Safety, Reliability and Risk of Structures, Infrastructures and Engineering Systems – Furuta, Frangopol & Shinozuka (eds)

Automated speed detection of moving vehicles from remote sensing images

Wen LIU & Fumio YAMAZAKI
Graduate School of Engineering, Chiba University, Chiba, Japan

Tuong Thuy VU
Royal Institute of Technology, Stockholm, Sweden

ABSTRACT: A new method is developed to detect the speed of moving vehicles from two consecutive digital aerial images and a pair of QuickBird (QB) panchromatic and multi-spectral images. Since two consecutive digital aerial images taken by UCD have an 80% overlap, the speed of moving vehicles in the overlap can be detected by the movement between the two images and the time lag (about 3 sec). Firstly, vehicles are extracted by an object-based method from each image. Then the vehicle matching is carried out to find the same vehicles from the two consecutive images and calculate their speeds. QB images also can be used to detect the speed of moving vehicles, using the time lag (about 0.2 sec) between the panchromatic and multi-spectral images. Since the resolution of QB's multi-spectral images is about 2.4 m/pixel, the area correlation method is introduced to detect the exact location of vehicle. The results can be extensively used to access road traffic.

1 INTRODUCTION

As the population in cities continually increases, road traffic becomes more congested than the level which city and infrastructure planning expects. As the first step to solve the problem, monitoring vehicles must be the important task. Normally, the field-based equipment, like cameras installed at fixed locations or weigh-in motion sensors on the pavements, are used to monitor road traffic. Presently remote sensing technique has been emerged as another option to collect the traffic information. Using this way, the wider range of information can be collected over a long time. Thus, vehicle detection by remote sensing can be extensively used to manage traffic, assess fuel demand, estimate automobile emission and also important for transportation infrastructure planning.

There have been several researches on vehicle detection using remote sensing data. They can be categorized into two groups: model-based extraction and data-based extraction. Model-based extraction is based on the vehicle models built from a learning process. Then the models are used to identify whether the target is a vehicle or not. For example, Gerhardinger et al. (2005) tested an automated vehicle extraction approach based on an inductive learning technique, which was implemented using *Features Analyst*, an add-in extension of ArcGIS software. Zhao & Nevatia (2001) combined the multiple features of vehicles in a Bayesians network for leaning prior to detecting vehicles.

In data-based extraction, the processing follows a bottom-up procedure to group pixels into objects and the vehicle objects are subsequently discriminated from the others. Hong & Zhang (2006) used an object-oriented image processing technique to extract vehicles. A detailed description, which requires a large number of models to cover all types of vehicles, is the key of the former approach. It takes time and cannot be widely applied. The latter is simpler and convenient to be widely used. Those recent researches mainly reported on the position of vehicles, and few of them went to speed detection as well as created a traffic information database.

It is known that the panchromatic (PAN) and multi-spectral (MS) sensors of QuikBird have a slight time lag, about 0.2 seconds, depending on the scanning mode of the instrument. Using this time lag between the two sensors of one scene, the speed of moving objects can be detected (Etaya et al. 2004, Xiong & Zhang 2006). The speed of vehicles can also be detected from aerial images. Aerial images are often taken along a flight line with an overlap among adjacent scenes. If a moving object captured in a scene is also captured in an adjacent image, the speed of the object can be detected.

In this research, a new method is developed for both vehicle extraction and speed detection. From high resolution aerial images, vehicles can be extracted by an object-based method. Then speed is detected by matching the results of vehicle extraction from two consecutive images. Since the resolution of QuikBird's MS image is not high enough to extract vehicles directly, an area correlation approach is introduced to search the best match position with the result of vehicle extraction from the PAN image. Then speed is detected by the distance of the vehicle's location in the PAN and MS images. The proposed approach is tested on both QuikBird and simulated QuikBird images.

2 MOVING OBJECTS IN QUICKBIRD IMAGE

Google Earth (http://earth.google.com/) recently provides high resolution optical images of urban areas, either from aerial images or pansharpened (PAN) QB images. For one scene, a PAN image can be produced by co-registering a PAN image and a MS image. But due to the slight time lag (about 0.2 sec) between a pair of PAN and MS images, the locations of moving objects displace after the short time interval. Even if the PAN and MS bands have been co-registered for still objects like roads and buildings, they cannot be co-registered for moving objects.

The time lag between the PAN and MS sensors of QB was investigated using bundle products of QB. Figure 1 shows the time lag for 36 scenes, which we have at hand. These images were taken for various parts of the world, e.g. Japan, USA, Peru, Thailand, Indonesia, Morocco, Iran, Turkey, Algeria, from March 2002 to July 2006. The time lag is either about 0.20 seconds or about 0.23 seconds.

Figure 2 shows a part of Ninoy Aquino International Airport, Metro Manila, Philippines, from Google Earth. Two airplanes are seen on the runway. The right plane is just at the moment of landing and the left one is standing still and waiting for take-off. A "ghost" is only seen in front of the moving airplane. Similar ghosts were observed in several airports in the world such as Narita/Tokyo International (Japan), Bangkok International (Thailand), and Hong Kong International. These ghosts were produced due to the time lag between the PAN and MS sensors of QB. The distance between the ghost and the airplane is about 18.1m in Figure 2. The speed of the airplane is evaluated as 326 km/h, assuming the time lag as 0.2 seconds.

This kind of ghosts are also seen in front of other moving objects, like trains, automobiles, ships, but due to limitation of the image resolution and the short time lag, ghosts are not so clear as those for airplanes. We simulated a higher resolution pansharpened image of an expressway with 0.25m resolution from an aerial image. By this resolution, the ghosts resulting from the time lag between PAN and MS sensors were clearly seen in front of moving vehicles.

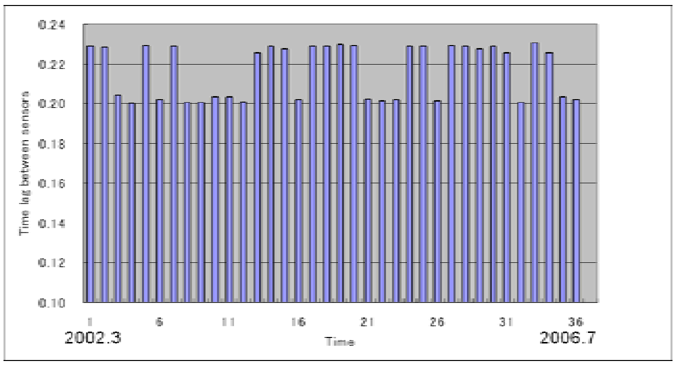


Figure 1. Time lag between the PAN and MS sensors of QB for 36 scenes in the world.

Figure 2. A ghost is generated in front of the just landing air-plane in a pansharpened QuickBird image from Google Earth.

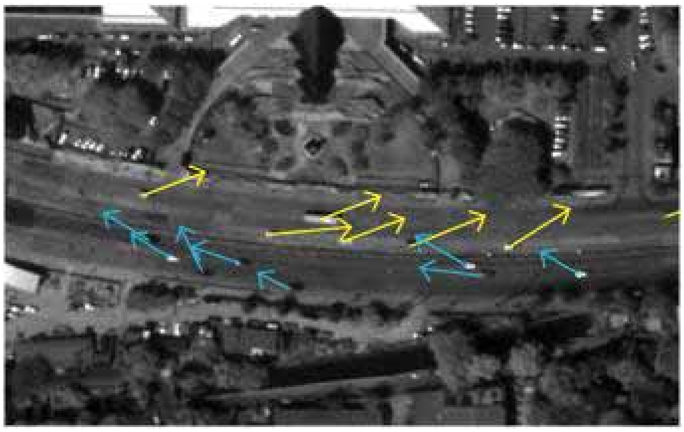


Figure 3. Result of visual detection of vehicle speed from the QB bundle product for central Bangkok, Thailand. The length of arrow represents the speed of vehicles.

Since the spatial resolution of a QuickBird multi-spectral image is 2.4m, rather coarse to figure out small cars, measuring the speed for smaller and slower objects is not so accurate. A part of QB image of central Bangkok, Thailand, was used to detect vehicle speed visually. Comparing the location of cars on the road in the PAN and MS images with 0.20s time lag, the speed and moving direction of the vehicles can be evaluated as arrows in Figure 3. In this investigation, we encountered some difficulty to locate vehicles in the MS image with 2.4m resolution. The result of visual detection also contains subjectivity and uncertainty. Thus, an automated detection method is sought.

3 OBJECT-BASED VEHICLE EXTRACTION

To detect the speed of a moving vehicle, the location of the vehicle in two images with a time lag is needed. From a QuickBird PAN image or an aerial image, a vehicle has a clear shape and can be easily extracted visually. Thus, we propose an automated object-based method to extract vehicles from high resolution remote sensing images and to record the information on a traffic database in this study. The procedure is tested using digital aerial images.

3.1 *Study area and data used*

The study area is located in Minato-ku, a central part of Tokyo, Japan. Two pairs of consecutive aerial images are used in this study. The images were taken by a digital aerial camera UltraCamD (Leberl & Gruberl 2003) on August 4, 2006, by Geographical Survey Institute of Japan.

The UltraCamD (UCD) offers simultaneous sensing of high-resolution panchromatic channel (pixel size is 9μm) as well as lower-resolution RGB and NIR channels (pixel size is 28μm). It has the ability to capture images with higher overlap, up to 90%, in the along track direction.

A panchromatic image has 7,500×11,500 pixels and a multi-spectral image 2,400×3,680 pixels. One image pair covers the area near Hamazakibashi Junction and another pair covers the area of Roppongi. Color images with resolution of about 0.12 m/pixel, obtained through a pansharpening process, were used in this study. Since the PAN image and MS image by UCD camera were taken at the same time, the “ghost” does not appear in the pansharpened image. Note that the two consecutive images have an overlap of about 80% (Figure 4).

3.2 *Methodology*

Since vehicles are moving on roads, road extraction should be the first processing step. Focusing on the extraction of vehicles and the detection of their speeds, we do not propose a new road extraction method. There have been a number of researches on road extraction from remote sensing images (Quackenbush 2004). Those can be easily employed to extract road objects here. Additionally, GIS road data can also be used to extract roads. However, to avoid errors involved in road extraction, which influences the final vehicle extraction results, the roads are extracted manually in this study. Then the areas out of the road areas are masked.

Figure 4. Two consecutive digital aerial images of Roppongi, Tokyo, with about 80% overlap.

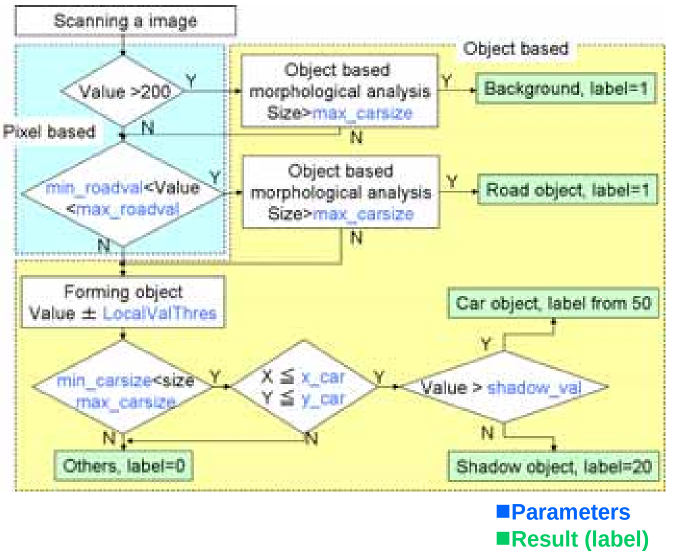


Figure 5. Flowchart of automated vehicle extraction

Prior to carrying out vehicle extraction, other irrelevant information such as lines on the road surface should be removed. Concerning the shapes and sizes of the objects, area morphological filtering was employed (Vu et al. 2005). This filter perfectly removes long and thin road lines and retains the shapes of vehicles. The window size used here was set as 5×5.

Since the vehicle extraction is based on the gray value, color images were converted to black-and-white images. The flowchart of the object-based vehicle extraction approach is shown in Figure 5.

Pixels were scanned and grouped into objects according to the criteria of the gray value. In this step, the image represents 4 kinds of objects: background, roads, vehicles (including their shadows) and the others treated as noise. The road extraction step assigns the background as black color. It can be easily discriminated by the lowest range of the gray value. Meanwhile, the road surfaces normally show another specific range of the gray value. Based on those two gray value ranges, the objects are formed. There might be vehicles which show very similar characteristics with the black background. Fortunately, the background and the road are often big objects compared to the others. Then, these two kinds of objects can easily be extracted based on a size threshold.

The remaining pixels are reformed into objects again based on a local threshold of the gray value. The fact is that all the pixels belonging to a vehicle should have a similar gray value. Vehicles and their associated shadows generally have a specific range of size. It is the criteria to distinguish them from the others. Consequently, the initially extracted result was obtained, and the information on the vehicle position and size was stored in a database. The parameters of the value range and the object size were examined several times till the best result was obtained.

3.3 *Experiment and result*

The target of extraction is the vehicles on the expressway in the two study areas. As the result, the vehicles with light color were presented in white and the shadows or dark vehicles were extracted as gray (Figure 6). Additionally, the information of vehicle positions and sizes was stored in a database for speed detection. Then the results were compared with visual extraction results (Table 1).

There were 292 vehicles in the pair images of Hamazakibashi area, and 282 vehicles were extracted correctly by the process of vehicle extraction. Only 10 vehicles were missed, and the noises which are not vehicles but extracted as vehicles were 116. The producer accuracy is 96.5%, and the user accuracy is 71%.

In the image pair of Roppongi, 195 vehicles and 191 vehicles, respectively, were extracted correctly. Four vehicles were missed, and noises were 42. Thus, the producer accuracy of these images is 98%, and the user accuracy is 82%.

Overall, almost all the vehicle could be extracted. Not only light-color vehicles but also dark vehicles and some vehicles in shadow were extracted successfully. Because we extracted both vehicles and shadows, even the vehicle's gray value is similar to that of the road, the vehicle can be extracted by its associated shadow. There still exist a few commission errors due to a signboard, its shadow, and some lines on the road. The environmental condition around the target area influences the result of vehicle extraction. Accuracy gets higher as an environment becomes simpler.

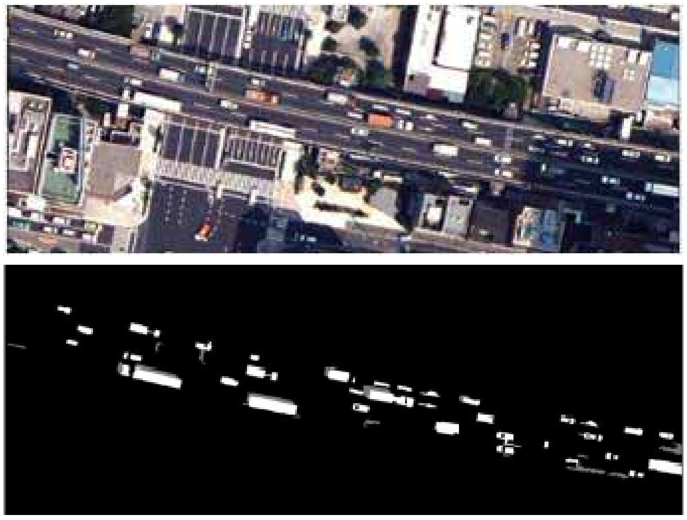


Figure 6. Original aerial image (up) and result of vehicle detection (down)

Table 1. Accuracy of object-based vehicle extraction

Image	Hamazakibashi	Roppongi
Vehicles in image	292	195
Extract result	398	233
Correctly extracted	282	191
Omission	10	4
Commission	116	42
Producer accuracy	97%	98%
User accuracy	71%	82%

4 SUB-PIXEL LEVEL VEHICLE EXTRACTION

From PAN images with 0.6m resolution, vehicles could be extracted by the object-based approach. But the resolution of MS images is about 2.4m, a vehicle appears in about only 1 or 2 pixels. Most vehicle pixels were mixed with roads, and it is difficult to extract the accurate edge and position of vehicles. The proposed object-based approach could not extract vehicles from MS images. To detect the speed of vehicles, the shift of the location in the PAN and MS images is needed. Thus, an area correlation method is proposed to estimate the location of vehicles from a MS image in a sub-pixel level.

4.1 *Methodology*

Area correlation is a method for designating Ground Control Point (GCP) in image-to-image registration (Schowengerdt 1997). A small window of pixels (template) from the image to be registered is statistically compared within a region of the reference image, which is bigger than the template image. From the distorted images, templates of M rows by N columns are selected. A bigger size window is selected for the reference image. The template is overlaid on the reference image and a

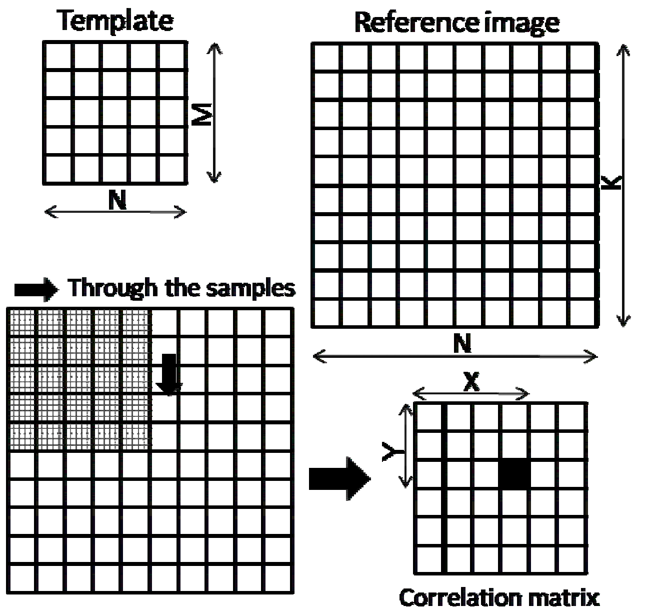


Figure 7. Shifting template overlaid reference image, and the correlation matrix is calculated by each shift.

similarity index is calculated. This procedure is carried out for every possible position in the reference image and the similarity matrix is obtained. The content of the similarity matrix is the value of the statistical comparison between the template and the reference image. The position in the similarity matrix where the similarity index is the maximum represents the necessary offset that the template has to move horizontally and vertically to match the reference image. This process is shown in Figure 7. Note that if there is a relative rotation between the template and the reference image, a rotational angle should be introduced for matching.

One of the similarity indexes is the cross-correlation coefficient between the template and the reference images (Eq. 1). The cross-correlation coefficient is a scalar quantity in the interval [-1.0, 1.0]. The cross-correlation coefficient can be used as a similarity index since it gives a measure of the degree of correspondence between the reference and the template or can be seen as a coefficient that is a direct measure of how well two sample populations vary jointly (Brown 1992).

$$r_{ij} = \frac{\sum_{m=1}^{N}\sum_{n=1}^{N}(T_{m,n} - \mu_T)(S_{i+m,j+n} - \mu_S)}{K_1 K_2} \qquad (1)$$

where

$$K_1 = \left[\sum_{m=1}^{N}\sum_{n=1}^{N}(T_{m,n} - \mu_T)\right]^{1/2}$$

$$K_2 = \left[\sum_{m=1}^{N}\sum_{n=1}^{N}(S_{i+m,j+n} - \mu_S)\right]^{1/2}$$

First, a vehicle is extracted from PAN image with 0.6m resolution by the object-based approach. A database is obtained after vehicle extraction with the information of vehicle location. Using the location information, a vehicle and the surrounding road is selected as a template. Since the time lag between PAN and MS images is 0.2s, the maximum moving distance is about 7m (when the maximum speed is 120km/h). The reference image is selected with the same center as the template but 7m bigger than that in the two directions. The cross-correlation coefficient between the two areas is calculated by sliding the template over the reference image, multiplying the two arrays pixel by pixel. The point of maximum correlation indicates the position of the vehicle in the MS image with the highest probability. To raise the accuracy of correlation, the template and the reference image are transformed to 0.24m/pixel by cubic convolution. Then the template and the reference image can be matched in a sub-pixel level.

4.2 *Experiment and result*

One QB scene with the PAN and MS bands covering the central Bangkok, Thailand was used to test the area correlation method for vehicle extraction.

Since the MS image has 4 bands as R, G, B and Near-Infrared, it needs to be transformed to one band image before the area correlation analysis. The Principal Component Analysis (PCA) was employed to transform a MS image to a new 4 band image. The first component image with the highest variance was used to calculate the cross-correlation coefficient with the PAN image.

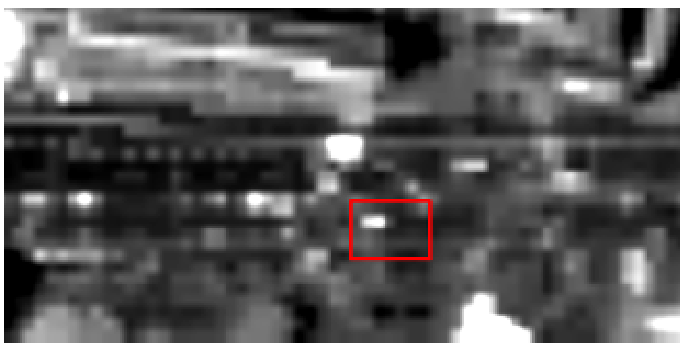
A vehicle in the MS image was mixed with the road. Thus, the template selected from the Pan image includes not only a vehicle but also road around it. Then a bigger reference image was selected around the location of the template from the MS image (Figure 8). The cross-correlation coefficient of each shift was calculated as a matrix, shown in Figure 9. The location of the maximum correlation (8, 14) is the upper left point of the template in the reference image.

From the PAN image, several vehicle templates were selected and they were statistically compared with the reference area extracted from the MS image. Comparing the results visually, the vehicle templates were accurately matched with the reference extracted from the MS image. But it is difficult to access the accuracy of sub-pixel level vehicle extraction only by visual comparison. Thus, simulation was carried out to verify the accuracy.

The digital aerial images used in the previous section, were employed to the simulation. Since the time lag between the two consecutive aerial images is about 3s, the shift of a vehicle is large. Thus, we

PAN image with 0.6m resolution

First component of the MS PCA image with 2.4m resolution



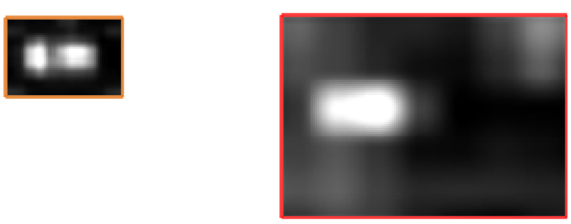


Figure 8. Template selected from the PAN image (orange) and reference image selected from the MS PCA image (red)

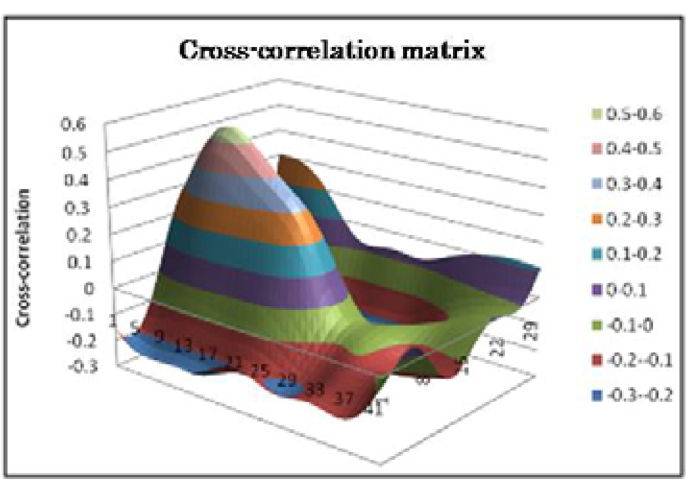


Figure 9. Cross-correlation matrix obtained by shifting the template over the reference image.

selected the templates of vehicles and the reference images manually. The reference image selected from the second image should include the target vehicle and the surrounding road. To compare with the QuickBird study, the pixel size of the original PAN images were also resized from 0.12m to 0.24m. The template from the first PAN image was overlapped on the reference image extracted from the second PAN image. The cross-correlation matrix obtained is

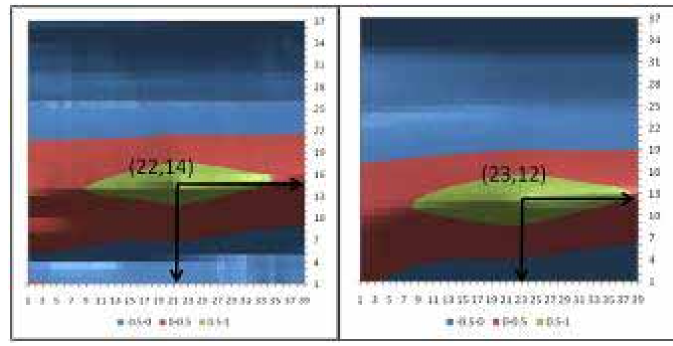


Figure 10. Comparison of the cross-correlation matrix for the 0.24m resolution image (left) and that for the simulated QB image (right).

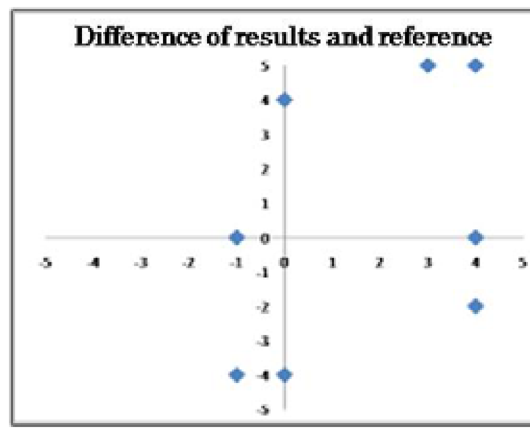


Figure 11. Difference between the extracted results and the reference data to the x-axis and y-axis

shown in Figure 10 (left). Since the digital aerial images are high spatial resolution, the result is considered to be accurate. The upper left point in the template is located at (22, 14) in the reference image.

Then the resolution of the PAN image from the first aerial image was converted to 0.6m/pixel, simulating a PAN image of QuickBird. The resolution of the MS image from the second aerial image was also converted to 2.4m/pixel, simulating a MS image of QuickBird. The first component of the MS PCA image was used to calculate the cross-correlation matrix with the simulated PAN image. To register the two images in a sub-pixel level, the pixel sizes of the two images were resized to 0.24m. The template and the reference image were selected from the simulated PAN and MS images, the same location with the original high-resolution data. The cross-correlation matrix was obtained by shifting the simulated PAN image over the MS image. The result is shown in Figure 10 (right), where the upper left point of the template locates at (23, 12) in the reference image. This location represent the most probable position of the vehicle object in the MS image.

Comparing the result with the original data, the standard deviation of the difference to the x-axis is about 2 pixels (0.48m), and that to the y-axis is about 3 pixels (0.72m), as shown in Figure 11. Since the width of vehicles is less than 2.4m, the difference in width is bigger due to a mixed-pixel effect. However, the area correlation method could still extract a vehicle from a MS image with 2.4m resolution in a sub-pixel level.

5 SPEED DETECTION

Speed detection uses the time lag between two images. Generally, it can be performed using two consecutive aerial images or one scene of QuickBird image. The proposed vehicle extraction approach can be extended to speed detection.

5.1 *Speed detection from aerial images*

The vehicle and shadow database of each image was developed after the object-based vehicle extraction process from the aerial images of Tokyo. Then, the vehicles in the two databases (two time instants) were linked by the order, moving direction, size and distance. If a vehicle in the second image is in the range of the possible distance from the one in the first image and if they have a similar size, they are linked as the same vehicle. Subsequently, using their positions stored in the databases, the speed can be computed.

To detect the speed of vehicles, two images, covering the same area with a time lag, are needed. Firstly, an overlap area from two consecutive images was extracted to obtain two images over the same area. Because of the perspective projection of an aerial camera, geometric distortions between two images exist. Hence registration for the pairs of images was conducted using 8 ground control points. After registration, the two images in a pair have different pixel sizes. Thus the images were arranged to the same pixel size by image mosaicing.

Visual speed detection was first carried out to obtain reference data by overlapping the second image to the first one. The speed can be detected by measuring the difference of vehicle's outline, as shown in Figure 12.

Then the speed and moving direction of vehicles were detected automatically by matching the databases for the two consecutive images using the parameters of order, direction, size and distance (Figure 13). About 71% of vehicles' speeds were detected automatically for the Roppongi area. The standard deviation for the difference of speed between the automated and visual detections is 0.83km/h, and the standard deviation for the difference of direction is 0.38 degrees.

For the Hamazakibashi area, only a part of images were used for speed detection since the accuracy of vehicle extraction was low (71%).

Figure 12. Visual speed detection by overlapping two images

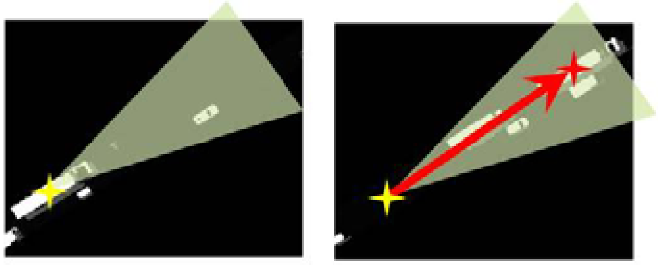


Figure 13. Condition of matching the same vehicle

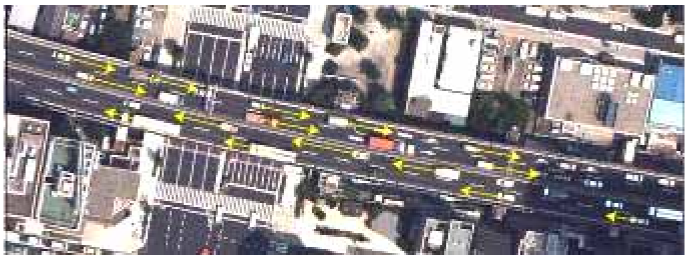


Figure 14. Result of automated speed detection from two consecutive aerial images of Roppongi.


Figure 15. Comparison of visual detection (yellow and blue arrows) and the sub-pixel based automated detection (red arrows) for the QB image of Bangkok.

Vehicle matching depends on the order of vehicles, and matching error occurs when many noises influencing the order of vehicles exist. From the part of the images, 64% of vehicle speeds were extracted.

The standard deviation for the difference of speed between the automated and visual detections is 1.01km/h, and the standard deviation for the difference of direction is 0.59 degrees.

Since the rules for vehicle matching are very severe, not all the vehicles could be matched from the image pairs. The order changes by noise, and the size changes in vehicle extraction are the main reasons for matching error. But the result of speed detection for the matched vehicles showed high accuracy.

5.2 *Speed detection from QuickBird*

The location of vehicles in a QB's PAN image can be extracted by the object-based method. By shifting the template of a vehicle extracted from the PAN image over the MS image of the same scene, the most possible location of a vehicle in the MS image can be obtained. Then speed of vehicles is computed by the location change between the PAN and MS images with the time lag about 0.2s.

From the QuickBird image of central Bangkok, vehicles were extracted and their speeds were calculated, as shown in Figure 15. Comparing with the result of visual detection, the moving direction of vehicles looks better than the visual detection result, but still has some errors to the transverse direction of the road due to the mix-pixel effect.

6 CONCLUSIONS

Methods to extract moving vehicles and measure their speeds from high-resolution satellite images and aerial images were proposed. First, an object-based approach was employed to extract vehicles on an expressway automatically from high-resolution remote sensing images, such as by digital aerial cameras. The method was applied to two consecutive aerial images of central Tokyo. Comparing the location of extracted corresponding vehicles in the image pair, the speed and azimuth direction of moving vehicles were obtained with high accuracy.

From Google Earth, "ghosts" of moving objects in pansharpened QuickBird images were demonstrated. The slight time lag, about 0.2s, between panchromatic and multispectral sensors of QuickBird was shown to be responsible for the ghost and they can be used to measure the speeds of moving objects using only one scene of QuickBird's "bundle product". Due to limitation of the short time lag and the resolution (2.4m for MS bands), high accuracy cannot be expected by visual inspection.

An area correlation method to detect the accurate vehicles' location from 2.4m MS image in a sub-pixel level was also proposed. A template including vehicle was selected from a PAN image, and a reference image was selected from a MS image. From the cross-correlation matrix, the position of the maximum correlation could be obtained. The test result showed that vehicles were detected with a sub-pixel level accuracy (1/3 pixel of the MS image).

The accuracy of vehicle extraction and speed detection from QuickBird will be improved by introducing a rotation between PAN and MS images to the area correlation method.

The result of this study is useful for better understanding the traffic dynamics. Images of a large road network can, for instance, be used to acquire information from a whole region at one time. Such a snapshot of the entire network can give more insights into the distribution of vehicles in a region, and can also provide valuable information for areas not covered by traditional traffic counters.

ACKNOWLEDGEMENT

The digital aerial images used in this study were provided from Geographical Survey Institute of Japan.

REFERENCES

Brown, L. G., 1992. A survey of image registration techniques. *ACM Computing Survey*, Vol. 24, No. 4, pp.325-376.

Etaya, M., Sakata, T., Shimoda, H., & Matsumae, Y. (2004). An experiment on detecting moving objects using a single scene of QuickBird data, *Journal of the Remote Sensing Society of Japan*, Vol. 24, No. 4, pp. 357-366.

Gerhardinger, A., D. Ehrlich, & M. Peseresi, 2005. Vehicles detection from very high resolution satellite imagery. *CMRT05. IAPRS*, Vol. XXXVI, Part 3/W24.

Hong, G., Zhang, Y. & Lavigne, D.A., 2006. Comparison of car extraction techniques for high resolution airborne images. *First Workshop of the EARSeL Special Interest Group on Urban Remote Sensing*.

Leberl, F., & Gruber1, M., 2003. Economical large format aerial digital camera, *GIM International*.

Schowengerdt, R. A., 1997. *Remote Sensing, Models and Methods for Image Processing, Second Edition.* Arizona: Academic Press.

Quackenbush, L. J., 2004. A review of techniques for extracting linear features from imagery. *Photogrammetric Engineering & Remote Sensing* Vol. 70, No. 12, pp. 1383-1392.

Vu, T.T., M.Matsuoka & F. Yamazaki, 2005. Preliminary results in development of an object-based image analysis method for earthquake damage assessment. *Proc. of 3rd International workshop Remote Sensing for Post-Disaster Response, Chiba, Japan*.

Xiong, Z., & Zhang, Y., 2006. An initial study of moving target detection based on a single set of high spatial resolution satellite imagery, *Proc. ASPRS 2006 Annual Conference*, Reno, Nevada, May 1-5, 2006.

Zhao, T., & Nevatia, R., 2001. Car detection in low resolution aerial image. *International Conference on Computer Vision*.