

This document is published in:

2012 IEEE Intelligent Vehicles Symposium (IV) IEEE, 2012, pp. 117 - 122.

DOI: [10.1109/IVS.2012.6232242](https://doi.org/10.1109/IVS.2012.6232242)

© 2012 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

# Contrast Invariant Features for Human Detection in Far Infrared Images

Daniel Olmeda, Arturo de la Escalera, and Jose Maria Armingol

**Abstract**—In this paper a new contrast invariant descriptor for human detection in long-wave infrared images is proposed. It exploits local information histogram of orientations of phase coherence. Contrast in infrared images depends on the temperature of the object and the background, which makes gradient based descriptors less robust, especially in daylight conditions. The objective is to obtain a scale, brightness and contrast invariant descriptor that can successfully detect pedestrians in images taken with a cheap, temperature-sensitive, uncooled microbolometer. The descriptor, packed into grids is feed to a Support Vector Machine classifier. The algorithm has been tested in night and day sequences and its performance is compared with a day only descriptor: the histogram of oriented features (HOG).

## I. INTRODUCTION

Onboard perception systems provide human pilots information relevant to the task of driving. These systems typically operate sensors that expand what the driver is able to perceive. There are a number of reasons why these sensors exceed the capabilities of the driver. The point of view, inside the vehicle may be incomplete due to occlusion of the road, or because their attention is needed in several places at one time. Another benefit of using Advanced Driver Assistance Systems (ADAS) is that the sensors can be designed to acquire information not available to the driver's senses. In this article, the authors present an ADAS module designed to automatically detect the presence of pedestrians in the vehicle's path using a far-infrared camera. This approach is especially useful in low light conditions, such as night driving. The driver is warned if the detection system determines a hazardous condition.

The features should be invariant to illumination but also to scaling to successfully identify small objects. In this case, pedestrians that are far away from the camera. The theory of phase congruency [1] 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.

Fig. 1 is an example comparing the resulting magnitude image of points with high phase congruency and the edges obtained with a Sobel filter, both applied to a long wave

infrared image. The most prominent edge in fig. 1c is that between the buildings and the sky. Intensity gradients also depend on magnification, making it difficult to identify small objects. In figure 1b symmetric areas have the same importance, despite of their contrast.

In this article, a detection scheme based on the histogram of oriented phase congruency to detect pedestrians in infrared images is presented. In section II a state of the art of pedestrian detection, specially on infrared images, is presented. The phase congruency theory, on which our Histogram of Oriented Phase Energy (HOPE from now on in this article) descriptor is based is briefly explained in section III. The procedure for the descriptor extraction and the classification scheme are further described in sections III and IV. The impact on the classification performance of the different descriptor parameters is detailed on section V, as well the results of applying the HOPE descriptor to a visible images database. Finally, results and future work are presented in section VI.

## II. STATE OF THE ART

In this paper the authors focus on the detection of a special type of obstacles: pedestrians. Most of the traffic accidents involving pedestrians takes place in urban scenarios, because vehicles and pedestrians share the same ground, so there is greater likelihood of a traffic collision than in highway traffic.

Far infrared images have a very valuable advantage over the visible light ones. They do not depend on the illumination of the scene. 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 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

D. Olmeda, A. de la Escalera and J.M. Armingol are with the Department of Systems Engineering and Automation, Universidad Carlos III de Madrid, Leganes, Madrid, 28911 Spain e-mail: {dolmeda, escalera, armingol}@ing.uc3m.es.

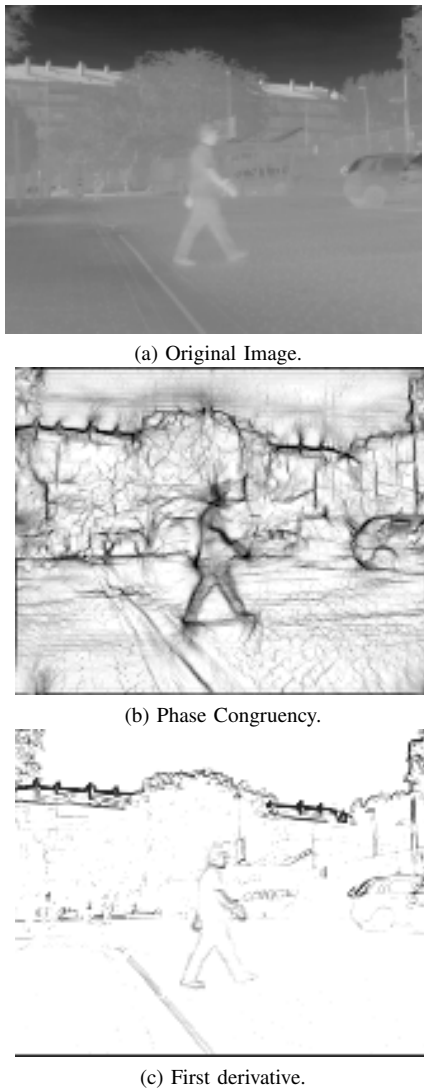


Fig. 1: Examples of phase congruency and derivate results from an infrared image.

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. However, HOG features are based on the first derivative and, as such, unable to cope with a wide range of temperatures, for this kind of images.

In [13] a through evaluation of pedestrian detection methods is presented.

### III. PEDESTRIAN FEATURES

#### A. Method

The pedestrian descriptor presented here is partially inspired by the Histogram of Oriented Gradients approach,

based on the idea that a local gradient distribution can characterize the general shape of an object. Local information is extracted by dividing the image into sets of small spatial regions, called cells. Each cell contains a number of contiguous pixels of an image. For each cell a histogram of the orientation of borders is calculated. The final feature is a concatenation of all the histograms in the portion of the image being analyzed.

#### B. Phase Congruency

Features of high phase congruency are those in which a wide range of their Fourier components is in phase. That is, we are looking for points in the signal that have an order, but without worrying about the shape or amplitude of it. In a one-dimensional signal those correspond to points in the signal with a high slope or at peaks. Decomposition of smooth areas has its frequencies spread over a wider range, thus being its phase congruency score lower.

In an image, regions with a high phase coherence are those in which a wide range of Fourier components are in phase. That is, we are looking for points in the signal which has an order, but without worrying about the form or amplitude of it.

A set of frequencies is extracted from the signal, which represents the information of the features. The solution is the convolution of the Fourier transform of the signal with a set of filters. Each of these filters is able to extract a narrow frequency range. The filters used in this case are complex Gabor filters.

Fig. 2 represents the magnitude and orientation of the phase congruency of an image containing a pedestrian. In fig. 2c the orientation of each pixel, from 0 to  $2\pi$  radians is scaled from black (0) to white (1).

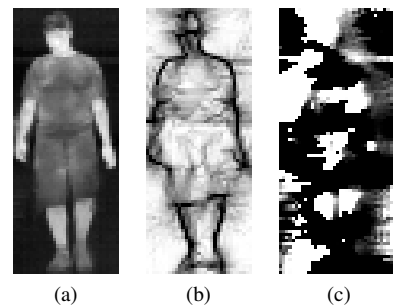


Fig. 2: (a) Original image. (b) Magnitude of phase congruency. (c) Orientation of phase congruency. (d) Representation of the descriptors packed into grids.

The descriptor for the window being analyzed is represented in Fig. 3, a dense grid of histograms of orientation of each window of the image.

#### C. Features Parameters

The final descriptor depends on the parameters selected to create the image of congruence phase, the window size and the number of orientations of the histogram. The performance of the classifier has been studied for the following values of these parameters:

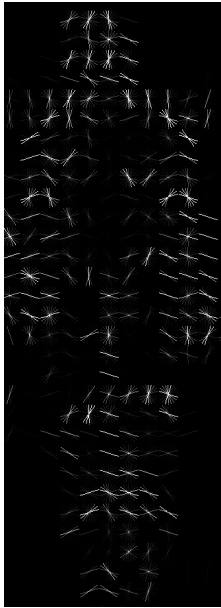


Fig. 3: Representation of the descriptors packed into grids.

- Number of scales: indicates the range of frequencies to be extracted by the filters.
  - $N_{scales} = \{2, \dots, 6\}$
- Number of orientations: for each scale the filters are rotated to a number of angles to extract information at different rotations.
  - $N_{orientation} = \{2, \dots, 6\}$
- Spread of two-dimensional Gabor filter.
  - $\sigma = 2 + 2 \cdot N_{scale}$
- Cell Size: the size of the cell determines the scale at which the features are extracted.
  - $S = \{3, 6\}$
- Histogram orientations binning: number of bins of the histogram of orientations for each cell.
  - $B = \{4, 9, 12\}$
- Normalization of blocks of adjacent cells.
  - $Norm = \{None, L_2\}$
- Size of rois: The image regions in which to search pedestrians for are resized. The chosen size affects the performance of the classifier as well as the computation time. The authors propose dividing the training set based on the distance of the pedestrian to the camera.
  - $Crops = \{\{32, 16\}, \{64, 32\}, \{128, 64\}\}$

#### IV. CLASSIFICATION

A Support Vector Machine classifier takes the decision of whether or not an analysis window contains a pedestrian. SVMs maximize the hyperplane separating two classes by projecting the training set into a possibly infinite dimension. For SVM training is necessary to have a representative sample of feature vectors from both positive and negative classes. In this case, pedestrians and non-pedestrian images.

To better represent the shape of a pedestrian, the training set contains samples of persons at different temperatures, as well as pedestrians at different scales. The number of samples is approximately the same for each class.

The training set  $x_k$  is mapped in a high dimensional space defined by a function  $\phi$ . The decision function in equation 1 is optimized so that  $y(x)$  maximizes the distance between the nearest point ( $x_i$ ) and the hyperplane.

$$y(x) = w^T \cdot \phi(x) + b \quad (1)$$

Where  $w$  is normal to the hyperplane and  $\frac{b}{\|w\|}$  is the perpendicular distance from the hyperplane to the origin.

The training set contains  $N$  samples  $x\{k\} = (x_1, \dots, X_n)$ , manually sorted and assigned a binary label  $l = -1, 1$ . Each of these vector samples is a concatenation of the cell histograms in the cropped image.

As indicated in section III there are several parameters on which depends the border image. The performance of the classifier will also rely on the initialization of the SVM. We first consider, as our base classifier, a one-norm, soft-margin support vector machine with a Gaussian Radial Basis kernel function. As the algorithm is supposed to run in real time, the final selection of the parameters also take into account the speed of processing.

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 DET (Detection-Error Tradeoff) curves. The set of parameters tested has been:

- Training set:
  - $N_{samples} = \{9000, 12000, 15000\}$
- Kernel:
  - Linear.
  - Quadratic.
  - Radial Basis Function.
- $C = \{10^{-2}, 1, 10^3, 10^6\}$



Fig. 4: Samples shot at different temperatures and illumination conditions.

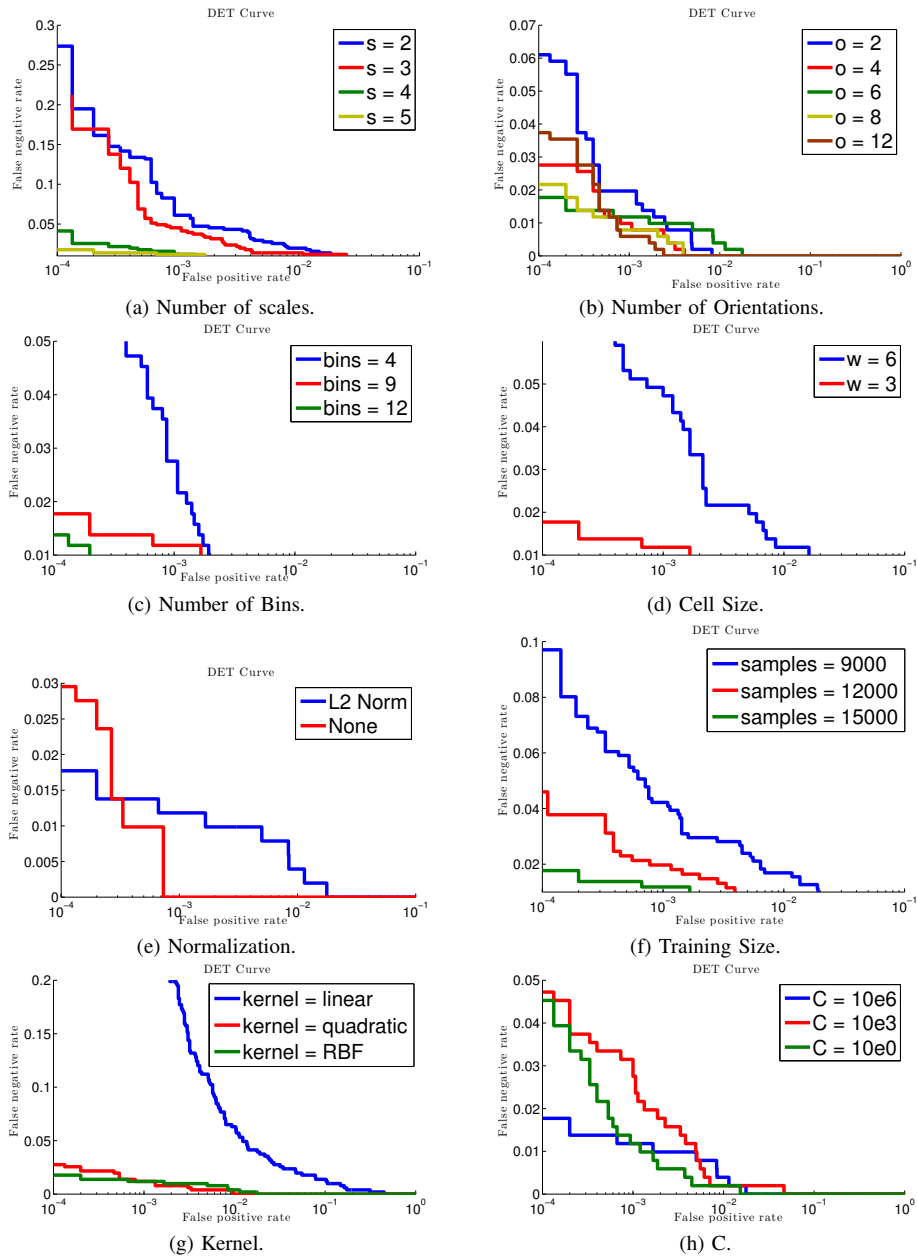


Fig. 5: DET curves of the tested parameters.

## V. RESULTS

### A. Dataset

The training dataset contains 127000 images recorded from a moving vehicle. They were evenly selected to represent a wide range of temperatures and illumination conditions. The camera used has a low resolution sensor of  $(164 \times 129)$  pixels and a gray level depth of 14 bits. Figures 4 contains far infrared images of pedestrians and background shot at different circumstances.

### B. Parameter Evaluation

Fig. 5 contains the performance curves of the classifier. The DET curves represents False Negative Rate against False Positive Rate. Lower values are better.

From those curves we can learn that the number of scales and orientations affects the performance. Specifically, it is recommended to use a minimum of 4 scales. With each scale more information is available to the classifier, thus improving the performance. However, adding a fifth scale improved the classification by only 2.66% at  $10^{-4}$  FPPW. The peak performance is at 6 orientations. For 8 orientations the classification deteriorates by 0.4% at  $10^{-4}$  FPPW.

The parameter that affect the most to the calculation of the histogram of orientations are the number of bins and cell size. Twelve bins over the default nine bins improves marginally the classification, just 0.39% at  $10^{-4}$  FPPW, but this means a longer descriptor and more memory allocation. We found the best cell size to be  $\{3 \times 3\}$  for our  $\{64 \times$

32} dataset, for a total of 210 cells per sample. It is worth to remark that the normalization of the cells doesn't have a significant impact on performance. This is because, the phase congruency map is in itself a normalized magnitude, as opposed to the gradient used in the HOG descriptor.

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.

Finally, a large and diverse training set improves classification.

Our default parameters are:

- Scales = 4
- Orientations = 6
- Cell Size =  $\{3 \times 3\}$
- Crop Size =  $\{64 \times 32\}$
- No cell normalization.
- Quadratic SVM kernel.

### C. HOPE vs. HOG

Fig. 6 represents the DET curves of the HOPE and standard HOG descriptors for the Intelligent Systems Laboratory (ISL) infrared database. The HOG descriptor achieves a 54% recall rate at  $10^{-4}$  false positives per window (FPPW) for a linear SVM kernel and 91% for a radial basis function (RBF) kernel. With the approach here presented the recall rate goes up to 98% at  $10^{-4}$  FPPW.

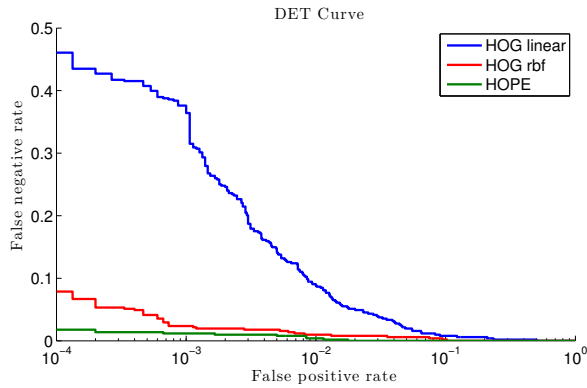


Fig. 6: HOPE vs. HOG performance for the ISL Database.

## VI. CONCLUSIONS AND FUTURE WORK

### A. Conclusions

In this article, the authors have presented a new descriptor for the classification of pedestrians in far infrared images. This approach is based on information from low resolution, uncalibrated, uncooled microbolometer sensors to detect the presence of people in the path of a moving vehicle. The main application of the system is to be used while driving at night, although our test demonstrate good performance in a wide range of temperature and illumination.

The best combination of parameters achieves 98.5% at  $10^{-4}$ FPPW.



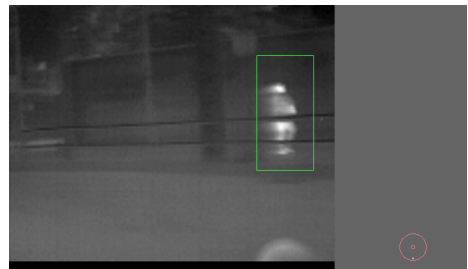
(a)



(b)



(c)



(d)



(e)

Fig. 7: Examples of processed images with pedestrians being tracked.

Fig. 7 contains some examples of pedestrians being detected from the experimental vehicle while driving through

urban environments. For each pedestrian found an Unscented Kalman Filter is created to track its position. The driver is then warned if the pedestrian's trajectory intersects that of the vehicle. False positives are more prone to appear in images shot at very high temperatures. In fig. 7e a light post is misclassified as a pedestrian due to its high vertical symmetry.

The pedestrian detection algorithm has been tested on the IVVI vehicle, an experimental platform for driving assistance systems (fig. 8)



Fig. 8: Intelligent Vehicle based on Visual Information.

### B. Future Works

The scheme here presented requires a deep scan of each image. This is computationally expensive. In order to reach real time processing, the authors are deploying a boosting scheme based on this descriptor to quickly ignore large regions of the image, analyzing only a small set of sub features and lowering computation requirements of the classification.

Once a pedestrian has been found, the priority of the algorithm should be to find it again in the next image. This descriptor has been designed as a grid of features, but it is still possible to modify it to work as a small feature that can be uniquely identify. This way, once the algorithm has a high certainty of the presence of a pedestrian, this unique features could be tracked with very little computational demands.

Though the dataset is rather complete, new images are constantly being added to ensure that the training samples are as heterogeneous as possible.

The proposed method of pedestrian has been tested also on the INRIA pedestrian database, which contains 2416 positive samples of pedestrians in color images for training. A random set of 12180 negative samples were cropped from 1218 images without pedestrians. The performance of both our implementation of HOG and our HOPE descriptor is shown in figure 9.

Though the images in the database are not thermal and cannot benefit from all the characteristics of phase congruency maps, our descriptor out-performs the original HOG.

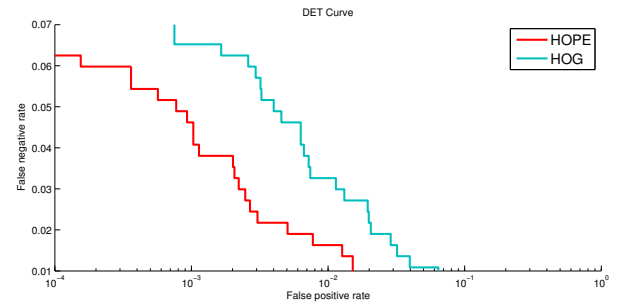


Fig. 9: HOPE vs. HOG performance for the INRIA Database.

## VII. ACKNOWLEDGMENTS

This work was supported by the Spanish Government through the Cicyt projects FEDORA (GRANT TRA2010-20225-C03-01) and VIDAS-Driver (GRANT TRA2010-21371-C03-02), and the Comunidad de Madrid through the project SEGVAUTO (S2009/DPI-1509).

## REFERENCES

- [1] P. Kovese, "Image features from phase congruency," *Videre: Journal of Computer Vision Research*, vol. 1, no. 3, pp. 1–26, 1999.
- [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. PACIA 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 (ICCSM), 2010 International Conference on*, 2010.
- [9] N. Dalal, B. Triggs, and I. Rhone-Alps, "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. IEEE Conference on*, pp. 1–8, may 2007.
- [13] P. Dollár, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: A benchmark," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 304–311.