IJCTA, 9(14) 2016, pp. 6477-6484

Human Detection in Infrared Imagery Using Support Vector Machine and Curvelet Features

Abhilash Chandan[*] and S. Prabakaran[**]

ABSTRACT

This paper presents a two phase method for detection of humans in infrared imagery. Here we obtain the region of interest first by using the high brightness property of the human's pixels in accordance to Plank's law. One of the major setbacks in the first phase is the failed detection of body parts covered with cloths which is solved by using the principle of region growing. Now in the second phase this approach then uses the curvelet entropy features to perform the validation of the detected ROIs. From all the identified ROIs, those having humans are classified, which is achieved using the well known methodology of support vector machine (SVM). Major advantage of this approach as compared to the older approaches is the significant improvement in detection time, lower requirements in terms of compute power and memory as a result of which it can also be run on constrained platforms.

Keywords: Support Vector Machine, Human detection, Curvelet Transform

INTRODUCTION

1.1. General Introduction

Human detection in infrared images now a days is gaining more and more attention [6-9]. The major problems encountered while detecting humans are the pose variability of the humans, which becomes more varied in outdoor environment. The appearance of the humans further increases the complexity of the problem due to the texture, light conditions in both day and night, color. Thermal infrared images are the answer to these kinds of problem as they remove all the color, texture, and light conditions from the subjected images.

In thermal infrared images the intensity of the object in the thermal image is determined by the temperature and radiated heat of the object which is independent of the light conditions. As a result of which the detection of humans can be performed both in the day as well as in the night.

But even the thermal images are not perfect. The infrared images are affected by the surface property of the object link emissivity, reflectivity, transmission etc. The wavelength of the object also affects the thermal images. Many objects apart from humans also produce bright areas like animals, cars, lamp post, electric pole, buildings etc.

Another big challenge faced while performing analysis of thermal images is the limitations in the quality of camera used, most of the thermal images have very low spatial resolution as compared to the visible images and even the sensitivity is also very less as compared to the visible images. As a result of this the final product images which we get has very low image quality with noise and blurring. Thus performing detection in such quality of images is a very challenging task.

[*] Post Graduate Student,

[**] Professor, Department of computer Science and Engineering, SRM University, *E-mail: abhilashchandan@gmail.com*

The difficult problem of detecting humans has been addressed using vision sensors, image processing, and pattern recognition. But these approaches work only with daylight vision [9], [5].

The approaches used earlier is template matching [11], projection of the intensity orientation [12-13] and difference in moment [2], [6]. But this whole process is very much time consuming and in order to improve upon it the approach we are using here is a two step phase approach.

In the first phase we simply create region of interests around all the high intensity regions which are above a calculated threshold value. These regions are our probable humans. Now taking those regions as a seed we will implement region growing [15] to get the approximate body of our human. This phase finally completes our human detection phase. Now in the first phase we may have a lot of false detection. In order to remove them we will move on to our second phase which is our human validation phase. In this phase we extract curvelet entropy features [16], [17], [18] from our probable human region of interests and check whether they are humans or not using the classifier support vector machine [19], [20].

2. PROBABLE HUMAN REGION SELECTION

During the selection of the probable regions where all the humans may be present we make use of the brightness intensity oriented projection method, where we first convert our image to a binary image with a flexible threshold $\boldsymbol{Th}$. In the binary image all the intensities above the threshold value will be white whereas all the intensities below the threshold value will be black. We set the threshold value $\boldsymbol{Th}$ in between the maximum intensity $\boldsymbol{I_{max}}$, the mean image intensity $\boldsymbol{I_{mean.}}$ The threshold is expressed as:

$$\boldsymbol{Th = w_1 I_{max} + w_2 I_{mean} + (1 - w_1 - w_2) P_{mean}}$$

Where $\boldsymbol{w_1}$ and $\boldsymbol{w_2}$ are weights, satisfying the property $0 \leq \boldsymbol{w_1}, \boldsymbol{w2} \leq 1$. Values of $\boldsymbol{w_1}$ and $\boldsymbol{w_2}$ are varied in order to get the best segmentation results through tests. The value of $\boldsymbol{Th}$ is also varied with respect to the changing image intensity. The head of humans have the highest intensities as compared to rest of the body. Now from the segmented binary image we obtain the region of interests containing the probable head of the humans in the original image using the following steps:

1. All the pixel values which are greater than the threshold value $\boldsymbol{Th}$ in the original image are marked as white in the binary image and all values less than $\boldsymbol{Th}$ are marked as black in the binary image
2. All those high intensities pixels forming are then made smooth by applying Gaussian blur in the binary image.
3. Then we scan the binary image column by column and create region of interests around all the white spaces encountered taking all the length of the ROI as the number of rows in the image.
4. Now we will take the white spaces in the binary image to predict the approximate body of the human by making an increasing the width of the ROI using the following expression:

$$\mu = \frac{\Sigma_{i=0}^{n} x_i f(x_i)}{\Sigma_{i=0}^{n} f(x_i)}$$

$$\sigma = \sqrt{\frac{\Sigma_{i=0}^{n} (i - \mu_i)^2 f(x_i)}{\Sigma_{i=0}^{n} f(x_i)}}$$

Where μ is the mean of all the white pixels within the region of interest obtained by scanning them column by column, taking one ROI at a time. σ Is the variance of those pixels.

5. Now our new ROIs will have the left top corner at $\boldsymbol{(0, \mu_i - 5\sigma_i)}$ and right top corner at $\boldsymbol{(0, \mu_i + 5)}$. Which is an approximated value obtained after a lot of tests and experimentations. Using those coordinate points from the binary image we will finally draw the ROIs in the original image.

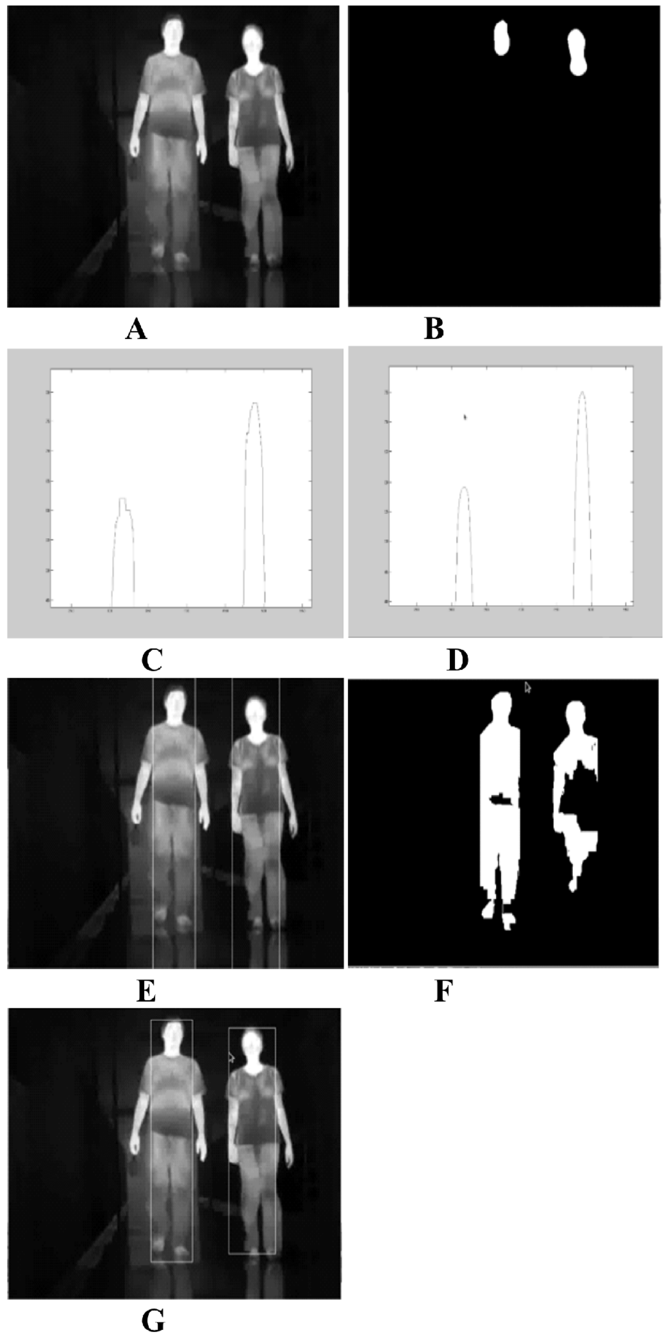


Figure 1: Results of human detection phase: (A) Original image, (B) Binary image after thresholding and applying Gaussian blur, (C) Intensity vertical projection curve before applying Gaussian blur, (D) Intensity vertical projection curve after applying Gaussian blur, (E) Original image after performing body prediction, (F) Binary image after implementing region growing, (G) Original image after completion of human detection

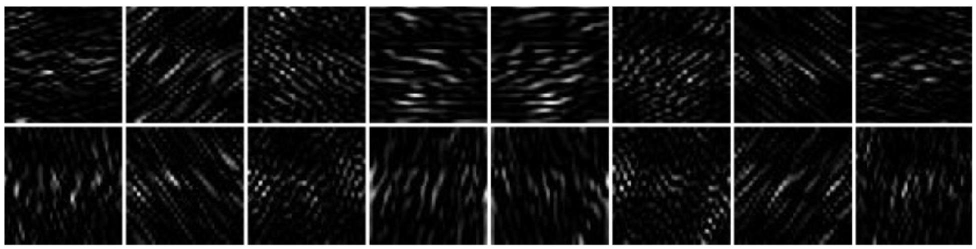
Now many times objects which are not connected to the human but are in the foreground are also detected. In order to reduce the false detection and remove some of the non-human foreground objects we will perform region growing, taking the probable head of the pedestrian as a seed.

A threshold T value is calculated taking the average of all the pixels within the ROI in the original image and all pixels values greater than T and connected to the seed, those pixel's coordinate values are taken and in the binary image those coordinate values are made white finally when we get our fully connected human body in the binary image we draw an ROI around our detected pedestrian. Using the coordinate values from the binary image we finally draw ROIs in the original image.

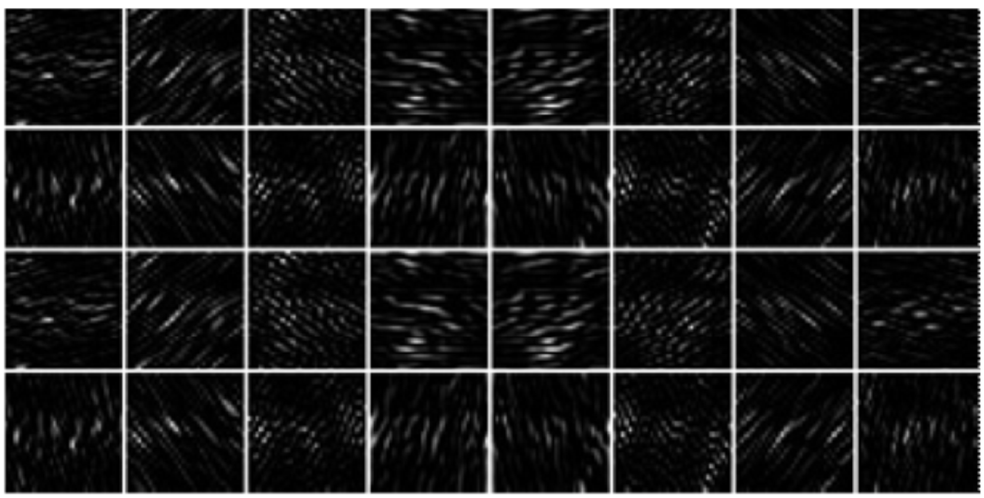
3. FEATURE EXTRACTION

Our approach is based on thresholding the curvelet coefficients. We extract image features by applying curvelet transform. Images are then partially reconstructed by applying inverse curvelet transform to the coefficients after thresholding. These partially reconstructed images. form the feature vector.

We then transformed this feature vector into the basis space of PCA and ICA for dimensionality reduction. Trained face images are represented as points in this space. In order to identify, test images are also projected into this basis space. Euclidean distance measure has been used to estimate the similarity. We then compared the performance in both PCA and ICA subspaces.



A



B

Figure 2: Subbands at various scales: (A) Subbands at scale 2, (B) Subbands at scale 3

We resized the images to size 128×128. The feature extraction using curvelet is applied to each database image. For image size of 128×128, the maximum number of levels possible are 4. Hence each image is decomposed into 4 levels of scales using curvelet transform. The numbers of subbands at different scales are different. For levels of decomposition, there are 1, 16, 32 and 1 subbands at decomposition level 1, 2, 3, and 4 respectively. Therefore, 4 levels decomposition creates 50 (=1+16+32+1) subbands of curvelet coefficients. However, because a curvelet oriented at an angle θ produces the same coefficients as a curvelet oriented at an angle $\pi + \theta$, only half of the subbands at level 2 and 3 may be used. Figure 7.1, 7.2 and 7.3 shows curvelet coefficients of a sample image for all the 50 subbands.

3.1. Curvelet entropy features

So, the structural activity extracted from the Curvelet transform of the image can be analyzed statistically to generate fingerprint features vector. Thus, we have applied the Curvelet transform on the ROI of the binarized & skeletonized fingerprint. The Curvelet decomposition generates several sub-bands images. Then, we calculate the statistical features from all these sub-bands to generate the fingerprint features vector. In the Basque language for example words are formed by adding the affixed to the dictionary entries. More specifically, the affixed corresponding to the determined number and declension case are taken in this order, independently of each other order. As prepositional functions are realized by case suffixed inside word forms Basque presents a relatively high power to generate inflected word forms which makes morphosystematic analysis very important to be able to extract information from text fragments.

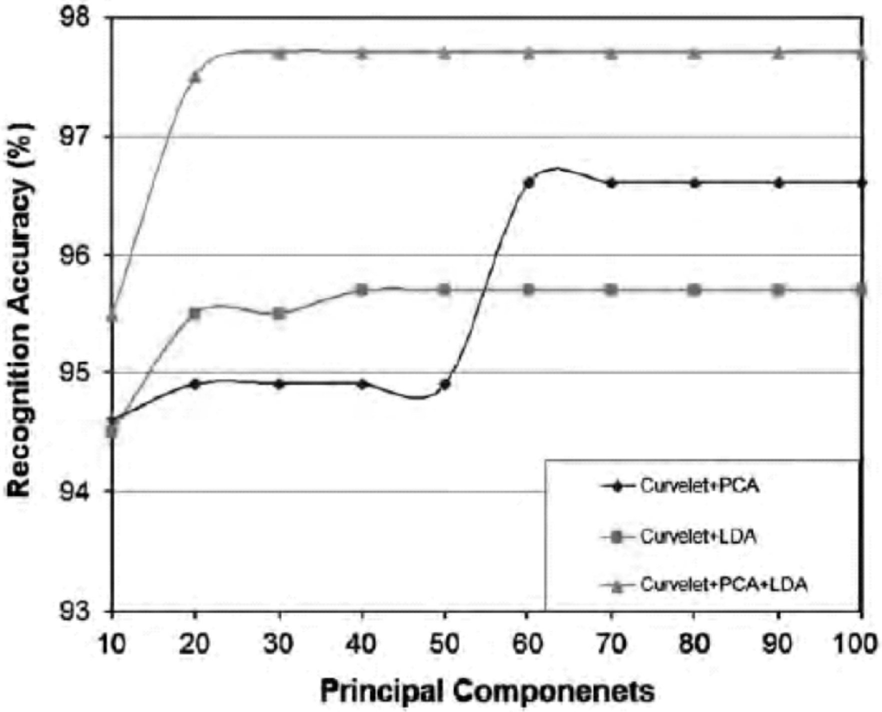


Figure 3: Curvelet –based recognition accuracy for ORL database [Mandal *et al.*, 2009]

4. SUPPORT VECTOR MACHINES

Support Vector Machines (SVMs) implement the following idea : they map input vectors into a high dimensional feature space, where a maximal margin hyperplane is constructed [6]. It was shown that when training data are separable, the error rate for SVMs can be characterized by

$$h = R^2/M^2, \quad (1)$$

where **R** is the radius of the smallest sphere containing the training data and **M** is the margin (the distance between the hyperplane and the closest training vector in feature space). This functional estimates the VC dimension of hyperplanes separating data with a given margin **M**.

To perform the mapping and to calculate R and M in the SVM technique, one uses a positive definite kernel **K(x, x2)** which specifies an inner product in feature space. An example of such a kernel is the Radial Basis Function (RBF),

$$K(x, x2) = e^{-\|x-x'\|2/2\sigma 2}$$

This kernel has a free parameter σ and more generally, most kernels require some parameters to be set. When treating noisy data with SVMs, another parameter, penalizing the training errors, also needs to be set. The problem of choosing the values of these parameters which minimize the expectation of test error is called the model selection problem.

It was shown that the parameter of the kernel that minimizes functional (1) provides a good choice for the model : the minimum for this functional coincides with the minimum of the test error. However, the shapes of these curves can be different.

5. COMPLETE SYSTEM

When analyze an infrared image for detection of humans the first step involved is the detection of humans followed by the validation of humans.

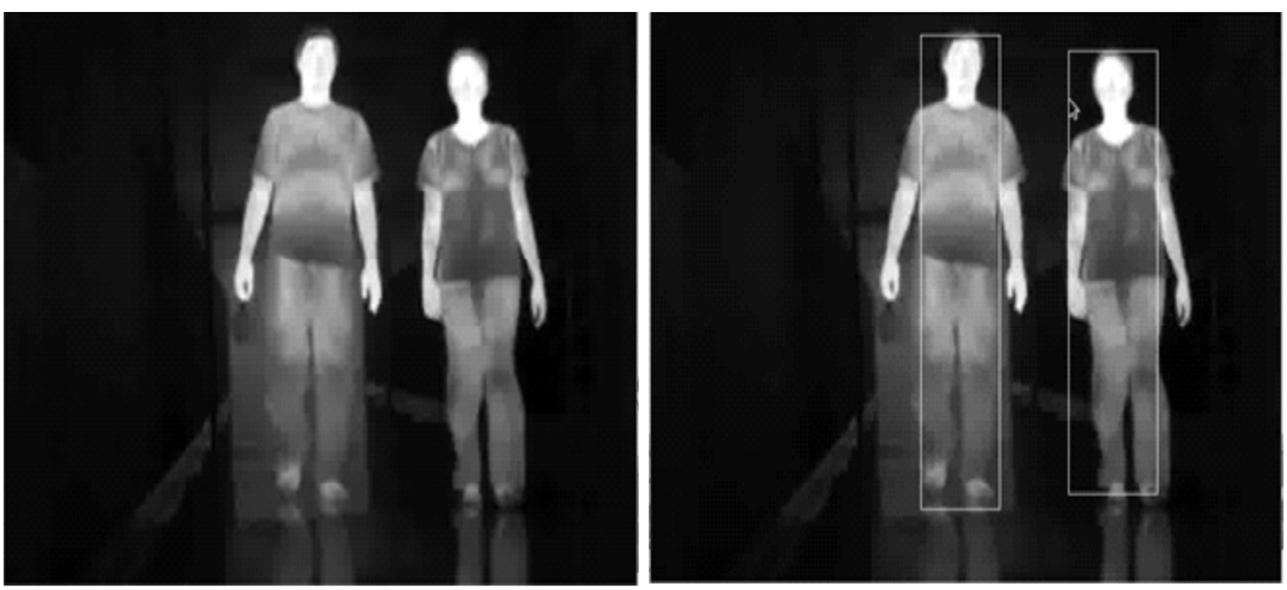
The whole process is divided into two phases,

Phase 1:

1. Search for all the probable humans in the given image using the intensities of their head. As humans head has the high intensities of heat thus they are easier to detect.
2. Create ROIs around the detected probable human's head with the height of the ROIs as the image rows.
3. Now we predict the probable body of the pedestrian by calculating the mean and variance of all the high intensity pixels identified in the above step.
4. We then use the probable head of the pedestrian as a seed and implement region growing to detect all the components which are connected to the head of the human.
5. Finally we create the ROI around the detected pedestrian .

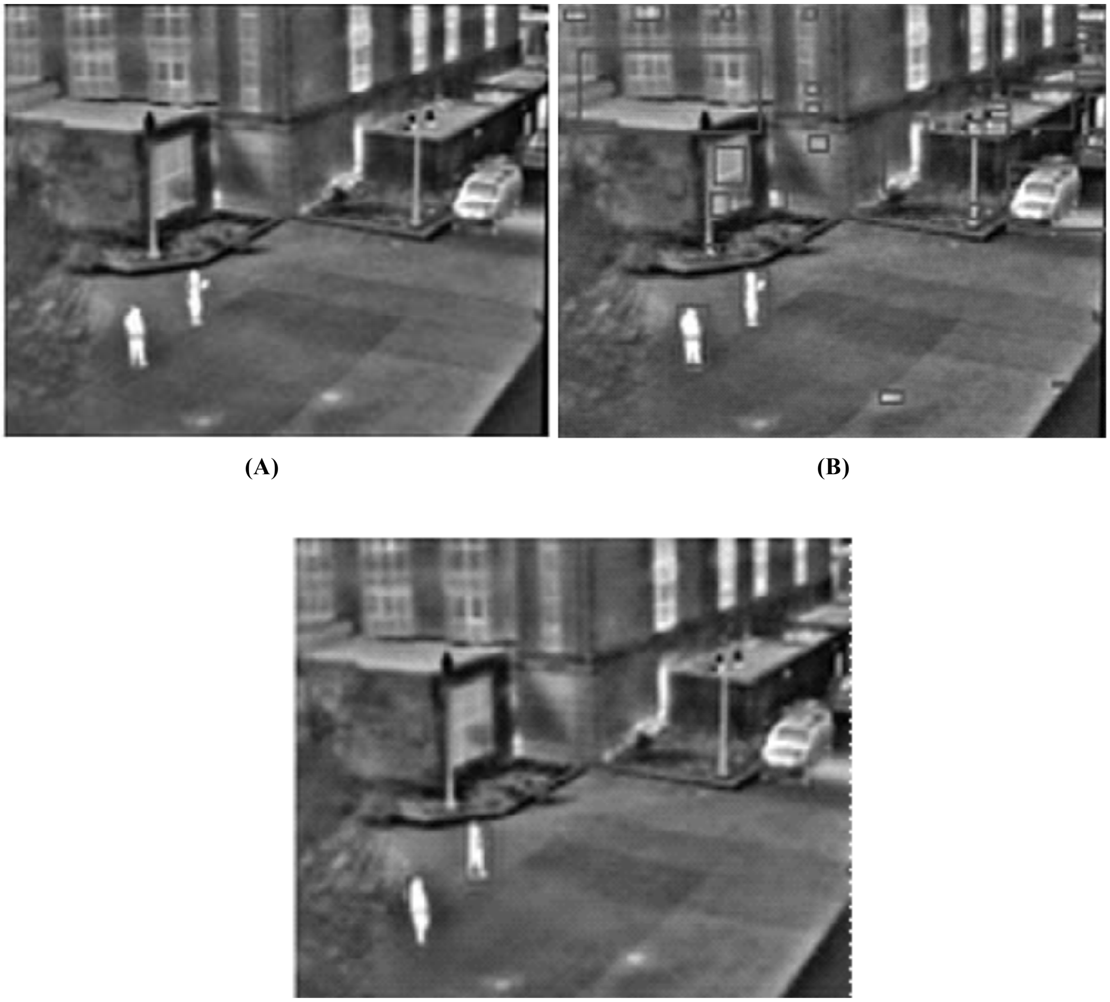
Phase 2:

1. In the second phase we train the support vector machines with plenty of examples so that it can efficiently classify between human and non-human.
2. The features we use are the curvelet entropy features, we extract these features from the training images to train the support vector machine.
3. The training images we used contains a variety of near and far humans, different variations of the humans poses are also considered while training the support vector machine.
4. In this step we perform the curvelet transform of the ROIs we got in the detection phase and extract their curvelet entropy features.
5. The support vector machine then classifies that whether the given ROI is of a human or not. Thus reducing a lot of false detection detected in the previous step.



(A) (B)

Figure 4: Human Detection: (A) Original image, (B) Image after performing human detection



(A) (B)

(C)

Figure 5: Human Detection: (A) Original image, (B) Image after performing pedestrian detection, (B) Image after performing pedestrian validation

REFERENCES

[1] A. Broggi A. Fascioli P. Grisleri T. Graf M. Meinecke. Modelbased validation approaches and matching techniques for automotive vision based pedestrian detection. In Intl. IEEE Wks. on Object Tracking and Classification in and Beyond the Visible Spectrum, San Diego, USA, page in press, June 2005.

[2] Massimo Bertozzi, Alberto Broggi, Alessandra Fascioli, Thorsten Graf, and Marc-Michael Meinecke. Pedestrian detection for driver assistance using multiresolution infrared vision. IEEE Trans. On Vehicular Technology, 53(6):1666-1678, nov 2004. ISSN 0018-9545.

[3] Y Fang K. Yamada Y Ninomiya B.K.P. Horn and I. Masaki. A shape-independent method for pedestrian detection with farinfrared images. 53(5), September 2004.

[4] C. Papageorgiou and T. Poggio. Trainable pedestrian detection. In Proceedings of the 1999 International Conference on Image Processing, pages 35-39, 1999.

[5] D. Gavrila and J. Geibel. Shape- based pedestrian detection and tracking. In Proceedings of IEEE Intelligent Vehicles Symposium, pages 215-220, 2000.

[6] M. Shah, O. Javed, K. Shafique, Automated visual surveillance in realistic scenarios, IEEE Multimedia 14 (2007) 30–39.

[7] A.E. Maadi, X. Maldague, Outdoor infrared video surveillance: a novel dynamic technique for the subtraction of a changing background of IR images, Infrared Phys. Technol. 49 (2007) 261–265.

[8] M. Bertozzi, A. Broggi, A. Fascioli, T. Graf, M. Meinecke, Pedestrian detection for driver assistance using multiresolution infrared vision, IEEE Trans. Vehicle Technol. 53 (2004) 1666–1678.

[9] T. Tsuji, H. Hattori, M. Watanabe, N. Nagaoka, Development of night-vision system, IEEE Trans. Intell. Transport Syst. 3 (2002) 203–209.

[10] P. Viola, M. Jones, and D. Snow. Pedetrian using patterns of motions and appearance. In IEEE Int. Conf on Computer Vision, pages 734-741, 2003.

[11] F.L. Xu, X. Liu, K. Fujimura, Pedestrian detection and tracking with night vision, IEEE Trans. Intell. Transport Syst. 6 (2005) 63–71.

[12] M. Bertozzi, A. Broggi, C. Caraffi, M.D. Rose, M. Felisa, G. Vezzoni, Pedestrian detection by means of far-infrared stereo vision, Comput. Vision Image Understan. 106 (2007) 194–204.

[13] Y.J. Fang, K. Yamada, Y. Ninomiya, B. Horn, I. Masaki, A shape-independent method for pedestrian detection with far-infrared images, IEEE Trans. Vehicle Technol. 53 (2004) 1679–1697.26) D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," Int'l J. Computer Vision, vol. 60, no. 2, pp. 91-110, 2004.

[14] Candes, E. J.; Demanet, L.; Donoho, D. L. & Ying, L. (2007). Fast discrete curvelet transform, *SIAM Multiscale Modeling and Simulations*, 2007.

[15] A Semantic Region Growing Approach in Image Segmentation and Annotation T.Athanasiadis, Y.Avrithis and S.Kollias 9, Iroon Polytechniou St., 157 73 Zographou, Greece.

[16] Candes, E. J. & Donoho, D. L. (2000). *Curvelets—A surprisingly effective non- adaptive representation for objects with Edges*, Vanderbilt University Press, Nashville, TN, 2000.

[17] Chien, J. T. & Wu, C. C. (2002). Discriminant waveletfaces and nearest feature classifiers for face recognition, *IEEE Transactions on PAMI,* vol. 24, pp. 1644–1649, 2002.

[18] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.

[19] T. S. Jaakkola and D. Haussler. Probabilistic.

[20] V. Vapnik. Statistical Learning Theory. Wiley, 1998.