Pedestrian Detection in Infrared Images based on Local Shape Features

Li Zhang, Bo Wu and Ram Nevatia
University of Southern California
Institute for Robotics and Intelligent Systems
Los Angeles, CA 90089-0273
$\{zhang11|bowu|nevatia\}$@$usc.edu$

Abstract

Use of IR images is advantageous for many surveillance applications where the systems must operate around the clock and external illumination is not always available. We investigate the methods derived from visible spectrum analysis for the task of human detection. Two feature classes (edgelets and HOG features) and two classification models(AdaBoost and SVM cascade) are extended to IR images. We find out that it is possible to get detection performance in IR images that is comparable to state-of-the-art results for visible spectrum images. It is also shown that the two domains share many features, likely originating from the silhouettes, in spite of the starkly different appearances of the two modalities.

1. Introduction

Video surveillance systems in most applications must operate on an around the clock basis. Normal cameras that produce visible spectrum images (called VS images from here on) are not very effective without the presence of external illumination. In day time, this illumination can come from the sun (though deep shadows may still be a problem) but at night, artificial illumination is required. Important areas of interest could be lit with bright lights but undesirable activities are more likely to occur in darker areas. Infrared (IR) cameras are ideally suited to image under these conditions, as they sense emitted radiation from objects of interest such as humans and vehicles. They can be used to image in the day time as well though in the day time, one may also want to supplement the data with VS images. The quality of IR images has been improving and the prices have been declining rapidly. While they still remain rather expensive for deployment on a large scale, the prices are affordable enough to start experimenting with them to develop techniques that work with IR data.

In this paper, we focus on the task of detecting humans, in standing or walking poses (*i.e*. pedestrians), from IR images. Detection of humans is of fundamental importance as they are the principal agents in carrying out activities of interest in a surveillance environment. Pedestrian detection has been a subject of intense study for the VS images. An interesting question is whether, and to what extent, do the techniques developed for VS images transfer to IR images and whether the same set of methods is superior for IR images as is for the VS images.

The VS and IR images share several common characteristics. The appearance of objects changes with viewpoint in similar ways. However, VS images are highly sensitive to external illumination changes. Also, the texture pattern details, such as clothing worn by humans, are imaged in fine detail (depending on the imaging resolution). In IR images, however, the fine texture details are lost as the human body temperature is relatively constant over the entire body though the measured intensity does depend on clothing to some extent. Not only are many details lost in the interiors of humans but also in the background. Thus, IR images have appearance that can be described as blob-like or "blobby". Fig.1 shows some examples that make the differences apparent.

In spite of the many differences between the VS and IR images, and their starkly different appearances, one can see that the silhouettes of human objects are similar (some differences may occur due to different resolutions and contrast). Thus, we expect that methods for pedestrian detection that emphasize silhouettes should function well for both classes of images though the importance of different features may be different for the two classes of images.

In particular, we explore the use of two classes of features that have been used in previous works: one is the set of edgelet features introduced in [7]; the other is the HOG feature used in several recent papers [8, 1]. We investigate the use of these features with two types of classification models: those based on AdaBoost methods introduce in [16] and SVMs. The features and classification models are then combined to perform detection: AdaBoost serves as a fea-

1-4244-1180-7/07/$25.00 ©2007 IEEE

(a) (b)

Figure 1. Examples of data in visible and infrared spectrum:(a) humans in visible spectrum images; (b) humans in infrared images

ture selection procedure for the SVM classifier to set up a tractable feature space, and the SVM classifier is also organized as a cascade, which is novel to the best of our knowledge.

In this paper, we do not make use of motion properties as we want to be able to detect standing humans and also not have to rely on the assumption that the cameras are static and a reliable model of the background can be learned. Our methods could be supplemented by motion information where available.

Experiments are conducted on a variety of VS and IR image datasets. Our main observation is that the VS methods do translate to IR images to a large extent; however, the methods can benefit from training on IR images and that the performance of different types of methods are not the same on IR images as they are on VS images.

The rest of this paper is organized as follows: section 2 contains description of related work, section 3 provides an overview of our approach, 4 and section 5 describe the image features and learning methods we compared respectively; section 6 shows the experimental results; and some discussions and conclusions are given in the last section.

2. Related Work

The literature on pedestrian detection in VS images is abundant. Mainly two types of image cues are used, motion and shape. The motion based methods, *e.g*. [12, 15], rely on the output of a motion segmentation preprocess. The motion segmentation methods compute the difference between the current image and a background model of the scene and output some motion blobs. Further processing is required to distinguish between categories of moving objects, such as vehicles and humans. Motion blobs also do not necessarily correspond to single objects; further processing is needed to split and merge blobs to find humans; this process usually requires some shape analysis. A basic limitation of motion-based segmentation is sensitivity to sudden illumination changes, the requirement of a static camera or situations where background motion does not have strong parallax, and inability to detect static objects.

Shape analysis can be applied either directly to the image or to the detected motion blobs. Several types of shape features have been used. Some of these are global features, *e.g*. the edge template of Gavrila *et al*. [20, 19]. Others are local features, *e.g*. the Haar wavelet of Mohan *et al*. [18], the SIFT like orientation feature of Mikolajczyk *et al*. [13], the Histogram of Oriented Gradient (HOG) descriptor of Dalal and Triggs [8], the edgelet feature of Wu and Nevatia [7], and the motion enhanced Haar feature of Viola *et al*. [16].

To build an object model from the features, some methods use graphical models, *e.g*. the Markov Random Field model of Wu *et al*. [6], and the Implicit Shape Model of Leibe *et al*. [5]. Other methods use learning based approaches, *e.g*. the SVMs [24] classifiers in [18, 8], and the ensemble cascade structured classifiers learned by the AdaBoost algorithm [23] in [13, 7, 1]. To detect partially occluded objects, some methods use a part based representation, *e.g*. [7], where a detector is built for each individual part. Compared to motion based methods, shape based methods are more general, because they can detect both stationary and moving target objects, even in present of other moving objects, and they do not rely on motion segmentation results.

Compared to the methods for data from visible spectrum, less attention has been paid to the infrared data. Most of the human detection methods, *e.g*. [11, 10, 9, 2, 3] introduced for IR data are based on background modeling and pixel-level segmentation.

[9, 10] are both based on a contour saliency map. [9] used template matching with a set of edge templates, while [10] did a pixel-level segmentation combining both thermal and VS image information. [11] models the foreground/background layers through EM algorithm and detected human based on a set of statistical criteria. [2, 3] combined background subtraction with blob tracking. Most of these methods do not model human shape explicitly, and the assumption of a static background constrains the performance of these methods.

3. Overview of our Approach

We treat human detection as a general object classification problem. Our approach consists of three stages (as shown in Fig.2): local feature extraction, feature selection and object classification. In feature extraction part, local spatial information of the image is transformed into a feature space that is descriptive of the shape and appearance of the object but less sensitive to noise and other varia-

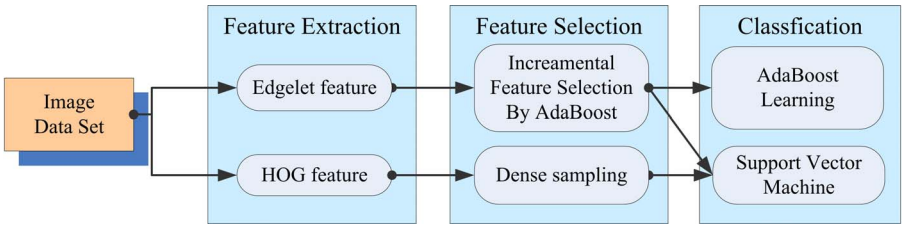


Figure 2. Outline for our approach

tions. Usually a large number of local features are needed to represent an object globally. For local feature such as edgelets[7], the size of the entire feature set can be in the millions. It is not practical to compute and use all of these features so a selection procedure is necessary to trim their number. For some algorithms, such as AdaBoost, the feature selection procedure is intrinsically combined with classifier learning, while for others such as an SVM, the feature selection needs to be performed explicitly. After the features have been selected, all the training samples of object and background are represented in this feature space. A classifier is optimized to separate the two classes of samples, usually by minimizing an error-measurement function.

The two types of local shape features we use are the HOG descriptor [8] and edgelets [7]. We combine them with two classification models: AdaBoost cascade classifiers [17] and cascade of SVM classifiers[24]. These techniques have demonstrated good performance on the pedestrian detection task in VS images [8, 7]. Three feature-learner combinations are evaluated: HOG based SVM, edgelet based SVM, and edgelet based AdaBoost.

For pedestrian detection in VS images, there are several publicly available data sets, *e.g*. the MIT pedestrian set [22], the USC pedestrian set [7], and the INRIA pedestrian set [8]. For pedestrian detection in infrared images, OTCBVS Benchmark Dataset Collection is a large data set of video and images in non-visible spectrum, which has been used in several previous works, *e.g*. [11, 10, 9, 2, 3]. The OTCBVS data has limited background scenes, and size of pedestrians is relatively small. To provide a bigger variety, we also collected our own IR data as desribed later in the paper.

We perform a quantitative evaluation of the different combinations of features and classifiers on the IR images (also on a small set of VS images for comparison). We also apply the classifiers learned on one modality to the other to evaluate the shared features between them.

4. Image Features

We now describe the two features used in this work.

4.1. Edgelet Feature

An edgelet is a short segment of a line or a curve. Denote the positions and normal vectors of the points in an edgelet, E, by $\{\mathbf{u}_i\}_{i=1}^k$ and $\{\mathbf{n}_i^E\}_{i=1}^k$, where k is the length of the edgelet. Given an input image I, denote by $M^I(\mathbf{p})$ and $\mathbf{n}^I(\mathbf{p})$ the edge intensity and normal at position $\mathbf{p}$ of I. In practice, the edge orientation is quantized[7]. Denote by $\{V_i^E\}_{i=1}^k$ and $V^I(\mathbf{p})$ the quantized edge orientations of the edgelet and the input image I respectively. The affinity between the edgelet E and the image I at position $\mathbf{w}$ is calculated by

$$f(E;I,\mathbf{w}) = \frac{1}{k}\sum_{i=1}^{k} M^I(\mathbf{u}_i+\mathbf{w}) \cdot l\left[V^I(\mathbf{u}_i+\mathbf{w}) - V_i^E\right] \tag{1}$$

where $l[\cdot]$ is an approximation of the cosine function.

Note, $\mathbf{u}_i$ in the above equation is in the coordinate frame of the sub-window, and $\mathbf{w}$ is the offset of the sub-window in the image frame. The edgelet feature values within one sub-window is normalized by the gray scale standard deviation of the sub-window concerned. The edgelet affinity function captures both intensity and shape information of the edge.

In practice, the length of one single edgelet is between 4 pixels and 12 pixels. The edgelet features we use consist of single edgelets, including lines, $\frac{1}{8}$ circles, $\frac{1}{4}$ circles, and $\frac{1}{2}$ circles, and their symmetric pairs. A symmetric pair is the union of a single edgelet and its mirror.

4.2. HOG descriptor

Histogram of Oriented Gradients(HOG) feature[8] is a grey-level image feature formed by a set of normalized gradient histograms.

For each spatial rectangle cell an oriented gradient histogram is calculated by weighted votes. Each pixel within the cell votes for its gradient orientation weighted by its gradient magnitude. The sign of the orientation is omitted, because the contrast of object and background is unreliable due to changing appearance(and also polarity changes in infrared). A fine quantization in orientation is suggested for good performance in [8]. In our implementation, $[0,\pi)$ are evenly divided into 9 bins. Neighboring cells are grouped into a block in which the histogram bin values are normalized using the L2-normal. Blocks are sampled through the

image with 1/2 overlap with each other, and by all the normalized histograms from each block. A single HOG feature is formed as a high-dimensional feature vector.

The computation cost of the original HOG feature [8] is relatively high. However, by employing an integral image, the histogram for each separate orientation can be calculated efficiently but loose the sub-pixel accuracy of the original method. We find that the effect of this approximation on detection accuracy is negligible.

5. Learning Methods

We compare two classifiers: the boosted cascade classifier and the SVM classifier. We briefly describe the two methods below.

5.1. Boosted Cascade Classifier

We use an enhanced version [14] of the original boosting method of Viola and Jones [17] to learn pedestrian detectors. We evenly divide the range of feature values into n sub-ranges; in our experiments, we set $n = 32$. For object detection, a sample is represented as a tuple $\{\mathbf{x}, y\}$, where $\mathbf{x}$ is the normalized image patch and y is the class label whose value can be $+1$ (object) or -1 (non-object). According to the real-valued version of AdaBoost algorithm [21], the optimized weak classifier $h^{(w)}$ based on an image feature f can be calculated as

$$h^{(w)}(\mathbf{x}) = \frac{1}{2}\sum_{j=1}^{n} \ln\left(\frac{\bar{W}_{+1}^{j} + \varepsilon}{\bar{W}_{-1}^{j} + \varepsilon}\right) B_n^j\left(f(\mathbf{x})\right) \qquad (2)$$

where ε is a smoothing factor [21],

$$\bar{W}_c^j = P\left(f(\mathbf{x}) \in [\frac{j-1}{n}, \frac{j}{n}), y = c\right), c = \pm 1, j = 1 \ldots n \qquad (3)$$

and $B_n^j(u)$ is an indicator function

$$B_n^j(u) = \begin{cases} 1, & u \in \left[\frac{j-1}{n}, \frac{j}{n}\right) \\ 0, & \text{otherwise} \end{cases}, j = 1 \ldots n \qquad (4)$$

For each feature, one weak classifier is optimized. Then the real AdaBoost algorithm [21] is used to learn strong classifiers, called layers. The strong classifier $h^{(s)}$ is a linear combination of a series of weak classifiers selected from the weak classifier pool:

$$h^{(s)}(\mathbf{x}) = \sum_{i=1}^{T} h_i^{(w)}(\mathbf{x}) - b \qquad (5)$$

where T is the number of weak classifiers in $h^{(s)}$, and b is a threshold.

The learning procedure of one strong classifier is referred as a *boosting stage*, *i.e.* one layer in the cascades structure. At the end of each boosting stage, the threshold b is set so that $h^{(s)}$ has a high detection rate (99.8% in our experiments) and reject as many negative samples as possible. The accepted positive samples are used as the positive set for the training of the next boosting stage; the false alarms obtained by scanning the negative images with the current detector are used as the negative set for the next boosting stage. Finally, the strong classifiers of all boosting stages are connected sequentially as a cascade to perform the classification. An input is accepted as object if and only if it passes all the layers in the cascade.

5.2. SVM Cascade Classifier

Support Vector Machines are a well-known statistical learning method, proposed by Vapnik[25]. Especially, it is effective for learning with small sampling in high-dimensional spaces.

The objective in SVM learning is to find a decision plane that maximizes the inter-class margin. The feature vectors are projected into a high dimensional space by the "kernel trick". The final SVM classifier has the following form

$$f(x) = \sum_i w_i K(x, x_i) \qquad (6)$$

where x_i are support vectors and $K(x, y)$ is the kernel function. In our experiments, we use polynomial kernels.

There are two main issues in applying the SVM method for object detection. One is that the feature set may be too large to be included in entirety for SVM learning, and a feature selection method is needed; in our case, the set of all possible edgelet features is too large to be used directly. We run AdaBoost as the feature selection method. All the edgelet features selected in the boosting procedure, are used to form the feature descriptor for SVM learning.

The other issue is that usually the training data for detection are highly unbalanced between the object and background examples; *i.e.* negative samples are easy to collect and we have many more of them than positive examples. To overcome this bias, we apply the cascade structure to SVM classifiers as for the AdaBoost classifiers. Bootstrapping is used to re-collect the negative samples after each layer of cascade is trained. With the cascade structure, we are able to organize the SVM classifiers in a novel way, so that they can be trained using one or two thousand positive samples against much more negative samples collected from thousands of background images. The final cascade SVM classifier accepts an input if and only if it passes through all the classifiers in the cascade.

6. Experimental Results

In this section, we describe our experimental setup and the results. In our experiments, we used two dataset, one for

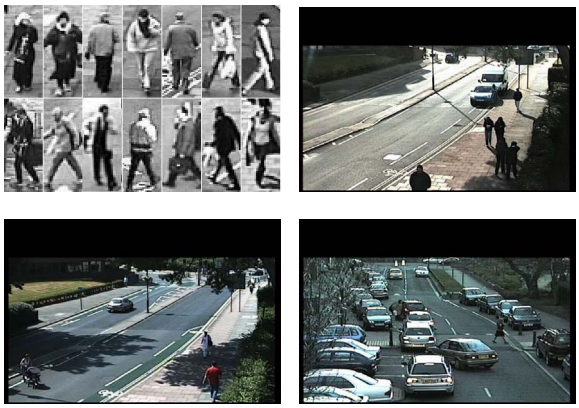


Figure 3. Images in our VS dataset: top-left are the training samples; others are the testing images

Figure 4. Images in our collected IR dataset: top-left are the training samples; others are the testing images

the VS images and the other for IR images. Eeach feature-leaner combination was trained and tested on the two image data sets. On each of the two types of data, the training and test sets are the same for every combination, so that the results are directly comparable.

6.1. Data set

We now describe the datasets.

VS image data: For VS images, we used a public data set [4]. The data set consists of 61 street surveillance video sequences that include pedestrians with different view points and sizes between 25 and 60 pixels wide (those whose with size less than 24 pixels were omitted for both training and testing). The training set includes 2326 pedestrian image samples collected through 30 of the sequences as the positive sample set, and 2311 background images as the negative sample set. In the testing set, there are 250 images sampled randomly from the other 31 sequences in the data set, which include 222 human objects in total. Some of the training and testing images are shown in Fig.3.

IR image data: For IR images, we collected ourselves a set of 13 video sequences containing pedestrians from multiple view points and sizes, using a Thermal-Eye 250D Camera (spectral response in 7-14 microns). Images were captured at both day and night and the width of humans in the images range form 40 to 200 pixels. We then divided the data into training and testing set. The training set includes 1097 pedestrian samples as the positive set, and 937 background images as the negative set, from seven of the video sequences. The testing set is constructed with 100 images sampled randomly from the other six video sequences, which include 216 human objects in total. Some of the training and testing images are shown in Fig.4.

6.2. Training

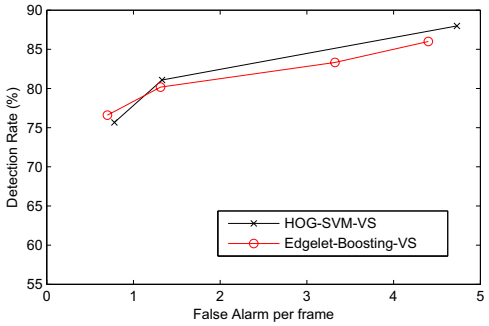
There are three feature-learner combinations evaluated in our experiment. In all these cases, the training sample size is fixed to 24×58; test images are resized by various factors so as to detect humans in multiple scales.

Edgelet based AdaBoost cascade Using AdaBoost method, edgelets are selected automatically from a large feature pool containing 108499 features. Weak classifiers are learned and combined as weighted-sum to form a strong classifier as one layer of the detector. Then we use existing layers to scan the non-human image set and collect negative samples again for the training of the detector in the next layer, until the criteria of a set false alarm rate is reached.

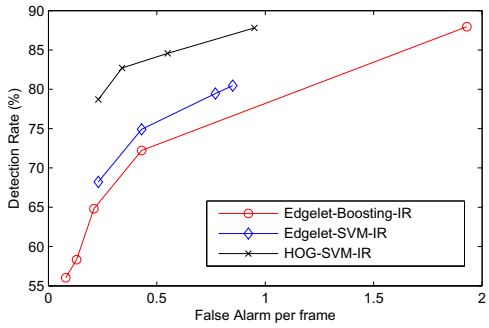
For Edgelet based AdaBoost cascade, two detectors were trained on VS images and IR images respectively. For VS images, the final cascade detector contains 277 edgelet features; for IR images, the training converged with 117 edgelet features. We denote the two detectors by the Edgelet-Boosting-VS and Edgelet-Boosting-IR detectors.

Edgelet based poly-kernel SVM cascade: The set of edgelet features in this combination is obtained by using AdaBoost as a feature selection method. Then, unlike the AdaBoost detector, all these edgelet features are combined into a single feature descriptor, by conjoining all affinity response of the edgelet features. The SVM detector with 3-degree polynomial kernels is learned as a two-class classifier of the feature descriptors on the training sets. Similar to AdaBoost, a cascade structure of detectors and a bootstrapping strategy are employed. This detector was trained only on the IR images (as our goal is not to study performance primarily on VS images). We refer to this detector as the Edgelet-SVM-IR detector.

HOG based poly-kernel SVM cascade: As recommended in [8], we sample HOG features densely by a 3 pixel interval, with a cell size of 6×6 and a block size of



(a) ROC curves on VS images



(b) ROC curves on IR images

Figure 5. Detection rate vs. false alarm per frame on the VS and IR test sets

2×2. This results in a 972-dimensional HOG feature descriptor for each sample. With HOG feature descriptors calculated from all training samples, SVM detectors are then trained using 3-degree polynomial kernels in the same way as that for edgelet features. We applied training to both VS and IR images, and refer to the detectors learned as the HOG-SVM-VS and HOG-SVM-IR detectors respectively.

6.3. Tests on VS Images

First, we briefly describe the performance of Edgelet-Boosting-VS and HOG-SVM-VS on the VS image test set, to show that both of the methods achieve state-of-the-art performance. The VS image test set includes 250 images, size of 720×480. The image width of human ranges from 24 to 60 pixels as described in Sec.6.1. For each image, 251939 detection windows are scanned at different scales and positions, which means that 2.5 false alarms per frame(FPP) is equivalent to 1e-5 false alarms per window(FPW).

The curves of detection rate vs. FPP are shown in Fig.5(a). The performances of both Edgelet-Boosting-VS and HOG-SVM-VS achieve a detection rate of about 80% with 1e-5 FPW(2.5FPP). These results are similar to the performance given in [8] and [7]. Some examples of the detection results on VS images for both methods are shown in the first two rows of Fig.6; the first row shows the results of the HOG-SVM-VS detector, the second row shows the results of the Edgelet-Boosting-VS detector. The top blanked areas are marked as "don't care" in the original dataset and hence are not processed (to eliminate small-sized pedestrians).

6.4. Test on IR Images

In this experiment, we compare the performance of Edgelet-Boosting-IR, Edgelet-SVM-IR and HOG-SVM-IR on the IR test set.

The IR test set contains 100 images, with a size of 720×480. The image width of humans ranges widely from 40 to 200 pixels as described in section 6.1. For each image, 116758 detection windows are scanned at different scales and positions, which means 1.17 FPP is equivalent to 1e-5 FPW.

The curves of detection rate vs. false alarms per frame are shown in Fig.5(b). From the results, It can be seen that on IR images, HOG features outperform edgelet features with either AdaBoost or SVM. Edgelet features based SVM are slightly better than edgelet based AdaBoost. The second observation is consistent with intuition because the SVM makes use of all the features at the same time. For the weaker performance of Edgelet-Boost-IR against HOG-SVM-IR, we investigated the learned edgelet features selected by the AdaBoost on both VS and IR training dataset. We find that on VS dataset, some of these features are from the interior texture of the object; they get used mainly to reject the negative samples. The weights learned by HOG-SVM-VS also have similar property as discussed by [8]: HOG features along boundary have positive weights while HOG features inside/outside the boundary have negative ones. On IR dataset, however, as there is less texture, many edges corresponding to the edgelet features inside the object do not exist anymore, compared to VS data. Therefore, those "negative" edgelets do not work as well as in VS images, which increases the false alarm rate. The histogram-type HOG feature is less affected by IR data. Some examples of results of the HOG-SVM-IR detector on IR dataset are shown in In the third and fourth row of Fig.6.

We also test our methods on the OSU Color-Thermal data set[10] which is publicly available. However, as most of the previous works tested on this data set base their methods on background subtraction, segmentation and blob tracking, the performance is not directly comparable. Some examples of the detection results on this dataset are shown in the last row of Fig.6.

6.5. Cross-modality Evaluation

In this experiment, we apply our detectors learned from VS data directly to the IR dataset and evaluate the cross-

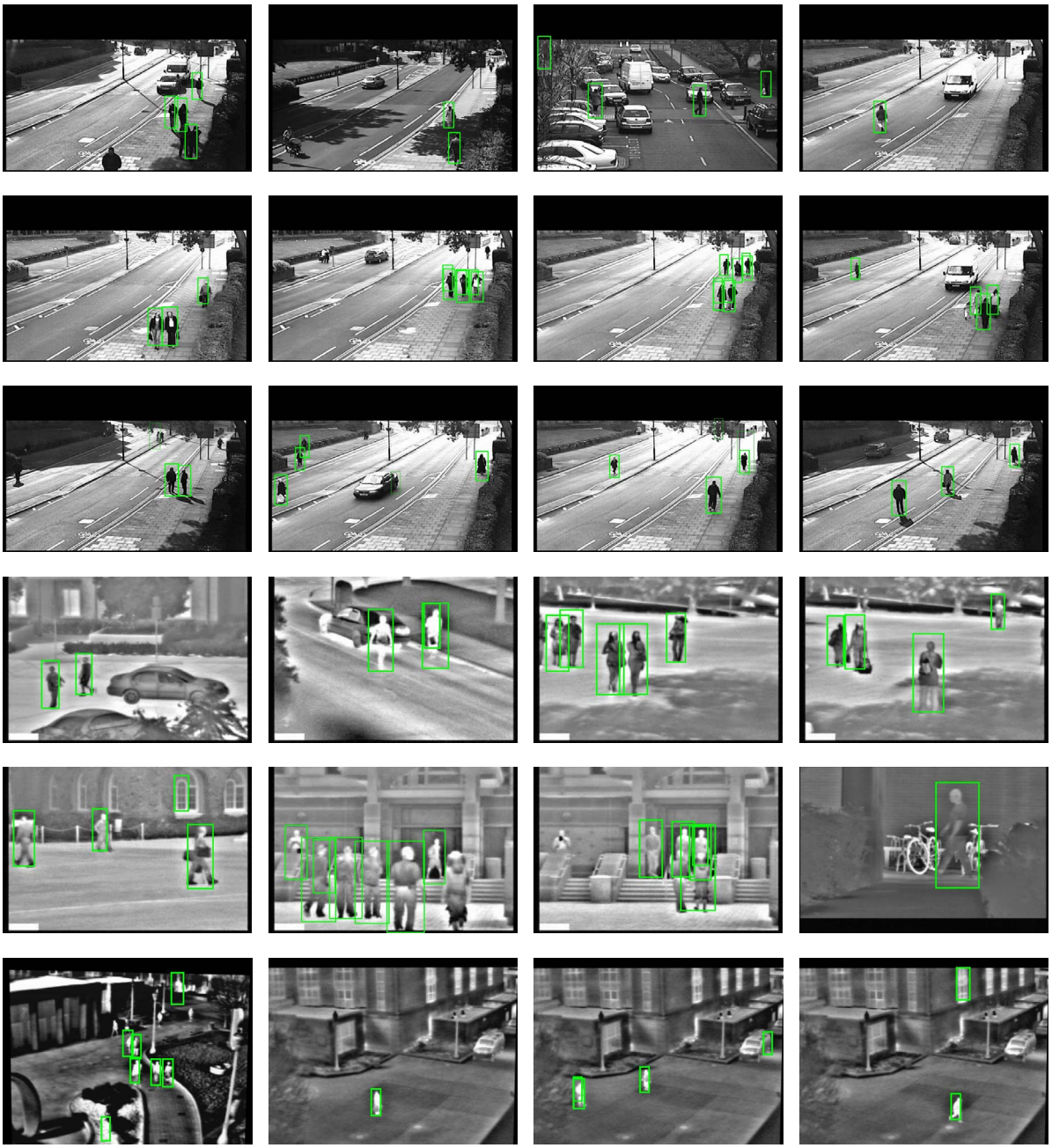


Figure 6. Examples of detection results: On VS image dataset, the first row is for HOG-SVM-VS; the second row is for of Edgelet-Boosting-VS. On IR image dataset, the third and fourth rows are for HOG-SVM-IR on our own IR dataset; the fifth row is for HOG-SVM-IR on the OSU Color Thermal dataset

modality performance. Table.1 gives the performance of Edgelet-Boosting-VS, HOG-SVM-VS and HOG-SVM-IR on both VS and IR test sets. For each detector in the table, the same setting of parameters is used for testing on the two datasets.

In the first two rows of Table.1, the detectors are learned on VS dataset and tested on both VS and IR dataset. We achieve similar detection rates on these two types of data, but the false alarm rate on VS data is much lower than on IR data. One possible explanation is that this is due to the relatively large differences in the background texture between VS and IR images whereas the object properties are domi-

nated by silhouette features in both cases.

In the third row of the table, the detector is learned on IR dataset and tested on both types of data. From the result it can be seen that the detection rate drops by 10% when we apply the detector trained on IR images to VS images. We speculate that this is because the appearance of humans in VS images has much larger variation than in IR images and this variation is not modeled by the IR trained detector. However, a detection rate of 68% still gives a hint that many features are shared among VS and infrared images.

Overall, the experimental results show that the two imaging modalities share many features.

Table 1. Testing on data unmatched with training set: EBV is Edgelet-Boosting-VS, HSV is HOG-SVM-VS, HSI is HOG-SVM-IR, DR is detection rate, FFW is false alarm per window

	VS Image dataset		IR Image dataset	
	DR(%)	FFW	DR(%)	FFW
EBV	79	0.67e-5	79.68	1.67e-5
HSV	75.67	0.22e-5	79.16	0.66e-5
HSI	68.46	1.12-5	78.70	0.19e-5

7. Conclusion and Discussion

We have presented several combinations of image features and classification models and compared their performance on a significant number of examples. One main conclusion is that by applying the state-of-the-art methods developed for VS data to IR data, we can achieve a detection accuracy on IR images comparable to to that on VS images. The choice of features and classifiers may be different for the two modalities but it is also clear that the two share some common properties, primarily those captured by the silhouettes. This suggests that we do not need to invent radically different methods for the IR domain. We hope that these results and observations will help promote further research into use of the IR images which have some inherent advantages for around the clock usage in environments that can not be conveniently lit by artificial illumination.

8. Acknowledgement

This research was supported, in part, by the Office of Naval Research with Contract #N00014-06-1-0470.

References

[1] Q. Zhu, S. Avidan, M. Yeh, K. Cheng. Fast human detection using a cascade of Histograms of Oriented Gradients. CVPR 2006. 1, 2

[2] A. Leykin, R. Hammoud, Robust Multi-Pedestrian Tracking in Thermal-Visible Surveillance Videos. CVPR Workshop OTCBVS, 2006. 2, 3

[3] J. Wang, G. Bebis, R. Miller, Robust Video-Based Surveillance by Integrating Target Detection with Tracking. CVPR Workshop OTCBVS, 2006. 2, 3

[4] Human Detection and Tracking Test Sequences of CLEAR-VACE Eval, 2006. http://isl.ira.uka.de/clear06/?Evaluation_Tasks 5

[5] B. Leibe, E. Seemann, and B. Schiele. Pedestrian Detection in Crowded Scenes. CVPR 2005. 2

[6] Y. Wu and T. Yu and G. Hua. A Statistical Field Model for Pedestrian Detection. CVPR 2005. Vol I: 1023-1030 2

[7] B. Wu, and R. Nevatia. Detection of Multiple, Partially Occluded Humans in a Single Image by Bayesian Combination of Edgelet Part Detectors. ICCV 2005. Vol I: 90-97 1, 2, 3, 6

[8] N. Dalal, and B. Triggs. Histograms of Oriented Gradients for Human Detection. CVPR 2005. Vol I: 886-893 1, 2, 3, 4, 5, 6

[9] J. Davis, M. Keck, A Two-Stage Template Approach to Person Detection in Thermal Imagery. WACV, 2005. 2, 3

[10] J. Davis, V. Sharma, Fusion-Based Background-Subtraction using Contour Saliency. CVPR Workshop OTCBVS, 2005. 2, 3, 6

[11] C. Dai, Y. Zheng, X. Li, Layered Representation for Pedestrian Detection and Tracking in Infrared Imagery. CVPR Workshop OTCBVS, 2005. 2, 3

[12] T. Zhao and R. Nevatia. Tracking multiple humans in crowded environment. CVPR 2004. Vol II: 406-413 2

[13] C. Mikolajczyk, C. Schmid, and A. Zisserman. Human detection based on a probabilistic assembly of robust part detectors. ECCV 2004. Vol I: 69-82 2

[14] C. Huang, H. Ai, B. Wu, and S. Lao. Boosting Nested Cascade Detector for Multi-View Face Detection. ICPR 2004. 4

[15] T. Zhao, and R. Nevatia. Tracking Multiple Humans in Complex Situations, PAMI, 26(9): 1208-1221, 2004. 2

[16] P. Viola, M. Jones, and D. Snow. Detecting pedestrians using patterns of motion and appearance. ICCV 2003. 1, 2

[17] P. Viola, and M. Jones. Rapid Object Detection Using a Boosted Cascade of Simple Features. CVPR 2001. 3, 4

[18] A. Mohan, C. Papageorgiou, and T. Poggio. Example-based object detection in images by components. PAMI, 23(4):349C361, 2001. 2

[19] D. Gavrila. Pedestrian detection from a moving vehicle. ECCV 2000. 2

[20] D. Gavrila and V. Philomin. Real-Time Object Detection for "Smart" Vehicles. ICCV 1999. 2

[21] R. E. Schapire and Y. Singer. Improved Boosting Algorithms Using Confidence-rated Predictions. Machine Learning, 37: 297-336, 1999. 4

[22] C. Papageorgiou, T. Evgeniou, and T. Poggio A Trainable Pedestrian Detection System. In: Proc. of Intelligent Vehicles 1998. pp. 241-246 3

[23] Y. Freund and R. E. Schapire. Experiments with a New Boosting Algorithm. Machine Learning 1996. pp. 148-156 2

[24] C. Cortes, and V. Vapnik. Support-Vector Networks, Machine Learning, 20, 1995. 2, 3

[25] V. Vapnik, The Nature of Statistical Learning Theory. Springer-Verlag, 1995. 4