

VEHICLES DETECTION FROM VERY HIGH RESOLUTION SATELLITE IMAGERY

A. Gerhardinger^{a*}, D. Ehrlich^a, M. Pesaresi^a

^a IPSC, Joint Research Centre, 21020 Ispra, Italy - andrea.gerhardinger@jrc.it

KEY WORDS: Remote Sensing, Spatial Information Science, Urban, Calibration, Detection, Automation, High resolution

ABSTRACT:

This paper evaluates an inductive learning approach classification technique for vehicle detection and enumeration on very high resolution imagery. It tests pre-processing procedures applied to different images with different atmospheric conditions and automatic detection algorithms for detection and enumeration. This work contributes to the longer term objective that is beyond the scope of this paper to use vehicles counts to derive indicators of societal activity; this for a number of applications including situation assessments in conflict areas. The work uses Ikonos and Quickbird images collected over Baghdad in pre-war situation and during the Iraq conflict of 2003. Very high resolution imagery is an excellent information source to detect vehicles. The enumeration can be carried out using photo-interpretation techniques. However, this is impractical and relatively expensive. Automatic detection is possible. Image pre-processing is needed.

1. INTRODUCTION

The fleet of vehicles in cities continues to increase and monitoring vehicles fleet is an emerging necessity. A number of cities especially in developed countries use field based equipment such as cameras installed at fixed locations or weigh-in motion sensors on the pavements to monitor traffic. Recently video imaging and airborne imaging is tested especially to provide more synoptic view of traffic and to monitor traffic trajectories (Zhang et al. 2003). Aerial photography has also been evaluated (Hinz 2003, Schlosser et al. 2003) and the newly available imagery at very high resolution provides a new opportunity and starts also to be considered (Rocio Alba Flores 2004). The comparison of optical, infrared and SAR data (Stilla et al. 2004) shows obviously they are both advantages and limitations. For example SAR sensor is independent from day/night and weather condition but its oblique view causes distortion, the presence of buildings hides some roads and its analysis is very much affected by the experience of the interpreters (Lohmann et al. 2004).

In the developing countries vehicle fleet is increasingly very rapidly and hardly measurable since field based equipment is rarely in place. Vehicles enumeration is required for a number of reasons including management of traffic, fuel demand assessments, traffic emissions estimates in crowded metropolitan areas prone to accumulate air pollutants. Vehicle enumeration over time is also required to estimate future vehicle fleets and thus important for planning transport infrastructure. In crisis affected areas vehicle enumeration at different moments in time may provide insight on the severity of a crisis and/or how an area is affected and is recovering from a disaster.

Satellite imagery provides complementary information to ground measurements. Satellite imagery is a snapshot in time that it covers a relatively vast area. The measurement is robust and repeatable over time even though can not be considered for

continuous monitoring. For many cities in the developing world satellite imagery may be the only information that can be obtained with a given amount of time and budget. Satellite imagery is also not intrusive and can be used to enumerate vehicles or estimate vehicles fleets in countries for which there is not established ground based traffic monitoring systems or that have experienced a disaster that precludes access to area or measurement.

Vehicle detection and enumeration from satellite imagery requires pre-processing. Very high resolution satellite sensors provide imagery with different viewing angles, different atmospheric conditions that in urban areas can be affected by pollution and illumination conditions. In addition, when multi-temporal imagery is analyzed scene spectral characteristics may also have changed.

The aim of this work is to (1) assess the potential of very high resolution satellite imagery for vehicle detection and enumeration from two satellite imagery – Ikonos and QuickBird and compare the two data sources, (2) develop procedures of image pre-processing that can increase the performance of the classifiers, (3) test automated feature extraction procedures over two satellite images of Baghdad. This work is part of a study aiming to detect and enumerate vehicle fleets over large metropolitan areas and compare them through time to derive a societal activity index.

2. DATA

The study site is the city of Baghdad. The 8 municipalities of the city extend over 719 km² (Figure 1). Three images were available over the area: Ikonos of 2000 and 2003 and Quickbird of 2003 covering respectively 637 km², 330 km² and 417 km². The classification procedures were tested on some sample areas inside a subset of the images.

The two Ikonos images were acquired on 31 May 2000 and 01 April 2003 at respectively 1 x 1 m in the panchromatic and 4 x

* Corresponding author.

4 m for the multi-spectral channels and the QuickBird image was acquired at 01 April 2003 with 0.68 x 0.68m resolution in panchromatic and 2.56 x 2.56m in the multi-spectral channels. Ancillary data included 1:25,000 and 1:100,000 Military Topographic Russian maps, and a NIMA city map at 1:25 000 scale used to provide context to the study.

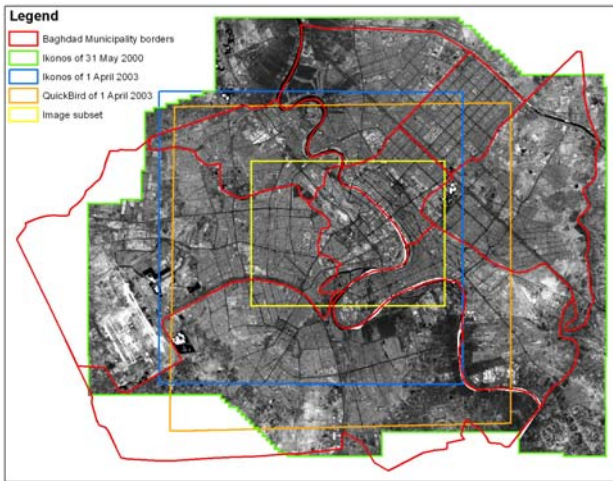


Figure 1. Outline of Baghdad city border with the footprint of the Ikonos and Quickbird imagery

Two digital vector data layers, the administrative borders of the municipalities and road network, were also available and provided by the Humanitarian Information Centre for Iraq coordinated by the Office for the Coordination of Humanitarian Affairs (OCHA).

Only the panchromatic bands of the respective Ikonos and Quickbird imagery were considered.

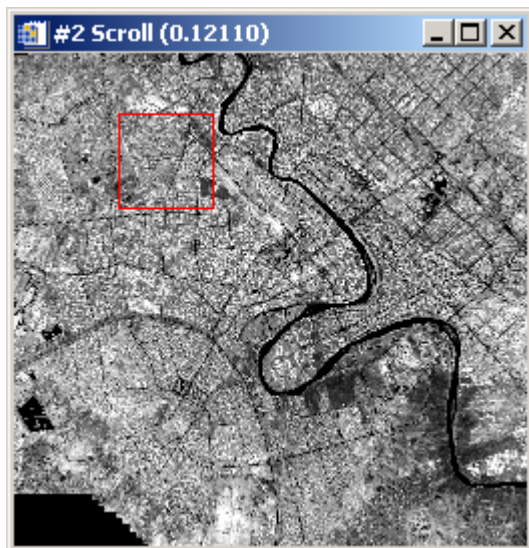


Figure 2. Ikonos image (31 May 2000)

The 1 x 1m Ikonos images were re-sampled at the higher resolution of QuickBird (0.68 x 0.68m). The two 2003 images are characterized by very peculiar atmospheric conditions with large portion of the city covered by smoke. In fact open air fires where ignited during the conflict to reduce visibility and reduce the ability to hit targets during conflict. Despite the smoke the

surfaces completely obscured by the plumes are a very small fraction of the images.

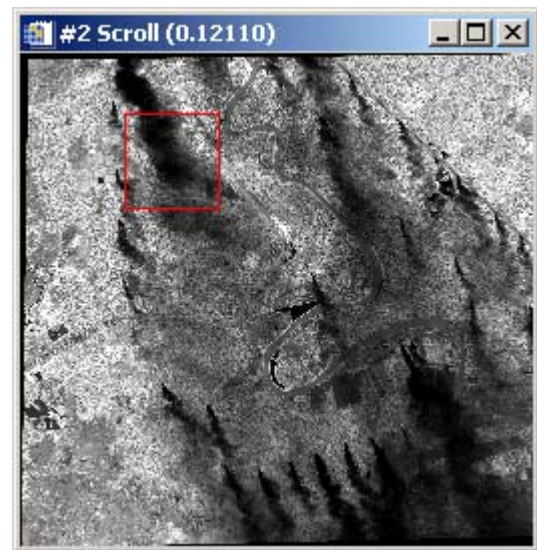


Figure 3. QuickBird image (01 April 2003) shows the smoke plumes generated from the oils on fire in the city

The analysis of the multi-temporal set of imagery required pre-processing that includes rectification and radiometric enhancement. Rectification was also performed on the road network digital layer. The road network was further processed to produce a digital layer of the road surfaces. The algorithms for vehicle detection and enumeration were applied within the portions of the images corresponding to main roads of the city only.

3. METHODOLOGY

3.1 Geometric rectification

In order to obtain homogeneous and comparable images, for the estimation of the vehicles on the entire city, the three data bases were rectified. The rectification was made using the Ikonos image of 2000 as reference. About 200 ground control points for each image of 2003 were retained. The Ikonos image of 31 May 2000 was clipped and re-sampled on the base of the QuickBird image of 03 April 2003 with 0.68 meters resolution. The road information layer was also produced accordingly.

3.2 Radiometric enhancement

Smoke plume in the 2003 images reduce the ability to detect ground features. However, the detection of image structures is still possible by selectively modifying the dynamic range of the image as shown in figure 4. The challenge for the automatic feature extraction procedure is to apply a correction factor in order to obtain a radiometrically homogeneous image for both the regular city and the smoke affected image.

A continuous local stretch was considered the most appropriate technique for radiometric enhancement. The procedure is based on the calculation of the mean and variation value of the cells in an appropriate neighbourhood. The procedure was applied on a re-sampled image for computational efficiency purposes. First

we re-sampled the original panchromatic data to 10 x 10 meters pixel size using a pixel aggregate algorithm (the output value is calculated by averaging the input pixel values).

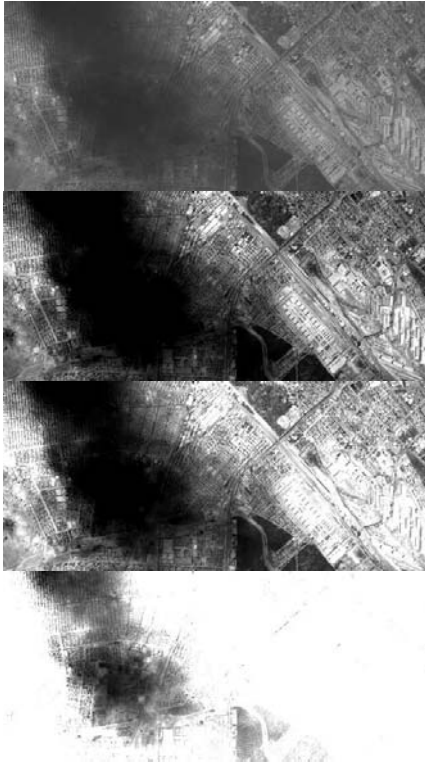


Figure 4: Radiometric enhancement. Original image, enhancement based on bright, mid range and dark areas (from top to bottom).

Then we calculated the mean and standard deviation by the neighbourhood focal statistics over the re-sampled image. The running window is therefore of 100 meters in size.

These derived data were re-sampled back at the 0.68 meter resolution and used in (1) in order to apply the stretch at the original image.

$$y = \frac{x - \text{mean}}{\text{std}} * \text{scale} \quad (1)$$



Figure 5: Image radiometrically adjusted using the continuous local stretch



Figure 6: Subset image covered by the smoke plume before (left) and after (right) the radiometric enhancement.

3.3 Geo processing

To reduce the data to be processed and increase the efficiency of the classification we created a polygon corresponding to the part of the city covered by roads (road mask). Baghdad is a well planned city with a well structured transport infrastructure including a lot of expressways. A first approach was based on the creation of a 12 meters standard buffer using the road centre line network available as a vector layer. The 12 meter buffer covered a good part of the roads but was not considered adequate use on all roads having different size or number of lanes and including squares and parking lots.

We create a new polygon layer digitalizing the road surfaces directly using the Ikonos image as reference. The coverage of this layer coincides with the extracted sample areas (Figure 8).

The classification test was applied just on the effectively surface of the roads.

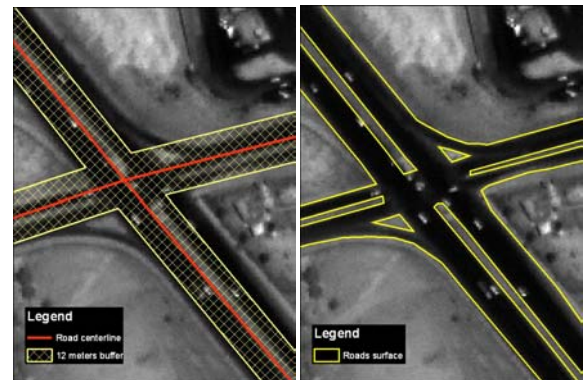


Figure 7: Road center line and derived buffer (left) and digitalized road surfaces (right).

3.4 Image analysis

The images were analyzed through visual inspection and using automatic routines. The analysis was conducted on six 1 km² samples (Figure 8). The visual inspection and enumeration, within the six samples, was encoded as a separate digital layer that was used as the ground truth. We produced a vector layer for each images.

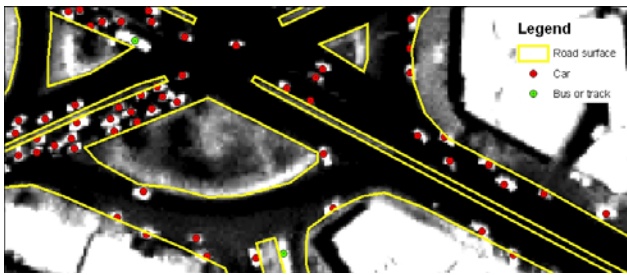


Figure 8: Vector layer of vehicles enumerated by image interpretation.

The six samples were randomly selected from a regular 1 km grid and are localized in several Municipalities of Baghdad. The samples cover areas characterized by different kind of roads (with different number of lanes, localization, density) and, regarding the image of 2003, by presence of smoke plumes.

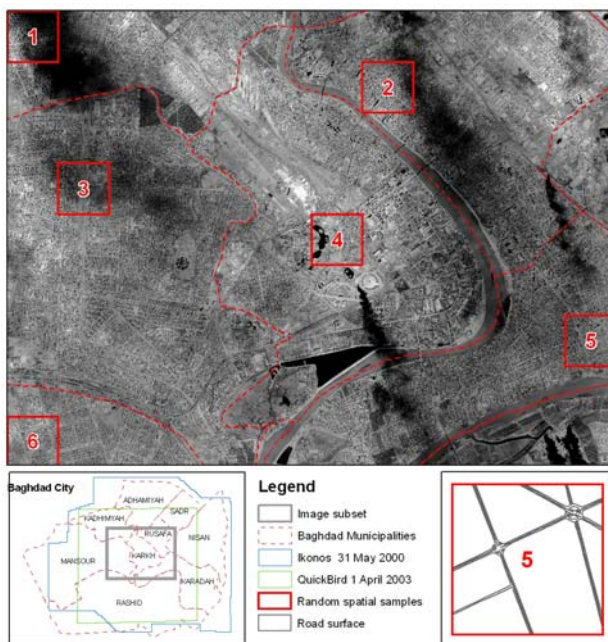


Figure 9: Samples extracted for the test analysis over the Quickbird image with smoke plumes

The main pixel based image classification techniques, also if use pattern recognition or spatial analysis processes beyond the spectral signature analysis, result ineffective for some features of interest as small object, in our case the vehicles.

In order to solve this limitation, we tested an automated classification approach joining inductive learning techniques (Maloof et al. 1998) at the traditional methodologies.

Inductive learning technique is based on process of learning by examples that define general rules from a set of observed instances. This involves classification assigning a particular input at the class to which it belongs. A learning system is able to evolve its own class descriptions also if the initial model definition may not be adequate because it is not well understood or it changes. The task of constructing class definitions is called induction or concept learning.

The inductive learning technique, by an interactive and sequential multi-step process, improves the model of the feature to be extracted and enhances the accuracy and the robustness of the system.

Differently from the traditional supervised classification methodologies, based mainly on statistical methods, the inductive learning doesn't require a statistical model (not Gaussian) or extraction of pure spectral signature resolving the problems about the classification of small objects and edges. The inductive learning is applied iteratively and accepts also negative input (correct and not correct examples or training samples)

In the present test the main input information used in the classification process are radiometry (images filtered/not filtered), size (area, width), position (road surface mask) and spatial (convolution neighbouring, pattern). The algorithm was calibrated for small objects, ($5 < \text{pixels} < 20$) localized within the road mask and without a predefined orientation (free rotation). The pixel pattern used for these objects is described in the following figure where the white squares represent ignored pixels while the dark squares are those used in the classification. A single vehicle is represented by the central pixels meanwhile the pixels at the corners overlap other contiguous vehicles or other objects at the border of the road

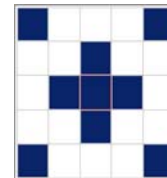


Figure 10: Pixel pattern representation for small object extraction.

The process starts with the creation of samples adequate to define the features to be detected. Then the learning algorithm, using this information, implements a model that describes the targeted object. All available data, as spectral, spatial, temporal and ancillary included vector layers, help to develop the model. The following step is then the automatically extraction by the image of the remaining features that coincide with the learned model. The detection analysis was implemented in *Features Analyst*® for Arc/Gis.

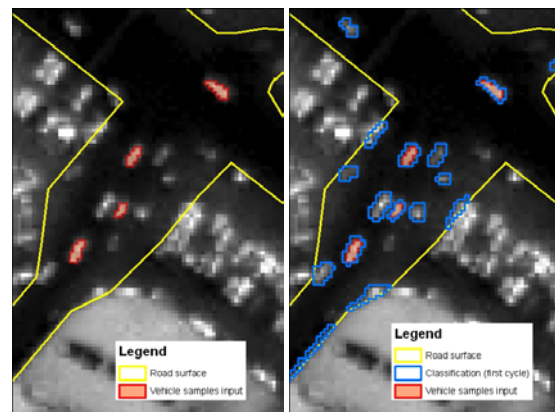


Figure 11: Iterative process; samples input (left) and first classification (right).

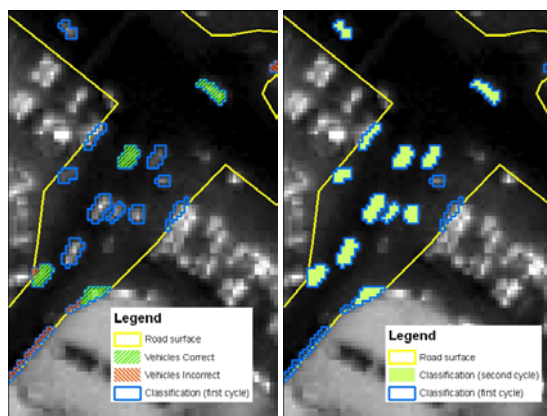


Figure 12: Iterative process; learning step (left) and second classification (right)

The classification was applied just on the selected area defined by the samples and by the road mask as above described, and for each image: Ikonos of 31 May 2000 and at two different version of the Quickbird of 1 April 2003 (the original, with smoke, and the filtered).

4. RESULTS AND DISCUSSION

The Ikonos and Quickbird imagery collected over Baghdad are adequate for the detection and enumeration of vehicles. Limitations are due to the atmospheric effects present in the Quickbird collected in 2003. For the darkest part of the plumes the detection is more difficult. Otherwise vehicles visual detection and enumeration is therefore possible but not practical if to be carried out on a large area such a city.

The objective of the geo-processing was to derive a thematic layer related to roads. This needs to improve the efficiency of processing and reducing influence of other parameters in the classification process.

A first thematic layer of the road derived by buffering the road centerline was substituted by a more precise layer created by digitalization. In fact also using a selectively buffering, based on the kind and size of each road, the precision of this layer is not enough satisfactory because the available network road vector data doesn't match with the images.

The image enhancement step appeared necessary particularly for the image collected in 2003 resulting degraded by the presence of smoke. These atmospheric effects inevitably reduce the ability to automatically detect vehicles but the application of simple stretch doesn't produce homogeneous results.

The radiometric enhancement technique based on filtering referred as continuous local stretch provides an improvement. The local stretch is a compromise providing results comparable with those obtained with not degraded images. This improvement is demonstrated analyzing the results concerning zones with and without smoke plumes as listed in the following tables.

The initial aim of the test was the detection of vehicles by the two images of Baghdad, first and during the conflict, and their comparison, but the presence of degradation caused by smoke on the second image gives us the possibility to compare it with

its enhanced version too.

The test has regarded three images: Ikonos of 31 May 2000 and original and filtered Quickbird of 1st April 2003. The availability of an Ikonos image of 1st April 2003 will be used in future to analyze difference with Quickbird of the same date.

The evaluation of the procedures was simply carried out comparing the classified objects with those detected by image interpretation and digitalized ground truth. The intersection of the respective layers permitted to select the vehicles matching with the ground truth and to individuate those correct, incorrect and not classified.

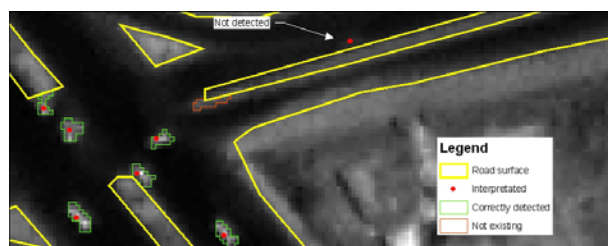


Figure 13: Classification check by intersection with ground truth.

The best results are obtained with Ikonos of 2000 and Quickbird of 2003 filtered (Table 1). Vehicles correctly detected are the 95% of those counted but there is a great difference between the two images. For Ikonos of 2000 the vehicles not detected are more than those detected but not existing instead for Quickbird filtered are the opposite. This result can be explained by the fact the dark vehicles are less visible on the Ikonos image of 2000 because the roads appear homogeneously black, instead on the Quickbird image of 2003 there are more light noise, mistook for vehicles, particularly along the center line and the border of the roads.

	Ikonos 2000	Quickbird 2300 original	Quickbird 2300 filtered
Digitalized	1831	1130	1130
Classified	1772	1037	1163
Correctly detected	1735	939	1082
Not detected	96	191	48
Not existing	37	98	81

Table 1: Enumerated vehicles summary

Sample No.	1	2	3	4	5	6
Digitalized	102	918	201	102	430	78
Classified	99	878	199	98	418	80
Correctly detected	97	857	195	96	414	76
Not detected	5	61	6	6	16	2
Not existing	2	21	4	2	4	4

Table 2: Enumerated vehicles summarized per sample – Ikonos of 2000

Sample No.	1	2	3	4	5	6
Digitalized	108	462	188	49	222	101
Classified	63	398	180	61	226	109

Correctly detected	52	360	177	48	203	99
Not detected	56	102	11	1	19	2
Not existing	11	38	3	13	3	10

Table 3: Enumerated vehicles summarized per sample – Quickbird not filtered

Sample No.	1	2	3	4	5	6
Digitalized	108	462	188	49	222	101
Classified	100	494	185	174	14	11
Correctly detected	87	459	174	46	216	100
Not detected	21	3	14	3	6	1
Not existing	13	35	11	3	6	13

Table 4: Enumerated vehicles summarized per sample – Quickbird filtered

In particular the samples 1 and 2 are the most interested by smoke on the Quickbird image not corrected (Figure 8) and they present a higher percentage of not detected vehicles (respectively 56 on 108 and 102 on 462).

Obviously the quality and condition of the images and the precision of the vector layers have a great influence on the results. Several problems are given for example by the different acquisition angle of the image (slant buildings, shadows, etc), the condition of the roads (kind of surface, lack of homogeneity, presence of dust or sand, etc), type of vehicles (colour respect the road surface, size and length, etc).

5. CONCLUSIONS

This work was centered on the analysis of very high resolution images for vehicle detection and enumeration. It is shown that Quick Bird and Ikonos imagery can provide excellent information on traffic in cities. Vehicles can be detected through visual inspection. However manual extraction and enumeration over large metropolitan areas is impractical and we show that automatic classification can also provide operational results.

We applied procedures based on the automatic classification of objects, using an iterative learning approach based on spectral signature and spatial context analysis. The analysis was conducted on two Ikonos acquired on (31 May 2000 and 1 April 2003), and one Quickbird acquired also on 1st April 2003. We tested pre-processing procedures in order to improve the Quickbird image collected over Baghdad during the Iraq conflict of 2003 that was degraded by presence of smoke plumes. The pre-processing procedure showed to be very effective and provides greatly improved classification results if compared with the non pre-processed image. This especially over the part of the imagery that is mostly contaminated by the smoke plume. In fact, the classification of the filtered Quickbird image of 2003 resulted qualitative comparable with that of Ikonos of 2000 that is not affected by smoke plumes.

The classification procedure based on inductive learning technique and on multi cycles sequential process appears accurate and robust. In fact, it works efficiently also on the smoke contaminated Quickbird image that even if corrected is of poorer quality than Ikonos. The information layers of vehicle

detected can be used for analyzing the variation in vehicular traffic, correlating it at the different time of acquisition of the data.

The good results of the classification are however dependent on precise vector data describing the road surfaces. The information layer corresponding to the road surfaces was derived manually by digitizing on screen. The digitizing work was carried out on sample zones only. Should the work have to be conducted on the entire city automatic procedures for extraction of road surface should also be developed.

This work sets the base for the enumeration of vehicles from very high resolution imagery. The use of multi-temporal images permits to analyze vehicular traffic dynamics. The number of vehicles in the time could be correlated at one precise moment or at other specific situations and could be used for traffic modeling and simulation (Toth et al.2003).

5.1 References

- Angel, A., Hickman, M., et al, 2003. Methods of Analyzing Traffic Imagery Collected from Aerial Platforms. IEEE Transactions on Intelligent Transportation Systems, Vol. 4, No.2, pp. 99-107.
- Hinz, S., 2003. Detection and Counting of Cars in Aerial Images. *Proceedings of the International Conference on Image Processing*, Barcelona, Spain, Vol. 3, pp. III-997-1000.
- Maloof, M.A., Langley, P., Sage, S., and Binford, T.O., 1998. "Learning to Detect Rooftops in Aerial Images." *Image Understanding Workshop*, Monterey, CA, pp. 835-845.
- Rocio Alba Flores, 2004. Evaluation of the use of high-resolution satellite imagery in transportation applications. Final Report, Department of Electrical and Computer Engineering, University of Minnesota Duluth 271 MWAH, 10 University Dr., Duluth, MN 55812.
- Schlosser, C., Reitberger, J., and Hinz, S., 2003. Automatic Car Detection in High Resolution Urban Scenes Based on an Adaptive 3D-Model. *Proceedings of the IEEE/ISPRS joint Workshop "Remote Sensing and Data Fusion over Urban Areas"*, Berlin, IEEE, Piscataway, pp. 167 - 171.
- Zhang, C., Chen, S-C., et al., 2003. Adaptive Background Learning for Vehicle Detection and Spatio-Temporal Tracking. *Proceedings of the Fourth IEEE Pacific-Rim Conference On Multimedia*, December 15-18, Singapore, pp. 1-5.
- Stilla, U., Michaelsen, E., et al, 2004. Airborne monitoring of vehicle activity in urban areas. *Proceedings of the XXth ISPRS Congress*, July 12-23, Istanbul, pp. 973
- Lohmann, P., et al. 2004. Comparative information extraction from SAR and optical imagery. *Proceedings of the XXth ISPRS Congress*, July 12-23, Istanbul, pp. 535
- Toth, C.K., et al. 2003. Supporting traffic flow management with high-definition imagery, *Joint ISPRS/EARSel Workshop "High Resolution Mapping from Space"*, October 6-8, Hannover
- Features Analyst* ® is an external module for Arc/Gis of Visual Learning System Inc. <http://www.vls-inc.com/index.html> (accessed 14 Jan. 2004)