IADIS International Journal on Computer Science and Information Systems
Vol. 8, No. 2, pp. 1-11
ISSN: 1646-3692

HUMAN DETECTION BY USING CENTRIST FEATURES FOR THERMAL IMAGES

Irfan Riaz. *Department of Electronics and Communication Engineering, Hanyang University, South Korea.*

Jingchun Piao. *Department of Electronics and Communication Engineering, Hanyang University, South Korea.*

Hyunchul Shin. *Department of Electronics and Communication Engineering, Hanyang University, South Korea.*

ABSTRACT

In this paper, we present a new human detection scheme for thermal images by using CENsus TRansform hISTogram (CENTRIST) features and Support Vector Machines (SVMs). Human detection in a thermal image is a difficult task due to low image resolution, thermal noising, lack of color, and poor texture information. For thermal images, contour is one of the most useful and discriminative information, so capturing it efficiently is important. Histogram of Oriented Gradient (HOG) is still the most proven way to capture the human contour. CENTRIST is a computationally efficient technique to capture contour cues as compared to HOG, but so far no one has implemented and tested the accuracy of CENTRIST descriptor for infrared thermal images. We developed CENTRIST based human detection system for thermal images and tested its variants. We also made a new dataset of thermal images, since there was no realistic dataset. Experimental results show that CENTRIST exhibits better detection accuracy than HOG, while reducing the training and the testing time significantly.

KEYWORDS

Human detection, vision, CENTRIST, HOG, thermal image.

1. INTRODUCTION

Human detection is of fundamental importance in computer vision due to its various applications that intersect with many aspects of human life. Computer vision based systems are becoming more feasible and affordable with recent advancement of technology. In near future, vision based systems will become an essential part of our lives e.g. driver assistance systems for smart cars, video surveillance, security, and robotics. Developing a reliable and robust human detection system is one of the most challenging tasks with lots of potential applications.

Pedestrian safety is an issue of global significance. In post-industrial countries, pedestrian fatalities are inevitable. The traffic fatality rate at night is about three times higher than that in day-time. In night-time, the poor lighting condition affects the driver a lot. There is an essential need for improvement of visibility under poor lighting and weather conditions e.g. at night, bad weather, under fog, and so on. To develop a splendid human detector under different lighting conditions and various weather conditions, thermal imaging can be one of solutions for relieving the problems of visible imaging techniques. Object detection in thermal domain has attracted more interests, as the infrared cameras become more affordable.

Only limited visual information can be captured by CCD cameras under poor lighting and weather conditions. Meanwhile, the brightness intensities of thermal images are representatives of the temperatures of object surface points. Pedestrians typically emit more heat than background objects, such as trees, road, etc. Image regions containing pedestrians or other "hot" objects will be brighter than the background. Hence theoretically, infrared thermal images can be a reliable source for human detection at night-time and in bad weather conditions. Thermal images, compared to visible images, lack several features, such as color and texture information, which plays a vital role for human detection and classification. The most discriminative and distinctive features of human beings from the background lie in their contours.

In this paper, we implement CENTRIST [1] feature based human detection method for thermal images. The goal is to make an appropriate feature descriptor which will be a part of our efficient human detection system. From our dataset, we extract HOG and CENTRIST features and use them to train two separate linear Support Vector Machines (SVMs). Based on that, a comparative analysis of both methodologies is carried out.

This paper makes four major contributions.

1. We implemented CENTRIST feature extraction on pedestrian thermal image dataset, and trained a linear SVM classifier to do per window based binary classification.
2. We compared the computational efficiency and detection accuracy of CENTRIST and its variants with HOG on our pedestrian thermal image dataset.
3. We highlight the extent of training dependency of the both techniques.
4. We have extended our pervious thermal image dataset [2] by including pedestrian scale and position variations. It make our dataset more realistic and results more reliable.

2. RELATED WORKS

Bertozzi [3] mentioned a human detection method as a part of the Advanced Driver Assistance System (ADAS). Various feature descriptors have been applied to human detection. However HOG is probably the most popular feature descriptor in human detection [4], [5], [6], [7], [8]. Recently the Local Binary Pattern (LBP) also shows high potential [8], [9]. A new trend in pedestrian detection is to combine multiple sources, e.g., color, local texture, motion, etc. [7], [8], [10], [11]. Wu and Nevatia [12] proposed a detection method that detects body parts by a combination of edgelet features and combines the responses of the part detectors to compute the likelihood of the presence of a person.

Kai and Arens [13] proposed a local-feature based human detector on thermal dataset. In the training phase, they used Speed Up Robust Features (SURF) [14]. Then a codebook is created by clustering these features and building Implicit Shape Model (ISM) to describe the spatial configuration of features relative to the object center. Wu et al. [1], [15] believe that contour is the most useful and discriminative information for pedestrian detection. Therefore they designed the CETRIST descriptor to detect human contour. Using the CENTRIST descriptor, they proposed the C4 algorithm that can accurately detect pedestrians. The phases of the method are suitable for parallel processing.

In terms of classifiers linear SVM is widely used for its fast testing speed. Recently SVM has been used in many application domains. It provides a supervised learning approach for object recognition such as faces [16], [17], face components [16], and pedestrians [18].

3. THERMAL IMAGE BASED HUMAN DETECTION

In this section we present our dataset and explain implementation of CENTRIST features and its variants. For the sake of completeness, a brief introduction to HOG features is presented. Training and testing methodologies are also discussed in detail.

3.1 Thermal Pedestrian Dataset

For making our dataset, we collected approximately 5 hours of 25fps night-time video scenario along the road which contains various types of pedestrians, including along-street, across-street, and bicyclists. The ambient temperature at the time of recording was about 21^{o}C. The NEC-C200 infrared thermal camera with 320*240 pixel image resolution was used for recording. After recording, the video was split into 5 frames per second and was up-sampled to 640x480 using cubic kernel. From the extracted frames, we cropped 2000 pedestrians as positive images; 1000 for testing and 1000 for training; number of positive images were doubled i.e. 2000 for testing and training each, by flipping the image about vertical axis. For each of the 2000 images in *testing dataset* eight slightly rescaled and translated images were generated; making 16000 images in testing dataset. Total 26000 negative images were also extracted out of which; 10000 were at random scale; 2000 were high mean, 2000 were high variance and 6000 were randomly selected. This is extension of our pervious thermal image dataset [2] by including pedestrian scale and position variations. It make our dataset more realistic and results more reliable. Based on careful observation of the pedestrian size and aspect ratio in the data set, we chose image window size of 64x128 for our experiment as shown in Fig 1.

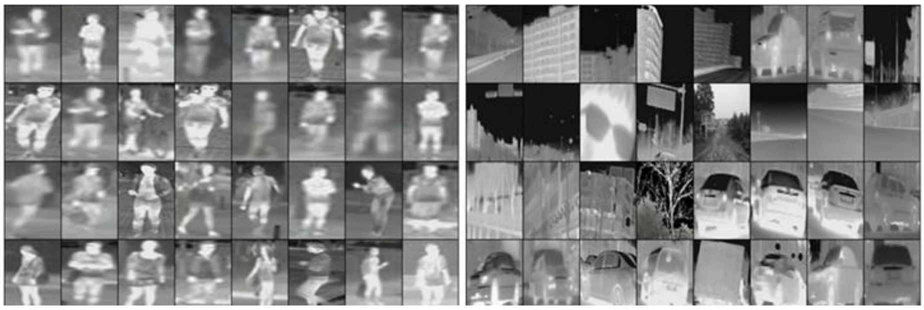

Figure 1. Pedestrian and non-pedestrian samples in the thermal image data set

3.2 HOG Features

Histogram of Oriented Gradients (HOG) is a popular feature descriptor for object classification, especially for human detection. The working philosophy behind HOG is local object appearance and shape can often be characterized rather well by the distribution of local intensity gradients or edge directions [4].

It basically counts occurrences (histogram) of gradient orientations in localized cells of an image, as shown in Fig 2. Gradient G_x and G_y is computed by applying [-1, 0, 1] and $[-1, 0, 1]^T$ in horizontal and vertical directions of image. Using this information gradient magnitude and orientation is calculated. Gradient information is collected from local cells into histograms using tri-linear interpolation. On the overlapping blocks composed of neighboring cells, as shown in Fig 3, normalization is performed. Extracted HOG features are robust to changes in lighting conditions and small variations in pose, thanks to the process of interpolation, local normalization, and histogram binning.

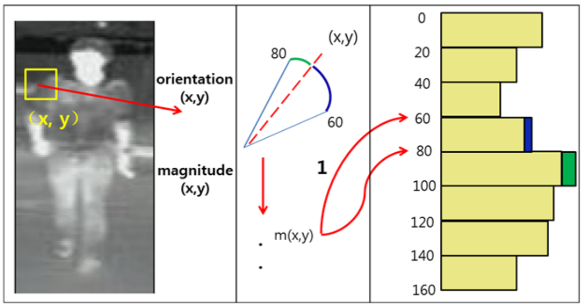

Figure 2. Histogram of Oriented Gradients.

3.3 CENTRIST Features

Wu et al.[1] presented a fast and efficient method of detecting humans by emphasizing on the human contour using a cascade classifier and the CENTRIST visual descriptor. The author claims that CENTRIST is particularly suitable for human detection, as it concisely encodes the sign information, and is able to capture large scale structures or contours. Also it detects humans at 20 fps speed on 640x480 resolution using only one processing thread and achieves accuracies comparable to the state-of-the-art.

In [1], Census transform is computed on the Sobel edge map of the visible images. During our experiments, we have found that computing Census Transform directly on the IR images (skipping the edge extraction stage) improves the detection accuracy.

A toy example for computing Census transform [9] is shown in Fig 3. For the pixel under consideration, in the neighboring 8 pixels, census transform finds pixel intensities that are greater than the intensity value of pixel under consideration and replace them with ‘0’; otherwise it replaces them by ‘1’. The CT value is computed by collecting bits from left to right and top to bottom and then converting them to base-10 value. CENTRIST descriptor, for the image I, is the concatenation of the histogram of CT values on the overlapping blocks composed of neighboring cells, see Fig 4.

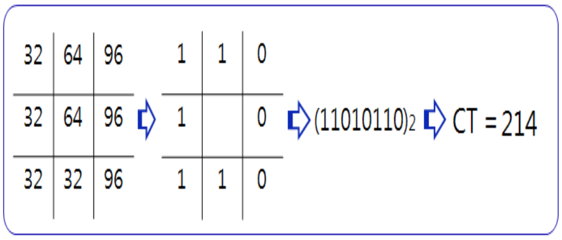

Figure 3. Census Transform

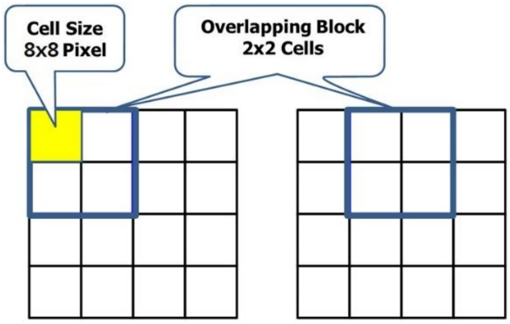

Figure 4. Cells and overlapping blocks

However, there are two major limitations of the conventional Census transform method. Firstly, Census transform produces unevenly distributed histograms and low frequency pattern types. Secondly, higher dimensional data may be very sparse, which makes it difficult to find any effective structure. It is therefore natural to ponder whether these feature vectors can be reduced without suffering great information loss. There would be many potential benefits due to reduced storage requirements, reduced computational complexity and possibly improved classification performance.

Inspired from [19], we learn the most frequently occurred Census patterns from our FIR dataset. The main idea is to learn the *dominant patterns* that occupy foremost proportions in terms of occurrence among all patterns. The assumptions here are

– The real images are of smooth nature and these dominant patterns are sufficient to reflect almost all textural structures in them.
– Noise cause the occurrence of “rare patterns” and using them as features will possibly mislead the classification.

Dominant patterns are extracted by a very simple approach. For FIR dataset, we compute the Census transform and then sort the patterns in descending order by their occurrence probability. We choose top ‘n’ patterns such that the sum of their occurrence probabilities reaches around 95 percent. Experimentally we have found that choosing 60 dominant patterns fulfills the above criteria. For computing the feature vector, by giving only dominant patterns a bin and discarding rare patterns, the number of bins for the histogram is reduced from 255 to 60. We refer this sparse version of CENTRIST as ‘CENTRIST-60’ features. Except the highly discriminant patterns “00000000” and “11111111”, by assigning a dominant pattern and its binary complement a single bin, the number of bins for the histogram is further reduced to 35, this even more sparse version of CENTRIST is referred by us as ‘CENTRIST-35’ features.

We use 64x128 pixels as the detection window size, 8x8 pixels as the cell size, and 2x2 cells as block size. We take any adjacent 2x2 cells as a block and extract a CENTRIST descriptor from each block. The overlap is kept 50% so there are 15x7 = 55 overlapping blocks, thus the feature vector for a candidate image patch has 15x7xn (feature vector length) dimensions. We don’t compute Census Transform for the border pixels of the detection window as the Census Transform requires at least 3x3 region hence borders are excluded while computing the CENTRIST descriptor.

The parameters used for implementation of CENTRIST and HOG descriptor are presented in Table 1.

Table 1. Parameter Settings

Parameters	**CENTRIST**	**HOG**
Window Size	128x64 pixels	128x64 pixels
Preprocessing	Histogram Equalization	Gamma Normalization
Cell Size	9x9 pixels	8x8 pixels
Block Size	2x2 cells	2x2 cells
Number of Bins (n)	256,60,35	9
Orientation Range	0^o-360^o	0^o-180^o
Overlap	50%	50%
Block Normalization	Nil	L2-Hys Norm
Overlapping Blocks	15x7	16x8
Feature Vector Length	15x7xn	4608
SVM	Linear	Linear

3.4 SVM Classifier

From a set of labeled training images, we extract HOG and CENTRIST features and use them to train two separate linear SVM’s. We have used the MATLAB’s `svmtrain` and `svmclassify` functions with their default settings for training and binary classification of testing data respectively. Details regarding training and testing methodology are presented in the following sections.

3.5 Training Methodology

For both feature vector descriptors, we have employed a similar training methodology as used in [4]. For initial training of SVM, we have used 2200 positive and 6000 (3000 at fixed scale and 3000 at random scale) negative sample windows of resolution 64x128. Our dataset consists of total 17,000 far infrared (FIR) images of up-sampled resolution 640x480, out of which 1500 are negative training images that are resampled to create hard examples for retraining of linear SVM. Retraining of SVM with hard examples reduces of false positive rate by almost 10%. Table 2 shows details of our dataset used for training and testing.

Table 2. FIR Data Set

Parameters	
IR Dataset Resolution	640x480 pixels
Negative Training Images	1500
Total images	17000
Sample window size	64x128 pixels
Initial Training	
Positive sample windows	2200
Negative sample windows	3000+3000=6000
Testing	
Positive sample windows	16000
Negative sample windows	20000

The training methodology consists of the following steps, shown in Fig 5.

1. Take initial positive and negative window examples from training dataset and generate label vector.
2. Generate a feature vector set by encoding all positive and negative windows with the selected feature vector descriptor.
3. Generate a linear SVM model, using feature vector set and label vector.
4. Using selected the descriptor and the SVM model, search 1500 negative training images exhaustively for false positives ('hard examples').
5. Augment the initial training data with collected hard examples and retain the SVM.

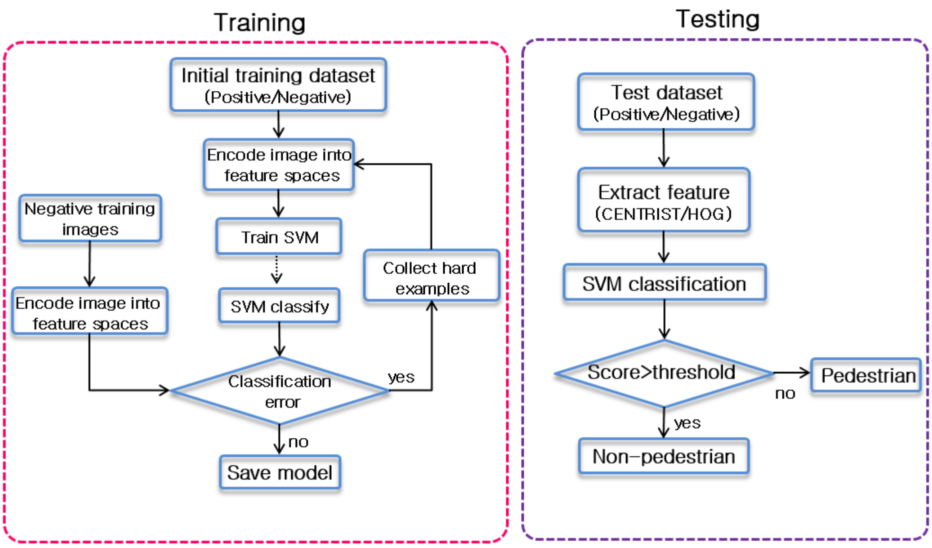

Figure 5. Training and Testing Methodology

3.6 Testing Methodology

A pedestrian can be detected in a scene by using brute force searching and testing of scale space in FIR camera based pedestrian detection systems. For example, all sliding window based models involve feature extraction, dense multi-scale scanning of detection windows, and binary classification, followed by non-maximum suppression [20]. The other way can be to use some simple tests to generate possible candidate locations and then verify them by using more sophisticated methods [21], [22].

For detecting pedestrians in a scene, depending upon the technique used, the number and the values of parameters (other than descriptor and classifier parameters) are quite diverse and their erroneous selection can make even a well-trained state of the art detector perform badly. Therefore, for an accurate evaluation of a descriptor, we strip it down to its basic functionality, as shown in Fig 5, for a given fixed-sized normalized window, the descriptor will extract important features which are fed to a pre-trained binary classifier. The classification accuracy is the measure of discriminative power of the descriptor under question.

4. EXPERIMENTAL RESULTS

For testing both detectors we have adopted per-window evaluation methodology. We measure the per-window (PW) performance on cropped positive and negative image windows based on equally trained binary linear SVM classifiers. To quantify detector performance we plot *Detection Error Tradeoff (DET)* curves on a log-log scale *i.e*. miss rate or false negative rate versus False Positive Rate (FPR). Lower values are better. They present the same information as Receiver Operating Characteristics (ROC's) but allow small probabilities to be distinguished more easily. We will use miss rate at 10^{-4} FPR as a *reference point* for results. For better readability in figure legends the detectors are sorted by their miss rate at the reference point. Lower positions are better.

Figure 6 presents a comparison of training dependencies of CENTRIST variants and HOG. They were trained on the fixed scale and tested the multiscale dataset. It seems HOG is more resilient to scale variations compared to CENTRIST and its variants. It also highlights the significance of training methodology and dataset in the performance of a detector.

Figure 7 presents a comparison when all detectors were trained and tested on multiscale dataset. This means negative windows were of random scale and in positive windows 20 percent scale variation of pedestrian size was allowed. All detectors were also tested against randomly introduced small translations in the position of pedestrians in the detection window. It can be seen that, if properly trained, all CENTRIST variants perform throughout better than HOG. CENTRIST-60 performs even better than its original version. It seems irrelevant and noisy features in higher dimensional data make it difficult for SVM to find any effective structure which misleads the classification results. CENTRIST-35 is better than HOG and least among other CENTRIST variants as far as accuracy is concerned but its lower computational cost makes it a noteworthy candidate for real-time applications.

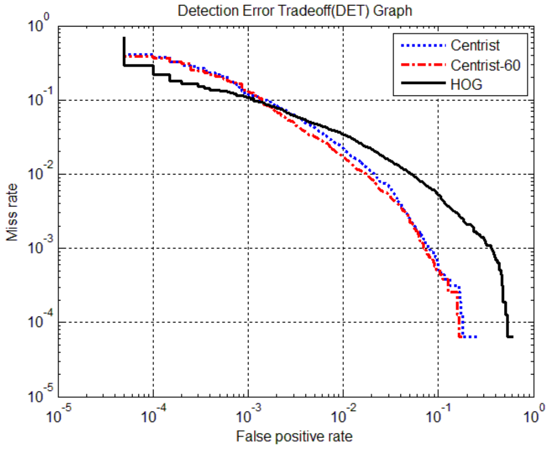

Figure 6. Comparison of CENTRIST variants and HOG. Trained on fixed scale and tested multiscale dataset.

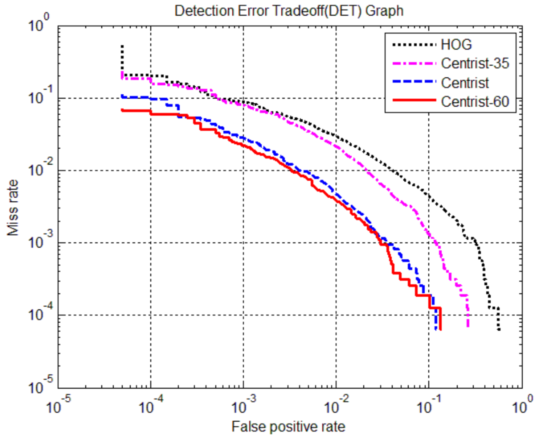

Figure 7. Comparison of CENTRIST variants and HOG on FIR pedestrian dataset. Trained and tested on multiscale dataset.

Table 3 presents the comparison of the computational complexities of both algorithms in terms of training and testing time. The reported training time only takes into account the time taken by the initial training and It does not include SVM retraining phase. It can be seen that CENTRIST variants reduce the testing and training CPU time quite significantly and delivers better detection accuracy than HOG. As CENTRIST captures the contour information more efficiently than HOG, we can add other complementary features, such as self-similarity and motion, to get better detection accuracy at lower complexity when compared to HOG based method.

The experiments were carried out on a single core Intel i5-2400 3.1 GHz CPU with 16 GB memory. The coding was done in MATLAB.

Table 3. Comparison of CPU time

Time	Training	Testing
HOG	326 sec	1407 sec
CENTRIST	61 sec	463 sec
CENTRIST-60	40 sec	256 sec
CENTRIST-35	34 sec	211 sec

5. CONCLUSIONS

In this paper, we have implemented CENTRIST and its variants based visual descriptor for pedestrian detection for thermal images. We have created pedestrian thermal image dataset, and trained a linear SVM classifier using CENTRIST features. An experimental comparison with HOG feature descriptor was carried out regarding the human detection rate and computation time. The comparative analysis shows that CENTRIST and its variants exhibits better detection accuracy than HOG also they cut the training and the testing time significantly. Hence for pedestrian detection in thermal images, CENTRIST and its variants seems to be better choice than HOG in detection accuracy and computational complexity. Furthermore, we can improve the recognition rate significantly by combining complementary information sources, e.g. motion and self-similarity with CENTRIST.

ACKNOWLEDGEMENT

This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korean Government (MOE) (NRF-2013R1A1A2004421). Moreover, the author Irfan Riaz was sponsored by 'Higher Education Commission (HEC) of the Government of Pakistan' under the scholarship program titled: Overseas Scholarship Scheme for PhD in Selected Fields Phase-I.

REFERENCES

[1] Wu, J. et al, 2011. Real-Time Human Detection Using Contour Cues. *In Proc. ICRA*, Shanghai, China, pp. 860-867.

[2] 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, 2013.

[3] Bertozzi, M., 2003. Pedestrian detection in infrared images. *In Proc. IEEE Intelligent Vehicles Symp.*, Columbus, OH, pp. 662-667

[4] Dalal, N. and Triggs, B., 2005. Histograms of oriented gradients for human detection. *In CVPR*, USA, pp. 886–893.

[5] Felzenszwalb, P.F. et al, 2008. A discriminatively trained, multiscale, deformable part model. *In CVPR*, Anchorage, Alaska. USA, pp. 1-8.

[6] Maji, S. and Berg, A.C., 2009. Max-margin additive classifiers for detection. *In ICCV*. Kyoto, Japan, pp. 40-47.

[7] Schwartz, W.R. et al, 2009. Human detection using partial least squares analysis. *In ICCV*, Kyoto, Japan, pp. 24-31.

[8] Wang, X. et al, 2009, An HOG-LBP human detector with partial occlusion handling. *In ICCV*, Japan, pp. 32-39.

[9] Mu, Y. et al, 2008. Discriminative local binary patterns for human detection in personal album. *In CVPR*, Anchorage, Alaska. USA, pp. 1-8.

[10] Doll´ar, P. et al, 2009. Integral channel features. *In BMVC*, London, England, pp. 1-11.

[11] Leibe, B. et al, 2005. Pedestrian detection in crowded scenes. *In CVPR*, San Diego, CA, USA, pp. 878–885.

[12] Wu, B. and Nevatia, R., 2007. Detection and tracking of multiple, partially occluded humans by bayesian combination of edgelet based part detectors. *International Journal of Computer Vision*, vol. 75, No. 2, pp. 247–266.

[13] Arens, M. and Jungling, K., 2009. Feature based person detection beyond the visible spectrum. *In IEEE CVPR Workshops*, pp. 30-37.

[14] Tuytelaars, T. et al, 2006. Surf: Speeded up robust features. *In Proc. 9th ECCV*,Graz, Austria, pp. 404-417.

[15] Wu, J. and Rehg, J.M., 2011. CENTRIST: A visual descriptor for scene categorization. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 33, pp. 1489-1501.

[16] Heisele, B. et al, 2001. Face recognition with support vector machines: Global versus component-based approach. *In*

ICCV, Vancouver, BC, Canada, pp. 688-694.

[17] Osuna, E. et al, 1997. Training support vector machines: An application to face detection. *In CVPR,* San Juan, Puerto Rico, pp. 130–136.

[18] Mohan, A. and Poggio, T., 2001. Example-based object detection in images by components. *IEEE Trans.Pattern Anal.Machine Intell.,* vol. 23, pp. 349–361.

[19] Yimo Guo, Guoying Zhao, and Matti PietikäInen, 2012. Discriminative features for texture description. in *Pattern Recognition*., vol. 45, pp. 3834-3843.

[20] Timo Ojala, Matti Pietikäinen, David Harwood, 1996. A comparative study of texture measures with classification based on featured distributions, *Pattern Recognition*, Volume 29, Issue 1, Pages 51-59.

[21] Dollar, P. et al, 2012. Pedestrian Detection: An Evaluation of the State of the Art. *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol.34, no.4, pp.743-761.

[22] Walk, S. et al, 2010. New features and insights for pedestrian detection. *In CVPR,* San Francisco, USA, pp.1030-1037.

[23] Miron, A. et al, 2012. Intensity self similarity features for pedestrian detection in Far-Infrared images. *Intelligent Vehicles Symposium,* pp.1120-1125.