ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Indian Journal of Science and Technology, *Vol 9(24), DOI: 10.17485/ijst/2016/v9i24/96057, June 2016*

Vehicle Detection using Images taken by Low-Altitude Unmanned Aerial Vehicles (UAVs)

Chang-Jin Seo*

Department of National Defense Intelligence Engineering, SangMyung University, Korea; cjseo@smu.ac.kr

Abstract

Background/Objectives: This paper discusses the detection of vehicles in a dense area using low-altitude aerial images produced by unmanned aerial vehicles (UAVs). Vehicles in a dense area are difficult to detect accurately in images because of the narrow distance between parked vehicles. **Methods/Statistical Analysis**: This paper proposes a method that detects vehicles by applying a Histogram of Oriented Gradients (HOG) feature-extraction method to obtain information about vehicles found in images. Images used in the experiment were shot using a Phantom 3 Professional UAV developed by DJI Corp. **Findings:** Aerial images can be collected to measure traffic via a variety of methods. In recent years, studies on traffic volume using UAVs have been actively conducted. Until now, satellites, aircrafts, and helicopters have been used to obtain aerial images; however, the cost of such images is high, and these methods cannot respond to changes over time and weather in real time. As UAV technology has advanced in recent years, the methods of obtaining cost-effective and ultrahigh definition (UHD) aerial images have become available. The experimental results show the proposed approach can be used to detect vehicles that are densely packed in an area more effectively than the method using the conventional histogram of oriented gradients (HOG), which employs information about the brightness of images. **Application/Improvements:** In this paper, the detection of vehicles found in a densely packed area is performed using the HOG, which exhibits good adaptability to environmental changes.

Keywords: Feature Extraction, HOG, ITS, Unmanned Aerial Vehicle, Vehicle Detection

1. Introduction

Intelligent transport systems are next-generation transport systems that employ intelligent technologies in electronics, information, communication, control, etc. to improve their efficiency and stability. Intelligent transport systems control traffic flow in terms of improving the road environment by analyzing the road environment in real time[1,2]. Various studies are underway to establish these intelligent transport systems that should be preceded by the collection of transport data to enable the identification of road conditions. The current method of using fixed loop detectors and wireless sensors installed on roads has the disadvantage that it can only detect information about a limited area[3]. In recent years, the development of Unmanned Aerial Vehicles (UAVs) and Ultrahigh Definition (UHD) cameras has made it possible to obtain aerial images at a low cost, and studies are being conducted on the measurement of traffic volume using them. A review of the existing studies shows that areal images have been taken using tools such as satellites, aircrafts, and helicopters, but they have the disadvantage of high costs and sensitivity to time and weather changes[4]. In recent years, studies have been actively conducted on the detection and tracking of objects on the ground using low-altitude UAVs to resolve this problem. Of particular interest are a number of studies investigating the detection and location tracking of vehicles using UAVs in the field of intelligent transport systems[5]. The resolution of cameras installed in UAVs has recently improved from Full High Definition (FHD) to UHD, which has enabled the measurement of traffic volume in a large area using composition techniques[6]. While this technical advance has the advantage of information processing for a wide

Author for correspondence

area using high-resolution images, it also has limitations, such as the reduced ability for real-time analysis due to large-scale datasets. As this shows, aerial images require a lot of computing time for arithmetic operations to detect objects due to the use of large-scale datasets. Therefore, the performance of object detection algorithms significantly affects the performance of detectors and trackers. While the adoption of a complex algorithm to detect objects on images improves detection performance, it reduces the ability for real-time analysis. However, the adoption of a simple algorithm improves the ability for real-time processing but reduces detection performance[7]. This study suggests a method of detecting the final location of vehicles by selecting the candidate locations of vehicles on images by employing the pre-processing technique to reduce the time for object detection and then applying the Histogram of Oriented Gradients (HOG) to the selected locations.

This paper proceeds as follows. Section 2 discusses studies related to object detection. Section 3 discusses the research methods. Section 4 consists of the experimental findings, and Section 5 consists of the conclusion and suggestions for future research.

2. Relevant Studies

A review of the recent literature on the detection and tracking of objects using aerial images obtained by UAVs shows that various studies are underway to find a variety of objects—such as humans, cars, bridges, power lines, aircrafts, and military targets—that appear in aerial images[8-10]. This approach affects the performance of detectors depending on the method of extracting and interpreting feature information about various objects present in such images. Target detection using images is the technology that detects targets by extracting, analyzing, and processing various pieces of feature information that exist in images through the analysis of image information. The detection of targets on images is the technique of interpreting information about targets projected into an image plane. Different methods of extracting feature information result in differences in performance. The methods of extracting, analyzing, classifying, and interpreting feature information about targets are largely divided into single-image feature analysis and multi-image analysis techniques. In terms of detailed methodologies, various studies are underway exhibiting various features depending on their respective analysis methods. The Scale Invariant Feature Transform (SIFT) method of analyzing single images suggested by Lowe is applied to a number of application fields[11]. As shown in Figure 1, this method selects feature points that are easily identifiable on an image and then extracts a feature vector related to the local patch for each feature point. The SIFT is characterized by its strength against changes in size, shape, and rotation, and numerous feature points can be matched for a single feature point. Therefore, the SIFT's overall performance can vary depending on the type of matching method selected rather than on the performance of the SIFT itself. The HOG suggested by Dalal is another technique applied to various application fields. As shown in Figure 2, the HOG uses geometric information about edges and directions using histogram blocks and therefore is strong against geometric changes and exhibits good performance in facial recognition[12]. The HOG is a sliding window-based feature detection algorithm to detect objects on a single image. It uses information about the distribution of the brightness and direction of colors that appear in a target area for detection. Information about the gradient of an image is determined by how rapidly colors change, and information about the direction of an image is determined by the direction of color changes. The HOG is divided into three stages: image segmentation, the creation of histograms, and the combination of histograms. The first stage involves the process of segmenting an entered image. In this stage, the features that would be extracted from an image are divided into N windows. In the second stage, histograms are created using information about the gradients and directions within the windows. This information is measured using changes in brightness on the x and y axes. In the third stage, the histograms composed of R vectors produced from the N windows created in the previous stages are combined via connection. In the process of combining the histograms, HOG feature vectors composed of N×R vectors are produced and then used for target detection. The HOG is suitable when an object does not change significantly and has simple internal patterns and the object is detected using its edge information. The SIFT is more useful when an object's shape changes slightly and its internal patterns are complex. In addition to these techniques, techniques such as the Harr and Local Binary Pattern (LBP) suggested by Viola and Liao are now frequently used to detect objects on single images[13,14]. Different object detection methods are applied depending on the application fields.

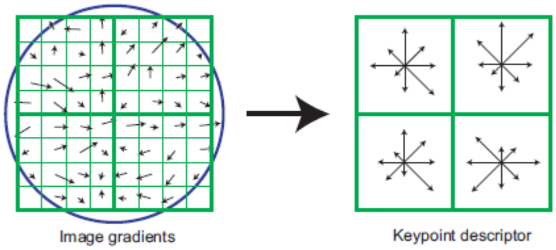


Figure 1. SIFT features[11].

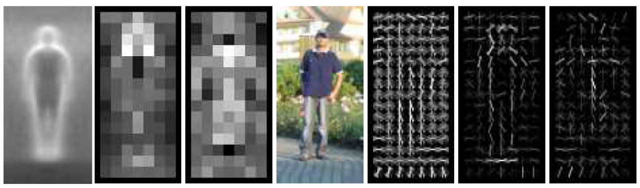


Figure 2. HOG features[12].

The object detection techniques using multiple images identify the location of a target by analyzing the complexity of the target's route on the images by determining the target's features on multiple frames. In general, these techniques are divided into various methods according to the types of analysis employed to analyze feature information on images. While the object detection techniques using changes in multiple frames exhibit outstanding performance in tracking the target's location and path by measuring changes between the frames with simple arithmetic operations, they fail to detect the target if the object does not show movement between the frames. Background frames have been created and applied to target detection to resolve this problem, but it is difficult to create perfect background frames based on environmental changes. The analysis of single images cannot detect a target perfectly when changes in information on brightness between the target and the background on an image, the image's complexity, and the target's shape exhibit large variations. However, the algorithms that analyze multiple images are unlikely to detect a target easily when the target on multiple frames exhibits small changes in movement and the target and the background present similar information[15]. As the method of analyzing features using single images performs complex arithmetic operations targeting an image's entire area, it cannot be easily applied to real-time analyses given that object detection is time consuming. However, as it uses various types of feature information, it has recently shown outstanding results in object detection. This study suggests an effective vehicle detection method in which unnecessary arithmetic operations are reduced as much as possible by applying the pre-processing technique to an entered image to overcome the weakness exhibited in the single-image analysis method.

3. Proposed Methods

Target detection using the existing HOG has the disadvantage of requiring a lot of computing time because its object detection targets the entire area of an image. To resolve this problem, the method suggested in this study selects candidate locations by employing the pre-processing technique to set the candidate locations of vehicles on an image and then detects the vehicles by focusing on these candidate locations. Therefore, this method can detect vehicles more quickly and effectively than the existing method of using the entire area of an image.

Figure 3. DJI Corp's Phantom 3 Professional (4K) UAV.

The images used in this study's experiment were taken using DJI Corp's Phantom 3 Professional UAV, shown in Figure 3. In terms of its major features, the perceived distance is 3.5 km based on the CE and 5 km based on the FCC, and the maximum flight time is 23 minutes. It also prevents tremble with a resolution of 4k (4096*2160) and includes a global positioning system (GPS). The images were taken from a height of 50 m above the ground, targeting vehicles parked on our university campus.

3.1 Pre-processing

The application of pre-processing to an image for object detection within a limited area has multiple advantages. As an object targeted for detection on an image is indeed not present in every area of the image, the speed of an object detector can be improved by dividing the image into an area in which the object exists and an area in which the object does not exist by employing the pre-processing technique. In particular, as with the HOG technique, the object detection methods using the sliding technique perform a substantial amount of arithmetic operations and thus can hardly perform real-time processing. Therefore, this study suggests an effective detection method by dividing an image into an area in which objects exist and an area in which objects do not exist by employing the pre-processing technique and applying it to vehicle detection to reduce the complexity of arithmetic operations involved in the existing HOG technique. Images (a) and (b) in Figure 4 show the results of applying the pre-processing technique, and Images (c) and (d) show the candidate areas for object detection as synthetic images obtained from their original images. This study applied a Gaussian filter to remove the noise of the entered images[16]. The filter was 5×5 in size, and the sigma was set at 0.87. Equation (1) is the Gaussian equation applied to the experiment.

$$G(x,y) = \frac{1}{2\pi\sigma^2} exp^{-\frac{(x^2+y^2)^2}{2\sigma^2}} \quad (1)$$

3.2 Detection of Edge Areas

This study detected edge information by applying the Canny edge detection method to detect vehicles using road images. The Canny edge detector is a leading edge detector often used when detecting edges on images[17]. Non-maximum suppression was performed after calculating the sizes and gradients of edges by applying the Canny edge detector to the images to which a Gaussian filter had been applied. When non-maximum suppression is performed, all gradients except a maximum value become zero. Even if non-maximum suppression is performed, wrongly detected edges may exist due to the noises of images. For this reason, strong edges were maintained and weak edges were removed by employing double thresholding.

3.3 Determination of A Final Target Area using Region Growing

When initially selecting the candidate locations of vehicles using the edge information detected using an edge detector, the volume of information may not be sufficient. This study strengthened the candidate target areas of vehicles using the region growing technique based on 3×3 blocks to determine the candidate regions of vehicles after edge detection[18]. The region growing technique is shown in Equation (2). *m* and *n* are the sizes of the filters, and

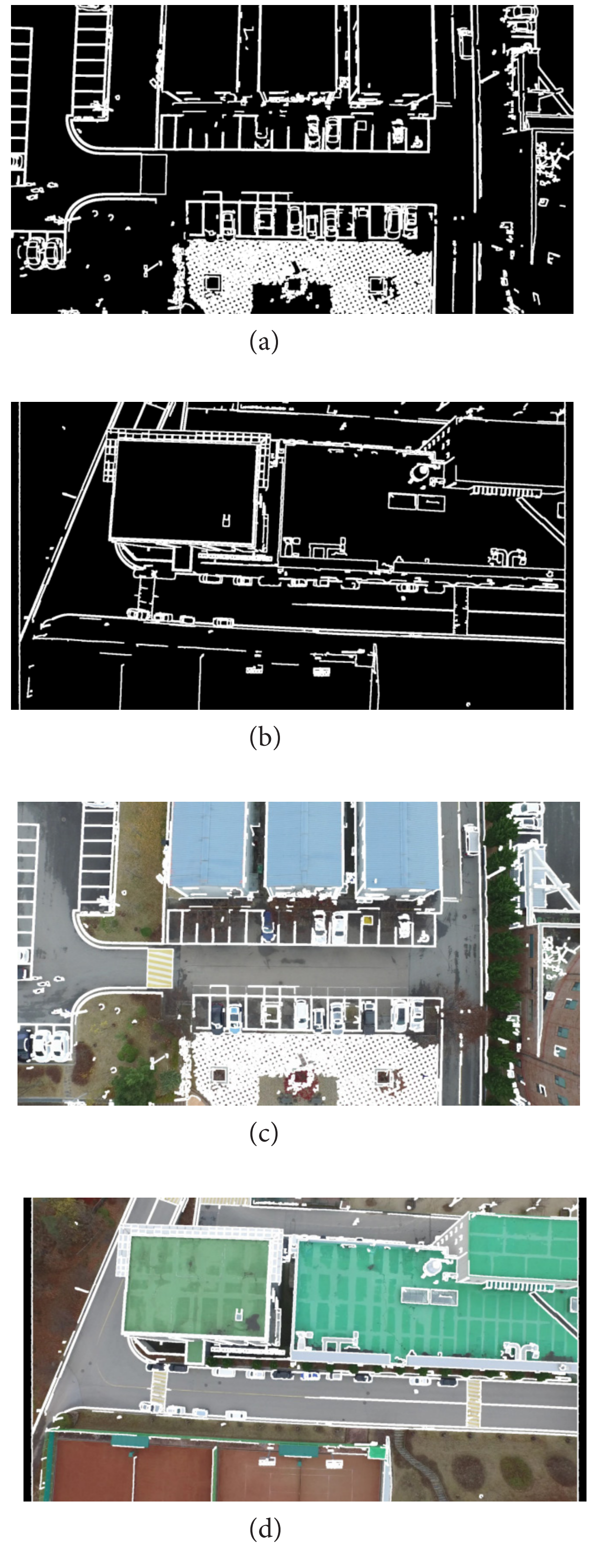


Figure 4. Results of pre-processing (a), (b) and images of predicted areas (c), (d).

BITCOUNT, which is the effective resolution threshold, confirms whether a region is the edge of a target.

$$\sum_{i}^{m}\sum_{j}^{n} I(x+i, y+j) \geq BITCOUNT \tag{2}$$

There are some advantages in using region growing. First, it does not respond sensitively to the noises of an image. Second, the processing speed can be increased. The eight-direction region growing technique is shown in Figure 5.

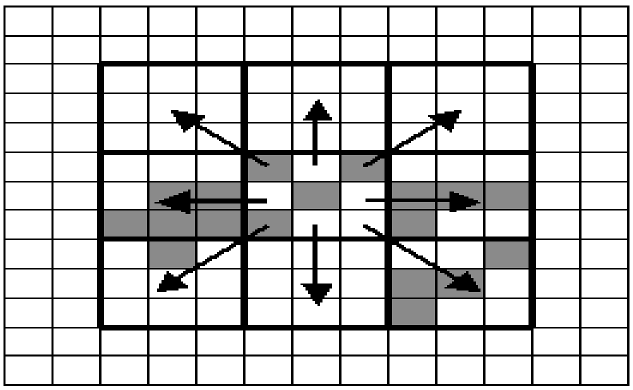


Figure 5. Eight-direction region growing.

3.4 HOG-based Vehicle Detection

The HOG indicates the amount and direction of brightness on local areas as feature information vectors based on the information in the form of histograms. It uses information about the shapes of objects. As the HOG uses the histograms of local areas, it is less affected by changes in brightness and is also resistant against geometric changes. To extract feature information based on the HOG, θ , which is information about the gradient, and *G*, which is information about the size, regarding changes in the *x* and *y* axes on *I(x,y)*, which is the entered image, are extracted using Equation (3).

$$\theta(x,y) = tan^{-1}\frac{\Delta I_y}{\Delta I_{yx}}$$

$$\Delta I_x = I(x+1,y) - i(x-1,y) \quad \mathbf{\Delta} I_y = I(x,y+\mathbf{1} \tag{3}$$

An 8×8 pixel cell was produced using information about the direction and size of gradients calculated from an entered image. A histogram of changes in brightness was then produced using information within the cell. The directional components here were composed by converting 0 to -360º to 0 to -180º, and the histogram was produced regarding nine directions for each 20º. Here, the histogram of the calculated sizes and gradients was divided into cells, and their values were accumulated to identify an area in which the image's edge components were distributed. The HOG of the brightness composed in each cell was normalized into 2×2 blocks. This reduces the influence of partial lighting differences and various noises generated from the image. This study divided feature information into nine pieces by producing 105 blocks. When the entered image contains 64×128 pixels, seven blocks are produced horizontally and 15 blocks are produced vertically, and a total of 105 blocks are normalized. An Ls-norm was applied for normalization. As this study set the overlap rate at 50% based on a 64×128 pixel image, a total of 105 blocks were produced. For each cell, bins were produced, meaning the areas resulting from dividing the cell's angles into nine areas. Therefore, a feature vector with 105×2×2×9 = 3,780 dimensions could be obtained. The feature vector obtained in this manner was taught based on vehicles and non-vehicles using a linear support vector machine (SVM). This study took the images of vehicles that were parked and moving on the university campus from a point approximately 50 m above the ground using a UAV. The composition of data used in the experiment is shown in Table 1.

The HOG adopts the method of producing histograms of the amount and direction of brightness in a local area, thereby composing feature vectors. As the HOG uses the histograms of a local area, it is less affected by changes in brightness and is also resistant against geometric changes. With these characteristics, the HOG can generally show strong performance in the detection of objects whose shapes and colors are consistent, such as vehicles. This study applied the HOG to detect vehicles present in a dense area.

4. Experimental Results

The images used in the experiment were of parked or moving vehicles on the university campus taken from a point 50 m above the ground using DJI Corp's Phantom 3 Professional UAV. The size of the images for the experiment was adjusted to 1980×1080 FHD. The system environment for the experiment included Windows 10, Visual Studio 2013, an Intel i5-3570k processor, 8 GB of memory, a Radeon HD 7800 graphics card, and the OpenCV library

Table 1. Experimental Dataset

	Dataset	
Train Data	Car image	868
	Non-car image	545
Test Data	Car image	436
	Non-car image	378

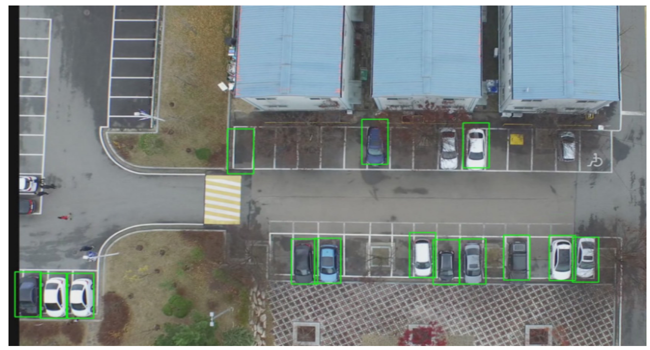
of programming functions. The experimental results are given in Table 2 and show that the method of combining preprocessing and the HOG is around three times faster than the method of using only the existing HOG in the speed of vehicle detection. While a lot of information was removed using pre-processing, the experimental results show that this technique does not significantly affect the performance of vehicle detection. This may be because the aerial images are of high resolution, and the areas in which vehicles were present were clearly distinguished from the areas in which vehicles were not present. In general, areas containing vehicles are clearly distinguished from roads, and even information about black vehicles is not entirely deleted and some areas containing them still remain. Vehicles generally have simple feature structures, single colors, and standardized structures. In this regard, the experimental results also suggest that among the vehicle detection techniques, the HOG is effective because it detects vehicles using information about the edges, colors, and patterns of images. In some cases, the suggested technique could not detect vehicles. When an object on a road or the road's color was similar to information about the shape of a vehicle, it was misrecognized as the vehicle. In addition, when a vehicle overlapped with an object around it, the vehicle was not properly recognized. The precision rate was measured using Equation (4), and the recall rate was measured using Equation (5). As information about the images was reduced during pre-processing, both false positive and false negatives were reduced, as were true positives.

$$precision\ rate = \frac{true\ positives}{true\ positives + false\ positives} \tag{4}$$

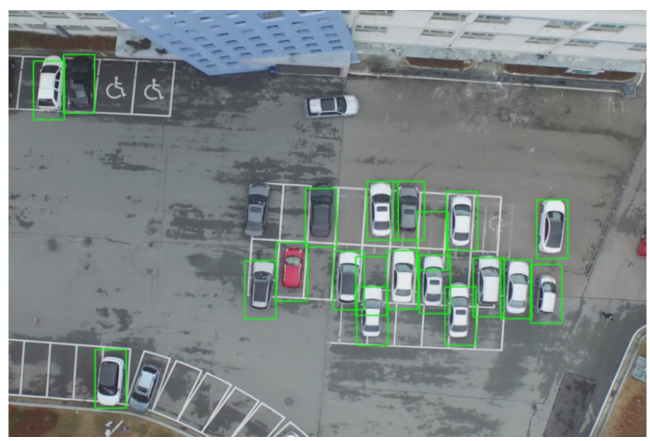
$$recall\ rate = \frac{true\ positives}{true\ positives + false\ negatives} \tag{5}$$

5. Conclusion and Future Work

This study detected vehicles on roads quickly using the pre-processing technique and the extraction of HOG features, which shows good adaptability to environmental changes. It could detect vehicles effectively without detecting the entire area of an image by applying the pre-



(a) Test video #1



(b) Test video #2

Figure 6. Results of detection with different videos.

Table 2. Performance comparison

Extraction Method	#video1		#video2		Average detection speed (#video 1, 2)
	Precision rate	Recall rate	Precision rate	Recall rate	
HOG	90.04%	92.85%	89.34%	93.16%	6.74 fps
Pre-processing+HOG	93.33%	95.14%	91.58%	94.23%	16.82 fps

processing technique to detect vehicles parked on the ground using low-altitude aerial images. While this study succeeded in detecting vehicles on the ground using a low-altitude UAV, based on the study's findings, future studies should research the speed and route of vehicles running on roads. Moreover, if an UAV uses both GPS information and electronic maps, it is likely to increase the system's performance by enabling the more accurate detection of road areas. While the experiment used 1980×1080 FHD images, the detection of vehicles still required a lot of computing and detection time. Therefore, studies on the optimization of detection algorithms and the pre-processing technique to resolve the above problem should continue to be conducted.

6. References

1. Cheung S-CS, Kamath C. Robust techniques for background subtraction in urban traffic video. Visual Communications and Image Processing. 2004.
2. Yao W, Zhang M, Hinz S, Stilla U. Airborne traffic monitoring in large areas using LIDAR data. International Journal of Remote Sensing. 2012; 33(12):3930–45.
3. Chen X, Meng Q. Vehicle Detection from UAVs by Using SIFT with Implicit Shape Model. IEEE International Conference on Systems, Man, and Cybernetics. 2013; 3139–44.
4. Kim NV. Automated Decision making in road traffic monitoring by on-board unmanned aerial vehicle system. Indian Journal of Science and Technology. 2015 Dec; 8(S10):1–6.
5. Elmiktay M, Stathaki T. Car Detection in High-Resolution Urban Scenes Using Multiple Image Descriptors. International Conference on Pattern Recognition. 2014; 4299–304.
6. Leitloff J, Hinz S, Stilla U. Vehicle Extraction from very high resolution satellite images of city areas. IEEE Transaction Geoscience Remote Sensing. 2010; 48(7):2795–806.
7. Moon H, Chellappa R, Rosenfeld A. Performance analysis of a simple vehicle detection algorithm. Image and Computer Vision. 2002; 20(1):1–13.
8. Kozempel K, Reulke R. Fast vehicle detection and tracking in aerial image bursts. ISPRS City Models, Roads and Traffic(CMRT). 2009; 38(3):175–80.
9. Moranduzzo T, Melgani F. Detecting Cars in UAV Images with a Catalog-Based Approach. IEEE Transactions on Geoscience and Remote Sensing. 2014; 52(10):6356–67.
10. Tuermer S, Kurz F, Reinartz P, Stilla U. Airborne vehicle detection in dense urban areas using HoG features and disparity maps. IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing. 2013; 6(6):2327–37.
11. Lowe DG. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision. 2004; 60(2):91–110.
12. Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection. IEEE Computer Society Conf. Computer Vision and Pattern Recognition (CVPR). 2005; 1:886–93.
13. Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. Computer Vision and Pattern Recognition. 2001; 1:501–18.
14. Liao S, Zhu X, Lei Z, Zhang L, Li SZ. Learning Multi-scale Block Local Binary Patterns for Face Recognition. International Conference ICB. 2007; 1:828–37.
15. Cheung S-CS, Kamath C. Robust Background Subtraction With Foreground Validation For Urban Traffic Video. EURASIP Journal on Advances in Signal Processing. 2005.
16. Nixon MS, Aguado AS. Feature Extraction and Image Processing. Academic Press, 2008; 86–8.
17. Canny JA. Computational Approach to Edge Detection. IEEE Transaction. Pattern Analysis and Machine Intelligence. 1986; 8(6):679–98.
18. Gonzalez RC, Woods RE. Digital Image Processing 2nd Edition. Prentice Hall. New Jersey. 2002.