Real-Time detection and tracking moving vehicles for video surveillance systems using FPGA

Mohammed Abdulraheem Fadhel[1*], Omran Al-Shamaa[2], Bahaa Husain Taher[3]

University of Information Technology and Communication [1,2], University of Sumer[3]
Corresponding author E-mail: Mohammed.a.fadhel@uoitc.edu.iq

Abstract

With the growth of the electronic and communication devices, computer vision has become an significant application of smart cities. A smart city is controlled by smart autonomous systems. Many algorithms have been developed to satisfy these smart cities. This paper concerned with addressing the moving objects (vehicles) by using morphological techniques. For computational cheapness. The simulation has been built by a MATLAB 2012a and its implementation was done using Xilinx-ISE 14.6 (2013) XC3S700A-FPGA board that provides an exceptional tool for mixing between two platforms, the ISE 14.6(2013) and the MATLAB (2012a) platforms. MATLAB provides components for FPGA that invoke Verilog code of Xilinx platform, to avoid the size weakness of XC3S700A-FPGA board.

1. Introduction

The recent significant growth in the electronic, communication devices, and the complexity of urban civilization have encouraged the creation of a smart city by smart autonomous systems [1].
Computer vision has become a significant application of intelligent systems applied in a wide range of fields varying from human computer interaction to robotics [2].On the other hand; these systems must have a high reliability in all situations [3].
One of the important applications of these systems is the automated video surveillance which is now widely used in real-time traffic monitoring; it can analyze traffic flows, track vehicles, classify, identify and detect accidents by video cameras [4, 5]. The main challenges are to develop fully automatic systems that require limited processing time and storage capacity; they also do not require task-specific thresholds and tuning. These challenges underline the importance of algorithms that are computationally efficient, task, operator, and threshold-independent and are capable of detecting and tracking activities in a scene of observation [6].
All the previous requirements focus on the motion estimation, which is the process of analyzing successive frames in a video sequence to identify objects that are in motion [7]. The main objective of the visual tracking algorithm is to perform fast and reliable matching of the target from frame to frame by image analysis techniques [8], In other words, it identifies (detects) the pixels that are different from the background and are thus suspected to be a part of a physical object new to the scene (foreground) [9, 8].

2. Literature review

Ricardo et al. [10] propose an approach for vehicle tracking, using the Extended Kalman Filter (EKF), which can simultaneously integrate into the motion model both the position and the viewpoint of the object observed. They publicly release the dataset, with the ground truth annotations, to provide a common framework for evaluating the performance of vehicle detection and tracking systems within the context of smart city applications.
K. Shanmugapriya [11] proposes object segmentation and object detection, on a multi-object moving background based on morphological technique and cellular automata based segmentation. Motion segmentation is done to segment an object from a video. A morphological operation is proposed to remove unwanted object motion and enhance the segmentation result. This result is then used for object identification. Cellular automata based segmentation is performed to detect a particular object from a video.
Shih [12] has presented an automatic foreground object detection method for videos captured by freely moving cameras. He focused on extracting a single foreground object of interest throughout a video sequence, based on the scale-invariant feature transform (SIFT) correspondence across video frames. He also constructed a SIFT trajectory in terms of the calculated foreground feature point probability. While for fast camera a consensus foreground object template (CFOT) was proposed for detecting the foreground object in videos captured, even if the contrast between the foreground and background regions is low.
Pritam et al. [4] present a technique for object identification and tracking based on background subtraction with optimized threshold binarization. Also, mapping techniques have been developed to relate image with the real world. The algorithm is also capable of working with bad lighting conditions using histogram equalization approach.

3. Real time of object detection and tracking

The early stage and the task in the detection is the segmentation. It is the first step of computer vision applications, such as video-surveillance, object tracking, autonomous navigation, and

monitoring; it detects changes and relevant changes for further analysis and qualification [13]. It is worth noting that, in such applications, the low-cost and real-time requirements should be satisfied. [14].

Background Subtraction represents one of the active or efficient methods of segmentation [9]. The segmentation could be employed by these systems to identify all the objects in the scene and then detect anomalous situations [13].

As in any processing operation, there is an error in the segmentation process; this error can affect the segmented video quality in two ways: statically (spatially) and dynamically (temporally) [14]. Spatial video segmented errors are defined by the amount of mis-segmented pixels that can be easily estimated by direct comparison between the reference and resulting segmentation mask, for a given frame. By taking into account the number of mis-classified pixels and by estimating these error pixels the principal algorithm for object segmentation can be evaluated. The sum of the distance between pixels that have been assigned to the wrong class, and the nearest pixels that belong to the correct class is usually used as an evaluation criterion [14]. Then, the foreground blobs are the segmentation results, and the blobs are a collection of connected pixels [9]. Then, the object tracking becomes suitable here.

Object tracking

Object's tracking data comprises its morphological properties (area, Euler number, etc.), pixel values, and path point at each position along the tracking route. This technique eliminates the need for object registration stage and, hence boosts the speed of the overall tracking process. Object tracking is the process of segmenting a moving object of interest from a video scene and keeping track of its motion, orientation, occlusion, to extract useful information [15]. Then, by using simple data association techniques in combination with adaptive background subtraction or frame differencing, a fixed camera can effectively track the moving objects in real-time [16].

Traditionally, tracking and identification has been considered separately, the target is firstly identified and then tracked kinematically to sustain the identification. The signature of the target represents the identification information; it is needed for tracking the target. Accordingly, the identification and tracking perform together as a unit [17]. The other important task is to determine the object tracking in image parts. In this context, and with fixed cameras, background modeling can often be addressed in this task [18]. It is worth noting that, the tracker discrimination of objects from the background is directly related to the features that are used. Most of the tracking applications are conducted using a fixed set of features, determined a priori [19]

Background subtraction

The three primary methods for foreground detection are temporal differencing, optical flow, and background subtractions. The most popular method of them is the background subtractions. In this context, it is used as a preprocessing step for foreground detection in a particular scene and tracking in a vision system [4]. It consists of the differentiation of moving objects from a maintained and updated background model. Also, it can be considered as flexible and fast, for a stationary camera and fixed background scene [20], [21]. Usually, each pixel is searched and compared step-by-step with a predefined dataset for detection and tracking off an object. Therefore, the vision system performance has been significantly affected by background modeling. Accordingly, it is important to employ a best background subtraction method [22].

One main challenge of the background modeling is the occlusions of the foreground objects. Other challenges produce changes in the statistical background representation; these changes come from the environment illumination change and motion of the background objects. To develop real-time background maintenance and motion detection algorithm along with

compression of frames for reducing size of storage is a real challenging task [23].

Foreground

Foreground detection is an essential procedure in video analysis, such as object detection and tracking. Also, it can be considered as a desired object, which is the aim of the processing for extracting the relevant information under consideration [4]. It is often difficult to optimize the foreground representation due to light, weather, clutter interference and shadow. One of the simplest and most common methods for foreground detection is the background subtraction, by which, foreground can be calculated by generating a binary difference between background and current image using a global threshold [25]. There are two methods to get a background image, one is to appoint an image as a background artificially, while the other uses a model to train the background, such as the Gaussian background model (GBM). The second method is more accurate.

The background subtraction method has the robustness to light changing and slight movement, but when using this method to deal with long image sequence there may be much accumulated error in the foreground. Optical flow covers long distance, and the noise due to brightness change is less which results in less accumulated error percentage [25].

Motion detection

Motion detection is a fundamental step in the lineage of information regarding objects in motion [26] and in simply counting the number of different pixels from the comparison between the reference image with the current one, for a fixed camera [20, 21]. Therefore, more motion can be seen when there is a vast difference in pixels between two consecutive frames, and then we can sum up the absolute differences to get the Sum Average Difference (SAD). Accordingly, for an increase in the sensitivity of motion detection, a threshold must be set to filter out the camera noise [23]. Then, the implementation of these comparisons has been represented as segmentation part in motion detection algorithm. Objects can be segmented out using the image subtraction technique for each pixel [26].

Error in detection

The changes of the light affect the pixel values, and then a mistake in detection happens. Accordingly, to solve this problem, the image can be disintegrated into small blocks. In this context, the objects will be separated, and then for each block, the mean value can be calculated and subtracted from each pixel value [23]. Also, the distance of the moving objects to the camera (close or far away), will affect the foreground and background consideration, where a nearby moving object is considered as foreground, while the far away one is considered as background [4].

Morphological operation

A morphological operation is usually done to decrease noise regions and to filter out smaller regions. This technique is performed to enhance the result of segmentation. In this work, two operations were used, they are closing and thickening. Therefore, the morphology is a tool for description and representation of the region shape, by extracting image components, such as skeletons, boundaries, and other components. This methods is used for both pre and post processing. Opening operation, for example, smoothes the contour of an object and eliminates thin protrusions. By opening operation, extra added pixels are removed. Whereas, Erosion operation shrinks objects and is used to remove image components. Closing operation is done to eliminate small holes and filling gaps in the contour. It simply fills holes in the background region [15].

Time constraint

There are two aspects regarding the time constraint. The first one is the recording of the detected motion as soon as it happened. The second one is the storage part which must be fast enough to prevent the freezing. The anticipated behavior of these two aspects depends on the efficiency of the algorithm and the data constraint [21]. There are two kinds of data constraints. One of them is the real time processing, it needs a system with fast access memory. The other data constraint is that, data must be in a known picture type [21].

Implementation of object tracking using FPGAs

FPGAs are extremely used in image processing algorithm applications due to their ability to serve this field like handling huge amounts of incoming and outgoing data, and their ability to perform multiple operations in parallel on that stream of data.
One conventional method for implementing an object tracking algorithm is the use of particle filters, comparator, and morphological operations. These approaches assume that the general features of successive video frames can be approximated via a set of discrete samples taken over the entire area of each frame. An FPGA-based mplementation of this algorithm allows the required process values to be sampled and calculated as often as possible [34].

4. Resuilt

The morphological function "bwmorph(bw, 'fill')" fills all holes of a single pixel; while "imfill(bwImg)" function fills all holes of any size, causing annulment of the Euler number of segmented objects; Euler number is defined as the total number of connected components (objects) in the image minus the number of holes in those objects. Object by "imfill(bwImg)" function will have no holes anymore, losing in this way an important feature that may be used to recognize a selected object, and let other objects be erroneously detected as the wanted one. This function it excludes an important blob (connected area or object) feature from the decision making process, the most object properties left are related to its area, making recognition process, not as good as desired. But if we let tracked objects keep more of their holes, by using "bwmorph(bw,'fill')", good discrimination ability will be added to the algorithm and the whole process gets a substantial boost.

The improvement of recognition process can be seen obviously, as there will be less false detections, yet the hit number will not be enough to declare an active recognition process. The best results can be achieved by modifying the morphological operation sequence to be (close > thicken) combination rather than (close > fill), keeping this way, Euler number intact since thicken operation preserves Euler number.The errors in object's morphological properties due to environment (lighting) and processing (morphological operations) reasons are unavoidable, i.e., detecting and identifying an object by one hundred percent property matching in all cases and circumstances is not possible. Absolute error ratio of an object's property can be calculated as follows:

$$\text{error ratio} = \left| (\text{Current Value} - \text{Reference Value})/(\text{Reference Value}) \right| \qquad (1)$$

As a compromise, the properties may be accepted as matching ones within a *maximum error ratio* (**error**). This ratio is determined empirically by examining the hit, miss, and false curves in Fig. (4.11) choosing the ratio which corresponds to global best-results zone (where false and misdetections are at minimum, while hit number is at acceptably high value). In this context, experience shows that the hit-number is best when **error** equals to or more than (0.05), so is the miss-number; while false recognition is best when **error** is below (0.04). So, choosing the value (0.05) for **error** would give global best recognition results. The *maximum error ratio*(**error**) has strong impact on the object recognition process.

Hit, miss, and false-hit occurrences for all objects are measured using fixed optimized morphological operations throughout the video.
Object detection is a common step in all surveillance systems even though each requires different treatments. Detection is intended to produce the object-mask (also denoted object-frame) logical image, in which only the object pixels are shown in white; and black elsewhere [53]. Detection time of the algorithm is the time required to segment foreground from background; it is usually measured as the average of all video frames.
The object-frame is usually generated by background model subtraction; this technique is commonly used for motion detection in static scenes. It subtracts a reference background image pixel-by-pixel from the current frame (image). The absolute of resultant image pixels are then compared to a threshold; pixels of values above the threshold are classified as foreground. Object-frame then undergoes some morphological operation such as erosion, dilation, and closing to reduce the effect of noise and to enhance the detected regions. Shadows, segmentation is a significantly difficult problem, it needs to be well handled for an efficient and robust surveillance system.
Object recognition is an optional step of object detection; it identifies the wanted or interesting object, depending on its properties. Object properties can be morphological or/and chromatic; both property types are adopted in the proposed system to perform object identification (object recognition).
On the other hand, object tracking is the process of establishing a correspondence between objects in successive frames and to produce temporal data about moving objects, like speed, path, direction, and position.
An object level tracking algorithm has been used, i.e., we did not track parts of objects, such as car's license plates, instead whole objects were tracked from frame to frame. The proposed approach adopts the use of features like size, mean intensity, the bounding box, and Euler number, which are extracted from previous parts of the system, to detect a matching between objects in successive frames.
When an object passes in front of the second camera, the algorithm compares between object's measured properties and that received from the previous camera site. When each object's measured property error ratio is within the maximum error ratio error (0.05), see equation (1), the object is declared as a matching one and is bounded by a red box, its path is shown in dotted red line and the tracking information region will also appear in red, as is shown in (Fig.1 a-f). This treatment will continue until the object exits from the second camera viewing area (being outside the region of interest (ROI)) then the red color disappears, as shown in (Fig.1g).
Tracking process is performed in a synchronous manner, since there is no pre-tracking data as the case with selectplay player; also, objects have no ID numbers since they are supposed to appear for the first time. Real-time tracking data and black-white image of each object on the screen are shown on the attached figure, one at a time. Also, a frame left-right flipper has been used, so that algorithm immunity against object orientation may be examined.
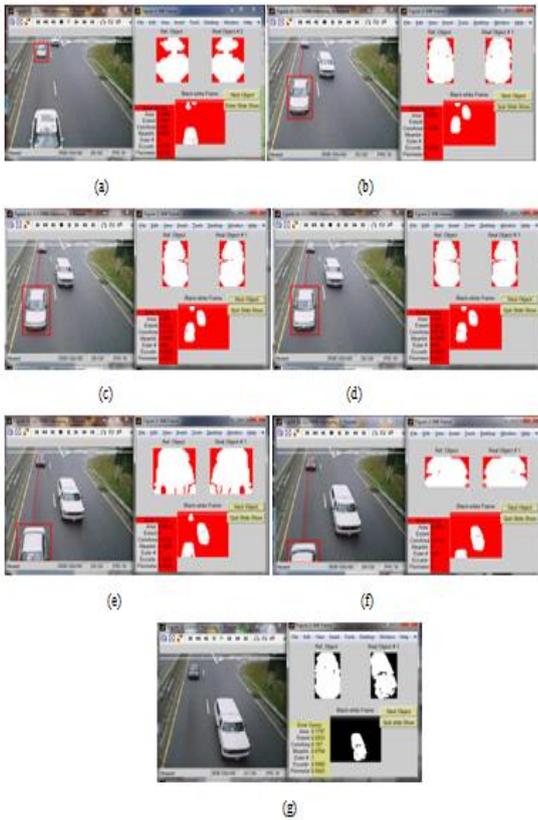
(a)   (b)

(c)   (d)

(e)   (f)

(g)

Fig. 1: Object detection and tracking steps

Table I: Statistical Information of Results

Scene	Area	Extent	Convex Area	Mean Intensity	Euler Number	Eccentricity	Perimeter and Solidity
A	0.0088	0.0041	0.008	0.0007	0	0.0036	0.0112
B	0.0052	0	0.0129	0	0	0.0002	0
C	0.001	0.0043	0.0119	0.0008	NaN	0.0029	0.0077
D	0.001	0.0043	0.0119	0.0008	NaN	0.0029	0.0077
E	0.0043	0	0.0194	0.0006	0	0.0263	0.0233
F	0.0056	0	0.0105	0	NaN	0	0
G	0.1797	0.2053	0.107	0.0794	1	0.0909	0.9503

Table I contained the total information of (fig.1. a-f) that meaning each scene has a detail of vehicles. Scene A has a small number of area and G has the maximum number, it is depend on the distance between the object and the camera. All other properties have the same previous idea.

Table II: Statistical Information of Results

Scene	Time Calculation
A	1.93 ms
B	1.02 ms
C	1.43 ms
D	1.43 ms
E	1.20 ms
F	0.93 ms
G	2.06 ms

Table II illustrated the time calculation of each frame using Xilinx platform. Scene G exhausted more time for processing because the size of detected pixel.

5. Conclusion

The main points that are drawn from this work can be summarized as follows:

1. The fixed background subtraction technique is straightforward and widely used for motion segmentation with a stationary camera.
2. An object may split if a part of its gray image is close in value to the background, and this part is wide enough to separate two or more parts of the object's image.
3. The dropping of the image mean value has compensated for the action of the camera's white-balance algorithm. This technique has greatly improved the segmentation operation.
4. Euler number can be used in object recognition as a connected area feature and can be considered as one of the morphological power points in the recognition process.
5. The best improvement in the recognition process can be achieved by carefully modifying the morphological operation to be (close > thicken) sequence combination; "thicken" operation preserves Euler number, rather than (close > fill).

References

[1] Cuevas C & García N, "Improved Background Modeling For Real-time Spatio-temporal Non-parametric Moving Object Detection Strategies", *Elsevier, Image and Vision Computing*, Vol.31, (2013), pp.616-630.

[2] Kropatsch WG & Bischof H, "Digital Image Analysis Selected Techniques and Applications", *Springer-Verlag New York*, (2001).

[3] Nieto M, Unzueta L, Barandiaran J, Cortés A, Otaegui O & Sánchez P, "Vehicle Tracking and Classification in Challenging Scenarios Via Slice Sampling", *EURASIP Journal on Advances in Signal Processing*, (2011).

[4] Das P, Ghoshal R, Kole DK & Ghosh R, "Measurement of Displacement and Velocity of a Moving Object from Real Time Video", *International Journal of Computer Applications*, Vol.49, No.13, (2012).

[5] Gao D & Zhou J, "Adaptive Background Estimation for Real-time Traffic Monitoring", *IEEE Intelligent Transportation Systems Conference Proceedings*, (2001).

[6] Ramezani R, Angelov P & Zhou X, "A Fast Approach to Novelty Detection in Video Streams Using Recursive Density Estimation", *IEEE 4th International Conference "Intelligent Systems"*, (2008).

[7] Marques O, *Practical Image and Video Processing Using Matlab,* John Wiley & Sons, Inc., Hoboken, New Jersey, (2011).

[8] Angelov P, Ramezani R & Zhou X, "Autonomous Novelty Detection and Object Tracking in Video Streams using Evolving Clustering and Takagi-Sugeno Type Neuro-Fuzzy System", *International Joint Conference on Neural Networks*, (2008).

[9] Prabhakar G & Ramasubramanian B, "An Efficient Approach for Real Time Tracking of Intruder and Abandoned Object in Video Surveillance System", *International Journal of Computer Applications*, Vol.54, No.17, (2012).

[10] Aghajan H & Cavallaro A, "Multi-Camera Network Principle and Application", *Academic*, (2009).

[11] Schneiderman H & Kanade T, "A statistical method for 3D object detection applied to faces and cars", *Proceedings. IEEE Conference on Computer Vision and Pattern Recognition,* (2000), pp.746-751

[12] Guerrero-Gómez-Olmedo R, López-Sastre RJ, Maldonado-Bascón S & Fernández-Caballero A, "Vehicle tracking by simultaneous detection and viewpoint estimation", *International Work-Conference on the Interplay Between Natural and Artificial Computation*, (2013), pp.306-316.

[13] Shanmugapriya K & SreeVidhya T, "An Efficient Detection Approach for Visual Surveillance System Using Morphology", *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, Vol.2, No.12, (2013).

[14] Cucchiara R, Onfiani P, Prati A & Scarabottolo N, "Segmentation of Moving Objects at Frame Rate: A Dedicated Hardware

Solution", *Proceedings of 7th IEE Conference on Image Processing and its Applications*, (1999).

[15] Gelasca ED & Ebrahimi T, "Application Dependent Video Segmentation Evaluation - a Case Study For Video Surveillance", *EURASIP, 14th European Signal Processing Conference*, (2006).

[16] Sun SW, Wang YCF, Huang F & Mark Liao HY, "Moving foreground object detection via robust SIFT trajectories", *Elsevier, J. Vis. Commun. Image R.,* (2012).

[17] Elisa DG, "Full-reference Objective Quality Metrics for Video Watermarking, Video Segmentation and 3d Model Watermarking", *PhD thesis, dottore in ingegneriaelettronica, UniversitàdegliStudi di Trieste, Italie, et de nationalitéitalienne*, (2009).

[18] Martnez SV, Knebel JF & Thiran JP, "Multi-object Tracking Using the Particle Filter Algorithm on the Top-view Plan", *12th European Signal Processing Conference*, (2004), pp.285-288).

[19] Nguyen DH, Kay JH, Orchard BJ & Whiting RH, "Classification and Tracking of Moving Ground Vehicles", *LINCOLN Laboratory Journal*, Vol.13, No.2, (2002).

[20] Leibe B, Schindler K, Cornelis N & Gool LV, "Coupled Object Detection and Tracking from Static Cameras and Moving Vehicles", *IEEE Transactions, Pattern Analysis and Machine Intelligence*, Vol.30, No.10, (2008).

[21] Collins RT & Liu Y, "On-line Selection of Discriminative Tracking Features", *IEEE ICCV03*, (2003).

[22] Cucchiara R & Piccardi M, "Vehicle Detection Under Day and Night Illumination", *Proc. of ISCS-IIA, Special session on vehicle traffic and surveillance*, (1999).

[23] Jagdale VB & Vaidya RJ, "High Definition Surveillance System Using Motion Detection Method Based on FPGA DE-II 70 Board", *International Journal of Engineering and Advanced Technology*, Vol.2, No.2, (2012).

[24] Jeisung Lee and Mignon Park, "An Adaptive Background Subtraction Method Based on Kernel Density Estimation", Journal of Sensors, Vol. 12, 2012.

[25] Badnerkar S & Kshirsagar Y, "Real Time Motion Detected Video Storage Algorithm for Online Video Recording", *IJCA Proceedings on International Conference in Computational Intelligence*, (2011).

[26] Rajvi Shah and P. J. Narayanan,"Interactive Video Manipulation Using Object Trajectories and Scene Backgrounds", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 23, No. 9, September 2013.

[27] Suganya Devi K, Malmurugan N & Sivakumar R, "Efficient Foreground Extraction Based on Optical Flow and SMED for Road Traffic Analysis", *International Journal of Cyber-Security and Digital Forensics (IJCSDF): The Society of Digital Information and Wireless Communications (SDIWC)*, (2012).

[28] Sahu AK & Choubey A, "A Motion Detection Algorithm for Tracking of Real Time Video Surveillance", *International Journal of Computer Architecture and Mobility*, Vol.1, No.6, (2013).

[29] Zúñiga MD, Brémond F & Thonnat M, "Real-time Reliability Measure-driven Multi-hypothesis Tracking Using 2D and 3D features", *EURASIP Journal on Advances in Signal Processing*, (2011).

[30] Chen Z, Soyak E, Tsaftaris SA & Katsaggelos AK, "Tracking-Optimal Error Control Schemes for H.264 Compressed Video for Vehicle Surveillance", *20th European Signal Processing Conference, EUSIPCO,Romania*, (2012).

[31] YasiraBeevi CP & Natarajan S, "An efficient Video Segmentation Algorithm with Real-time Adaptive Threshold Technique", *International Journal of Signal Processing, Image Processing and Pattern Recognition,* Vol.2, No.4, (2009).

[32] Cai L, He L, Xu Y, Zhao Y & Yang X, "Multi-object detection and tracking by stereo vision", *Elsevier, Pattern Recognition*, Vol.43, (2010), pp.4028–4041.

[33] Barth A & Franke U, "Tracking Oncoming and Turning Vehicles at Intersections", *University of Bonn, Germany, Institute of Geodesy and Geoinformation*, (2010).

[34] Mariano VY, Min J, Park JH, Kasturi R, Mihalcik D, Huiping L, Doermann D & Drayer T, "Performance Evaluation of Object Detection Algorithms", *Proceedings. 16th International Conference on Pattern Recognition,* (2002).

[35] Snidaro L & Foresti GL, "Real-time thresholding with Euler numbers", *Elsevier, Pattern Recognition Letters*, Vol.24, (2003).

[36] Marschner AR, "An FPGA-based Target Acquisition System", *M.Sc. Thesis in Computer Engineering, Blacksburg*, (2007).