Pedestrian detection and vehicle type recognition using CENTROG features for nighttime thermal images

Martins E. Irhebhude
Deparment of Computer Science
Loughborough University, UK
Email: m.e.irhebhude@lboro.ac.uk

Mohammad Athar Ali
Deparment of Computer Science
Loughborough University, UK.
Email: m.a.ali2@lboro.ac.uk

Eran A. Edirisinghe
Deparment of Computer Science
Loughborough University, UK.
Email: e.a.edirisinghe@lboro.ac.uk

Abstract—This paper proposes a feature-based technique to detect pedestrians and recognize vehicles within thermal images that have been captured during nighttime. The proposed technique applies the support vector machine (SVM) classifier on CENsus Transformed histogRam Oriented Gradient (CENTROG) features in order to classify and detect humans and/or vehicles. Although thermal images suffer from low image resolution, lack of colour and poor texture information, they offer the advantage of being unaffected by high intensity light sources such as vehicle headlights which tend to render normal images unsuitable for nighttime image capturing and subsequent analysis. Since contour is the most distinctive feature within thermal images, CENTROG is used to capture this feature information and is used within the experiments. The experimental results so obtained were compared with those obtained by employing the CENsus TRansformed hISTogram (CENTRIST). Experimental results revealed that CENTROG offers better detection and classification accuracy for both pedestrian and detection and vehicle type recognition.

I. INTRODUCTION

Detection of humans and vehicle type recognition have always been popular application domains for computer vision techniques since they can be easily deployed in a number of scenarios and can be quite effective. These range from video forensics where the objective could be to analyze a crime scene to corporate bodies and military establishments, which might employ them for environment surveillance activities. Recently, human and vehicle detection have found use in application domains that incorporate such a requirement as part of their core functionality namely, intelligent transportation systems, smart vehicles and robotics.

This ever-increasing range of applications, especially in mission-critical situations or wherein human safety may be compromised, necessitates the development of a reliable and robust human detection and vehicle type recognition system. Consequently, a number of detection and recognition techniques have been developed and are already in use. However, techniques that were initially designed for daytime images fail when applied in their original form on nighttime images. The primary reason is that conventional nighttime images suffer either from low light conditions or from bright and intense light sources that tend to flood the entire image such as the dazzle of headlights from oncoming vehicles. Consequently, thermal images tend to offer a better alternative for analyzing nighttime scenes than conventional nighttime images. Thermal images, on the other hand, have their own drawbacks such as lack of colour and texture information, which may be the very features required by the aforementioned techniques. This paper thus proposes the use of contour-related feature extraction from thermal images which are largely unaffected by widely varying lighting conditions. The proposed feature-based technique classifies vehicles and detects pedestrians at nighttime using thermal images and is termed as CENsus Transformed histogRam Oriented Gradient (CENTROG).

The rest of the paper is organized as follows. Section II discusses the state-of-the-art with regards to detection of pedestrians and recognition of vehicles types both in daytime images and in nighttime thermal images. Section III explains the CENTRIST and CENTROG descriptors along with their usage. Section IV provides an overview of the proposed system followed by experimental and performance results in Section V. Correlation-based feature selection (CFS) and its implication in order to improve computational efficiency is explained in Section VI. The paper is concluded in Section VII.

II. RELATED WORKS

This section reviews a number of reported techniques that can detect pedestrians and recognize vehicle types within both, daytime and nighttime thermal images. Benezeth et.al. [1] proposed the use of a Gaussian-based segmentation method with Haar-like features using a cascade of boosted classifiers to detect humans in a room. Two contributions was made by Yun et. al. in [2]; segmentation based on histogram cluster analysis using k-means and a feature extraction technique based on histogram of maximal oriented energy map using log-Gabor wavelets for selecting orientation. An evaluation of the efficiency of nighttime midrange infrared sensor and its application in human detection and recognition was done by Bourlai et. al. in [3]. Local oriented shape context feature was used by Li et.al. in [4] to detect pedestrians in

a nighttime scenario by adding orientation information to shape context feature, thereby capturing appearance and shape information. In [5], Liu et. al. proposed a technique based on entropy weighted HOG as a feature detector and SVM as a classifier. They sped up the classification phase by reducing the number of support vectors and filtered false alarms by introducing a validation phase that examined the gray-level intensity of pedestrians heads. For thermal images, Chang et. al. [6] used HOG features and Adaboost to detect and classify pedestrians. Their feature extraction method included image segmentation and Region-of-Interest (ROI) generation. In [7], Riaz et.al. detected pedestrians within thermal images by using CENTRIST features and compared the performance of their technique with HOG-based techniques. Both above-reported techniques proved that CENTRIST-based approaches exhibit better detection accuracy with lesser computation time when compared to other methods. In [8], a feature combination of HOG and contour was proposed for pedestrian detection. The authors also proposed a foreground segmentation technique for smart region detection. In [9], Wang et.al. proposed a shape context descriptor (SCD) based on the Adaboost cascade classifier framework. The technique was applied to thermal images and the results were compared with the rectangle-based detection feature. The authors claimed that their technique outperformed the rectangle-based feature in terms of detection accuracy but suffers in terms of higher computation intensity.

A look at the state-of-the-art reveals a lack of techniques to recognize vehicle types in nighttime thermal image sets, though quite a few techniques have been reported for visible images. For instance in [10], Iwasaki et. al. reported a vehicle detection mechanism within thermal images using the Viola-Jones detector. The technique involved detecting the thermal energy reflection area of tires as a feature. This paper therefore aims to contribute towards bridging this research gap.

III. CENTRIST AND CENTROG DESCRIPTORS

Census Transformed Histogram for Encoding Sign Information (CENTRIST) is a visual description technique that was proposed by Wu et. al. [11] that is used to detect topological sections or scene categories. It extracts the structural properties from within an image while filtering out the textural details. It employs the Census Transform (CT) technique in which an 8-bit value is computed in order to encode the signs of comparison between neighbouring pixels. According to [12], CT is a non-parametric local transforms which can be described as follows:

Let P be a pixel, $I(P)$ its intensity (usually an 8-bit integer), and $N(P)$ the set of pixels in some square neighborhood of diameter d surrounding P. All non-parametric transforms depend upon the comparative intensities of P versus the pixels in the neighborhood $N(P)$.

Define $\xi(P, P')$ to be 1 if $I(P') < I(P)$ and 0 otherwise.

$R\tau(P)$ maps the local neighbourhood surrounding a pixel P to a bit representing the set of neighbouring pixels whose intensity is less than that of P. Therefore, census transform compares the intensity value of a pixel with its eight surrounding neighbours; in other words, CT is a summary of local spatial structure given by equation (1) [12]:

Let $N(P) = P$, where $\oplus$ is the Minkowski sum and D is a set of displacements, and let $\otimes$ be concatenation.

$$R\tau(P) = \bigotimes_{[i,j] \in D} \xi(P, P + [i,j]) \tag{1}$$

Example:

26	75	65		1	0	0
26	**46**	22	$\Rightarrow$	1		1	$\Rightarrow (10010111)_2 \Rightarrow CT = 151$
26	40	65		1	1	0

From example above, it can be seen that if the pixel under consideration is larger than (or equal) to one of its eight neighbours, a bit 1 is set in the corresponding location; else a bit 0 is set. The eight bits generated from intensity comparisons can be put together in order and converted to a base-10 value. This is the computed CT value for the pixel under consideration. The CENTRIST descriptor therefore is the histogram of the CT image generated from an image.

In order to compute the CENTROG features, after the image structure has been captured using CT, the Histogram of Oriented Gradients (HOG) is computed from the transformed image. The HOG works by counting the occurrences of gradient orientation in localized portions of an image. The HOG captures local object appearances and shape, which can often be characterized rather well by the distribution of local intensity gradients, or edge directions as reported in [7]. Gradient is computed by applying [1,0,1] and [1,0,1]T in horizontal and vertical directions within an image. Gradient information is collected from local cells into histograms using tri-linear interpolation. On the overlapping blocks composed of neighboring cells, normalization is performed. CENTROG descriptor therefore is the HOG on the CT generated image. The resultant images are shown below in Fig 1. Parts (a), (b) and (c) shows the; original, edge, CT-edge, CT and HOG on CT-edge images respectively.

IV. PROPOSED SYSTEM DESCRIPTION

The proposed system consists of pedestrian detection and vehicle type recognition subsystems. These are described as follows:

A. Pedestrian detection

Nighttime thermal pedestrian images of resolution 360×240 pixels were obtained from [13] and out of these, human figures were manually extracted as rectangular regions of 20×40 pixels. Further, additional 20×40 pixels were extracted as background data. Canny edge detection [14] was applied on the extracted images followed by computation of the CT. HOG features were then extracted and employed to train the SVM classifier for pedestrian detection [15], [16]. The flow diagram for pedestrian detection is shown in Fig. 2.

(a) Pedestrian sample

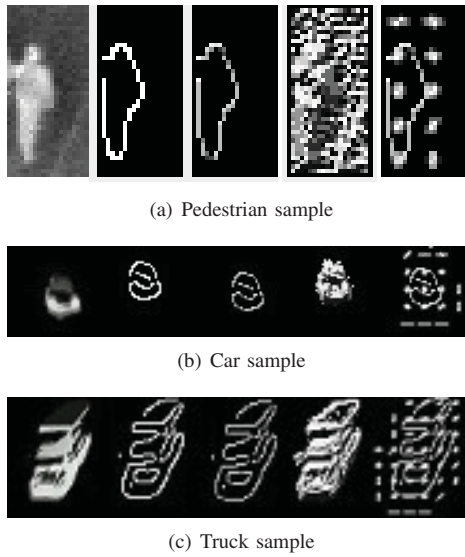(b) Car sample

(c) Truck sample

Fig. 1: Samples showing original image with processed images after CT, edge and HOG operations (a) Pedestrian sample, (b) Car sample, (c) Truck sample

Fig. 2: Proposed nighttime pedestrian detection technique

B. Vehicle type recognition

Nighttime thermal vehicle images were retrieved from a video dataset [17], which were then segmented using the Gaussian mixture model (GMM) based background subtraction technique [18]. The GMM technique uses a method to model each background pixel by a mixture of k-Gaussian distributions. The weight of the mixture represents the quantum of time for which the pixel values stay unchanged in a scene.

The resolution of the video dataset used was 720×480 pixels. Within these, a region-of-interest (ROI) from co-ordinate locations [127.5, 149.5, 401, 262] was extracted. These co-ordinate locations were selected since they represented a region of the image wherein the car was located when it was closest to the camera and hence offered the best view. This resulted in a ROI with a resolution of 401×262 pixels, which was then resized to 100×66 in order to maintain the aspect ratio. From the dataset used, two categories of trucks and cars were classified and used as a training set and a test set. Canny edge detection was applied on the extracted images followed by computation of the CT. HOG features were then extracted and employed to train the SVM classifier for pedestrian detection. The flow diagram is as shown in Fig. 3.
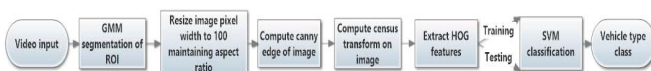

Fig. 3: Proposed nighttime vehicle classification technique

V. Experiment and Performance Results

A number of experiments were conducted to evaluate the performance of the proposed algorithm on pedestrian detection and vehicle type recognition within nighttime thermal images. The experiments were conducted on images retrieved from dataset given in [13] and used for pedestrian detection. Similarly, the video dataset given in [17] was used for vehicle type recognition. The results obtain from these experiments are discussed in the following subsections.

A. Pedestrian detection experiments

The following are the parameters associated with the image dataset used to capture images. The dataset comprises images captured under different environmental conditions.

TABLE I: Pedestrian Camera parameter

Name	Description
Sensor	Raytheon 300D thermal sensor core, 75 mm lens, Camera mounted on rooftop of 8-story building, Gain/focus on manual control.
Data	Pedestrian intersection on the Ohio State University campus, Number of sequences = 10, Total number of images = 284, Format of images = 8-bit grayscale bitmap, Image size = 360×240 pixels, Sampling rate = non-uniform, less than 30Hz.

Sections within these images consisting of humans were manually extracted. These were rectangular regions of 20×40 pixels. A total of 942 pedestrian image sections were extracted, half of which were used for training and the remaining half for testing. Similarly, a total of 2494 background image sections with dimensions of 20×40 pixels were also extracted and half of them were used for training and the remaining half for testing. Samples of extracted pedestrian and non-pedestrian image sets can be seen in Fig. 4.
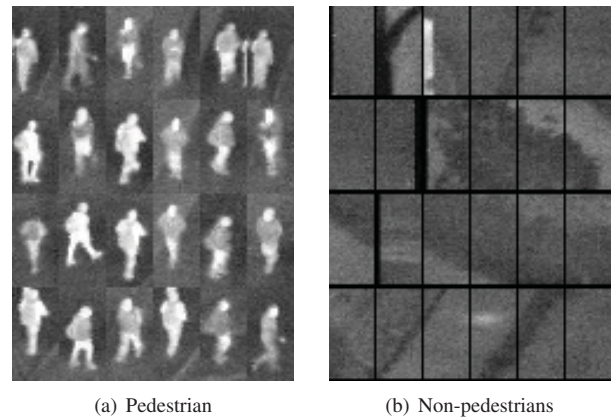

(a) Pedestrian   (b) Non-pedestrians

Fig. 4: Some examples of extracted 20×40 pixel pedestrian and non pedestrian images

The proposed technique uses a sliding window approach for annotating detected humans accordingly. To detect pedestrians in a given image sample, the whole image is scanned with a sliding window of width 20 pixels and a height of 40

pixels. Binary classification using SVM was conducted on feature sets of length 144. Experiments were conducted using CENTROG and compared with CENTRIST feature descriptor. Experimental results obtained showed that CENTROG (the proposed technique) outperformed CENTRIST by recording a detection accuracy of 97% versus 94%. Fig. 5 shows the results of detected pedestrians using the two approaches.

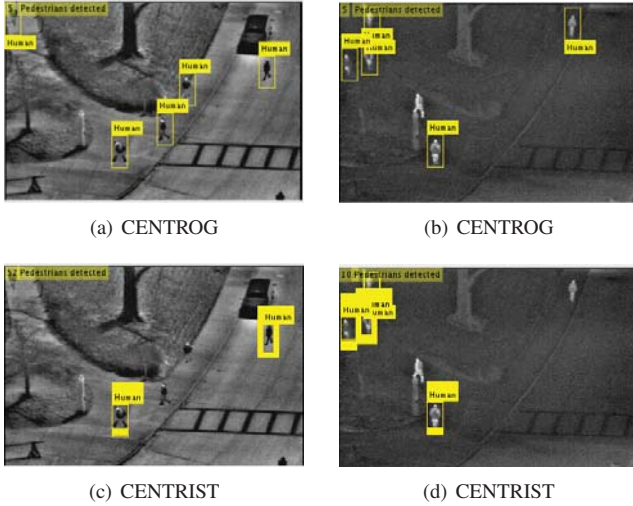(a) CENTROG   (b) CENTROG

(c) CENTRIST   (d) CENTRIST

Fig. 5: Detected pedestrians using (a), (b) CENTROG and (c), (d) CENTRIST

From Fig. 5, it can be observed that CENTRIST failed to detect some pedestrians and flagged a few false alarms, while CENTROG did not. However, CENTROG failed to detect one pedestrian due to an object that elongated the pedestrian in the image (see Fig. 5b, image on the right).

B. Experiments on vehicle type recognition

The following are specifications of the camera used to capture the video footage within the dataset provided:

TABLE II: Vehicle Camera parameter

Description
FLIR SR-19 Thermal Camera, White Box, Black Box, Total Video Footage Captured: 63 min of ROBB DRIVE and 1-80 OVERPASS

After segmentation using the GMM foreground/background subtraction technique, 650 truck and 650 car images were selected, half of which were utilized for training and the remaining half for testing (see Fig. 6a for an example of extracted vehicle image sets). Binary classification using SVM was conducted on feature sets of length 2772. A number of experiments were conducted using CENTRIST and CENTROG feature descriptor. Results from these experiments showed an accuracy of 100% for the CENTROG technique in contrast to 92.7% for the CENTRIST technique.

The CENTROG technique was tested on a number of randomly selected images, the results of which are depicted in Fig. 6b. As can be observed, CENTROG was able to

successfully recognize all vehicle types. This is in contrast to the application of the CENTRIST technique on the same data set which resulted in some cars being wrongly classified as a truck (see Fig. 6c).
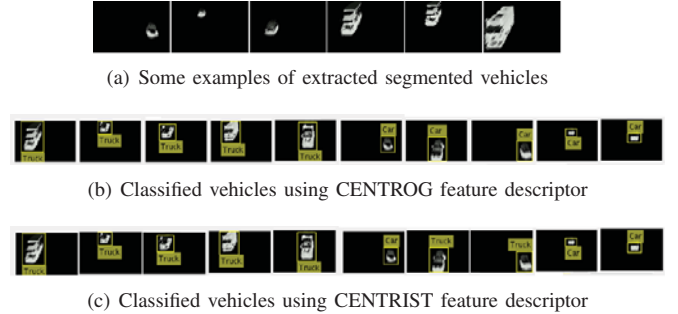

(a) Some examples of extracted segmented vehicles

(b) Classified vehicles using CENTROG feature descriptor

(c) Classified vehicles using CENTRIST feature descriptor

Fig. 6: (a) Segmented vehicles (b) Recognised Vehicles using CENTROG (c) Recognised vehicles using CENTRIST

C. Performance evaluation

A comprehensive performance evaluation was carried out of the proposed approach. The accuracy of the proposed technique was evaluated using:

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn} \qquad (2)$$

The accuracy is indicated by the Area Under the ROC Curve (AUC). An area of 1 represents the highest level of accuracy while an area of 0.5 represents the lowest. The most commonly used ranking system within published literature is as follows:

- 0.90-1 = excellent (A),
- 0.80-0.90 = good (B),
- 0.70-0.80 = fair (C),
- 0.60-0.70 = poor (D),
- 0.50-0.60 = fail (F).

In other words, the AUC is a measure of how well a parameter can distinguish between two contrasting groups of values.

Tables III and IV depict the performance comparison of CENTROG and CENTRIST feature descriptors whilst detecting pedestrians and recognizing vehicle types within the thermal image dataset. As can be observed, the CENTROG approach outperforms the CENTRIST technique in detecting both pedestrians and recognizing vehicle types.

TABLE III: Performance analysis on pedestrian detection

Technique	TP	FP	Precision	Recall	F-measure	AUC
CENTROG	97%	5%	97%	97%	97%	96%
CENTRIST	94%	8%	94%	94%	94%	93%

TABLE IV: Performance analysis on vehicle classification

Technique	TP	FP	Precision	Recall	F-measure	AUC
CENTROG	100%	0%	100%	100%	100%	100%
CENTRIST	92.7%	7.3%	92.9%	92.7%	92.7%	92.7%

Further experiments were performed on the pedestrian dataset by combining feature attributes. Experimental results proved that combining CENTROG and CENTRIST feature descriptors offered higher recognition accuracy in contrast to using these feature descriptors individually or combining with other feature descriptors. The results obtained are shown in Table V.

TABLE V: Recognition accuracies on pedestrian dataset using different feature combinations

Features	Recognition accuracy
CENTROG Edge	97%
CENTROG Image	86.6%
CENTRIST Edge	94%
CENTRIST Image	91.9%
CENTRIST Edge and Image	95.6%
CENTROG and CENTRIST	98.9%

VI. CORRELATION-BASED FEATURE SELECTION

In order to reduce the computation complexity and hence the computation time, a subset of discriminating features were chosen from the entire feature set and used within the experiments. It is noted generally that feature selection helps to improve machine learning. There are two approaches to feature selection; wrapper based and filter based approaches [19]. The method adopted here is the filter based approach; CFS. We chose CFS based approach as it performed better than the wrapper based approach and is not algorithm specific [19]. CFS filter algorithm helps to rank feature subsets according to the correlation based on the heuristic "merit" as reported by [20].

In [20] Lu et. al. reported CFS as:

$$M_s = \frac{k.\overline{r_{cf}}}{\sqrt{k + k.(k-1).\overline{r_{ff}}}} \quad (3)$$

where k is the number of features in current subset, r_{cf} is the mean feature-class correlation for each element of current subset, r_{ff} is the mean feature feature correlation for each pairwise of element. It begins with empty set and one at a time add features that holds best value. Best first search method is applied to get merit value.

After selecting discriminating features, table VI shows the; feature descriptors, total features, processing time required for testing, total number of features selected and recognition accuracies of various experiments.
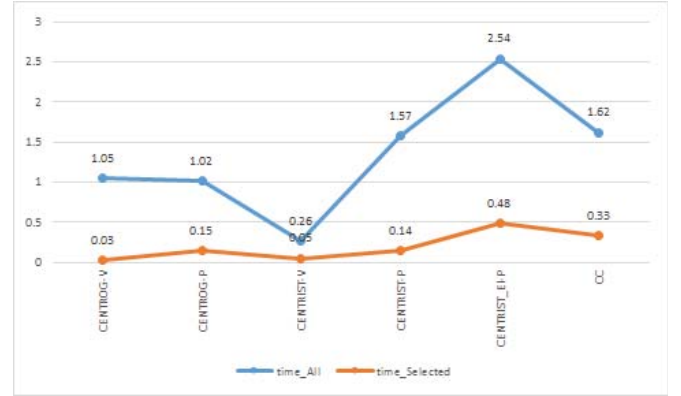
TABLE VI: Processing time and accuracy rates after feature selection

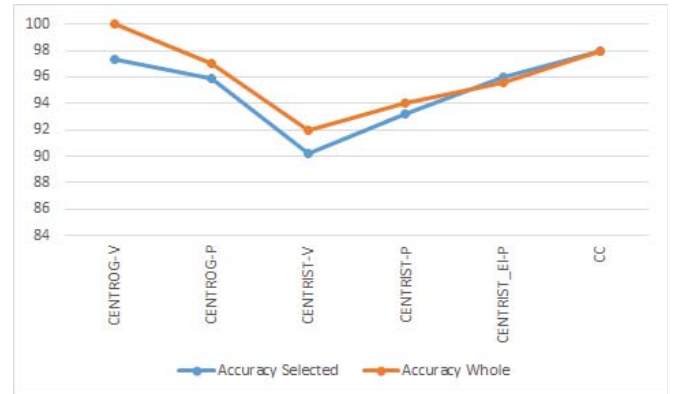| Features	TF	PT	Acc	SF	PT	Acc
CGV	2772	1.05 secs	100%	22	0.03 secs	97.3%
CGP	144	1.02 secs	97%	27	0.15 secs	95.9%
CTV	256	0.26 secs	92.7%	37	0.05 secs	90.2%
CTP	256	1.57 secs	94%	35	0.14 secs	93.2%
CTEIP	512	2.54 secs	95.6%	92	0.48 secs	96%
CGCTP	656	1.62 secs	98.9%	92	0.33 secs	98%

Notation used: CGV - CENTROG vehicle, CGP - CENTROG pedestrian, CTV - CENTRIST vehicle, CTP - CENTRIST pedestrian, CTEIP - CENTRIST Edge Image pedestrian, CGCTP - CENTROG and CENTRIST pedestrian, TF - total features, PT - processing time, Acc - accuracy, SF - selected features.

Fig 7 depicts the time required to build the model and the recognition accuracies before and after feature selection. As expected, there is a reduction in the time taken whilst maintaining the same level of accuracy. With the reduction in processing time, real-time implementation is realisable.

(a) Processing time using CENTROG and CENTRIST and their combinations descriptors.

(b) Accuracy rates of whole and selected features using CENTROG and CENTRIST and their combinations

Fig. 7: (a) Processing time using CENTROG and CENTRIST and their combinations descriptors. (b) Accuracy rates of whole and selected features using CENTROG and CENTRIST and their combinations descriptors. Notation used: v - vehicle, EI-P - edge+image pedestrian, CC - CENTROG+CENTRIST.

VII. CONCLUSION

This paper proposed a feature-based technique for pedestrian detection and vehicle classification in nighttime thermal images. The features were extracted by applying Histogram Oriented Gradient on Census Transformed images and hence termed as CENTROG. A linear SVM classifier was trained on the features obtained from the two datasets (pedestrian and vehicle). The proposed technique was implemented and compared with the CENTRIST technique. Experimental results

showed that CENTROG outperformed the CENTRIST approach in detecting pedestrians as well as recognizing vehicle types, thereby exhibiting a higher detection and classification accuracy. Further experiments revealed that combining CENTROG and CENTRIST feature descriptors offered the best performance. Finally the impact of the CFS on the processing time taken for detection and classification was also analyzed. Results indicated a significant reduction in time taken for detection and classification in contrast to employing the entire feature set. Reduction in processing time implies that the proposed technique can be employed in real-time detection and classification scenarios.

Future work could involve looking into identifying more categories, such as SUV, Sedan, truck and motorcycles etc.

REFERENCES

[1] Y. Benezeth, B. Emile, H. Laurent, and C. Rosenberger, "A real time human detection system based on far infrared vision," in *Image and Signal Processing*. Springer, 2008, pp. 76–84.

[2] T.-J. Yun, Y.-C. Guo, and G. Chao, "Human detection in far-infrared images based on histograms of maximal oriented energy map," in *Wavelet Analysis and Pattern Recognition, 2007. ICWAPR'07. International Conference on*, vol. 2. IEEE, 2007, pp. 933–938.

[3] T. Bourlai, J. Von Dollen, N. Mavridis, and C. Kolanko, "Evaluating the efficiency of a night-time, middle-range infrared sensor for applications in human detection and recognition," in *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2012, pp. 83 551B–83 551B.

[4] G. Li, Y. Zhao, D. Wei, and R. Cheng, "Nighttime pedestrian detection using local oriented shape context descriptor," in *Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering*. Atlantis Press, 2013.

[5] Q. Liu, J. Zhuang, and S. Kong, "Detection of pedestrians for far-infrared automotive night vision systems using learning-based method and head validation," *Measurement Science and Technology*, vol. 24, no. 7, p. 074022, 2013.

[6] S.-L. Chang, F.-T. Yang, W.-P. Wu, Y.-A. Cho, and S.-W. Chen, "Nighttime pedestrian detection using thermal imaging based on hog feature," in *System Science and Engineering (ICSSE), 2011 International Conference on*. IEEE, 2011, pp. 694–698.

[7] I. Riaz, J. Piao, and H. Shin, "Human detection by using centrist features for thermal images," in *International Conference Computer Graphics, Visualization, Computer Vision and Image Processing*. Citeseer, 2013.

[8] Y.-C. Lin, Y.-M. Chan, L.-C. Chuang, L.-C. Fu, S.-S. Huang, P.-Y. Hsiao, and M.-F. Luo, "Near-infrared based nighttime pedestrian detection by combining multiple features," in *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on*. IEEE, 2011, pp. 1549–1554.

[9] W. Wang, J. Zhang, and C. Shen, "Improved human detection and classification in thermal images," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*. IEEE, 2010, pp. 2313–2316.

[10] Y. Iwasaki, M. Misumi, and T. Nakamiya, "Robust vehicle detection under various environmental conditions using an infrared thermal camera and its application to road traffic flow monitoring," *Sensors*, vol. 13, no. 6, pp. 7756–7773, 2013.

[11] J. Wu, C. Geyer, and J. M. Rehg, "Real-time human detection using contour cues," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. IEEE, 2011, pp. 860–867.

[12] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in *Computer VisionECCV'94*. Springer, 1994, pp. 151–158.

[13] J. W. Davis and M. A. Keck, "A two-stage template approach to person detection in thermal imagery." *WACV/MOTION*, vol. 5, pp. 364–369, 2005.

[14] J. Parker, *Algorithms for Image Processing and Computer Vision*, 2nd ed. Indianapolis: Wiley Publishing, inc, 2011.

[15] C. Nakajima, M. Pontil, B. Heisele, and T. Poggio, "Full-body person recognition system," *Pattern recognition*, vol. 36, no. 9, pp. 1997–2006, 2003.

[16] S. Milan, H. Vaclav, and B. Roger, *Image Processing Analysis, and Machine Vision*, 3rd ed. Delhi: Cengage Learning, 2008.

[17] M. Smith, J. Gleason, S. Wood, and I. Beekun, "Vehicle location by thermal images features," Accessed 5th November 2014. [Online]. Available: http://cs426team11.github.io/vltif/

[18] Z. Zivkovic, "Improved adaptive gaussian mixture model for background subtraction," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, vol. 2. IEEE, 2004, pp. 28–31.

[19] M. A. Hall, "Correlation-based feature selection for machine learning," Ph.D. dissertation, The University of Waikato, 1999.

[20] Y. Lu, K. Boukharouba, J. Boonært, A. Fleury, and S. Lecœuche, "Application of an incremental svm algorithm for on-line human recognition from video surveillance using texture and color features," *Neurocomputing*, vol. 126, pp. 132–140, 2014.