Deer Detection in Thermal Images for Traffic Safety Using Contour Based Histogram of Oriented Gradient Method

Debao Zhou\textsuperscript{a} and Jingzhou Wang\textsuperscript{b}

Dept. of Mechanical and Industrial Engineering
University of Minnesota, Duluth, MN 55812, USA
\textsuperscript{a}dzhou@d.umn.edu and \textsuperscript{b}wang2511@d.umn.edu

Abstract- Car accidents due to Deer-Vehicle Crashes (DVC’s) are one of the major safety concerns for the driving on rural roads in Europe and North America. Many attempts have been made to reduce the occurrences of these accidents, but none of them has yet proved effective. With the development of infrared imaging technology, using thermal images to identify the presence of deer in night has become possible. A pattern recognition method, Histogram of Oriented Gradient (HOG), can be used to detect deer to reduce such accidents. However, the length of time required to process one image makes the method incapable of real-time deer identification and tracking. Based on the HOG method and a known classifier - Support Vector Machine (SVM), this research developed a contour based HOG + SVM method, called CNT-HOG method to process thermal images and to detect the presence of deer. Experimental results have demonstrated that this algorithm has the advantages of both high accuracy (up to 94.2\%) and short computation time (0.1s) when compared to traditional HOG + SVM method. In this research, further analysis has been performed to evaluate the influence of body postures and occlusions on detection accuracy. By achieving such computation speed and accuracy, this CNT-HOG algorithm proves that deer can be tracked in real-time. By using such a system, drivers on road can be warned of the presence of deer in real time and the frequency of DVC’s can be effectively reduced.

Keywords- Deer-vehicle Crashes; Thermal imaging; HOG; Contour

I. INTRODUCTION

It is estimated there are over 35,000 Deer-Vehicle Crashes (DVCs) yearly in the U.S., which results in about 200 deaths and close to 4,000 property damages of 1,000 dollars or more \textsuperscript{[1]}. These shocking statistics justify many attempts trying to detect deer on roads to avoid such accidents.

The attempts to reduce such accidents can be categorized into either passive or active methods. The passive method, which has been proven as either impractical or inefficient, is to keep deer off the road by means such as using electrical mats or fences \textsuperscript{[2],[3]}, while the active approach is to alert drivers the existence of deer around the roadside. Existing active approaches for deer detection can be further grouped into ‘Break-the-Beam’ method (BtB) \textsuperscript{[4]} and ‘area-cover’ method \textsuperscript{[5],[6]}. The key component in BtB methods is a beam which is formed by infrared light. The blockage of the beam means there is an animal crossing the road, which results in the issuance of a warning signal. This method can effectively detect passing deer. However, in terms of detection accuracy, it has the following drawbacks: (1) An object exiting from the road to roadside can activate the warning system, (2) Any objects which are big enough to block the beam can activate the warning system, (3) It cannot detect the existence of roadside deer, which may jump onto the roadway. In an area-cover system, a normal camera, radar or an ultrasound system is used to detect the presence of deer. Like BtB methods, once an object, such as deer, is detected, a warning signal will be issued. Different from the above methods, this research proposed using pattern recognition based thermal imaging method to detect the existence of deer around the road. This system cannot only detect deer at a wider range, it can also track the deer to indicate whether there is an immediate threat to the vehicle. In order to improve the detection speed, this paper proposed using contour based pattern recognition method to detect the presence of deer in the environment.

The main knowledge used in this research is contour finding and pattern recognition. In the following, the developments in these two areas are reviewed respectively.

In the earlier stage, contour finding methods were mainly based on edge detection via edge linking. In 1987, Michael Kass, et al. suggested a “snakes” method to find contours \textsuperscript{[8]}. This method has proved to be powerful and flexible for many image processing problems \textsuperscript{[9]}. This technique has been used in many fields and its accuracy and efficiency have been further improved \textsuperscript{[10],[11],[12]}. The detection of an object using pattern recognition approach needs to perform two tasks, finding descriptor and training classifier. The first task is to choose an appropriate descriptor to describe an object, i.e., the Histogram of Oriented Gradient (HOG) algorithms. The other work is to build a detection framework (called classifier) using the descriptor, such as the Support Vector Machines (SVMs) algorithm.

The fundamental idea in Histogram of Oriented Gradient (HOG) algorithms is that the existence of an object can be characterized by the distribution of local intensity gradients or edge directions, even without precise knowledge of the corresponding gradient or edge positions \textsuperscript{[13]}. This idea is realized by dividing the windows into blocks in which HOG vectors are calculated. Blocks are further divided into small spatial regions, called cells. For each cell, a one-dimensional
histogram of gradient directions or edge orientations over the pixels of the cell is accumulated. The normalized vectors are referred as the HOG descriptors. There are several HOG descriptors, including Rectangular HOG (R-HOG), Circular HOG (C-HOG), Bar HOG and Center-surround HOG. R-HOG descriptor uses square or rectangular grids of cells. C-HOG defines cells in each block into grids of log-polar shape. C-HOG allows fine coding of nearby cells compared to coarse coding of the outer cells by counting more pixels in the outer cells than in the inner cells. Instead of using first derivatives, Bar HOG uses oriented second derivative filters to compute a descriptor. Many objects can be modeled as connected bars and blob-like structures. Center-surround HOG allows the algorithm to avoid certain redundant computation as is evident in R-HOG and C-HOG algorithms. Moreover, Center-surround HOG algorithm can be optimized for fast computation.

Classifier is defined as a framework used to identify the targets from a set of data. Support Vector Machines (SVMs) and Adaboost are two of the most widely used methods to build classifiers. According to and , SVMs are a set of related supervised learning methods that analyze data and recognize patterns.

A significant number of studies apply the HOG and SVMs methods in objects detection, such as pedestrian recognition in traffic. However, there is no research about using HOG to detect deer with thermal imaging devices using contour based method, or CNT-HOG. In this research, the CNT-HOG is applied to identify deer on roadsides to provide a driver with the information about the existence of deer, resulting in the avoidance of DVC’s. In the following sections: (1) Section 2 introduces the CNT-HOG and SVM method. (2) Section 3 shows the image acquisition device for deer detection and basic image processing steps. (3) The results and discussion are presented in Section 4. (4) The conclusion is drawn and future work is illustrated in Section 5.

II. CONTOUR BASED HOG (CNT-HOG) PATTERN RECOGNITION METHOD

In this research, a contour based, multi-scale, multi-object HOG detection framework was constructed using the sliding-windows method. In the first step, in order to reduce the processing time, a contour finding algorithm was applied to an image to figure out the regions in which there are no deer. Then the rest parts of the image were named as the regions of interest, or ROI’s. Only the image portions in the ROI’s were processed and the portions in non-ROIs would be left untouched. The ROIs were then scanned with a window in a certain size, such as 144 by 104 pixels. HOG descriptors were generated in each window. All these HOG descriptors would be combined to represent the edge properties of the whole image. The final step is to use a classifier to specify whether there are objects in each window. To detect objects in different sizes, a multi-scale detection and fusion algorithm was applied.

A. Finding Contours

This research aims at using infrared thermal imaging method to achieve real time detection of the presence of deer in the surroundings. As a property of thermal images, the temperature difference is reflected as the intensity of the pixels in gray scale. As a matter of fact, deer are always shown in higher intensity than their surroundings in a thermal gray scale image. Based on the gray scale property of the IR thermal images, the contour based HOG method was pursued. In this method, the heated areas were identified first since the lower temperature objects would not be useful in the identification of deer. In image processing, this can be easily realized by adding an illumination filter to change the pixels with intensity lower than the threshold to zero. The illumination filter function is as follows:

\[
f(i, j) = \begin{cases} 
\frac{x}{\text{threshold}}, & x < \text{threshold} \\
0, & \text{otherwise}
\end{cases}
\]

After changing the lower intensity pixels to 0 through the illumination filter, the contours in the images were searched and the targets should be inside those contours where the intensity of the pixels was greater than the threshold. In this stage, however, there were still no deer in most of the high intensity areas. The problem then was to identify as many as possible portions in which there are deer. Since only the deer in side view were used in the training sets, two criteria were employed as a size filter to identify the possible areas (contours) with deer: (1) size of the contours and (2) ratio of width to height of the contours. In the available images, the targets (deer) have a height from 70 to 250 pixels in an image with a size of 640 by 480 pixels. Thus, some of the contours were eliminated from the candidate pool. The left contours were resized and reformed, and then were defined as the regions of interest (ROI’s). The HOG computation was applied only to the ROI’s. The whole process of finding contours is illustrated in Fig. 1.

![Fig. 1 Contour finding procedure](image-url)

B. Histogram of Oriented Gradient (HOG)

In this research, five stages were used to calculate HOG features of the image in ROI’s. Firstly, a gray scale normalization operation was performed to reduce the illumination influence, in which a gamma (power law) compression was applied by computing the square root of each color channel since images are acquired in full color. This compression helps to reduce the effects of local shadowing and illumination variations because image texture strength is typically proportional to the local surface illumination. Next, a Sobel mask was used to compute the gradients of the image. The gradients capture contour, silhouette and some texture information, while providing further resistance to illumination variations. Thirdly, original image was encoded. In this stage, the image window was divided into small spatial regions, called “cells”. Each cell contains several pixels. For each cell, a local 1-D histogram
of gradient or edge orientations over all the pixels in the cell was accumulated. These 1-D histograms formed the basic representation of the orientation histogram. Each orientation histogram divides the gradient angle range into a fixed number of predetermined bins. The gradient magnitudes of the pixels in the cell were used to vote into the orientation histogram. The obtained description of the histogram of the oriented gradient was taken as the default descriptor. The fourth step was to normalize the default descriptor, which introduces better invariance to illumination, shadowing, and edge contrast according to Navneet Dalal’s research\[13\]. It was performed by accumulating a measure of local histogram “energy” over local groups of cells that we call “blocks”. Typically each individual cell is shared between several blocks but its normalizations are block dependent and thus different. The cell thus appears several times in the final output vector with different normalizations. This step significantly improved the performance. The normalized block descriptors are referred to as histogram of oriented gradient (HOG) descriptors. Finally, the HOG descriptors were collected from all blocks of a dense overlapping grid of blocks covering the detection window into a combined feature vector for use in the window classifier.

C. Support Vector Machine (SVM) Algorithm

There are two phases in a SVM computation, the learning phase and the classification phase. In the learning phase, the main task is to construct a detector for a classifier, which will be applied in the classification phase, to judge the bench score of the input. In the first step of the learning phase, a preliminary detector is generated using the positive training data set and an initial set of negative windows. The positive training data set contains windows with at least one object located in the center of the window. The positive training set can help the algorithm to learn what the object looks like. The initial set of negatives is created by randomly sampling negative images, in which there are no objects. Compared to the positive dataset, it mainly helps the algorithm to learn what non-objects’ HOG characters are. Taking the positive data set together with the initial negative data set as the input of the classifier, a preliminary detector is thus trained. Secondly, the preliminary detector is used to exhaustively scan the negative training images for hard examples (false positives). The classifier is then re-trained using this augmented training set (user supplied positives, initial negatives and hard examples) to produce the final detector. The re-training process using large number of hard examples significantly improves the detection accuracy of the detectors.

The HOG features were taken as the input of the SVM classifier and they were multiplied the detector generated in the learning phase. The result was shown in the form of a bench mark--output of the SVM classifier. For each widow, a bench mark will be obtained either +1 or -1. +1 means there is an object in the window.

D. Multi-Scale Detection

Targets in the image are spread in different locations and are of varying sizes. To detect all the targets in different sizes and improve the detection accuracy, multi-scale detection skill was employed in this research. The algorithm started to scan an ROI with the original size first and then resized the ROI by dividing the scale factor, which is 1.05 in this research. This goes on until the size of the ROI reaches certain criteria. Then the results of the detection in all scales were fused to determine the final existence and location of the target. In this way, targets of all detectable sizes, with height from 70 to 250 pixels, can be detected and some false positives can be filtered. Thus the performance of the algorithm was improved.

The performance of the algorithm mainly depends on the classification and the HOG computation of the images. According to the study in [13], in step 2, mask [-1, 0, 1] gives the best performance in gradients computation. Meanwhile, to trade off between the speed and accuracy, cells with size 8 by 8 pixels, and blocks with size 3 by 3 pixels are selected. Based on the test result in [13], the HOG descriptor used here is R-HOG descriptor. Algorithm flow chart is shown in Fig. 2.

III. IMAGE ACQUISITION AND PROCESSING

A. Image Acquisition Device

Infrared deer images were collected to verify the algorithm in Section II. The following Fig. 3 shows the image acquisition devices and Fig. 4 demonstrates the connection of the components. The device consists of an image grabbing and processing system and a thermal camera. The camera’s direction is driven by two stepper motors incorporating a motion control system to realize two degrees of freedom of motion, i.e. the lateral and vertical rotation. The range of the lateral (yaw) motion is ±180° and the vertical (pitch) motion is ±45° relative to the front direction of the box.
As shown in Fig. 4, the camera connects to the frame grabber through a BNC connector. The frame grabber connects to the computer through a USB port. Two motors are connected to the computer via a motion controller card through a serial port.

The infrared camera used in this research is a Thermal-Eye™ TSCss-FF camera \[19\]. This camera is specifically designed for security purposes. The camera has a 17.5° field of view (FOV), but is adequate for this application due to the motion generated by the motors. The camera requires a mount that incorporates the two stepper motors in a way that they can drive the camera with two degrees of freedom (lateral and vertical rotation of the camera’s view).

The frame grabber used to capture the video from the thermal camera is a Sensoray 2255 image grabbing card \[20\]. The card can acquire and capture frames from up to four video inputs which ensure the ability to process signals from up to 4 cameras simultaneously. The acquisition speed of the card can reach 30 frames per second, which guarantees the fluency of the grabbed image flow. The captured images can be in color or monochrome.

B. Image Processing

Deer images were acquired in the Lake Superior Zoo in Duluth, MN USA. The date is January 20th, 2011. The outdoor temperature is -20°F and the distance from the camera to the deer is about 50m. A total of 3000 images were obtained with deer in different postures, from which 200 positive examples were generated while the negative and hard examples produced are 1000. One of the deer images obtained is shown in Fig. 5(a). It is evident that in a gray scale thermal image, deer show higher intensity than the most part of the background. Through the illumination filter, corresponding parts were changed to pure black, and the obtained contours are shown in Fig. 5(b). It can be seen that there are still no deer in some high intensity areas. After the size filter, the regions of interest (ROI’s) are shown in Fig. 5(c). Then the HOG computation is applied only to the ROI’s. ROI windows are further sent to SVM classifier to identify deer. The identified deer are shown in Fig. 5(d). Since the area to process was significantly reduced, there was an obvious speedup in the processing of an image. At the same time, the accuracy was also improved.
IV. RESULTS AND DISCUSSION

A. Accuracy Test

The detection accuracy is defined as

$$\eta = \frac{N - (n_1 + n_2)}{N} \times 100\%,$$

where $N$ is the total number of the targets in images, $n_1$ is the number of the detected targets which are not deer and $n_2$ is the number of deer which were not detected. A further study showed that among 1500 images, there were 54 deer that were not detected and 30 non-deer objects that were misidentified as deer. This results in an accuracy of 94.2%. It was also observed that when a deer’s posture was at certain angle with respect to the camera and/or when part of the body of deer was behind a tree (is blocked from the camera view), the deer cannot be detected. Additionally, it was determined that the detected non-deer objects were in the shape very similar to deer. These conclusions produced the following questions: what is the suitable range of the body posture angle and what is the suitable range of the body shown percentage? These two questions will be answered in the following sections.

B. Influence of Body Angle

The current research has mainly focused on the detection of side-view deer. Most of the training samples are the side view deer. As a result, the algorithm is good at detecting deer in this posture. Some of the body postures at certain angles are shown in Fig. 6 by supposing the body angle is 0° at right-head side view, 90° at back view, 180° at left-head side view and 270° at front view. The processing results are also shown in Fig. 6 where bigger (red) rectangles show the detected objects and the smaller (blue) rectangles are the ROIs. It is found that when the angle of the deer falls in the range of $[70^\circ, 110^\circ]$ and $[250^\circ, 290^\circ]$ (note that the angle values are estimated values), deer would be difficult to detect. This, however, doesn’t mean that deer with these postures cannot be detected. If more training samples with deer posture angles in the range of $[70^\circ, 110^\circ]$ and $[250^\circ, 290^\circ]$, the algorithm would have the capability to detect these deer.

(c) Deer detected at 115°  
(d) Deer not detected at 110°  
Fig. 6 Influence body posture
C. Influence of Body Shown Percentage

Body shown percentage means the percentage of the deer body shown in the images relative to its whole body in one posture. It is a common scenario that part of a deer’s body is behind certain obstacles, such as trees. In such a case, only part of the deer’s body is shown in image. Some of such images are shown in Fig. 7.

In this calculation, a total of 600 images have been used in different deer body shown percentage. The tests were performed in several scenarios where different parts of the deer body were occluded by trees, rocks and other large objects. Some of the identification results are shown in Fig. 7, where the blue rectangle shown the RIOs and the red rectangles shown the identified objects. It is observed that the head and the front body of a deer influence detection much greater than the rear part of the body. As is shown in Fig. 8, if the head and the front body are not covered, the deer can be easily detected. On the other hand, without the information of the front body, the identification accuracy will be low. Noted is that if the deer bodies are in perfect side view posture (0° or 180°), the identification rate can reach 100%.

![Fig. 7 Influence of body occlusion](image1)

![Fig. 8 Detection rate due to deer body occlusion](image2)
D. Speed Test

The system will be used in real-time deer identification and tracking, which requires fast image processing speed. When the R-HOG method is used without the identification of ROIs, it takes around 1.7s to process an image in the size of 640 by 480 pixels. If only the ROIs are processed, the processing time for one image is from 0.05s to 0.3s depending on the size and number of the ROIs and in the average of 0.1s. At this processing speed, the control loop can reach 10Hz refreshing rate, which is fast enough for real-time detection.

V. CONCLUSION AND FUTURE WORK

Thermal imaging method has shown promise with respect to the identification of deer from their surroundings to help to reduce the accidents of vehicle-deer crashes. In this research, based on the traditional HOG and SVM pattern recognition method, a contour based HOG+SVM method has been developed. Based on the testing results, it can be concluded that this method has considerably reduced the time for object identification while drastically improving the detection accuracy. Since the process time on one image is limited to 0.1s averagely, the method will be suitable to be used in the real-time deer tracking and identification.

To improve the capability of the CNT-HOG algorithm in detecting deer, further research will be carried out to study the influence of different seasons under different weathers, especially under the conditions of temperature, lightning, rain, fog, etc. Furthermore, the method will be used in the highway roadside to detect the presence of deer to reduce the accidents due to deer – vehicle crashes. The exact location of the deer will also be calculated and sent to the driver’s GPS by implementing a stereo camera system to measure the distance angle between the deer and the camera.

ACKNOWLEDGEMENT

This project is supported by the Northland Advanced Transportation Systems Research Laboratory at the University of Minnesota Duluth.

REFERENCES