Fast Pedestrian Detection from a Moving Vehicle

by

Shuang You

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2006

Author .....
Department of Electrical Engineering and Computer Science
May 26, 2006

Certified by .....
Trevor Darrell
Associate Professor
Thesis Supervisor

Accepted by .....
Arthur C. Smith
Chairman, Department Committee on Graduate Theses

Fast Pedestrian Detection from a Moving Vehicle

by

Shuang You

Submitted to the Department of Electrical Engineering and Computer Science
on May 26, 2006, in partial fulfillment of the
requirements for the degree of
Master of Engineering

Abstract

This paper presents a method of real-time multi-modal pedestrian detection from a moving vehicle. The system uses both intensity and thermal images captured from cameras mounted at the front of the vehicle to train cascades of classifiers, which results in a detector that is able to detect a large percentage of pedestrians with very few false positives. The system has also been tested with inputs of high-resolution intensity images along with low-resolution thermal images, showing that the addition of even a low-resolution thermal camera may return better pedestrian detection results than using only intensity information alone.

Thesis Supervisor: Trevor Darrell
Title: Associate Professor

Contents

1 Introduction **11**

1.1 Motivation 11

1.2 Goals 12

1.2.1 Detection Speed 12

1.2.2 False Positive Rate 12

1.2.3 Hardware Price 13

1.3 Past Related Work 13

1.4 Approach 16

2 Implementation **17**

2.1 Data Collection 17

2.2 Thermal and Intensity Alignment 20

2.3 Labeling 23

2.4 Training 26

2.5 Testing 27

2.6 Low-Resolution Thermal Images 27

3 Results **29**

3.1 Intensity Detector 29

3.2 Thermal Detector 30

3.3 Low-Resolution Thermal Results 31

3.4 Comparison of Various Detectors 34

4 Conclusion **37**

4.1 Summary 37

4.2 Future Work 38

A Sample Labels **41**

B Sample Detection Results **45**

B.1 Indoors–Comparsion of various resolution thermal-only detectors . . . 47

B.2 Outdoors–Comparsion of various resolution thermal-only detectors . . 54

B.3 Outdoors–Intensity-Only detector and Combined Intensity and Thermal detectors 61

B.4 Indoors–Intensity-only, thermal-only, and combined intensity and thermal detectors 64

List of Figures

2-1 Data Collection Equipment 18
2-2 Sample Indoor Frames 19
2-3 Sample Outdoor Frames 19
2-4 Corresponding Points Chosen for Alignment of Intensity and Thermal Frames 21
2-5 Intensity and the Corresponding Warped Thermal Frame 22
2-6 Examples of Positive and Negative Labels Within a Frame (green = positive, red = negative) 23
2-7 Examples of a Positive Label in labeler 24
2-8 Adjustment of a Positive Label Through labeler 25
2-9 Set of Basic Haar-like Features Used 26
2-10 Low-Resolution Thermal Images of Various Down-Scaling Factors . . 28

3-1 Intensity Detector Results 29
3-2 First Six Features of the Intensity Detector 30
3-3 Thermal Detector Results 30
3-4 First Six Features of the Thermal Detector 31
3-5 Thermal Images of Various Low-Resolutions and Their Downscaling Factors 32
3-6 Precision-Recall with Low-Resolution Images of Varying Downscaling Factors(Indoors) 32
3-7 Thermal Images of Various Low-Resolutions and Their Downscaling Factors 33

3-8 Precision-Recall with Low-Resolution Images of Various Downscaling Factors (Outdoors) 33
3-9 Precision-Recall Curve of Combined Detector (Outdoors) 34
3-10 Precision-Recall Curve of Combined Detector (Indoors) 35
3-11 Comparison of Detection Results of Various Detectors 36

A-1 Examples of Positive Intensity Labels 41
A-2 Examples of Negative Intensity Labels 42
A-3 Examples of Positive Infrared Labels 43
A-4 Examples of Negative Infrared Labels 44

B-1 Intensity-Only Detector Results 45
B-2 Thermal-Only Detector Results 46
B-3 720x480 Thermal-Only Detector Results 47
B-4 72x48 Thermal-Only Detector Results 48
B-5 36x24 Thermal-Only Detector Results 49
B-6 24x18 Thermal-Only Detector Results 50
B-7 18x12 Thermal-Only Detector Results 51
B-8 12x9 Thermal-Only Detector Results 52
B-9 10x6 Thermal-Only Detector Results 53
B-10 720x480 Thermal-Only Detector Results 54
B-11 72x48 Thermal-Only Detector Results 55
B-12 36x24 Thermal-Only Detector Results 56
B-13 24x18 Thermal-Only Detector Results 57
B-14 18x12 Thermal-Only Detector Results 58
B-15 12x9 Thermal-Only Detector Results 59
B-16 10x6 Thermal-Only Detector Results 60
B-17 Intensity-Only Detector Results 61
B-18 Combined Intensity and High-Resolution Thermal Detector Results . 62
B-19 Combined Intensity and Low-Resolution Thermal Detector Results . 63
B-20 Intensity-Only Detector Results 64

B-21 Thermal-Only Detector Results 65
B-22 Low-Resolution Thermal-Only Detector Results 66
B-23 Combined Intensity and High-Resolution Thermal Detector Results . 67
B-24 Combined Intensity and Low-Resolution Thermal Detector Results . 68

Chapter 1

Introduction

1.1 Motivation

Due to a recent public outcry for pedestrian safety, automakers are trying to design more pedestrian-friendly vehicles in order to reduce the frequency and severity of pedestrian injuries. One proposed plan involves having a vehicle automatically detect pedestrians that are in the projected path of the vehicle in real time. Then, the vehicle can reduce the severity of the possible impact by applying automatic brake, or launching a structure such as an external airbag towards the pedestrian.

To implement such a system, we need a pedestrian detection algorithm that is able to accurately and speedily detect pedestrians from a moving platform with a very low false positive rate. The majority of past pedestrian detection systems use only one mode of detection, such as detectors using only intensity images, detectors using only thermal images, or detectors using only stereo images. Furthermore, most of past detectors rely on images captured from a still camera. We propose a method of pedestrian detection that combines information from both intensity images and thermal IR images captured from two cameras on a moving vehicle to ensure fast and accurate detection of pedestrians.

1.2 Goals

In order for a pedestrian detection system to be useful to the automobile industry, we need the detection system to be fast, accurate with a very low false positive rate, and cheap enough to be a viable addition to the vehicle of an average consumer. Below, we discuss in detail the following three goals of the detection system.

1.2.1 Detection Speed

We assume the vehicle to be on a non-highway city street, traveling between 15 - 40 mph. Then, to accurately detect pedestrians that are 30 to 80 yards away and to allow the vehicle enough time after detection to respond accordingly, we need a image capture rate of at least two frames per second, and a corresponding detection rate of at most half a second per image. This detection rate gives the braking system of the vehicle around 1.5 seconds to come to a complete stop before reaching the pedestrian.

1.2.2 False Positive Rate

One main concern for the automobile industry regarding the pedestrian detection project is the possibility of having a detector with a relatively high false positive rate. There are many objects on the streets that a detector may falsely recognize as a pedestrian: a garbage can or a parked car on the side of the street, a small tree with dangling branches around the corner, etc. If a vehicle equipped with the pedestrian detection system jerked to a stop or launched an external airbag every time it passes by one of these falsely categorized items, then not only could the detector be a huge hassle for the owner of the vehicle, but would also cause new undeserved accidents to the passengers of the car.

Thus, the pedestrian detection system must ensure a very low false positive rate, while still ensuring a reasonable recall level.

1.2.3 Hardware Price

A good, thermal IR camera that is capable of capturing high-resolution images currently sells for a few thousand dollars a piece, which is too expensive of an addition to an average priced vehicle. However, with the current advance in technology, a low-resolution thermal camera can potentially be purchased for as low as a hundred dollars in the near future. If low-resolution images do indeed improve the performance of detection over intensity images alone, then the low cost of the low-resolution camera makes it an ideal choice as a practical addition to an automatic detection-braking system in any vehicle.

1.3 Past Related Work

Pedestrian detection has been a topic of interest for researchers for at least the past ten years. Pedestrian detection problems with fixed cameras, such as human detection on surveillance cameras, have generally been solved using background subtraction. However, this is not in general applicable to pedestrian detection from automobiles, since the background of a moving vehicle is constantly changing.

As a result of the complications that arise from a moving camera, motion-based pedestrian detection algorithms generally do not work well as a primary detection method when the camera is mounted on a moving automobile. [11] and [27] implemented motion-based approaches on moving vehicles by interpolating motion data from successive frames to predict the background in subsequent frames based on the motion patterns. This causes a delay of several frames before pedestrians can actually be detected.

Most recent approaches have used single-frame approaches to pedestrian detection. This has the advantage of allowing much quicker detection of pedestrians, at the expense of losing the motion data that makes pedestrian detection so much easier in surveillance environments. Many single-frame detectors are fast enough to work in real time.

Different approaches have used different types of sensors. Most work has been

performed on detecting pedestrians from visual intensity images. [3] used a shape-based approach that is very accurate except when the pedestrian is partially occluded. [19] used an approach based on wavelet features over the set of all shifted windows of various sizes over an image. This approach, while being very accurate, is not fast enough to run in real time.

Infrared (thermal) images have also been used in pedestrian detection. [2], [28], and [6] have all used various algorithms combined with infrared imagery to detect pedestrians from moving vehicles. Infrared imagery has the advantage that many objects which may resemble pedestrians visually can be easily distinguished from pedestrians because they give off no body heat. However, the levels in the thermal image may also be strongly affected by the weather and ambient conditions. Various machinery may also give off large amounts of heat, and be confused with pedestrians. [18] used a probabilistic template-based approach with infrared images, and achieved a reasonable detection rate, at 11 frames per second. However, that system averaged one false-detection per frame, which is unacceptable in a collision-avoidance system.

[29] and [15] used stereo images with visual and thermal images, respectively, and achieved a very high accuracy rate of detection. However, processing stereo images requires significant computation, and currently cannot be done in real-time on a video stream.

In 2001, Paul Viola and Michael Jones proposed a boosted cascade algorithm that is capable of rapidly processing images and achieving a high detection rate of 15 frames per second [26]. The training algorithm is based on AdaBoost, while the classifying algorithm uses a cascade structure to reduce computation time and increase performance. The cascade process can be thought of as a decision tree, where each image is tested against a series of classifiers. When an image tests true for the first classifier, it is then tested against the second classifier in the sequence, until either it has failed enough classifiers to be classified as "false", or it can be classified as "true". If for each stage of the cascade, the false negative rate is guaranteed to be below a certain small threshold, then the result of the cascade will correctly classify a very large percentage of the positive data, while rejecting the vast majority of the negative

data in an early stage of the cascade. Clearly, this process reduces the average number of classifying tests per data point. By choosing classifiers where the vast majority of the images fail as early stages of the cascade, we can significantly reduce the number of images that need to be processed by the second classifier, which in turn improves the overall performance.

Another main feature of the Viola and Jones paper is the introduction of a set of over-complete Haar-like features that can be computed in linear time using a method called integral image [26]. This allows for the very quick calculation of features over millions of different possible locations. In 2002, [14] extended this set of over-complete Haar-like features to also include features that are 45° rotated from the original Viola and Jones features. This new set of features, like the original basic features, can also be calculated very quickly, and may be useful in capturing certain domain knowledge that cannot be previously represented. For example, the arms of a walking pedestrian may be closer to 45° relative to the ground than completely vertical, and this information can be captured much better with rotated features than just the basic features. However, the additional features may also cause complexities in the training of the classifiers.

In 2003, Viola, Jones, and Snow implemented a new method of pedestrian detection that combines the intensity images with motion information for improved detection [27]. According to the authors, this was the first approach that combines multiple modes of information for pedestrian detection. This implementation used the same boosting cascade algorithm described above. In each round, Adaboost selects a set of filters including both the intensity filters and motion filters, and the output classifier is a linear combination of the selected features. The combined implementation of intensity with motion was able to detect pedestrians from low-resolution intensity images with a very low false positive rate.

Several researchers have tried to combine multiple types of sensors in order to achieve more robust systems to detect pedestrians. [12] combined infrared images with radar data, in order to more accurately detect pedestrians in all conditions. This method worked well, but radar equipment is still too bulky and expensive for

the automobile market. [8] used radar and infrared images also, and added a laser scanner, which allowed even more accuracy.

One notably interesting project [25] has used a combination of thermal and visual images for pedestrian detection, and achieves a very high accuracy. However, this implementation was for surveillance use, and used a background-subtraction motion-based algorithm.

1.4 Approach

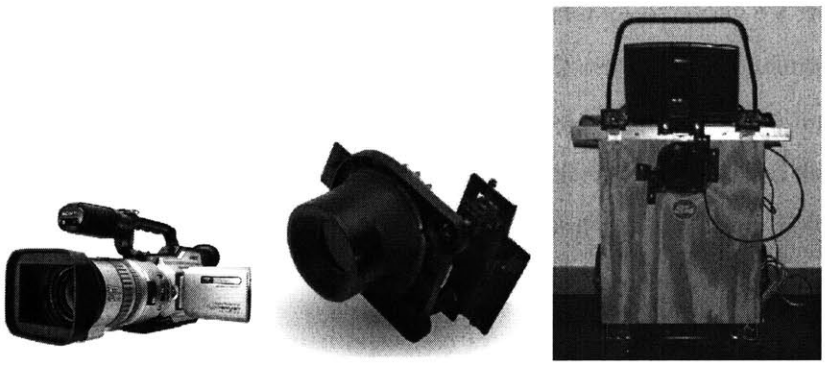
We will use the cascade classifier introduced by Viola and Jones to train two separate detectors, one over a set of intensity images, and another over a set of corresponding thermal images. Then, given inputs of corresponding intensity and thermal images, the detector can output the presence and the location of pedestrians in the image pair with a reasonable recall rate and a very low false-positive rate. Also, each thermal test image can be downscaled to various low-resolution thermal images, and the detection rate of the intensity and low-resolution thermal image pairs can also be tested.

Chapter 2

Implementation

2.1 Data Collection

The pedestrian detector is trained and tested using a combination of intensity images and their corresponding thermal IR images. The intensity images are gathered from frames of a video clip, while the thermal IR images are collected from a high-resolution Raytheon NIGHTDRIVER thermal IR camera. Both cameras are mounted to the front of a pushable cart in vertical alignment, so that we can simultaneously capture the thermal and intensity images of the same moving objects.



(a) Sony VX-2000 camera (intensity)

(b) Raytheon NIGHTDRIVER IR camera (thermal)

(c) Push-cart

Figure 2-1: Data Collection Equipment

The NIGHTDRIVER thermal camera does not have any zoom abilities. Thus, we adjusted the zoom level in the intensity camera so that both cameras captured approximately the same range of scenery. For a person standing upright to appear completely within the range of both cameras, the person must be at least 20 yards away from the front of the push-cart.

We took six indoor sequences and seven outdoor sequences with the push-cart. [1] Each sequence is from two to six minutes in length, and contains natural scenes of pedestrians walking about streets or corridors. Since we chose to leave the push-cart in place during the capture of each individual sequence, the background information remains approximately the same for each individual sequence. However, we made sure to vary the location of the push-cart between sequences, so that the background from one sequence onto the next is significantly different.

It is interesting to note that while the background of the thermal images remain largely an uniform level of gray in the indoor sequences, the background of the thermal images from the outdoor sequences contain much more variation. In the indoor scenes, the only significant sources of heat other than pedestrians are certain lamp posts and some surfaces underneath sunlight. In the outdoor scenes, however, locations of high heat range from underneath cars and trucks to factory chimneys, to heat leaks from building frames and asphalt pavements that have been heated up by sunlight.

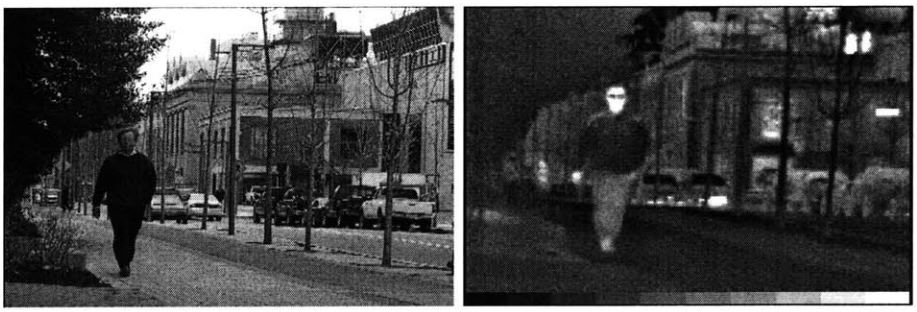
Eventually, we chose to focus mostly on the outdoor data sets. The indoor backgrounds contained so little noise that the detectors trained using the indoor data were able to pick out pedestrians just by picking out the lighter-shaded warm bodies from a uniformly dark background. With the outdoor data set, there are enough noise in the background that the detectors have to learn the actual shape and infrared properties of a human body.

In addition to the sequences that we collected using the push-cart, we were also able to use some of the intensity frames containing pedestrians from the online La-

[1] This thesis was part of a larger collaboration project. Jasper Vicenti trained the intensity-only and thermal-only detectors, C. Mario Christoudias wrote the matlab file to average the detected bounding boxes. I did most of the data collection and labeling, as well as combining the two detectors and testing the detectors on the low-resolution thermal images.

(a) An intensity indoor frame (b) A thermal indoor frame

Figure 2-2: Sample Indoor Frames



(a) An intensity outdoor frame (b) A thermal outdoor frame

Figure 2-3: Sample Outdoor Frames

belMe database [22]. The LabelMe database contains tens of thousands of annotated intensity images, and thus gives a much wider range of background variation. Each object in an image in the database contains an approximate bounding box and a short annotation describing the type of object. To pick out scenes containing a pedestrian, we first searched for terms such as *pedestrian* or *person walking*, and then hand selected a handful of images that satisfy our training needs. The only drawback is that the database contains only intensity images, and has no corresponding thermal images. Thus, we used the LabelMe database only to boost our background variation in the training stages of intensity classifier.

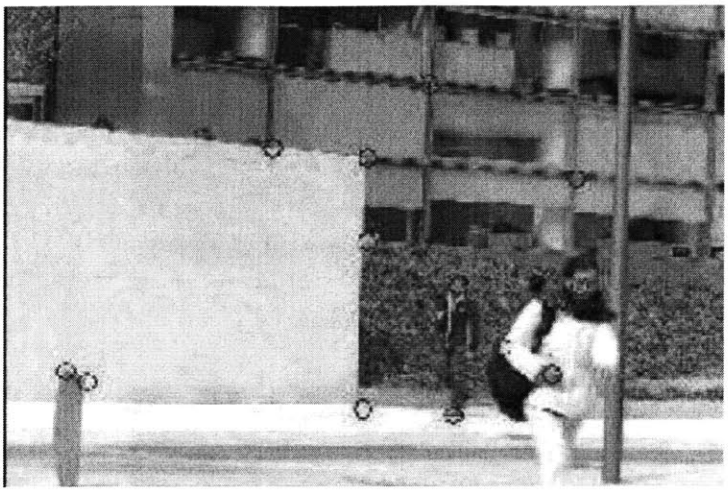
2.2 Thermal and Intensity Alignment

For the indoor clip sequences, we manually aligned the intensity video clip with the thermal video clip so that both clips begin and end with the same frame. Then, we exported the clips to frames at a rate of 30 frames/second. For the outdoor clip sequences, we followed the same procedure, but chose to export the clips to frames at a rate of 3 frames/second. This is mainly due to the fact that two frames that are separate by only one thirtieth of a second are very similar, and that the movement of the pedestrian during this amount of time is not enough for both frames to be useful.

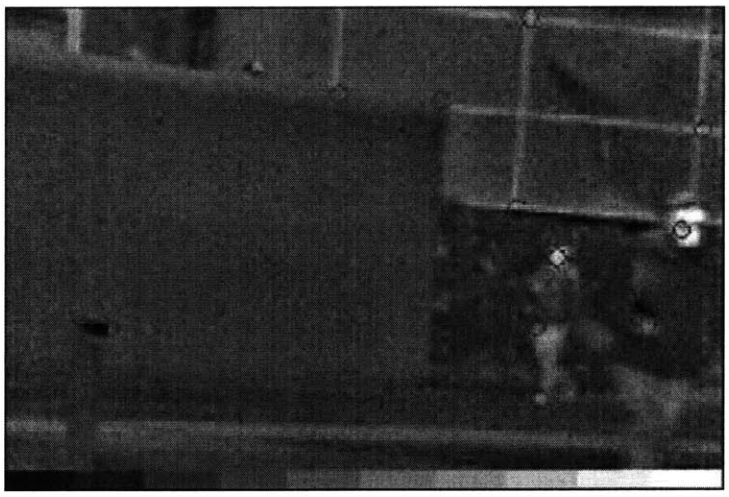
Though the intensity camera was zoomed to capture approximately the same field of view as the thermal IR camera in all the clip sequences, the frames exported from the two clips still needed to be aligned to achieve higher accuracy. For the alignment between intensity and thermal frames for each clip sequence, we manually chose 15 points from the intensity frame, and found the corresponding 15 points from the thermal frame. Then, we calculated the affine transformation that most closely match the transformation from the 15 intensity points to the 15 thermal points using the following equation:

```
[u v] = [x y 1] * Tinv
```

Where (u, v) are the coordinates of the intensity points, (x, y) are the coordinates of the thermal points, and $Tinv$ is the 3x2 matrix that is the inverse of the trans-

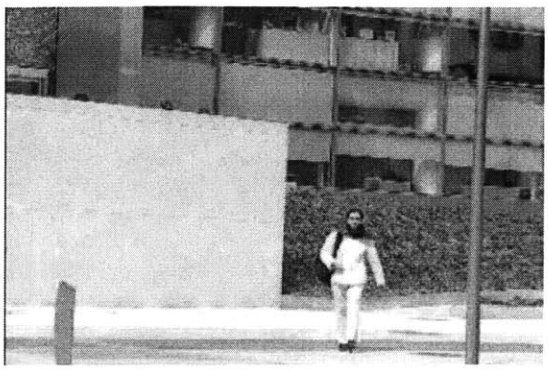


(a) Intensity points chosen for alignment



(b) Thermal points chosen for alignment

Figure 2-4: Corresponding Points Chosen for Alignment of Intensity and Thermal Frames



(a) Intensity frame

(b) Warped thermal frame

(c) Overlap of warped thermal with intensity

Figure 2-5: Intensity and the Corresponding Warped Thermal Frame

formation matrix that we want. Solving for the six elements of $Tinv$ gave us the corresponding affine transformation between the thermal and intensity images.

Since the push-cart was stationary within each individual clip sequence, we used the same affine transformation to warp all of the thermal frames into new thermal frames that align well with the corresponding intensity frames for each clip sequence. We also removed the stripe of gray scale that was present at all of the initial thermal frames. For any area in the intensity frame that was outside of the field of view of the thermal camera, we colored the corresponding area in the warped thermal frame completely black.

2.3 Labeling

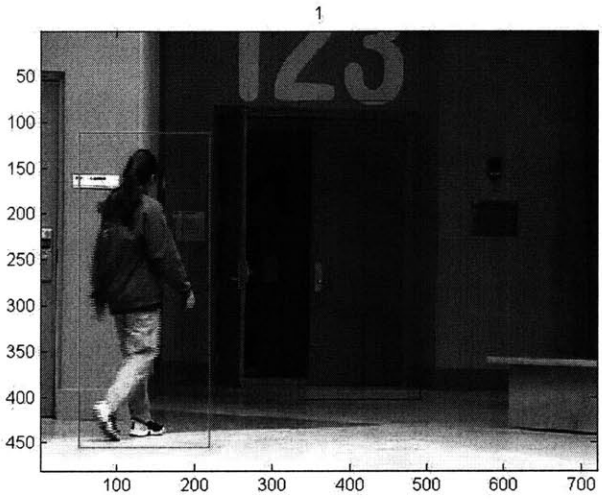


Figure 2-6: Examples of Positive and Negative Labels Within a Frame (green = positive, red = negative)

Once we obtained enough frames, we then labeled positive and negative pedestrian examples in both the intensity frames and the thermal frames. A positive label consists of the frame number and the coordinates of the tightest rectangular bounding box surrounding a walking or standing pedestrian in the frame. A negative label consists of the coordinates of a similar looking rectangular box, but the bounding

box contains no pedestrians. With the alignment of the intensity and the thermal frames, we can manually label the positive examples in the intensity frames, and automatically generate all of the positive labels for the warped thermal frames.

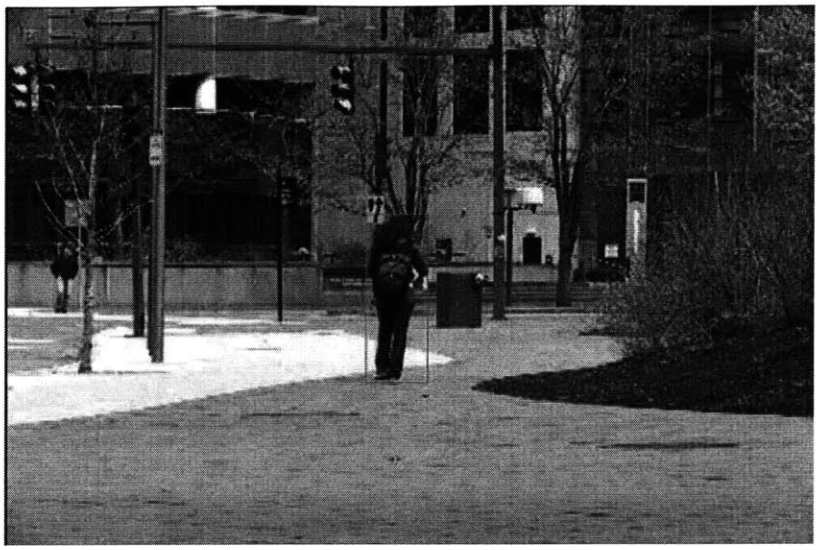


Figure 2-7: Examples of a Positive Label in labeler

Labeling all the pedestrians in the tens of thousands of frames by hand is a tedious and time-consuming task. To faciliate the labeling process, we use the fact that the location of a pedestrian in a given sequence does not change by very much from one frame onto the next, if the frames are taken within fractions of a second within each other.

Thus, we wrote a Matlab script, *labeler*, to help us with the labeling process. Each time a pedestrian first appears in a frame, the user of *labeler* can click on a few points on the outline of the pedestrian. Then, *labeler* will calculate the tightest rectangular bounding box of the pedestrian from those outline points inputted by the user.

For each subsequent frame, *labeler* will display the bounding boxes from the previous frame onto the current frame one by one, and ask the user for adjustments. Then, the user has the option of accepting the bounding box from the previous frame (if the pedestrian is standing still), stretch or squeeze the rectangular bounding box

(a) Previous frame

(b) Previous frame's bounding box on current frame

(c) Adjusted bounding box on current frame

Figure 2-8: Adjustment of a Positive Label Through labeler

vertically or horizontally, shift the bounding box up, down, left, or right, or discard the bounding box altogether, all through a sequence of keyboard commands. At all time during the labeling and adjustment period, *labeler* continues to display the most current bounding boxes of pedestrians to the user. The ability for the user to adjust current labels also results in labels of a much higher quality.

The number of pixels in the shifting and stretching adjustments can be decided by the user, depending on the resolution of the images and the number of pixels the average pedestrian takes up in each image. Then, an experienced user can label each pedestrian with just three keystrokes on average. For frames containing only one or two pedestrians, this correspond to a labeling rate of less than a second per frame.

The negative examples were randomly generated over the set of input images. For each image, a few rectangular boxes of 128x256 pixels were chosen randomly from the area outside of the bounding box of the positive label. Each of these rectangular boxes made up the coordinates of a negative label.

All in all, we collected 21782 outdoor intensity-thermal image pairs with around 7000 labeled pedestrians, and 48380 indoor intensity-thermal image pairs with around 20000 labeled pedestrians. To see some examples of positive and negative labels, for both the intensity and the thermal IR images, see Appendix A.

2.4 Training

We trained two separate pedestrian detectors: one for the intensity data, and another for the thermal IR data. Both detectors were implemented in C++, using the OpenCV libraries, following the boosting algorithm described by Viola and Jones and extended by Lienhart.

Each detector is composed of a cascade of weak classifiers, which in turn are each composed of a small number of features. For each stage of the cascade, we ran iterations of a varation of Adaboost until at least 50% of the negative data returns a negative result, while at most 0.5% of the positive data returns a negative result. Only data points that returns a positive result for the classifier at each stage can pass on to the next stage of the cascade. All other data points that return a negative result at a particular stage are immediately rejected by the cascade. Thus, the cascade is able to rapidly reject a very large percentage of negative data in the first two or three stages, which results in a very fast overall detector.

The features pool used for both classifiers are the over-complete set of basic Haar-like features. Integral image, as first mentioned in [26], was used to ensure fast calculation of all features. The set of rotated Haar-like features from [14] was not used for two reasons. First, the basic Haar-like features are already over-complete, and sufficient for our knowledge base of describing pedestrians of various upright poses. Also, adding the rotated features greatly increases the size of the feature pool, which results in a slower and over-fit training process.

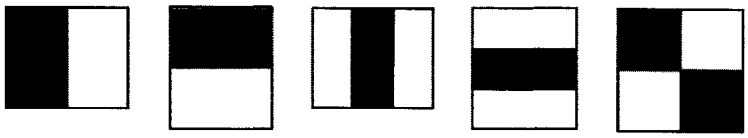


Figure 2-9: Set of Basic Haar-like Features Used

The training input to both classifiers are the scaled subimages within the bounding boxes of the labels. For each label, the rectangular region within the bounding box stated by the label is first cropped, and then resized to 32x64 pixels. Then, the resized

input image is large enough to still contain all the crucial features of pedestrians, while small enough to make the calculations of all possible features possible.

Finally, the intensity detector is trained using 1100 positive samples and 60000 negative samples, while the infrared detector is trained using 500 positive samples and 45000 negative samples.

2.5 Testing

To test the intensity-only and thermal-only detectors, we simply ran each detector over a set of 50 images. For each image, the detector loops through all the possible rectangular bounding boxes present in the image, and decides for each bounding box whether it indeed contains a pedestrian.

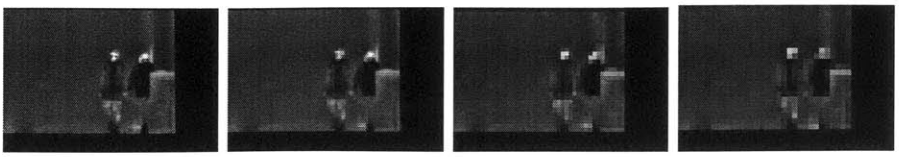
For each pedestrian present in a test image, the detector returns the number of bounding boxes found around the pedestrian. Bounding boxes below a certain size threshold are immediately discarded, since it is impossible to reliably detect the presense of pedestrians that are only a few pixels tall. The rest of the bounding boxes are averaged together to create one final bounding box around the outline of the detected pedestrian.

To create a detector that combines intensity and thermal information, we ran the intensity-only detector over the intensity images, and the thermal-only detector over the thermal images. Then, we combined the resulting bounding boxes of the intensity and the thermal only detectors, and found the final bounding box of the pedestrian.

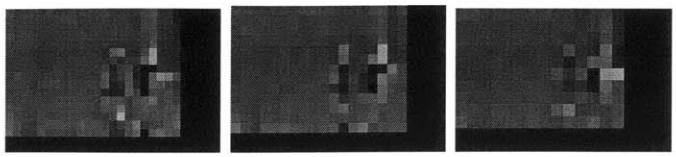
2.6 Low-Resolution Thermal Images

We also tested the performance of the thermal-only and the combined intensity and thermal detectors under inputs of low-resolution thermal images. The same testing procedure outlined in the previous section was used, replacing high-resolution thermal images with low-resolution thermal images.

The low-resolution thermal images used in the testing phase of the pedestrian



(a) 720x480 (1) (b) 72x48 (10) (c) 36x24 (20) (d) 24x16 (30)



(e) 18x12 (40) (f) 15x10 (50) (g) 12x8 (60)

Figure 2-10: Low-Resolution Thermal Images of Various Down-Scaling Factors

detector were obtained by downscaling the original high-resolution thermal images. From the original 720x480 pixel high-resolution test images, each image was downscaled in Matlab using bicubic interpolation to create low-resolution test sets of images of 72x48, 36x24, 24x16, 18x12, 15x10, and 12x9 pixels. A sample of the original high-resolution thermal image and its six downscaled low-resolution images is shown above.

Chapter 3

Results

3.1 Intensity Detector

The intensity detector contains a total of 30 stages of weak classifiers, using a total of 2868 features.

The first three stages of the intensity detector contains 19, 33, and 37 features, respectively. The number of features increases to 81 for stage 10, and 131 for stage 20. The small number of features in the first three stages allows the majority of the negative samples to be rejected quickly, while the larger numbers in the later stages allow for a more detailed description of pedestrians through a large number of features.

The first six features of the first stage of the intensity detector are shown in the figure below. Notice that out of the millions of possible features, the first two features

Figure 3-1: Intensity Detector Results

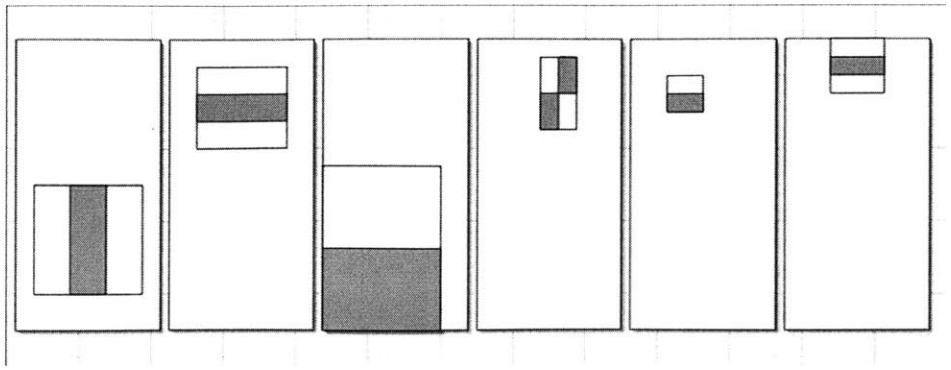


Figure 3-2: First Six Features of the Intensity Detector

chosen represent the legs and the head of an upright pedestrian, respectively.

3.2 Thermal Detector

The thermal IR detector contains a total of 18 stages of weak classifiers, using a total of 416 features.

Like the intensity detector, the thermal detector also had classifiers with relatively small numbers of features for the first few stages, resulting in very fast rejections of the majority of the negative subframes. The first two chosen features of the thermal detector highlight the fact that an upright pedestrian generates a lot of heat in his entire body compared to the surroundings, and that the face of the pedestrian usually appears as the most warm area in a given image.

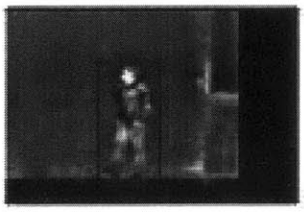


Figure 3-3: Thermal Detector Results

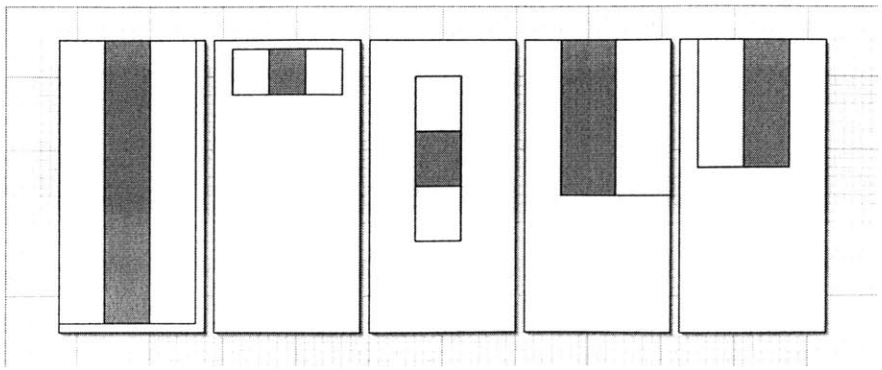
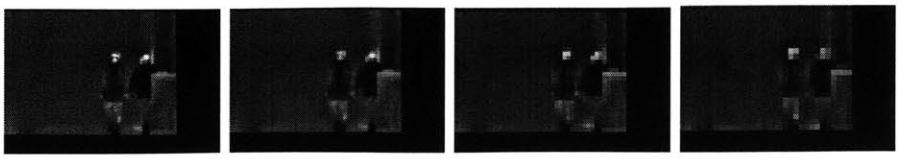


Figure 3-4: First Six Features of the Thermal Detector

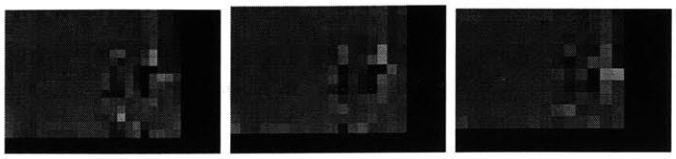
3.3 Low-Resolution Thermal Results

We took 50 indoors thermal test images and their downscaled corresponding test images, and tested each set of 50 thermal images of various resolutions with the thermal detector. Then, we plotted the precision-recall curve for each of the detectors. Due to the few number of samples, only local maxima points should be considered when reading the precision-recall curves. The dotted line in the precision-recall graph on the following page represents the detection results of the high-resolution thermal images, while each of the solid lines represents the detection results of using low-resolution thermal images of a specific downscaling factor. Note that on the precision-recall curve, there is no detection performance drop with low-resolution images with downscaling factors of 30 or less. Thus, a low-resolution image of 24x16 performs just as well as the high-resolution image of 740x480. With a downscaling factor of 40 or 50, there is a noticeable drop in the precision-recall curve. The detection performs drops even more with a downscaling factor of 60, which corresponds to thermal images of only 12x8 pixels.

We repeated the same experiment with 50 outdoors test images. This time, the detection did not drop significantly until a downscaling factor of 50, showing that low-resolution images of 18x12 still resulted in very good detection rates.



(a) 720x480 (1) (b) 72x48 (10) (c) 36x24 (20) (d) 24x16 (30)



(e) 18x12 (40) (f) 15x10 (50) (g) 12x8 (60)

Figure 3-5: Thermal Images of Various Low-Resolutions and Their Downscaling Factors

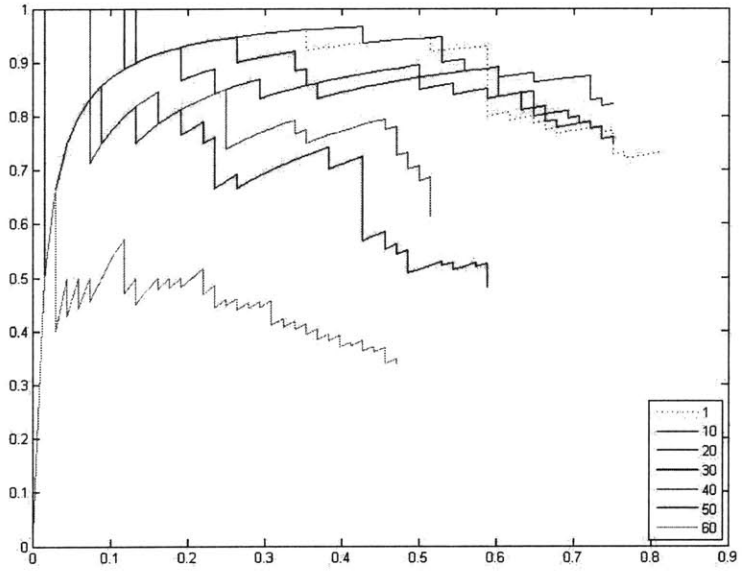
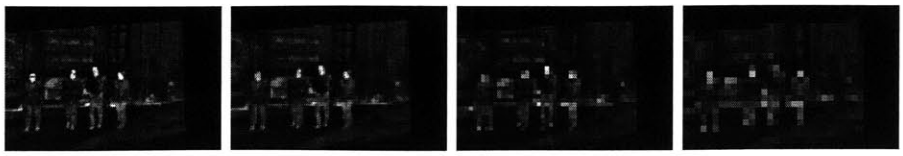
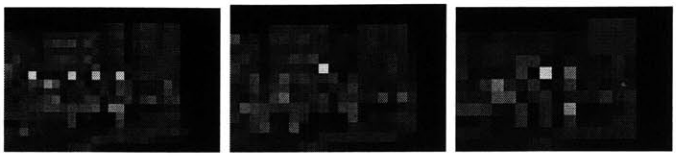


Figure 3-6: Precision-Recall with Low-Resolution Thermal Images of Varying Downscaling Factors(Indoors)



(a) 720x480 (1) (b) 72x48 (10) (c) 36x24 (20) (d) 24x16 (30)



(e) 18x12 (40) (f) 15x10 (50) (g) 12x8 (60)

Figure 3-7: Thermal Images of Various Low-Resolutions and Their Downscaling Factors

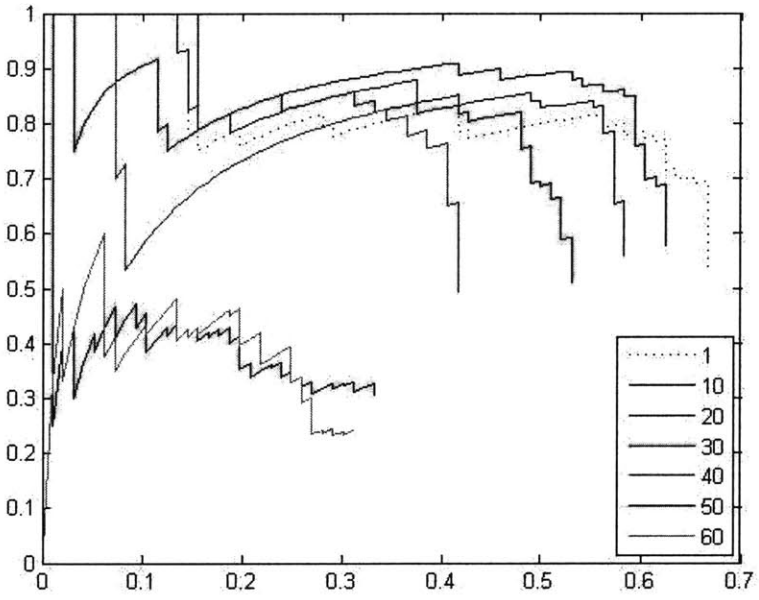


Figure 3-8: Precision-Recall with Low-Resolution Thermal Images of Various Downscaling Factors (Outdoors)

3.4 Comparison of Various Detectors

The graph below shows the precision-recall curves of the intensity-only detector, the thermal-only detector, and the combined intensity-thermal detector for the outdoors data. The x-axis shows the recall rate, while the y-axis shows the precision rate. Only subimages that survived past a certain number of stages were used in the calculation.

For outdoors data, the intensity-only detector outperformed the thermal detector, and thus, there is not a significant improvement to the intensity-only results when the thermal information is added to the detector. With a 0.88 precision rate, the recall rate for both the intensity-only detector and the combined detector is slightly more than 0.5, while the recall rate for the thermal-only detector is only at around 0.13.

Due to the high level of IR noise in outdoor images, intensity information is much more valuable than thermal information in the correct detection of pedestrians. Thus, any combined intensity and thermal detectors performed only about as well as the intensity-only detector, regardless of the resolution of the thermal images.

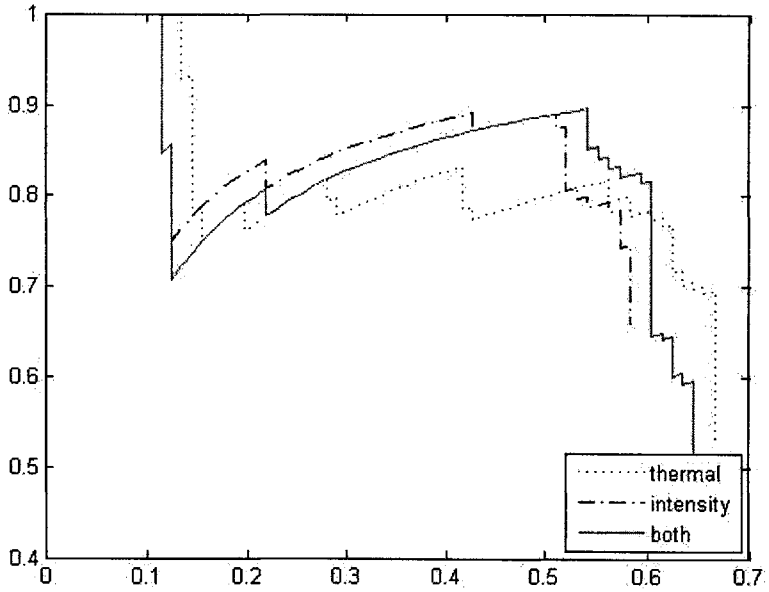


Figure 3-9: Precision-Recall Curve of Combined Detector (Outdoors)

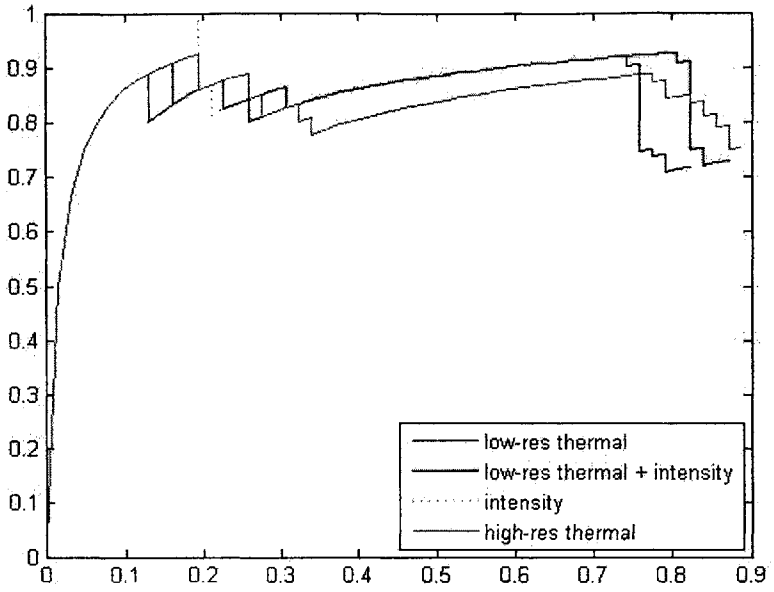


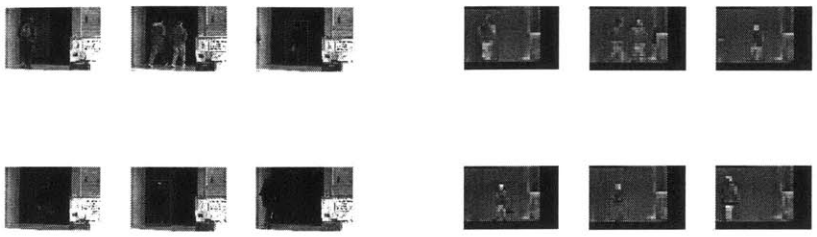
Figure 3-10: Precision-Recall Curve of Combined Detector (Indoors)

For detection indoors, however, the intensity-only detector did not perform very well on the test data (since the detector was trained using mostly outdoors data, and the outdoors and indoors features vary significantly for an intensity detector), and was not able to detect with a recall rate of higher than 0.2. The thermal detector, however, was able to detect pedestrians in high-resolution thermal images with a recall rate of 0.78 at a precision rate of 0.88. Low-resolution thermal images, in this case, outperformed even the high-resolution thermal images by a bit, since the down-scaling of the thermal image likely averaged out the slight noise around the thermal image of each pedestrian. With the low-resolution thermal-only detector, we were able to achieve a recall rate of 0.75 at a precision rate of 0.90.

The best detector, however, is the combined low-resolution (24x16 pixels) thermal and high-resolution intensity detector. By adding the additional high-resolution intensity information, the combined detector was able to achieve a recall rate of 0.81 at a precision rate of 0.90.

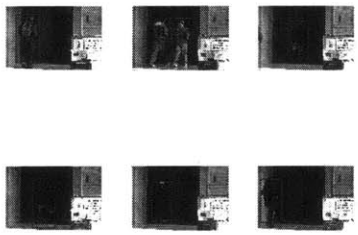
In the detection results below, note that the intensity-only detector missed quite a

few pedestrians, while both the thermal-only detector and the combined intensity and low-resolution thermal detector were able to detect many more pedestrians. Thus, the thermal information, even at a very low resolution, was able to help with the detection performance compared to an intensity-only detector.



(a) intensity only

(b) low-resolution thermal only



(c) both intensity and low-resolution thermal

Figure 3-11: Comparison of Detection Results of Various Detectors

More detection results can be found in Appendix B.

Chapter 4

Conclusion

4.1 Summary

This paper presents a pedestrian detector from a moving platform. The detector combines intensity information with thermal IR information, both captured from cameras mounted to the front of the moving platform. The information from the thermal images can easily narrow down the possible places where a pedestrian can be, while a high-resolution intensity image can be used to check for the exact outlines of the pedestrian.

We trained one detector using just intensity images of pedestrians and non-pedestrians, and a separate detector using just the thermal images. Both detectors were implemented using the rapid object-detection algorithm proposed by Viola and Jones, and each contains a cascade of weak classifiers. Each weak classifier is composed of a small number of features, where the feature pool consists of the over-complete set of basic Haar-like features. Then, each collection of weak classifiers can be calculated on all the possible bounding boxes of pedestrians of a test image to determine which bounding boxes are the ones that contain pedestrians. The combined intensity-thermal detector simply aggregates the collection of bounding boxes returned by both the intensity-only detector and the thermal-only detector at the post-processing stage.

The result of our detector shows that thermal information is very valuable in

pedestrian detection, especially in indoor scenes. A detector trained using thermal-only images significantly out-performs a similar detector trained using intensity-only images for indoor scenes, and is comparable in performance to that of an intensity-only detector in outdoor scenes. The combined intensity and thermal detector performs at least just as well as the best intensity-only detector, and often out-performs the intensity-only detector.

For the data sets that we tested, a 24x16 low-resolution thermal image is sufficient for containing all of the useful infrared information for pedestrian detection purposes. A thermal-only detector works just as well on low-resolutio 24x16 images as it does on high-resolution 720x480 images. Thus, for a combined detector using both thermal and intensity information, only a low-resolution thermal image of 24x16 is necessary.

With some more adjustments to the detection thresholds and post-detection processing, a detector utilizing a combination of high-resolution intensity and low-resolution thermal information can become a low-cost, reliable real-time pedestrian detector to be used by the automobile industry.

4.2 Future Work

The reliability of the detector cannot be verified until more training and testing inputs of intensity-thermal image pairs can be collected, preferably from an actual moving car rather than a push-cart. The motion-blur from a vehicle traveling at a high speed may present new challenges for the detector.

The detector's performance in a range of weather and temperature conditions will need to be examined. Most of the current thermal data was taken either indoors or outdoors in Massachusetts in the month of March. Since the ambient temperature varies significantly from January to July, and from a rainy day to a sunny day, more data from all seasons is needed to guarantee the performance of the detector. It could be the case that different parameters for the detector are needed in different seasons, in which case the detector may take in the outdoor temperature as an input value.

It would also be interesting to test is how well the detector performs at detecting

children. The current detector works very well for detecting upright adults in images. However, it is unclear how well the detector works at detecting child-sized pedestrians, or pedestrians that have fallen over on the ground or are bent over at the waist.

Currently, the thermal information and the intensity information are only combined after both detectors return their detection results. By combining the thermal and intensity features in the training stage, we could create just one cascade of classifiers, where each classifier can contain a number of both intensity features and thermal features. This would likely improve the performance of the detector, as only the best features out of the combined thermal and intensity feature pool would be picked.

Motion information could also be incorporated into the detector. Though motion-based detection methods are not very practical from a moving vehicle over a large time period, the background motion from one frame onto the next can be small enough for the foreground objects to still stand out. Like the addition of thermal information to an intensity-only detector, the addition of new motion information could also significantly improve the performance of the overall detector.

Finally, we need to build a cheap but reliable low-resolution thermal camera. Then we could test the benefit of the addition of low-resolution thermal information to high-resolution intensity images to detection performance.

Appendix A

Sample Labels

Here are some examples of both positive and negative intensity and infrared labels.

Figure A-1: Examples of Positive Intensity Labels

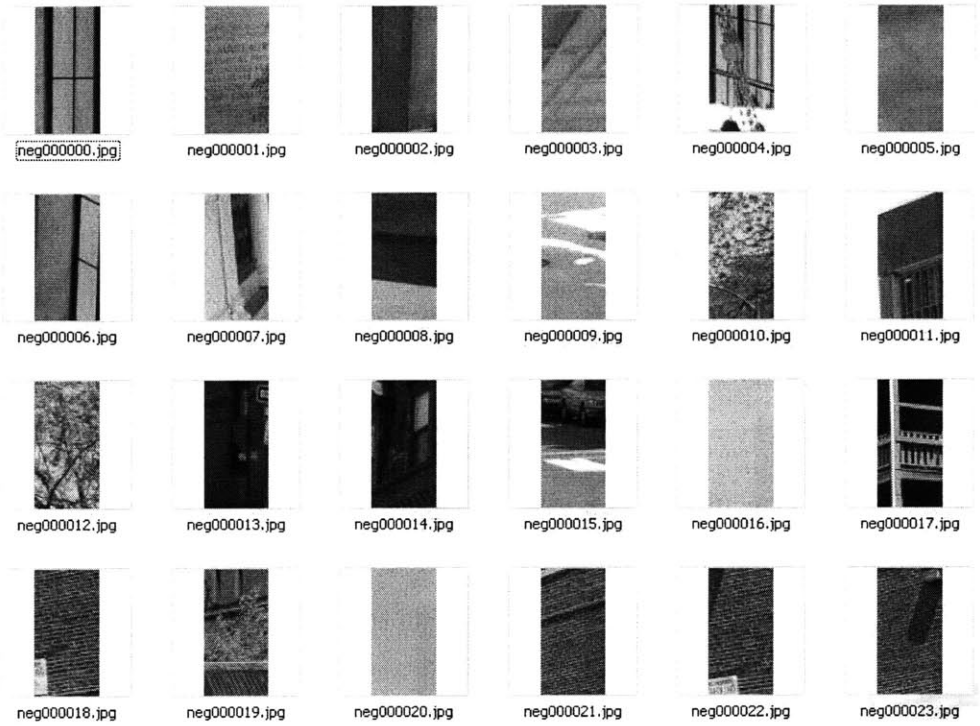


Figure A-2: Examples of Negative Intensity Labels

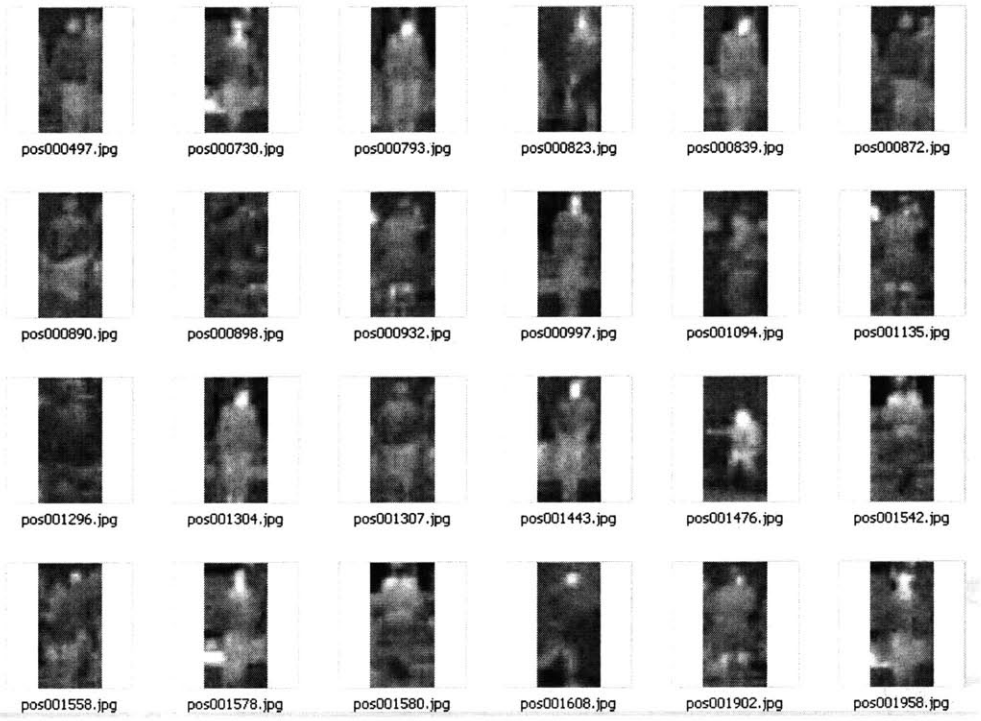


Figure A-3: Examples of Positive Infrared Labels

Figure A-4: Examples of Negative Infrared Labels

Appendix B

Sample Detection Results

Below are some more examples of detection results of various detectors

Figure B-1: Intensity-Only Detector Results

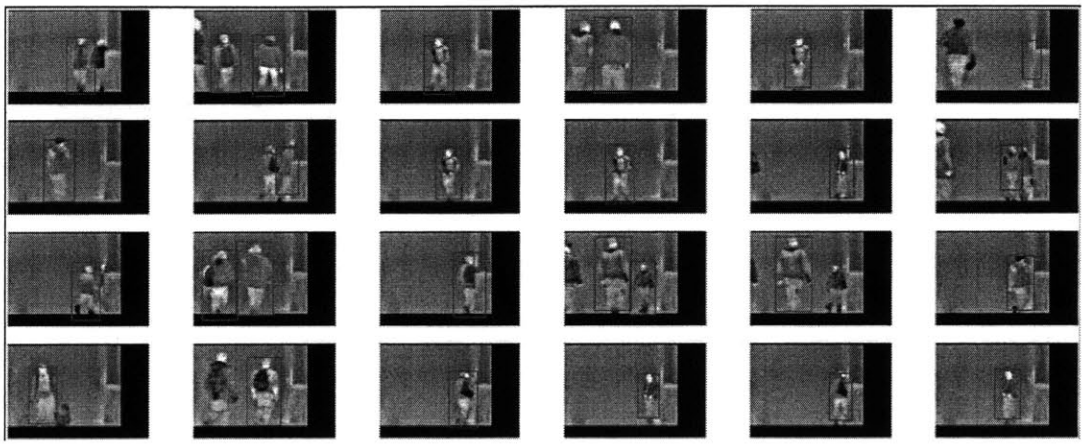


Figure B-2: Thermal-Only Detector Results

B.1 Indoors–Comparsion of various resolution thermal-only detectors

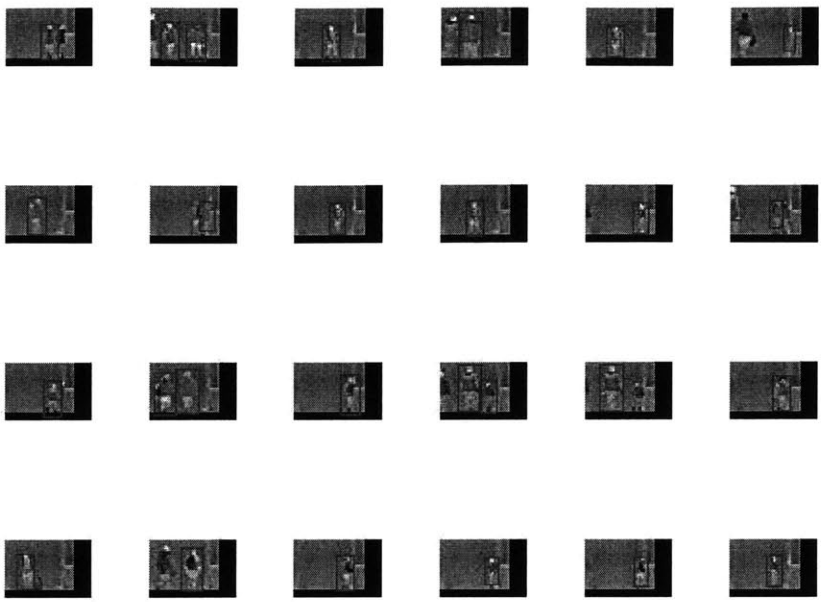


Figure B-3: 720x480 Thermal-Only Detector Results

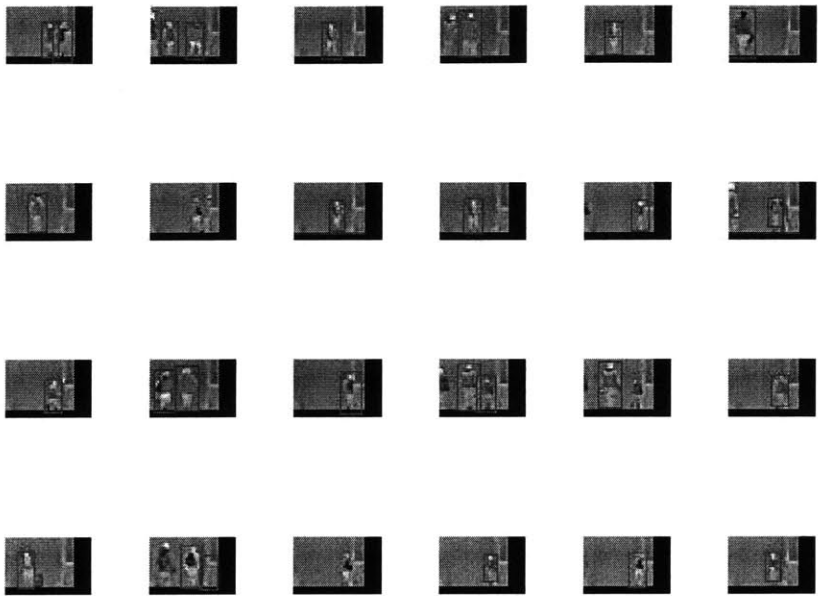


Figure B-4: 72x48 Thermal-Only Detector Results

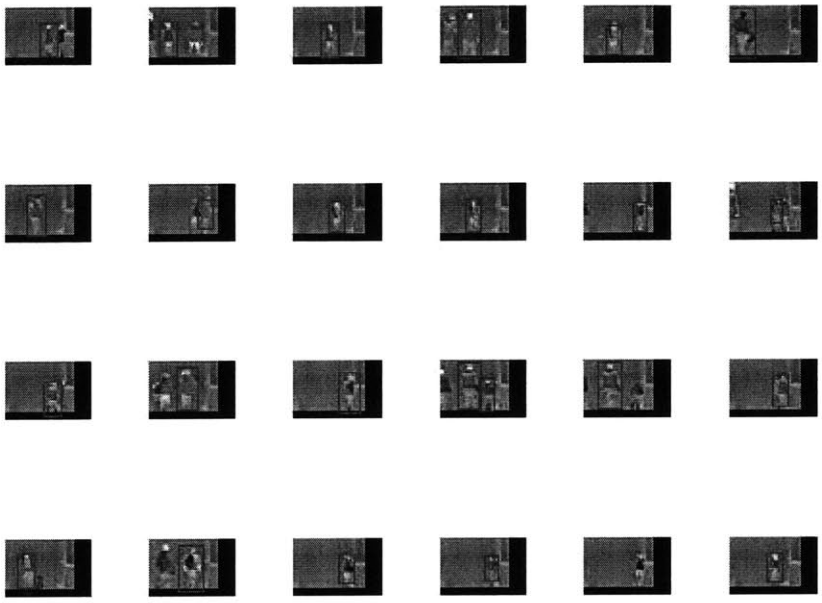


Figure B-5: 36x24 Thermal-Only Detector Results

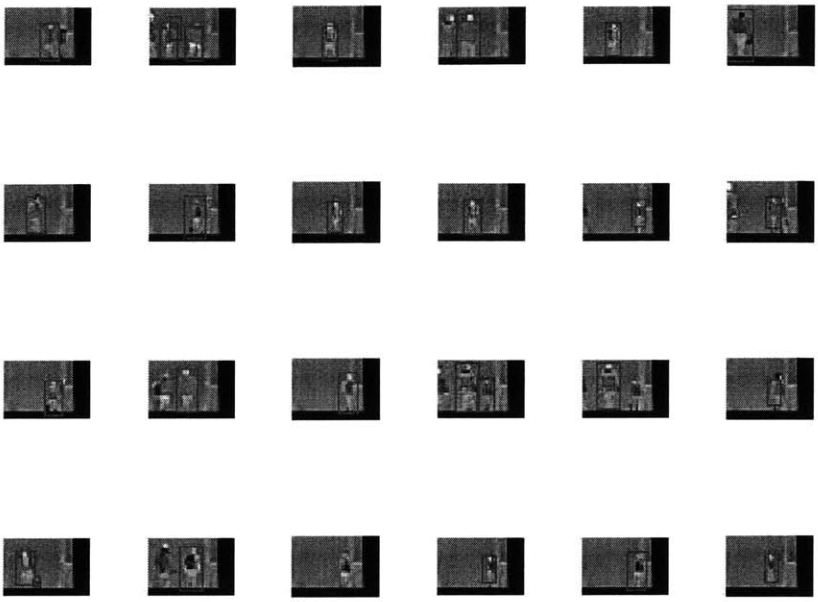


Figure B-6: 24x18 Thermal-Only Detector Results

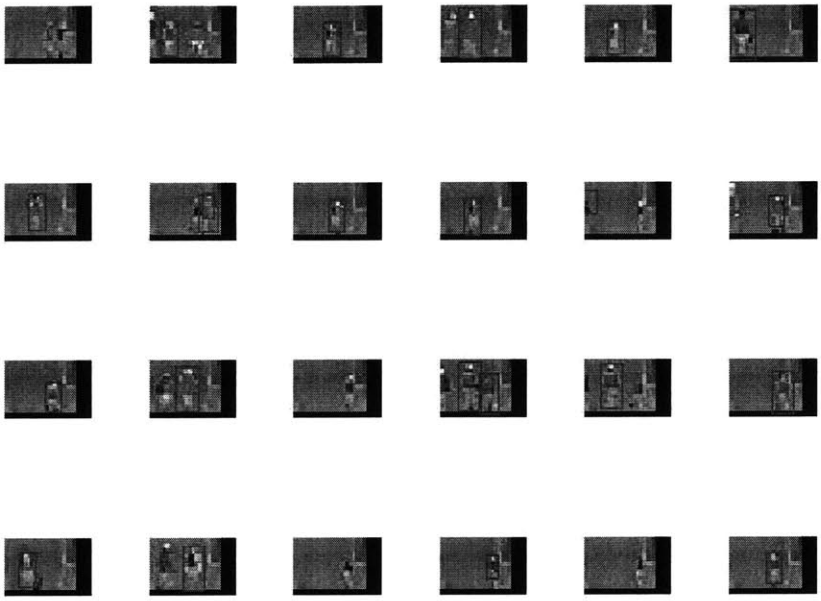


Figure B-7: 18x12 Thermal-Only Detector Results

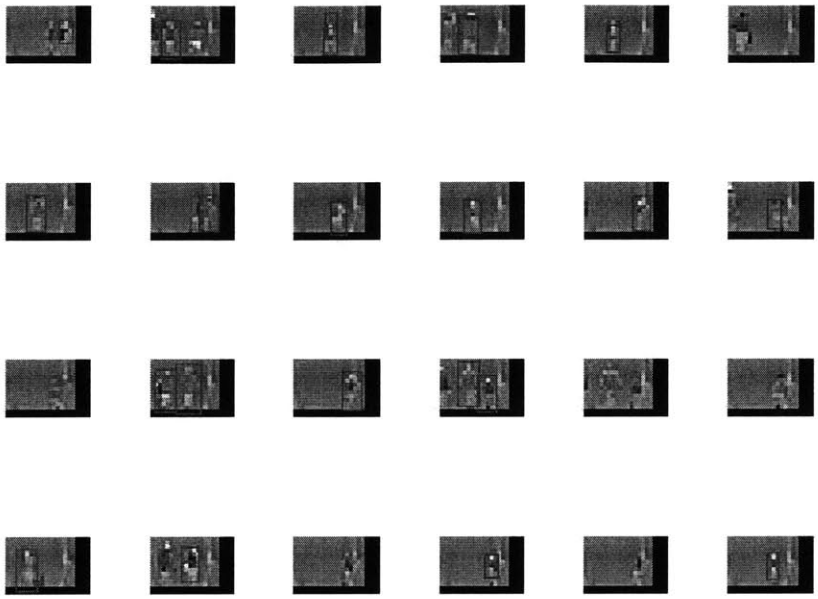


Figure B-8: 12x9 Thermal-Only Detector Results

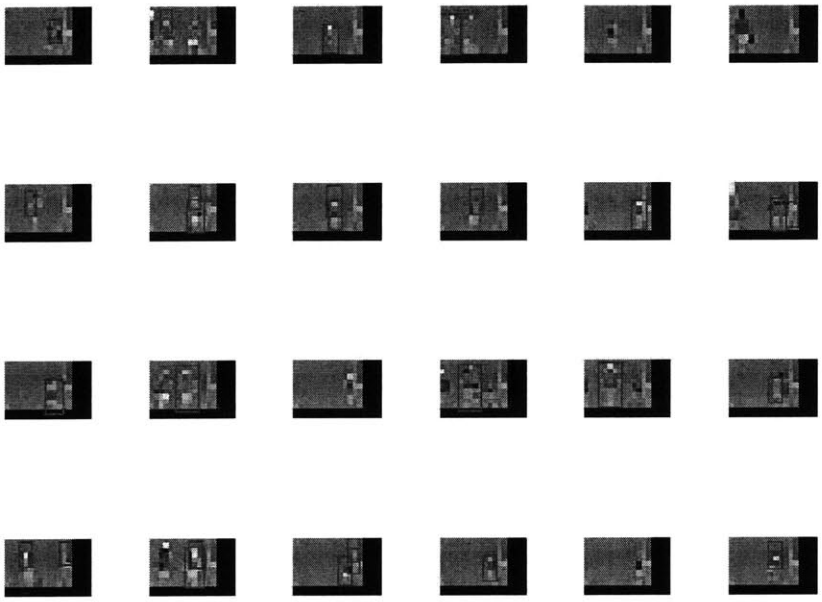


Figure B-9: 10x6 Thermal-Only Detector Results

B.2 Outdoors–Comparsion of various resolution thermal-only detectors

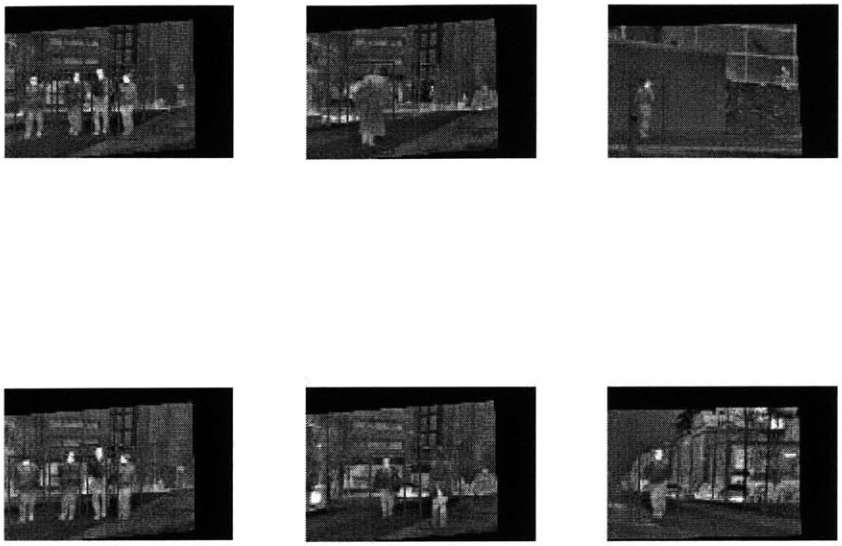


Figure B-10: 720x480 Thermal-Only Detector Results

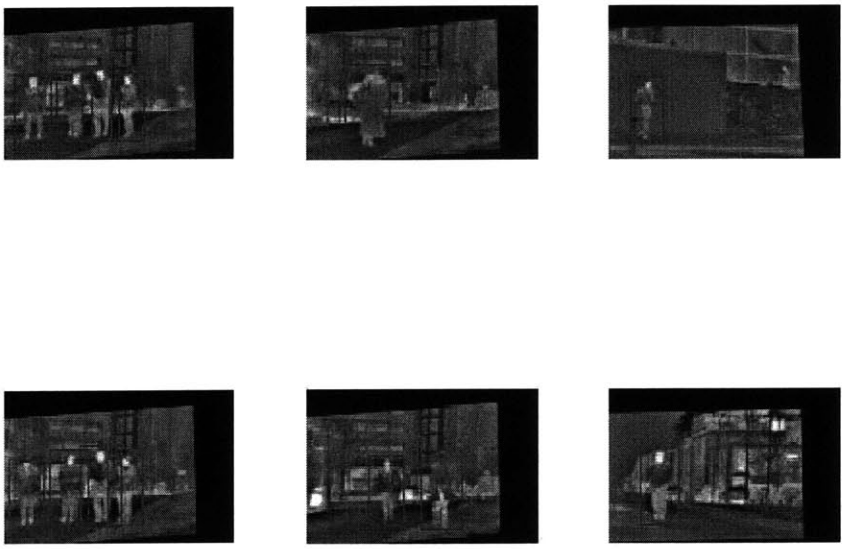


Figure B-11: 72x48 Thermal-Only Detector Results

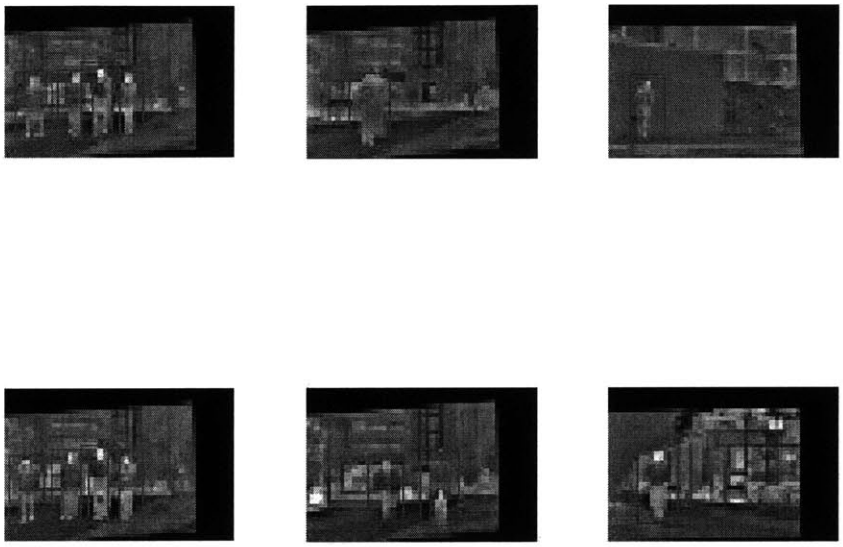


Figure B-12: 36x24 Thermal-Only Detector Results

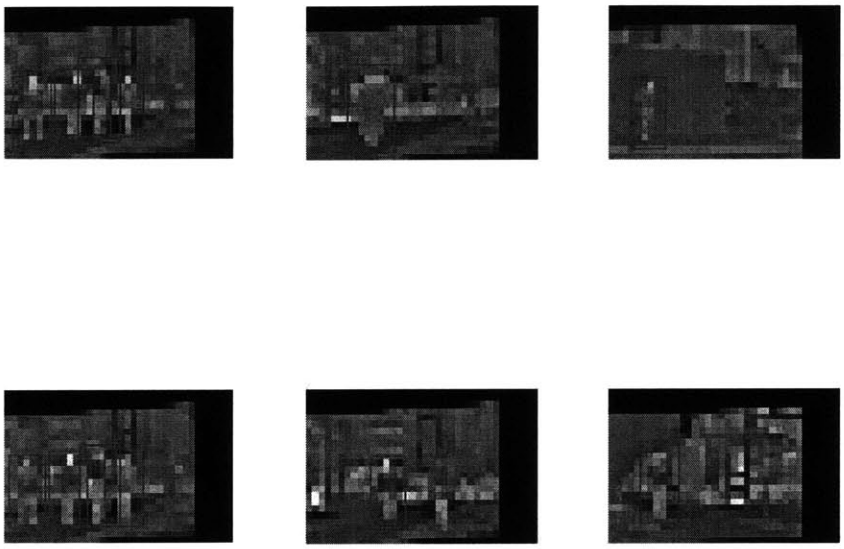


Figure B-13: 24x18 Thermal-Only Detector Results

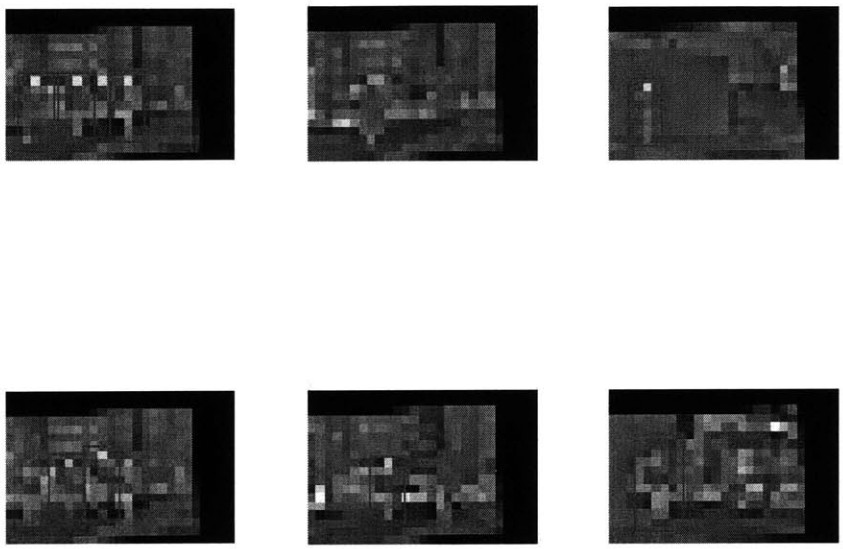


Figure B-14: 18x12 Thermal-Only Detector Results

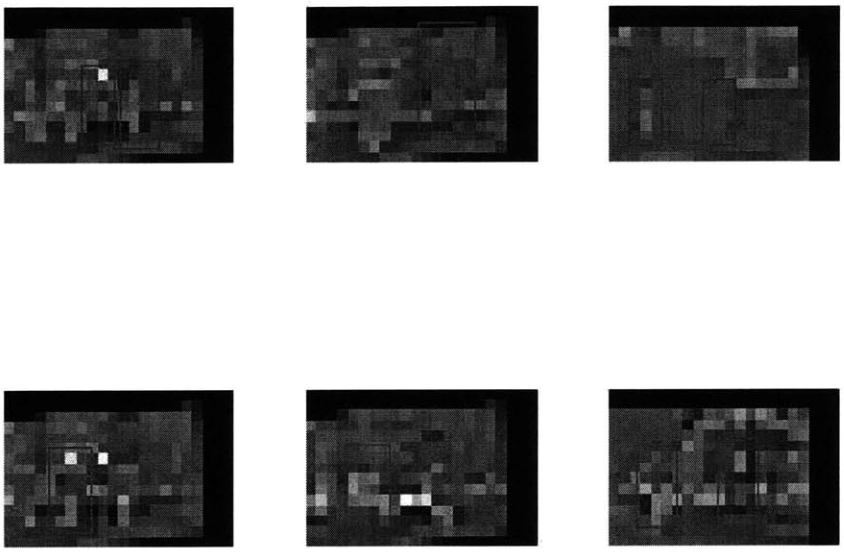


Figure B-15: 12x9 Thermal-Only Detector Results

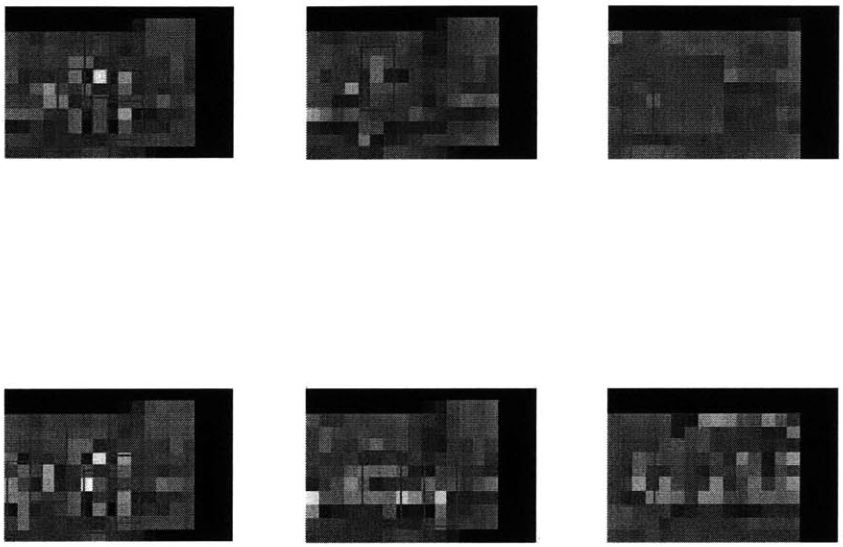


Figure B-16: 10x6 Thermal-Only Detector Results

B.3 Outdoors–Intensity-Only detector and Combined Intensity and Thermal detectors

Figure B-17: Intensity-Only Detector Results

Figure B-18: Combined Intensity and High-Resolution Thermal Detector Results

Figure B-19: Combined Intensity and Low-Resolution Thermal Detector Results

B.4 Indoors–Intensity-only, thermal-only, and combined intensity and thermal detectors

Figure B-20: Intensity-Only Detector Results

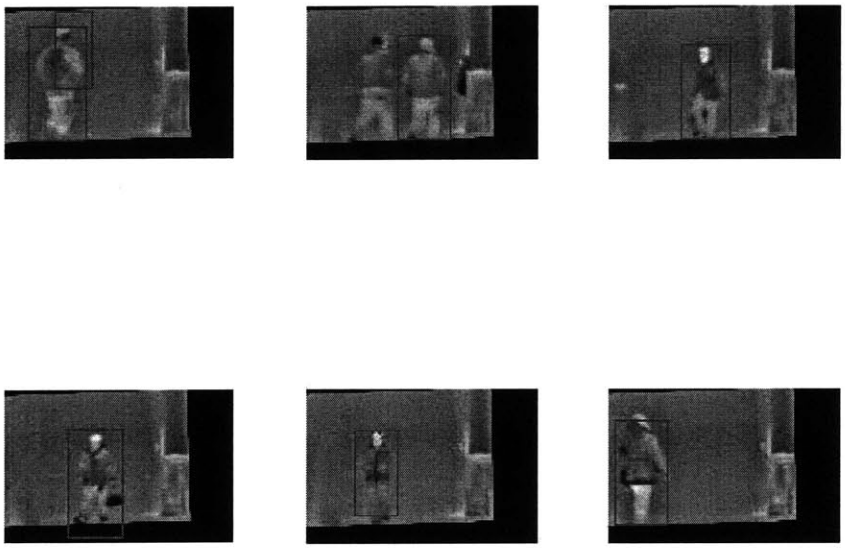


Figure B-21: Thermal-Only Detector Results

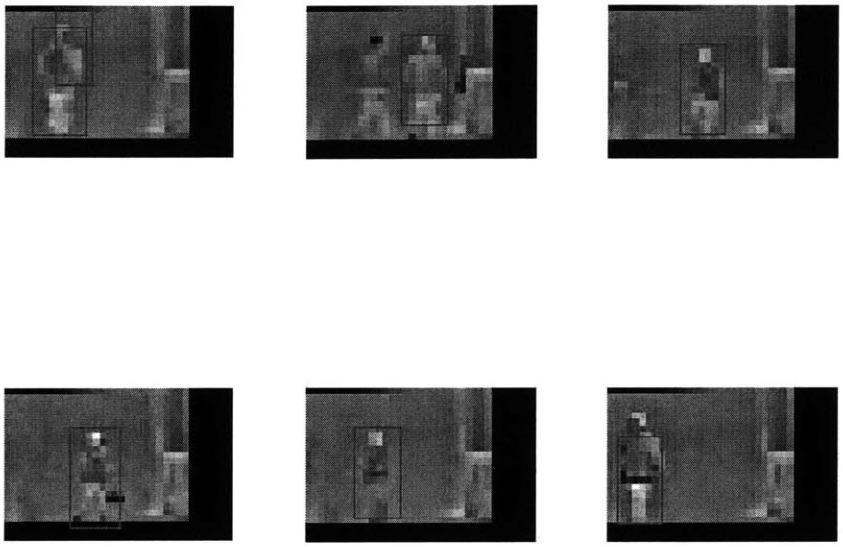


Figure B-22: Low-Resolution Thermal-Only Detector Results

Figure B-23: Combined Intensity and High-Resolution Thermal Detector Results

Figure B-24: Combined Intensity and Low-Resolution Thermal Detector Results

Bibliography

[1] M. Bertozzi, A. Broggi, A. Fascioli, P. Lombardi. "Vision-based pedestrian detection: will ants help?" *IEEE Intelligent Vehicles Symposium*, 2002.

[2] M. Bertozzi, A. Broggi, T. Graf, P. Grisleri, M. Meinecke. "Pedestrian detection in infrared images," *IEEE Intelligent Vehicles Symposium*, 2003.

[3] A. Broggi, M. Bertozzi, A. Fascioli, M. Sechi. "Shape-based pedestrian detection," *Proceedings of the IEEE Intelligent Vehicles Symposium 2000*, October 2000.

[4] N. Checka. "Fast Pedestrian Detection from a Moving Vehicle," *Proceedings of the 2004 Student Oxygen Workshop*, September 2004.

[5] Y. Fang, K. Yamada, Y. Ninomiya, B. Horn, I. Masaki. "Comparison between infrared-image-based and visible-image-based approaches for pedestrian detection," *IEEE Intelligent Vehicles Symposium*, 2003.

[6] Y. Fang, K. Yamada, Y. Ninomiya, B. Horn, I. Masaki. "A Shape-Independent Method for Pedestrian Detection with Far-Infrared Images," *IEEE Transactions on Vehicular Technology Vol. 53, No. 5*, September 2004.

[7] D. Gavrila. "Sensor-based pedestrian protection," *Intelligent Systems, IEEE Vol. 16, Iss. 6, pp. 77-81*, November-December 2001.

[8] D. Gavrila, M. Kunert, U. Lages. "A multi-sensor approach for the protection of vulnerable traffic participants - the protector project," *IEEE Instrumentation and Measurement Technology Conference*, 2001.

[9] D. Gavrila. "Pedestrian detection from a moving vehicle," *ECCV00*, 2000.

[10] D. Gavrila, V. Philomin. "Real-time object detection for 'smart' vehicles," *ICCV99*, 1999.

[11] B. Heisele, C. Wohler. "Motion-based recognition of pedestrians," *ICPR*, 1998.

[12] Y. Le Guilloux, J. Lonnoy, R. Moreira, C. Blanc, J. Gallice, L. Trassoudaine, H. Tattegrain, M. Bruyas, A. Chapon. "Paroto project : infrared and radar data fusion for obstacle avoidance," *AMAA 2004*, March 2004.

[13] R. Lienhart, A. Kuranov, V. Pisarevsky. "Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapid Object Detection," *DAGM03, pp. 297-304*, 2003.

[14] R. Lienhart, J. Maydt. "An exteded set of Haar-like Features for Rapid Object Detection," *ICIP02, pp. 155-162*, 2002.

[15] X. Liu, K. Fujimura. "Pedestrian detection using stereo night vision," *IEEE Transactions on Vehicular Technology Vol. 53, pp. 1657-1665*, November 2004.

[16] P. Lombardi. "A survey on pedestrian detection for autonomous driving systems," Technical Report, 2001.

[17] O. Masoud, N. Papanikolopoulos. "A novel method for tracking and counting pedestrians in real-time using a single camera," *IEEE Transactions on Intelligent Transportation Systems 50(5):1267-1278*, September 2001.

[18] H. Nanda, L. Davis. "Probabilistic template based pedestrian detection in infrared videos," *IEEE Intelligent Vehicles Symposium*, 2002.

[19] M. Oren, C. Papageorgiou, P. Sinha, E. Osuna, T. Poggio. "Pedestrian detection using wavelet templates," *CVPR97*, 1997.

[20] C. Papageorgiou, T. Evgeniou, T. Poggio. "A trainable pedestrian detection system," *IEEE Intelligent Vehicles Symposium*, 1998.

[21] V. Philomin, R. Duraiswami, L. Davis. "Pedestrian tracking from a moving vehicle," *IEEE Intelligent Vehicles Symposium*, 2000.

[22] B. Russell, A. Torralba, K. Murphy, W. Freeman. "LabelMe: a database and web-based tool for image annotation," *MIT AI Lab Memo AIM-2005-025*, September 2005.

[23] A. Shashua, Y. Gdalyahu, G. Hayun. "Pedestrian detection for driving assistance systems: single-frame classification and system level performance," *IEEE Intelligent Vehicles Symposium 2004 14-17 pp. 1-6*, June 2004

[24] C. Thorpe, J. Carlson, D. Duggins, J. Gowdy, R. MacLachlan, C. Mertz, A. Suppe, C. Wang. "Safe robot driving in cluttered environments," *Proceedings of the 11 International Symposium of Robotics Research*, October 2003.

[25] H. Torresan, B. Turgeon, C. Ibarra, P. Hébert, X. Maldague. "Advanced Surveillance System: Combining Video and Thermal Imagery for Pedestrian Detection," *Proceedings of SPIE Thermosense XXVI Vol. 5405, pp. 506-515*, April 2004.

[26] P. Viola, M. Jones. "Rapid object detection using a boosted cascade of simple features," *CVPR Vol.1, pp. 511-518*, December 2001.

[27] P. Viola, M. Jones, D. Snow. "Detecting pedestrians using patterns of motion and appearance," *ICCV03*, 2003.

[28] F. Xu, K. Fujimura. "Pedestrian detection and tracking with night vision," *IEEE Intelligent Vehicles Symposium*, 2002.

[29] L. Zhao, C. Thorpe. "Stereo and neural network-based pedestrian detection," *IEEE Transactions on Intelligent Transportation Systems, 1(1):148.154*, September 2000.