working Method

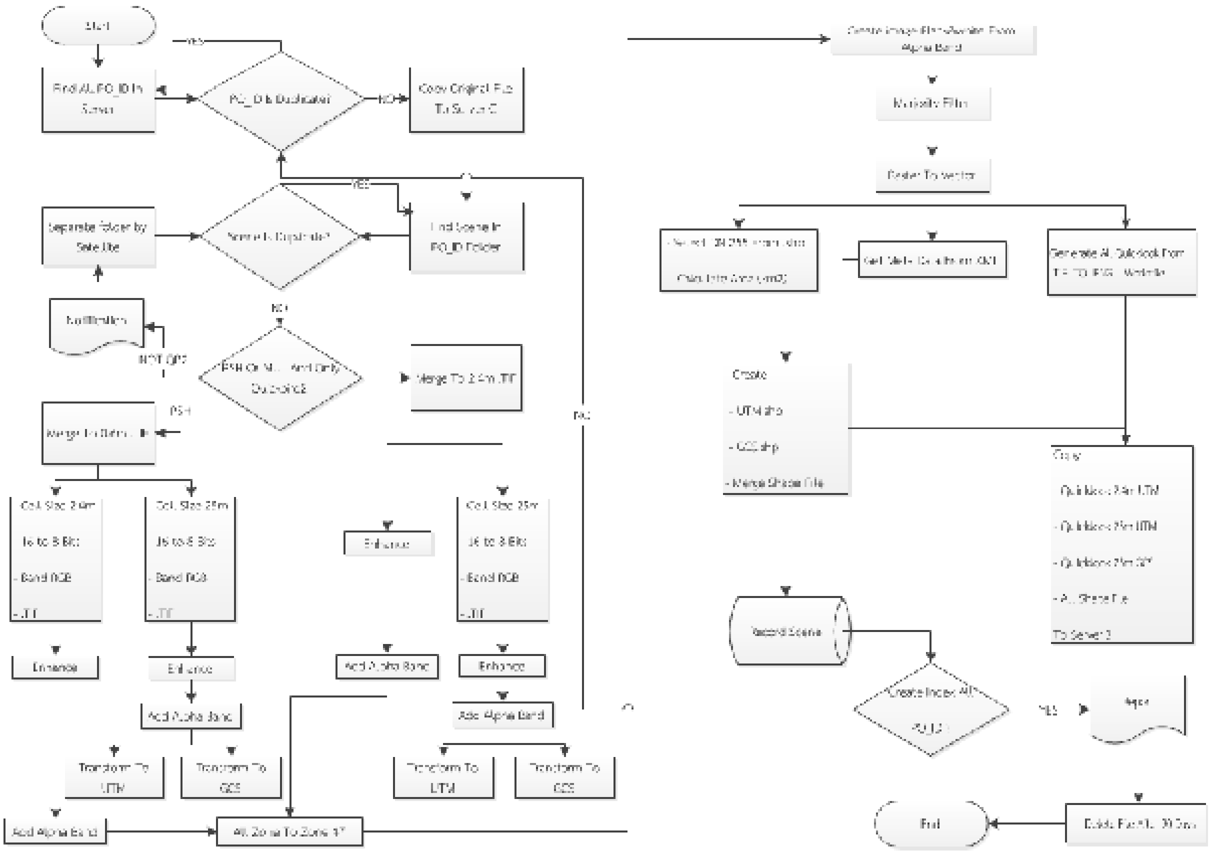


Figure 2 : Processing Work Flow

1.When new imageries are available in server A, the system will verify the data and index from the database. For the data that has not yet been duplicated or had an imagery index created, it will be duplicated to server C and then the index will be created in server A. The quick look image and vector data are then copied to server B. Moreover, the operator will be notified by an e-mail that all of this has occurred.

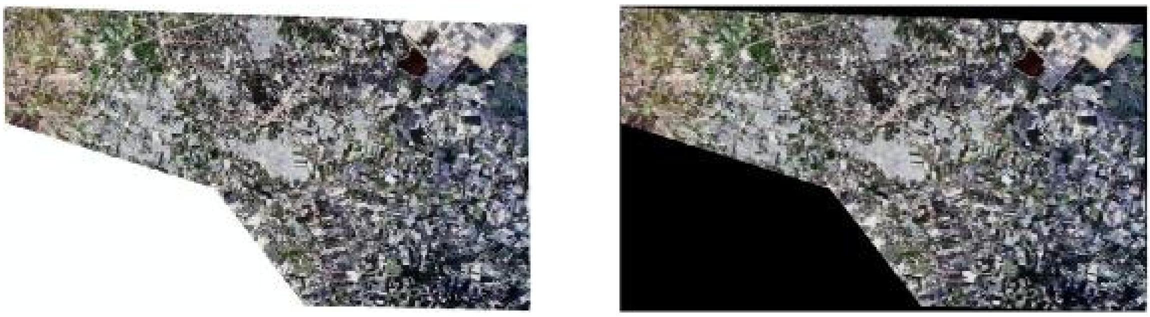
2. Index creation starts with verifying the level and type of product for quick look image generation. For 4-Band Pan-sharpened product, they will be mosaicked then resampled from the 0.61 meter GSD to the 2.44 meters and 25 meters GSD. In addition, the 4-Band multispectral of bundle product will be selected for mosaicking and resampling to 25 meters GSD.

3. 4-Band multispectral data with 2.44 meters and 25 meters GSD will be depth pixel converted form 16 bit pixel to 3-Band multispectral with 8 bit pixel, and Histogram Equalization enhanced. Both imageries will be processed simultaneously.

4. Each imagery will be used for generating as 1-band bit pixel, Alpha Band image, by accumulating the digital number in each band, then re-classing values to either 0 or 255. This alpha band image will be used for generating a transparent background.

$$\text{Alpha Band}_{(0,255)} = \text{Reclass}(DN_B + DN_G + DN_R)$$

5. Create the boundary and coordinate values (same as 3-Band multispectral image, RGB:321) for 4 Band multispectral with 8 bit pixel, and fill the transparent background with the alpha band image.



RGBA RGB

Figure 3 : 4 Band and 3 Band Multispectral

6. RGBA data of both derived imageries (2.44 meters and 25 meters GSD) will be coordinated system converted to UTM WGS 1984 zone 47N and Geographic WGS 1984. However, only 2.44-meters RGBA will be converted to UTM WGS 1984 zone 47N. After coordinate conversion, all imageries will be reduced in file size by changing a file name extension from TIF/TFW to PNG/PGW. Three Quick Look images will be generated in this step.

7. Convert the alpha band value (in matrix pattern, either 0 or 255) of RGBA at 2.4 meters GSD to an image. In this step, a majority filter (3×3) will be applied to get rid of salt and pepper (value of 0).

0	255	255		255	255	255
255	255	0	Majority Filter →	255	255	255
0	255	255		255	255	255

Figure 4: Majority Filter (3 x 3)

8. Create the boundary and coordinate values (same as 2.4-meter alpha band image), and specify it as 1-band bit pixel for the result from the previous step. Then convert from raster data to vector data.

9. Derived data is a combination of two values; 0 and 255. The value of 255 represents imagery, and the value of 0 represents the boundary. Majority filtering may not be capable of removing all of the salt and pepper, so it is necessary to check the condition and set new parameters at this step.

Conditions:
1. If there is no salt and pepper, set the parameter (Polygon = 255).
2. If salt and pepper appear, set the parameter for an outer ring (Polygon = 255).
3. If salt and pepper appear on some parts of the image, select the biggest polygon and set the parameter for the outer ring (Polygon=255).

10. Add 12 fields to the attribute data, calculate the area in square kilometers, and add aditional data from *.XML file of the original data to the attribute.

11. Combine all vector data into a single set for searching QuickBird imageries, and convert the coordinate system from UTM WGS 1984 zone 47N to Geographic WGS 1984.

12. Duplicate all three sample images and two vector data to server B, and duplicate the index to server A for verification prior to creating the new index.

13. Notify the operator after duplication and index creation, and delete the data from server A after 90 days.

RESULTS & DISCUSSION

The result from the developed system was compared to the previous manual operation's result.

Table 3 : Results Comparison.

Developed System	Existing System
1. Accuracy of duplicating original data to server C	
98%	95%
2. Accuracy of duplicating index to server B	
98%	95%
3. Color	
	
4. Object Contrast	
	
5. Size on disk	
3X	X
6. Boundary of imagery	
	

Developed System	Existing System
7. Accuracy of attribute data	
100%	95%
8.Average processing time (Quick Look image, Vector Data) per original scene	
X	2X

The accuracy of result duplication from server A to B and C via log in system, both new and existing system are not very different. The result from the new system is more accurate because the user can specify the location. However, if the hard disk of the server is full and disconnected from the network, then the data duplication cannot be completed. In this case, the system will notify the user by e-mail.

The color of imagery from the existing system looks better and more natural because the enhancement is calculated based on the proportion of the entire object on the image. Due to the fact that the new system will only do histogram equalization enhancement, and cannot adjust to the proportion of objects on the imagery, the color does not look natural.

The contrast of objects on the imageries from the new system is more distinguishable, due to the enhancement making imageries look brighter and redder. Thus, objects that represent forest, agriculture and water are distinguishable, while urban and barren lands are not.

The file size of Quick Look images from the existing system are one third smaller, due to the difference of band number. The existing system's images consist of 3 bands (RGB), while the new system's has 4 bands (RGBA), making the background transparent.

The boundaryof the imagery from the existing and new systems only differ by a small amount. The similarity of boundary from both systems is 95 per cent of the entire image. The existing system generates boundaries by manual digitizing, so the accuracy is dependent on the operator's skill. On the other hand, the new system generates boundaries by using digital number values (RGB=000) on the edge of imagery.

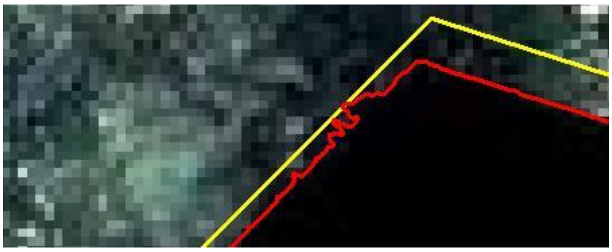


Figure 5 : Boundary from existing system (Top - Yellow) and new system (Bottom - Red)

The completeness of attribute table of the new system is more complete, since the developed tools will get information from an *.XML file and automatically fill the table. However, the existing system is filled by the operator, and that can cause some errors.

Comparing the average processing time from duplicating data through vector generation, the new system can perform two times faster than the manual operation. In the case of working with large amounts of data, the developed system can operate 24/7 which is difficult for human to do.

CONCLUSIONS

Overall, comparing the manual operation (existing system) and the new system, the new system is more effective in providing better object contrast, higher accuracy of attribute data, and it requires less processing time. However, the color of the result in the new system is not natural and the significantly larger file size of quick look images could be seen as disadvantages. In addition, data duplication from server to server and image's boundary are similar for the two systems. For better operation results, the prototype application will need to be adjusted in order to achieve a more accurate boundary.

Table 4 : Prominent Point

New System	Similarity	Existing System
Contrast	Accuracy of duplicating original data to the server C	Color
The accuracy of the attribute table	Accuracy of duplicating index to the server B	Quick Look image's file size
Average processing time	Boundary of imagery	

RECOMMENDATIONS

1. Resource management is necessary for processing imagery that is larger than 4 GB.
2. Some PO_ID of QuickBird imagery might have been combined with WorldView-1 and WorldView-2 data. The operator should store them separately.

ACKNOWLEDGMENT

Thanks to the Geo-Informatics and Space Technology Development Agency (GISTDA) for the opportunity to develop this project. Thanks are also due to Mr.Anusorn Rungsipanich, and Mr.Pisut Nakmuenwai for their suggestions and advices related to application development, equations, and techniques.

REFERENCES

Bassiou, N. and Kotropoulos, C., 2007. Color image histogram equalization by absolute discounting back-off. Computer Vision and Image Understanding, 107 (1-2), pp.108-122.

Drury, Stephen A., 2001. Image Interpretation in Geology (3rd). Routledge, New York, pp. 133-138.

Jensen, John R., 1996. Introductory Digital Image Processing : a Remote Sensing Perspective (2nd). Prentice-Hall, NJ, pp. 153-172.

Ji-Hee Han, Sejung Yang, Byung-Uk Lee, 2011. A Novel 3-D Color Histogram Equalization Method with Uniform 1-D Gray Scale Histogram. IEEE Trans. on Image Processing, 20 (2), pp. 506-512.

Laughlin, S.B., 1981. A simple coding procedure enhances a neuron's information capacity. Naturforsch, 9–10 (36), pp. 910–912.

Lillesand, Thomas M. and Kiefer, Ralph W., 1994. Remote Sensing and Image Interpretation (3rd). John Wiley and Sons, New York, pp. 585-618.

Naik, S. and Murthy, C., 2003. Hue-preserving color image enhancement without gamut problem. IEEE Trans. on Image Processing, 12 (12), pp. 1591–1598.

Trahanias, P.E. and Venetsanopoulos, A. N., 1992. Color image enhancement through 3-D histogram equalization. IAPR, 1, pp. 545–548.