This document is published in:

2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2012, pp. 3067 - 3072. DOI: 10.1109/IROS.2012.6385928

Discrete Features for Rapid Pedestrian Detection in Infrared Images

Daniel Olmeda, Jose Maria Armingol and Arturo de la Escalera

Abstract— **In this paper the authors propose a pedestrian detection system based on discrete features in infrared images. Unique keypoints are searched for in the images around which a descriptor, based on the histogram of the phase congruency orientation, is extracted. These descriptors are matched with defined regions of the body of a pedestrian. In case of a match, it creates a region of interest in the image, which is classified as a pedestrian / non-pedestrian by an SVM classifier. The pedestrian detection system has been tested in an advanced driver assistance system for urban driving.**

I. Introduction

Detecting people in images is applicable in areas such as surveillance and road safety in vehicles, whether manned or not. In the latter case, one faces a number of additional problems. Vehicles must be able to run both in night and day conditions. Moreover, in these applications, it is extremely important that the processing of the information contained in the images is done quickly, and on platforms with limited computing power. In this paper, the authors propose a method for pedestrian detection in low visibility conditions based on the evaluation of discrete phase congruency features in far infrared images. This approach allows focusing computational resources in the regions of the image most likely to contain pedestrians, greatly reducing processing time.

Usually, pedestrian detection algorithms in infrared images are based on temperature thresholding or edges information. The first approach allows for a preprocessing step, where regions of interest are extracted from the image. Only hot regions are further processed. On the second, images are densely scanned, using a sliding window approach [1], which achieves greater detection rates at the cost of computation time.

Far infrared cameras are useful for night driving safety systems, as the output of those cameras is a projection on the sensor plane of the emissions of heat of the objects, that is proportional to the temperature. Most systems take advantage of this characteristic and select the regions of interest based on the distribution of warm areas on the image [2] [3] [4].On systems that search for the temperature distribution, the discriminating feature of pedestrians would be the shape of the object. Regions of interest are correlated with some predefined probabilistic models in [5] and [6]. This approach requires a calibrated sensor to accurately threshold temperatures.

Another important feature is the intensity of the borders between pedestrians and their background. Those are used in systems that select regions of interest by the proximity of local shape features such as edgelets [7], or multi-block LBP [8], or Histogram of Oriented Gradients [9] (HOG) descriptor. This descriptor is thought to be used as a generic object classifier, but has achieved outstanding results in people detection in visible images. Their work, inspired by the SIFT descriptor [10], calculates a dense grid of histograms of orientations for each region of interest in the image. This grid approach allows the recognition of patterns of spatial characteristics for shape recognition. Their implementation classifies the candidates with a two-class linear Support Vector Machine, discriminating between pedestrians and non-pedestrians. Variations of this approach has been used in [3], [11] and [12] for pedestrian detection in infrared images. HOG features are based on the first derivative, which is then normalized within blocks. This makes it changeling to cope with a wide range of temperatures, for this kind of images.

The authors have found that simple gradients in long wave infrared images are not enough to satisfactorily define the shape of pedestrians. This is due to the much wider infrared spectrum, compared with visible light. Another difficulty is that the sensitivity curve of an uncooled microbolometer sensor changes very quickly with minimum changes of its temperature. To overcome this challenges, the authors propose a contrast invariant descriptor for object detection.

The features should be invariant to illumination, but also to scale, to successfully identify small objects. In this case, pedestrians that are far away from the camera. The theory of phase congruency [13] in signal analysis provides such as invariance. The resulting features are proportional to the local symmetry in a way that doesn't depend on the image contrast. As such, the resulting edges are not biased by the temperature difference between them and the background. Because these features doesn't depend on the contrast or the object temperature it is possible to achieve also some invariance to the sensor's temperature.

The method proposed in this paper consists of two steps. In first step, only regions with a high probability of containing a pedestrian are selected, avoiding a deep analysis of areas of the image without any interest. The approach followed in [14] pursues this goal through a cascade classification. This speeds up processing, since little time is spent in regions with little chance of containing the object sought. However, this article proposes that the analysis be reserved for image

*This work was supported by the Spanish Government through the Cicyt projects FEDORA (GRANT TRA2010-20225-C03- 01) and Driver Distraction Detector System (GRANT TRA2011-29454-C03-02), and by the Comunidad de Madrid through the project SEGVAUTO (S2009/DPI-1509).

The authors are with Department of Systems Engineering, Universidad Carlos III de Madrid, C/ Butarque 15, Leganes, Madrid, Spain. dolmeda at ing.uc3m.es

regions that have certain characteristics. In particular, discrete features are searched, from which larger regions of interest that may contain pedestrians are generated. In the second step, these regions are classified by a method inspired by the histogram of oriented gradients [9].

This paper is organized as follows. Section II states briefly the theory of phase congruency in images. Section III describes the selection of unique points in the images and extraction of its descriptors. The location of these discrete features is used to create regions of interest (ROIs) in areas most likely to contain a pedestrian. Section IV describes the classification of these ROIs. The results are described in section V. Finally, section VI discusses the conclusions and future work.

II. Phase Congruency Theory

Points of high phase congruency are those in which a wide range of their Fourier components is in phase.Decomposition of smooth areas has its frequencies spread over a wider range, thus being its phase congruency score lower.

A set of frequencies are extracted from the image, by convolving the fourier transform of the image with a set of filters. Each of these filter extract the information at a narrow range of frequencies. Because the filters have to be used over the fourier transform, a complex signal, they have to be complex too. In this case, a set of Gabor filters.

The real and imaginary parts of the one-dimensional Gabor filter are given by equations 1 and 2.

$$O_\theta = \sin(\theta) \cdot \frac{1}{\sigma \cdot \sqrt{2\pi}} e^{\phi} \tag{1}$$

$$E_\theta = \cos(\theta) \cdot \frac{1}{\sigma \cdot \sqrt{2\pi}} e^{\phi} \tag{2}$$

Where μ is the mean, σ is the standard deviation and ϕ is:

$$\phi = \frac{-(x-\mu)^2}{2\sigma^2} \tag{3}$$

The amplitude of the image at the frequency of the filter is the square mean of the convolution of the image with the odd and even filter (equation 4) ,

$$A_\theta(x) = \sqrt{(S(x) * O_\theta)^2 + (S(x) * E_\theta)^2} \tag{4}$$

where $S(x)$ is the image at point x, E_f the even Gabor filter and O_f the odd one.

The phase of the image is given by equation 5.

$$\phi_\theta(x) = \arctan(S(x) * O_\theta, S(x) * E_\theta) \tag{5}$$

Because a convolution in the space dominium is a product in frequency, the filters can be applied to the image once it is transformed to its Fourier decomposition. The filters only extract a fixed orientation of the image features, so it is necessary to apply a set of filters, each with a different orientation.

After applying all the filters the weighted mean of phase for each point in the image is calculated. This value maximizes equation 7 and determines the phase congruency score.

$$F(x) = \sum_n A_n \cos(n\omega x + \phi_n - \theta) \tag{6}$$

$$PC(x) = max_{\theta \in [0,2\pi]} \frac{F(x)}{\sum_n A_n} \tag{7}$$

Where $F(x)$ is the Fourier series expansion of the signal.

III. Discrete keypoints

A. Extraction of descriptors

For each pixel in the phase coherence image the maximum and minimum moment of covariance is calculated. The maxima in this minimum moment image correspond to corners. As a normalized image, and also invariant to the contrast, these points can be drawn thresholding the image with a fixed th. After a geometric operation "open", pixels adjacent to each other are grouped into blobs. The position of the keypoint is the center of gravity of the blob. Around these keypoints a descriptor based on the histogram of the phase coherence is extracted. In Fig. 1, an example of the thresholding of corner maxima is shown.

The orientation image O is decomposed into n subimages, corresponding to the division into equal angle ranges from 0 to π (eq. 8).

$$O_i = \begin{cases} 1 & \text{if } (i-1) \cdot \frac{\pi}{n+1} < O < i \cdot \frac{\pi}{n+1} \\ 0 & \text{in other case} \end{cases} \tag{8}$$

Where n is the number of bins in the orientation histogram and O_i are the orientation subimages from O_1 to O_n.

From each thresholded orientation image an equal number of magnitude images M_i are created, multiplying, element by element, the phase coherence image with the filters. Each contains the pixels with orientations falling within each range (eq. 9).

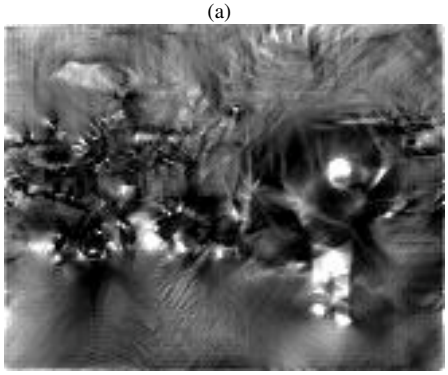
$$M_i(x,y) = \begin{cases} M(x,y) & \text{if } O_i(x,y) = 1 \\ 0 & \text{in other case} \end{cases} \tag{9}$$

For each M_i image, an integral image is computed, as shown in equation 10. This reduces the number of calculations in later stages of the algorithm.

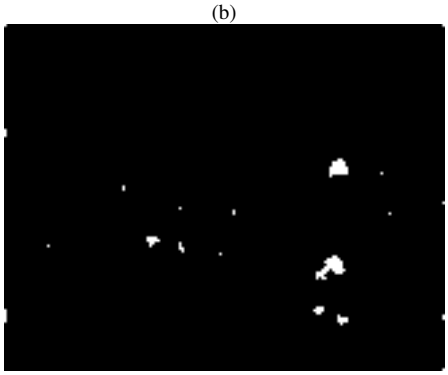
$$I_i(x,y) = \sum_{\substack{x' \le x \\ y' \le y}} M_i(x',y') \tag{10}$$

Each of the keypoints is associated with a descriptor at each scale. The descriptor is the concatenation of the histograms of phase coherence of surrounding cells, wherein each of the bins, corresponds to one of the orientations. Histograms block is a $3x3$ cell matrix. To accelerate the calculation of descriptors, rather than resizing the image at different scales, the size of the cell are rescaled. Fig. 2 is an example of three descriptors around the same keypoint, at different scales.

(a)



(b)



(c)

Fig. 1: a) Original image, b) Corners image, c) Thresholded corner image

The algorithm for the descriptors computation can be described as:

for each cell in block **do**
$w =$ width of the cell, in pixels
$j =$ column of the cell inside the block
$i =$ row of the cell inside the block
for each orientation k **do**
$descriptor(cell * n_{bins} + k) = (I_k(i+w, j+w) + I_k(i,j) - I_k(i,j+w) - I_k(i+w,j))$
end for
end for

B. Matching of descriptors with body parts

From a database composed of 2400 ROIs containing pedestrians in infrared images at different scales, a new classifier is trained for each block in the ROIs.

The dimensions of the blocks in each region of interest in the training set are selected so that all samples have the same number of features. Experimentally it has been found that the best results are obtained by dividing the region of interest in $(10x5)$ non-overlapping blocks. In figure 3 is shown each of the blocks in which the samples are divided, over the average magnitude of positive training set samples.

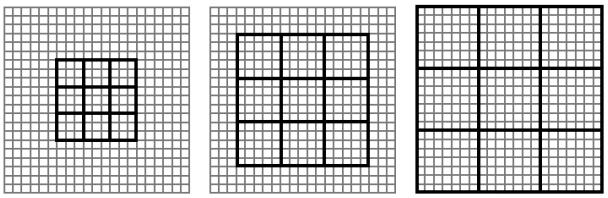


Fig. 2: Descriptors at different scales around the same keypoint

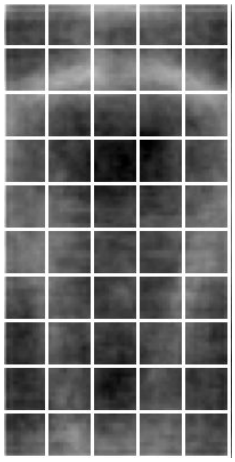


Fig. 3: Descriptor blocks of pedestrian parts. For each block an SVM classifier is trained.

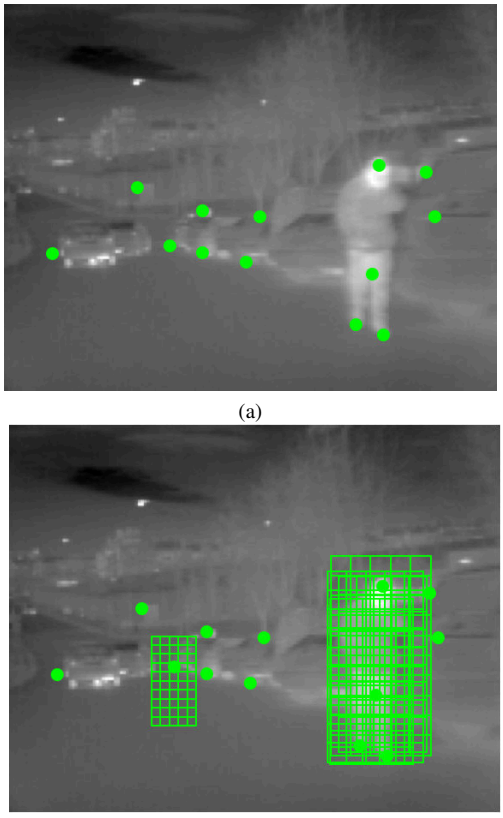
Each of the descriptors obtained in the previous step is evaluated by the 50 pedestrian-parts classifiers. The classifiers are ordered by their correct rate. The descriptor will be evaluated in descending order until a match occurs. For those with a positive result, a new region of interest is created, at the corresponding scale.

Figure 4a shows an example of unique keypoints in an infrared image. Figure 4b shows the descriptors that have a match with a pedestrian part, as well as the regions of interest generated. Those Rois that do not meet certain geometric constraints (i.e. regions that fall outside the image) are removed.

The resulting regions are fed to the next step of the algorithm, for a thorough classification.

IV. Classification of regions of interest

The decision about whether or not the regions contains pedestrians is done by an SVM classifier. The feature vector consists of the descriptors that make up each of the parts of the pedestrian described in section III, concatenated.



(a)

(b)

Fig. 4: Unique keypoints and the regions of interest they generate.

The final descriptor depends on the parameters selected to create the phase congruency image, the size and binning of the window and the number of orientations of the histogram. The following parameters, along with the values tested are those found to most affect to classification:

- Number of scales: indicates the range of frequencies to be extracted by the filters.
  - $N_{scales} = \{2, \cdots, 6\}$
- Number of orientations: for each scale the filters are rotated to a number of angles to extract information at different rotations.
  - $N_{orientation} = \{2, \cdots, 6\}$
- Spread of two-dimensional Gabor filter.
  - $\sigma = 2 + 2 \cdot N_{scale}$
- Histogram orientations binning: number of bins of the histogram of orientations for each cell.
  - $B = \{4, 9, 12\}$
- Normalization of blocks of adjacent cells.
  - $Norm = \{None, L_2\}$

The size of the roi is also an important factor. Pedestrians that are far away from the camera appear very small on the image. The smallest scale at which pedestrians can be successfully classified is $(30x15)$ pixels.

SVM classification calculates the boundary between the pedestrian/non-pedestrian classes by searching the hyper-plane that maximally separates the training set in a high-dimensional space. For training the SVM is necessary to have a representative sample of feature vectors of both classes. In this case, pedestrians and non-pedestrians images of the same size. To better represent the shape of a pedestrian the training set contains samples shot at different temperatures as well as pedestrians located both near an far from the camera. The number of samples if approximately the same for each season.

The training data set contains 5000 samples, manually classified and assigned a binary label $l = \{-1, 1\}$. Each of this vector samples is a concatenation of the histograms of all the bins in the cropped image.

Over the best sets of parameters, we search for the best classifier by varying the kernel function, as well as the C parameter of margin softness. Finally, the classifiers are evaluated based on the area under their ROC curves. The set of parameters tested has been:

- Kernel:
  - Radial Basis Function.
  - Linear.
  - Quadratic.
- $C = \{10^{-2}, 1, 10^3, 10^6\}$

It is possible that the same pedestrian is detected repeatedly. In this case, the overlapping regions are grouped to form a single detection.

V. Results

The results of the classification are divided into two parts. First, the results of the SVM classifier applied to the database, along with the best selection of parameters, are discussed. Then, the results of the search for pedestrians over full images, taking into account the error introduced by the selection of regions of interest step.

A. SVM classifier

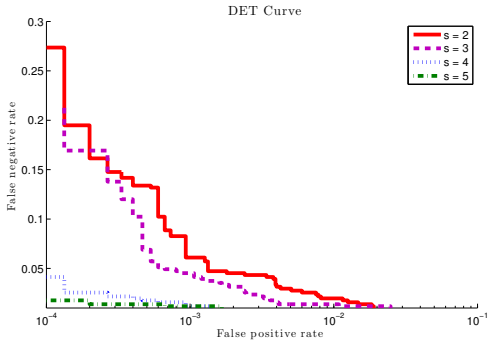
The parameter that affect the most to the calculation of the histogram of orientations are the number of bins and cell size. The windows are divided in $\{5 \times 10\}$ unnormalized cells. The normalization of the cells degrades the performance because phase congruency is in itself a normalized magnitude.

Several SVM kernels have been tested with the database. The best results have been achieved using a Radial Basis Function with scaling factor $\sigma = 25$, though a simple quadratic kernel performs almost as good and needs less time to compute. A linear kernel achieves slightly less performance, but can generalize better the problem, in images shot as extreme temperatures, such as in midday summer.

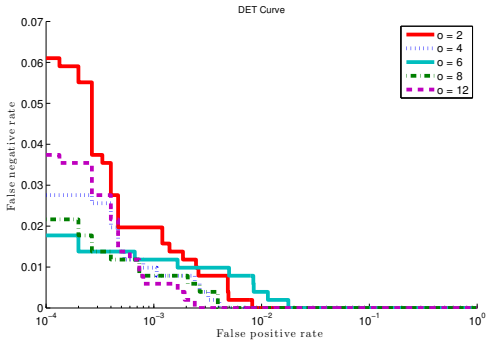
Figure 5 shows the DET (Detection Error Trade-Off) curves of the classifier for various scales and orientations of the phase coherence image.

B. Discrete keypoints results

To evaluate the result of pedestrian detection in an image, it is considered that detection is successful if the area of the detected pedestrian coincides by 80% with the manually annotated bounding box, and is in the same, previous, or



(a) Scales



(b) Orientations

Fig. 5: DET curves of the SVM classifier for various parameters

next scale. Figure 6 shows multiple detections for a single pedestrian (in green) and the correct bounding box (in red). From all positive detections, the final targets are selected by means of a mean shift non-maximum suppression algorithm as described in [15].

Fig. 6: Multiple positive detection for the same pedestrian (in green), and the manually annotated bounding box (in red).

Figure 7 shows the error rate introduced by the selection of regions at different scales. That is the percentage of pedestrians in the training sequence for which no region of interest has been generated. Note that the error obtained on the scale 6 is quite large. At this scale regions are of size $(30x15)$ pixels.

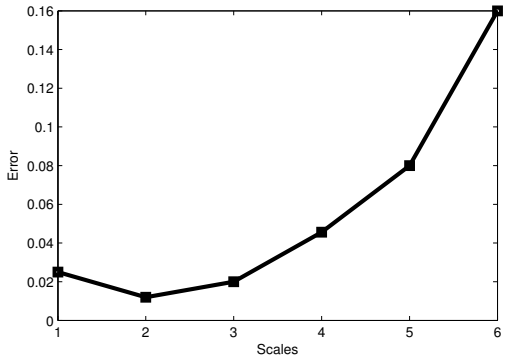


Fig. 7: Error rate introduced by the selection of regions at different scales.

VI. Conclusions and Future Work

In this paper, the authors have presented a new descriptor for classification of pedestrians in far infrared images. This approach exploits information from low resolution, uncalibrated, non-refrigerated microbolometer sensors to detect the presence of people in the trajectory of a moving vehicle. The main application of the system is to be used while driving at night, though our test prove a good performance in a wide range of temperature and illumination conditions.

In Fig. 8 the best parameters classifier of the approach presented is compared to the standard HOG algorithm as presented in [9]. Both classifiers have been trained on the same database and its performance tested on the same image set. Furthermore, the algorithm has also been tested on the OSU Thermal Imagery Database [16], with almost perfect results.

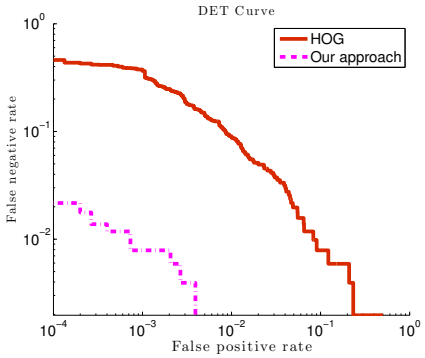


Fig. 8: Comparison of performance of our best classifier and standard HOG, both trained on the same database.

On live testing in urban environments, and on board a vehicle equipped with a Intel Core 2 Duo 2Ghz, the processing time has take an average of framerate of $60ms$, or 16fps.

Fig. 9 contains some examples of pedestrians being detected from the experimental vehicle while driving through

Fig. 9: Examples of processed images with pedestrians being tracked. In figures (a) and (b) it can be seen the ghosting effect derived from the lateral movement of the car. Figure (c) contains a critical situation in a pedestrian crossing. In figure (d) pedestrians are grouped and partially occluding one another.

urban environments. In figures 9a and 9b it can be seen the ghosting effect derived from the lateral movement of the car. For each pedestrian found an Unscented Kalman Filter is created to track its position. If there are multiple detection a single pedestrian, the filter merges them as a single location. The driver is then warned if the pedestrians trajectory intersects that of the vehicle. False positives are more prone to appear in images shot at very high temperatures.

Finally, the authors are working on feature fusion with other fast, low level descriptors, such as Haar, LBP, or more recently Image Self Similarity [17] to generate regions of interest via a cascade rejector approach.

REFERENCES

[1] C. Papageorgiou and T. Poggio, "A trainable system for object detection," *International Journal of Computer Vision*, vol. 38, no. 1, pp. 15–33, 2000.

[2] M. Bertozzi, A. Broggi, and P. Grisleri, "Pedestrian detection in infrared images," *Intelligent Vehicles Symposium, 2003. Proceedings. IEEE*, pp. 662–667, 2003.

[3] M. Bertozzi, A. Broggi, M. del Rose, M. Felisa, A. Rakotomamonjy, and F. Suard, "A pedestrian detector using histograms of oriented gradients and a support vector machine classifier," *IEEE Intelligent Transportation Systems Conference*, 2007.

[4] E. Binelli and Broggi, "A modular tracking system for far infrared pedestrian recognition," in *Intelligent Vehicles Symposium, 2005. Proceedings. IEEE*, 2005, pp. 759–764.

[5] D. Olmeda, A. de La Escalera, and J. M. Armingol, "Far infrared pedestrian detection and tracking for night driving," *Robotica*, 2010.

[6] H. Nanda and L. Davis, "Probabilistic template based pedestrian detection in infrared videos," *Intelligent Vehicle Symposium, 2002. IEEE*, vol. 1, pp. 15–20 vol.1, may 2002.

[7] J. Li, "Pedestrian tracking in infrared image sequences using wavelet entropy features," *Computational Intelligence and Industrial Applications, 2009. PACIIA 2009. Asia-Pacific Conference on*, vol. 1, pp. 288–291, 2009.

[8] D. Xia, H. Sun, and Z. Shen, "Real-time infrared pedestrian detection based on multi-block LBP," in *Computer Application and System Modeling (ICCASM), 2010 International Conference on*, 2010.

[9] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, 2005.

[10] D. G. Lowe, "Object recognition from local scale-invariant features," *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 2, pp. 1150–1157 vol.2, 1999.

[11] R. Miezianko and D. Pokrajac, "People detection in low resolution infrared videos," *Computer Vision and Pattern Recognition Workshops, 2008. CVPR Workshops 2008. IEEE Computer Society Conference on*, pp. 1–6, may 2008.

[12] L. Zhang, B. Wu, and R. Nevatia, "Pedestrian Detection in Infrared Images based on Local Shape Features," *Computer Vision and Pattern Recognition, 2007. CVPR '07. OTCBVS Workshop.*, pp. 1–8, may 2007.

[13] P. Kovesi, "Image features from phase congruency," *Videre: Journal of Computer Vision Research*, vol. 1, no. 3, pp. 1–26, 1999.

[14] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1, pp. I–511–I–518 vol. 1, 2001.

[15] D. Comaniciu, "An Algorithm for Data-Driven Bandwidth Selection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 2, pp. 281–288, 2003.

[16] J. W. Davis and M. A. Keck, "A Two-Stage Template Approach to Person Detection in Thermal Imagery," in *2005 Seventh IEEE Workshops on Applications of Computer Vision (WACV/MOTION'05)*. IEEE, 2005, pp. 364–369.

[17] A. Miron, B. Besbes, A. Rogozan, S. Ainouz-Zemouche, and A. Bensrhair, "Intensity Self Similarity Features for Pedestrian Detection in Far-Infrared Images," *IEEE International Conference on Intelligent Vehicles*, 2012.