Probabilistic Template Based Pedestrian Detection in Infrared Videos

Harsh Nanda & Larry Davis
Department of Computer Science
University of Maryland, College Park, MD-20742
{nanda,lsd}@cs.umd.edu

Abstract

In this paper we present a real time pedestrian detection system that works on low quality infrared videos. We introduce an efficient and effective idea of probabilistic templates to capture the variations in human shape specially for the case where the contrast is low and body parts are missing. The infrared videos provide the regions of interest, which considerably reduces the search space and the probabilistic template helps recognize the pedestrians. We present experimental results on infrared videos taken from a moving vehicle in an various urban street scenarios to demonstrate the feasibility of the approach.

1 Introduction

A huge amount of work is being done in the area of pedestrian detection, to be able to robustly detect pedestrians from moving platforms. The goal is especially hard because of a wide range of possible pedestrian appearances, cluttered backgrounds and moving cameras. Because the camera is not stationary and background is constantly varying, simple techniques of background subtraction cannot be used for getting the foreground objects.

Most of the current approaches to this task of pedestrian detection have treated this as a recognition task and hence used generic object detection and recognition techniques to solve the problem. The fact that humans emit heat and hence can be captured well by infrared cameras has not been put to use very well.

Recently there has been some interest in using infrared cameras for human detection because of the sharply decreasing prices of easy to use infrared cameras. Although infrared cameras do capture humans, they have problems of their own. The quality of video that we get from these cheap, uncooled infrared cameras is not very good (lot of noise and ghosting effects). Also, pedestrians are not the only hot objects. Lamps, cars and many other objects are also hot and hence are captured by infrared cameras. Last but not the least, human body does not emit heat uniformly. The amount of heat emitted and hence the intensity in the infrared video depends on the body part being captured, dress of the person, orientation of the person, time of day and many other factors. Thus infrared is not a complete solution to the task of pedestrian detection, but it definitely helps reduce the search space. What is needed is object detection and recognition technique which uses the cues provided by infrared to reduce the search space, but can then work in the domain at hand i.e. low quality videos, low contrast and missing body parts.

In this paper we discuss a system that integrates template matching based human detection and infrared videos. The problem is non-trivial because the notion of human templates in the domain of infrared videos is quite ambiguous, as all the body parts of a person are rarely completely visible. Thus [4] cannot be directly applied to this domain.

The outline of this paper is as follows. Section 2 describes the realted work in pedestrian detection. Our pedestrian detection system has been discussed in detail in section 3. In section 4 we discuss the experiments performed to demonstrate the feasibility and effectiveness of the approach, Section 5 gives the future direction of our research and finally section 6 & 7 provide the conclusion and acknowledgements for the project.

2 Related Work

A significant amount of progress has been made in the area of pedestrian detection from moving platforms in the past few years. Most of the vision-based pedestrian detection systems have taken a general learning-based approach, where the human appearance is described in terms of simple low-level features from a region of interest.

Most of the human tracking and motion analysis systems employ simple segmentation procedure such as background subtraction or temporal differencing to get the foreground region. Other than applications such as surveillance, where the camera is stationary, these techniques of extracting the foreground are not of much use. Some techniques such as Pfinder [5], W4 [6] and path clustering [6 liang], have been developed to compensate for small, or gradual changes in the scene. However they cannot deal with large sudden change in the background. Independent motion detection techniques can help [8, 9], but they are difficult to develop and are not feasible for non-rigid object extraction since different body parts move differently. A common drawback with all these approaches is the assumption that all detected objects are pedestrians. This limits the generalization and application of these techniques.

More sophisticated pedestrian detection techniques have a two-step process: foreground detection followed by recognition step to verify if the target object is a pedestrian or not. The recognition step can be motion-based, shape-based or multi-cue based. Motion based approaches use periodicity of human walk or learned gait for pedestrian detection [10, 11, 12, 13, 14]. A major drawback of all such approaches which use temporal information for pedestrian detection is that the procedure requires a sequence of frames, which delays the identification until several frames later and increases the processing time making it useless for critical applications like pedestrian detection from moving vehicles. Also such methods cannot detect pedestrians standing or doing something that does not contained the assumed periodic pattern.

Shape based approaches try to solve the harder problem of recognizing pedestrians in single images, hence taking care of both moving and stationary pedestrians. The biggest challenge that this problem offers is to model the huge amount of variations in the shapes, pose, size and appearance of humans and their backgrounds. [15, 16] use handcrafted human models for pedestrian detection. The main restriction of this approach is that it requires segmentation into body parts which itself is a very hard task. Lipton [17] uses an easy to calculate metric perimeter2/area to classify human and vehicle. The metric is rather fragile to many cases where group of people are walking together.

Another line of approach involves shifting windows of various sizes over the image at different resolutions, extracting low-level features, and using standard pattern classification technique to determine the presence of a pedestrian. [2] extract wavelet features and then use SVM to classify them. [4] extracts edges and then uses chamfer distance measure to compare with an hierarchy of templates of human shapes. However these systems have to search the whole image at multi-scales for pedestrians. This is a computationally expensive procedure and single targets might give multiple responses.

A powerful technique to establish regions of interest is stereovision. It is used in [18, 3] in combination with texture-based pattern classification. [2] uses stereo vision, but prefer to combine it with a verification technique based on symmetry properties. All these techniques have a common drawback that they use multiple cameras and hence the results depend highly on camera calibration. The setups are usually quite fragile and require calibration periodically.

Lately, there has been increased interest in video sensors, which operate outside the visible spectrum e.g. infrared, because of their sharply reducing prices and increasing easy of use. The fact that humans appear as bright blobs in infrared videos due to heat that their bodies emit[gavrila 18], helps establish robust regions of interest. But pedestrians are not the only source of heat (e.g. bulbs, lamps, cars etc.), hence pattern recognition techniques are required to classify them correctly. Also, due to the low quality of the video provided by un-cooled infrared cameras, and the fact that different body parts appear or not, depending on the clothes, time of day, pose and other factors makes the task quite challenging.

3 System Architecture

This section describes in detail our pedestrian detection system and the steps involved in it. In section 3.1 we justify our choice of our input features. In section 3.2 we describe the preprocessing, where we use the fact that humans appear brighter than surrounding objects, to get our target regions of interest. In section 3.3 we describe how we develop the probabilistic template that we later use for classification of target objects as pedestrians and non-pedestrians. The process of classification is discussed in section 3.4. Section 3.5 discusses the implementation details.

3.1 Input features

One of the key decisions to be made in designing a system for object classification is the features to be used for object representation. Better features increases inter-class distance and reduce intra class separation. Better is the ratio of inter and intra class separation, easier is the task of classification.

The examples of pedestrians in our database are shown in Figure 2, 3 & 4. All the images are gray level where intensity of the pixel corresponds to the temperature of the object being imaged at that point. The amount of heat radiated by humans depends highly on the body part, type of clothes, pose and last but not the least on the physical and mental state of the person. Thus different body parts have different amount of variations in their intensity. Hence, there is no reason to believe that there will be any sort of correlation in neighboring pixels. Thus region-based approaches are likely to fail in the current domain. Also because of the huge amount of noise in the data, low contrast and ghosting effects traditional fine scale edge-based representation is of no use. Due to low contrast most of the information is not captured by edges and a lot of spurious edges make the task of detection and recognition nearly impossible.

Thus we use a pixel-based representation. The raw intensity values at each pixel after some preprocessing are used to classify target objects obtained after pre-processing as pedestrians and non-pedestrians.

3.2 Detection of target objects

The target objects are extracted from the raw video by using simple intensity thresholding. As discussed above, the intensity of humans in the video is not fixed. It changes depending on a large number of factors. Based on the training data, which is a set of 1000 rectangular boxes containing pedestrians, we calculate the mean and standard deviation for the pixel belonging to pedestrian (μ_1 & σ_1) and pixels belonging to background (μ_2 & σ_2). Using Bayes classification and assuming equal priors for the pedestrian and non-pedestrian class and gaussian distribution, threshold is given by equation (1)

$$threshold = \frac{\sigma_1\sigma_2}{\sigma_1+\sigma_2}\ln(\frac{\sigma_1}{\sigma_2}) + \frac{\sigma_1\mu_2+\sigma_2\mu_1}{\sigma_1+\sigma_2} \tag{1}$$

Once the threshold is decided, the thresholding technique that is used is given by the equation (2)

$$\begin{aligned} th(x,y) &= 1\ if\ i(x,y) > threshold \\ th(x,y) &= 0\ if\ i(x,y) \leq threshold \end{aligned} \tag{2}$$

where th is the thresholded image and i is the raw input image. The pixels that belong to the background i.e correspond to objects that do not emit heat are given the value 0 and the pixels that correspond to objects emitting heat i.e. humans, cars, lamps etc. are replaced by 1's.

3.3 Probabilistic Template

The training data used for developing the probabilistic template consists of 1000 128x48 rectangular images all of which are known to contain humans in different poses and orientation but having the same height. Thresholding is performed on the templates as described above. This is done so that the model does not learn the intensity variations among the background pixels and intensity variations among the foreground pixels. Each template is then translated so that the centroid of the non-zero pixels matches the geometrical center of the image.

After this normalization step, for each pixel of the template, the probability p(x,y) of it being pedestrian is calculated based on the how frequently it appears as 1 in the training data. This 128x48 template in effect gives the probability of seeing a foreground at different pixel locations for pedestrians.

3.4 Pedestrian Detection

Now the problem at hand is, given the probabilistic template and a test window, which in this case will be an 128x48, what is the probability that the window contains a pedestrian? Our approach is as follows.

For each pixel, the probability that the pixel has the correct classification given that the window contains a pedestrian is p(x,y) if the pixel has a value 1 or 1-p(x,y) if the pixel has the value 0. We calculate the sum of these probabilities for all pixels and this gives us the combined probability of the given window containing the person given the prior. This is expressed in equation (3)

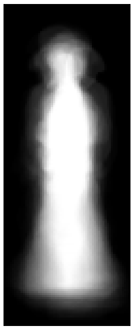

Figure 1: Probabilistic Template

$$combinedprobability(i,j) = \Sigma_{\substack{x=1..48\\y=1..128}}(th(x,y)*p(x,y)+(1-th(x,y))*(1-p(x,y))) \quad (3)$$

where th is a 128x48 window around a pixel (i,j). The implicit assumption we have made is that the intensity at a point is independent of its neighbors. In the current domain, as per the reasons discussed above in Section 3.1, for the lack of any better model, and to make the problem more tractable this seems to be a reasonable assumption and works quite well.

This 128x48 window is moved over the entire image and the *combinedprobability* calculated for each pixel (i,j). Once we have this probability map, the mean and standard deviation of the *combinedprobability* is calculated for 1000 training samples. Also the mean and standard deviation is calculated for 1000 128x48 windows that do not contain pedestrians. The probability map is then thresholded using equation (1). The value of the threshold is decided using Bayes classifier given by equation (2).

3.5 Implementation Details

Probabilistic templates were created for 3 different scales using 1000 different pedestrian templates. This idea of probabilistic template is quite robust to change of scale. A probabilistic template works well for pedestrians who are off in scale by upto 25%. Thus, 3 templates suffice to cover all scales that are of interest.

Coarse to fine approach is used to speed up the process. First the pedestrian detection is performed at a low resolution with a relatively low threshold. Only for regions that pass the threshold at the low resolution is the recognition done at higher levels. The same procedure is repeated for different scales. Three different probability maps are created and then thresholded using Bayesian classification as per equation (2). Local maximas are found on each and declared to be pedestrians. Some results of pedestrian detection can be seen in Figure 2', 3' & 4' (pedestrian heads are marked by blue contours).

The entire implementation has been done in Visual C++ and the system runs at 3 fps (640x480) or 11 fps (320x240).

4 Experimental Results

At the time of the implementation our dataset consisted of 6 infrared videos, 10 secs each clicked at 30 fps. The videos had quite a lot of variations in the scenes, sizes of people, amount of occlusions and clutter in the backgrounds as is clearly evident in Figures 2, 3 & 4 which are frames extracted from 3 different videos. The probabilistic template was created out of a 7th video clicked in a controlled environment so as to enable easy and robust extraction of human shapes. Thus none of the test data was used as training data. Preliminary experiments on this dataset of 6 videos showed detection rates ranging from 75%-90% with one false alarm per frame on an average.

The implementation is quite robust to noise and occlusions. In Figure 2', the pedestrian on the right is significantly occluded and is still detected by the system.

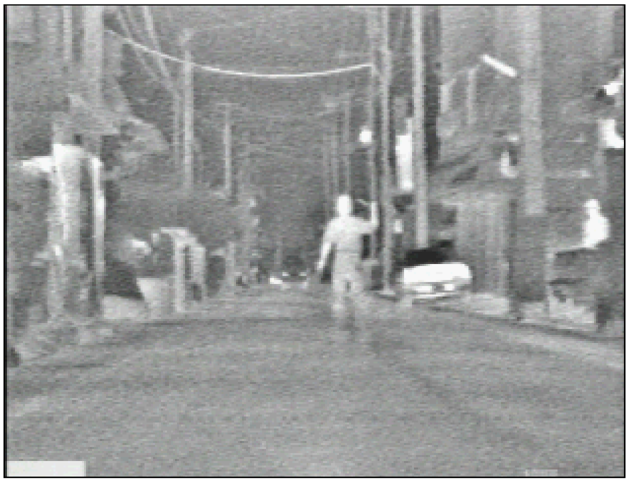

Figure 2: Input Frame 1

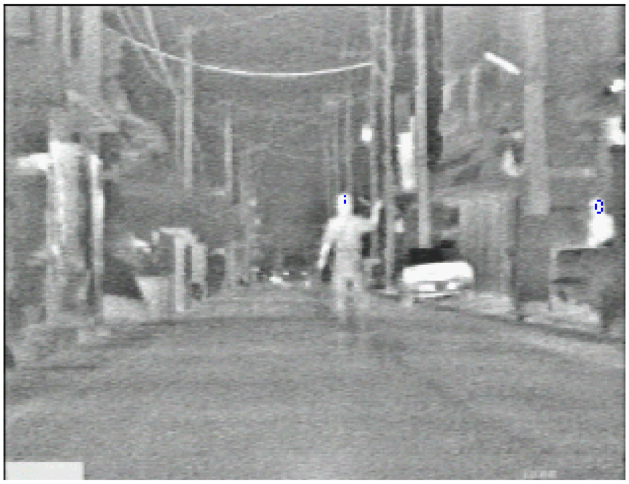

Figure 2': Output Frame 1

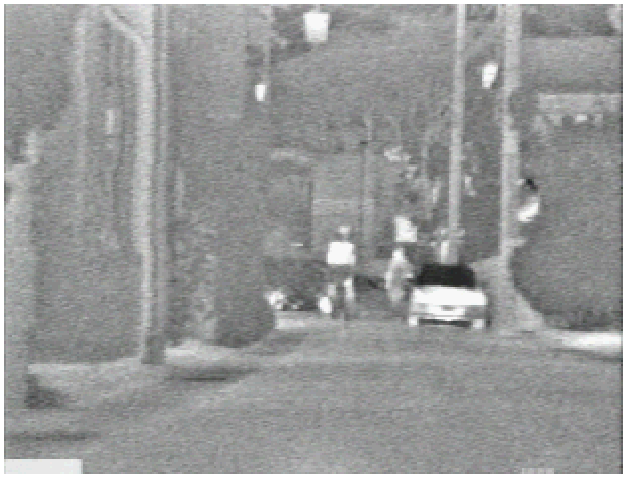

Figure 3: Input Frame 2

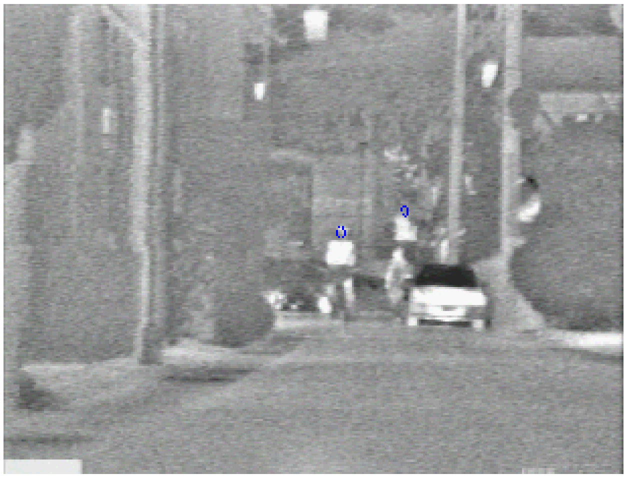

Figure 3': Output Frame 2

5 Future Work

In our current formulation the system works well when the learning and training are done on binary images. By converting into binary we are loosing some information. The benefit that we get out of this is that our decision is not based on small local variations due to noise, climate, orientation of pedestrian, and other factors. In future we need to explore how we can include the actual pixel values in our calculations and still be invariant to the factors mentioned above.

Many of the false alarms are due to lamp posts hanging high in the sky. Such false alarms that occur due to objects hanging from the sky or lying flat on earth should be easy to remove using the prior probability of occurance of pedestrians in the scene. Most of the false negatives are due to people walking in groups. This is because our training data contained only people walking by themselves and this created a bias towards better detection of people walking by themselves. We will be looking in this direction in our future work.

Also, more experimentation needs to done to benchmark the robustness and reliability of the system in the real world.

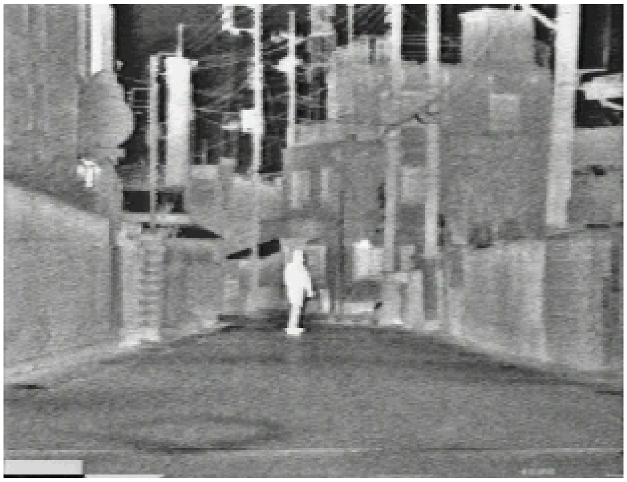

Figure 4: Input Frame 3

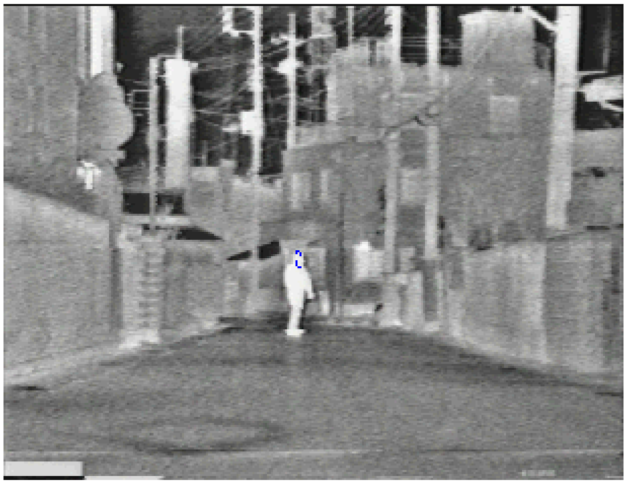

Figure 4’: Output Frame 3

6 Conclusion

In this paper we have presented a method of robustly detecting pedestrians in infrared videos taken from moving platforms. A simple yet effective idea of probabilistic templates has been introduced and used for pedestrian detection in cluttered backgrounds. The probabilistic templates contain information regarding probabilities of foreground and background at each pixel based on the training data. This encodes in itself the shape information of pedestrians and the variations that it can undergo. This technique proves to be quite robust to noise and occlusions as can be seen in Figure 2’ (pedestrian on the right is significantly occluded).

The implementation that we have developed is quite fast and with some optimizations depending on the hardware on which it will finally be used, it will be fast enough for most practical applications.

7 Acknowledgement

Support for this work and all the test infrared videos have been provided by Honda Research Labs.

References

[1] D. M. Gavrila. Protecting pedestrians in traffic: Sensor-based approaches. IEEE Intelligent Systems, 2001.

[2] C. Papageorgiou, T. Evgeniou, T. Poggio. A trainable pedestrian detection system. IEEE Int. Conf. on Intelligent Vehicles, pp. 241-246, Germany, Oct 1998.

[3] L. Zhao, C. Thorpe. Stereo- and neural network based pedestrian detection. In Proceedings ITSC, Tokyo, Japan, 1999.

[4] D. M. Gavrila. Pedestrian detection from a moving vehicle.

[5] C. Wren, A. Azarbayejani, T. Darrell, A. Pentland. Pfinder: Real-time Tracking of the Human Body, IEEE Trans. on Pattern Analysis and MAchine Intelligence, Vol. 19, No. 7, pp. 780-785, July 1997.

[6] I. Haritaoglu, D. Harwood, L. Davis. W4-Real Time Detection and Tracking of People and their Parts. Technical Report, University opf Maryland, Aug. 1997.

[7] J. Segen, S. Pingali. A camera-Based System for Tracking People in Real Time. Proc. of the 13th Int. Conf. on Pattern Recognition, pp. 63-67, 1996.

[8] P. J. Burt, J. R. Bergen, et al. Object Tracking with a Moving Camera: An Application of Dynamic Motion Analysis. Proc. of IEEE Workshop on Visual Motion, pp. 2-12, 1989.

[9] R. Polana, R. Nelson. Low level recognition of human motion. Proc. of IEEE Workshop on Motion of Non-Rigid and Articulated Objects, pg. 77-82, Austin, 1994.

[10] R. Cutler, L. Davis. Real-time periodic motion detection, analysis and applications. Proc. of IEEE Conference on Computer and Pattern Recognition, pg. 326-331, Fort Collins, USA, 1999.

[11] C. Wohler, J. K. Aulanf, T. Portner, U. Franke. A Time Delay Neural Netowrk Algorithm for Real-time Pedestrian Recognition. International Conference on Intelligent Vehicle, Germany 1998.

[12] H. Mori, N. M. Charkari, T. Matsushita. On Line Vehicle and Pedestrian Detection Based on Sign Pattern. IEEE Trans. on Industrial Electronics, Vol. 41, No. 4, pp. 384-391, Aug, 1994.

[13] A. A. Niyogi, E. H. Adelson. Analysing Gait with Spatiotemporal Surfaces. IEEE Workshop on Motion of Non-Rigid and Articulated Objects. pp. 64-69, Austin, 1994.

[14] S. A. Niyogi, E. H. Adelson. Analysing and Recognizing Walking Figures in xyt. IEEE Conference on Computer Vision and Pattern Recognition, pp. 469-474, 1994.

[15] D. Hogg. Model-based Vision: a Program to See a Walking Person. Imae and Vision computing, Vol. 1, No. 1, pp. 5-20, 1983.

[16] K. Rohr. Towards Model-Based Recognition of Human Movement in Image Sequences. CVGIP: Image Understanding, Vol. 59, No. 1, pp. 94-115, Jan, 1994.

[17] J. Lipton, H. Fujioshi, R. S. Patil. Moving Target Classification and Tracking from Real-Time Video. Workshop on Applications of Computer Vision, Princeton, NJ, Oct. 1998.

[18] U. Franke, D. M. Gavrila, et al. Autonomous Driving goes downtown. IEEE Intelligent Systems, 13(6):40-48, 1998.

[19] T. Tsuji, H. Hattori, N. Nagaoka, M. Watanabe. Development of night vision system. Proc. IEEE International Conference on Intelligent Vehicles, pg: 133-140, Tokyo, Japan, 2001.