An Effective Approach to Pedestrian Detection in Thermal Imagery

Wei Li, Dequan Zheng, Tiejun Zhao
MOE-MS Key Laboratory of Natural Language Processing and Speech
Harbin Institute of Technology
Harbin, China

Mengda Yang
Institute for Interdisciplinary Information Science
Tsinghua University
Beijing, China

***Abstract*—In this paper, an integrated algorithm to detect humans in thermal imagery was introduced. In recent years, histogram of oriented gradient (HOG) is a quite popular algorithm for person detection in visible imagery. We implement the pedestrian detection in infrared imagery with this algorithm by adjusting the parameters. Simultaneously, we have increased some other geometric characteristics, such as mean contrast, which is used as features for the detection. After analyzing the property of the infrared imagery, which is designed to meet the shortfall of the HOG in infrared imagery, the combined vectors are fed to a linear SVM for object/non-object classification and we get the detector at the same time. After that, the detection window is scanned across the image at multiple positions and scales, which is followed by the combination of the overlapping detections. At last, a pedestrian is described by a final detection, and we have detected the pedestrians in the thermal imagery. Experimental results with OSU Thermal Pedestrian Database are reported to demonstrate the excellent performance of our algorithms.**

Keywords-Pedestrian detection; histogram of oriented gradient (HOG); geometric characteristics; illumination difference; linear SVM.

I. INTRODUCTION

The detection of pedestrian in thermal imagery is a challenging task in the field of computer vision. In Fig. 1 we show outdoor surveillance images of the same scene captured using the same thermal camera. However considering different conditions, the thermal properties of people and background are quite different [5], which makes it hard for us to increase the accuracy of detecting the precise locations and shapes of people. How to detect the pedestrian from the video or image quickly and accurately is still a hotspot.

Recently, there has been a flurry of works on pedestrian detection and tracking in visible imagery and thermal imagery. Those can be divided into three categories: the first one is based on the construction of the human body model [8-11]; the second one is based on the template matching [1, 14]; and the third one is based on statistical classification, which classify the object through the usage of pattern recognition methods. The pattern recognition method follows the extraction of object features and this paper is based on this method. Pedestrian detection based on histogram of oriented gradient (HOG) is the most influential among these methods, since it have been shown to be efficient and robust and this method is initially introduced by Dalal et al [2].

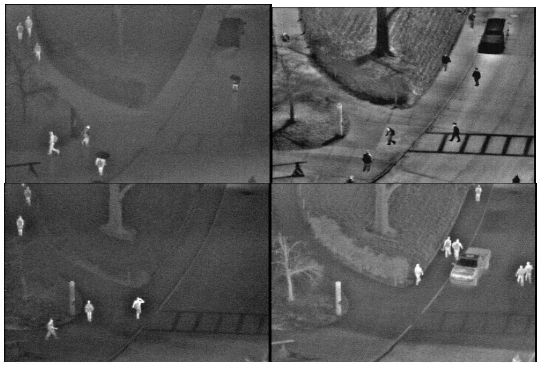

Figure 1. Thermal images show great sensitivity for different conditions.

HOG features have described the statistical information for the distribution of local intensity gradients or edge directions, which is a very rich response of the local object appearance and shape. Many methods have been raised to make it perfect, the typical example is the introduction of the boosted cascade algorithm used in face detection [12], which has increased the number of HOG features and improved the detection speed greatly. However，these studies are for the visible imagery and it seems that the statistical information for the distribution of local intensity gradients and edge directions is not well developed in thermal imagery [3, 7, 13]. In [6], a fuzzy inference system is used for target detection and classification, and some useful geometric characteristics have been introduced. In [5], Davis uses a two-stage template-based method to detect people in widely varying thermal imagery, but it is too restricted to build human body model, and the template of body may hardly cover all situations, it is not conducive to extend to other data sets.

This paper we present an improved HOG algorithm to detect pedestrian in thermal imagery. We adjust the parameters of HOG feature extraction process, and use a linear-SVM for imagery classification and detection; we find it useful for the pedestrian detection in thermal imagery. Meanwhile, in order to achieve higher detection rate, we add a series of geometric features because of the reason that the targets of thermal imagery appear around very hot or cold objects, which is also followed by using a linear-SVM for imagery classification and detection, then we get the final detector. The detection window is scanned across the image with the final detector, which is followed by the combination of the overlapping detections, and then we get the pedestrians in the imagery in the end.

978-1-4577-2133-5/10/$26.00 ©2012 IEEE

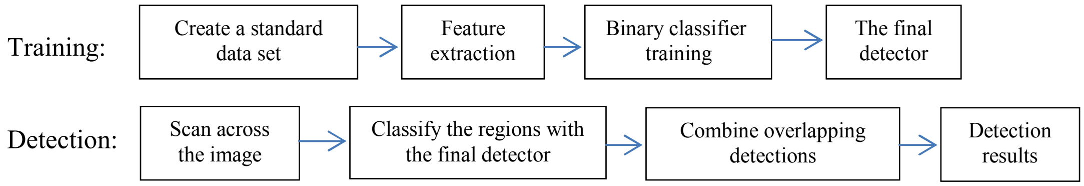
Training:
Create a standard data set
Feature extraction
Binary classifier training
The final detector
Detection:
Scan across the image
Classify the regions with the final detector
Combine overlapping detections
Detection results

Figure 2. an overview of our improved HOG algorithm chain

We give an overview of our improved HOG algorithm in section II, give a detailed description and experimental evaluation of the feature extraction in section III and describe our data sets and results of our experimental in section IV. The main conclusions are summarized in V.

II. Description of improved HOG algorithm

This section gives an overview of our improved HOG algorithm chain, which is summarized in Fig. 2. Classification-based framework for object detection is established to classify a single region of the image [15], which can be divided into two parts: training and detection. The purpose of training is to create a binary classifier which is used to distinguish whether a single region of the image is the target or not, and the task of detection is to classify the regions by scanning across the image at multiple scales and locations with the final detector, which is followed by the combination of overlapping detections. Finally we get the detection results.

Training and detection includes three steps, which constitute a common framework for object detection generally. And the final detection results not only depend on the accuracy of the classifier, but also on the combination of the overlapping detections. We introduce the process of pedestrian detection in thermal imagery with our improved HOG algorithm in details.

The first step of training is to create a standard data set. To create the positive examples a human expert manually delineates the rectangular outer boundaries of pedestrians in the imagery. The pedestrian centroids are computed from the bounding rectangle, which only contains a pedestrian in one; the negative samples are generated through scanning the imagery without pedestrian. Classifier is trained by these samples, and the detection window size is the same size with the example window.

The second step is to extract the features, this paper we choose HOG features and some geometric characteristics as our implement features. To extract HOG features, the object is divided into small equally-sized regions called cells, then accumulates a local 1-D histogram of gradient directions or edge orientations over the pixels of the cell which is followed by accumulating a measure of local histogram “energy” over somewhat larger spatial regions (“blocks”) and using the results to normalize all of the cells in the block [2]. Tiling the detection window with an overlapping grid of blocks and combing the feature vector are the HOG features. For the extraction of geometric characteristics, we choose Mean Contrast, Standard Deviation and Ratio Bright Pixels/Total Pixels as our features, and it is necessary to normalize the features to meet the reason that the target always appear around very hot or cold objects. And the details of the feature extraction are introduced in the next section.

The third step is to train the binary classifier with a linear SVM, because of the efficient and robust results with a low cost of run time while detecting pedestrian in visible imagery [2]. We use a soft (C=0.01) linear SVM trained with the LIBSVM package.

The next process is to detect pedestrians in the imagery with the final detector. First we need scan across the image at multiple positions and scales. Scaling is achieved by scaling the image, rather than scaling the detector because of the fixed-size of blocks. Calculate the HOG features and geometric characteristics of the regions in the scanned imagery, then detecting whether it is a pedestrian or not. If it is then calibrate out. If it is not, we give it up directly.

Because of that a number of detections will usually occur around each pedestrian while detecting and the changes of scale will also make a pedestrian corresponding to multiple detections. In practice it often makes sense to return a final detection per pedestrian. Toward this end it is necessary to combine overlapping detections into a single detection. Through those processes we get the detection results.

III. FEATURE EXTRACTION

A. *HOG Features*

HOG features are local region descriptors, which consist of many histograms of orientated gradients in localized areas of an image. The basic idea is that local object appearance and shape can often be characterized rather well by the distribution of local intensity gradients or edge directions [2]. Therefore, the human body can be well represented by the local region descriptors through calculating the histograms of orientated gradients.

To calculate the features we divide the image window into small spatial regions (“cells”), in our experiment we set the size of pixel cells to be 4*5 and our example pixel size is 20*25. Then each image window has been divided into 25 parts, we calculate their 1-D histogram of gradient directions each. The gradients at the point P(x, y) of image I can be found by convolving gradient operator with the image:

$$G_x(x,y) = [-1\ 0\ 1] * \mathrm{I(x,y)} \qquad (1)$$

$$G_y(x,y) = [-1\ 0\ 1]^T * I(x,y) \qquad (2)$$

The strength of the gradient at the point P(x, y) is:

$$G(x,y) = \sqrt{G_x(x,y)^2 + G_y(x,y)^2} \qquad (3)$$

The orientation of the edge at the point P(x, y) is:

$$\theta(x, y) = \arctan[\frac{G_y(x, y)}{G_x(x, y)}] \tag{4}$$

Then linear gradient voting into 9 orientation bins in $0^o - 180^o$ and denote the value of k_{th} bin to be:

$$\varphi_k(x, y) = \begin{cases} G(x, y) & \text{if } \theta(\text{x,y}) \in \text{bin}_k \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

For better invariance to illumination etc., it is useful to contrast-normalize the local responses before using them. The method is to group the cells into bigger normalized descriptor blocks which are the Histogram of Oriented Gradient descriptors. It is also useful to down weight pixels near the edges of the block by applying a Gaussian spatial window to each pixel before accumulating orientation votes into cells. We evaluated four different block normalization schemes for each of the above HOG geometries through L2-Hys. The final feature vector is combined by tiling the detection window with an overlapping grid of HOG descriptors.

B. *Geometric Characteristics*

HOG features represent the image gradient information while there is not that much gradient information contrast thermal imagery with visible imagery. As a result of that, it is necessary to add some other useful information for a better experiment result. This paper we choose the geometric characteristics of the detection window.

In Fig. 3, we show a thermal imagery contains a number of pedestrians, with a little drizzle in night, and the histogram associated with this imagery is also provided in the figure. It is easy to find that the average temperature of human body is higher than that of the background objects, so the pedestrian is the lighter object in the imagery. A thermal imagery with no pedestrian present is provided in Fig. 4, the histogram associated with this imagery is also provided in the figure. A review of the histograms provided in Figs. 3 and 4 show that in the case of the thermal imagery with no pedestrian present, there are no pixels with gray scale levels greater than 125 (Fig. 4). In the thermal imagery with pedestrians present there are a significant number of pixels with gray scale levels greater than 125 (Fig. 3). By this result, we can know that the pixels with gray scale levels greater than 125 in this case are due to the pedestrian in the thermal imagery. Similarly, the average temperature of human body is lower than that of the background objects, and the pedestrian is the dark object in the imagery.

We analyze the features of samples from the perspective of the histogram in order to get the geometric characteristics. As shown in Fig. 6, Pixels, gray scale levels greater than 125, in positive example are far greater than the negative example. Specially, there are no pixels with gray scale levels greater than 125 in the negative example. Therefore, it can be seen that the pixels of pedestrian are mainly the lighter regions where gray scale levels greater than 125. Similarly, the pixels of pedestrian are mainly the dark regions of samples in the daytime.

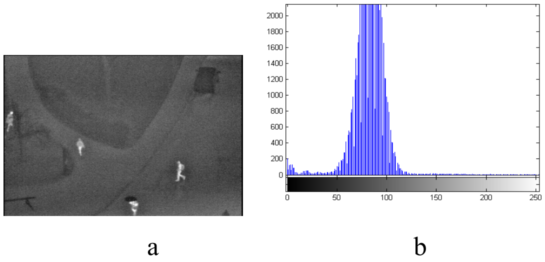

Figure 3. (a) Imagery (with pedestrian present), (b) The histogram of the imagery

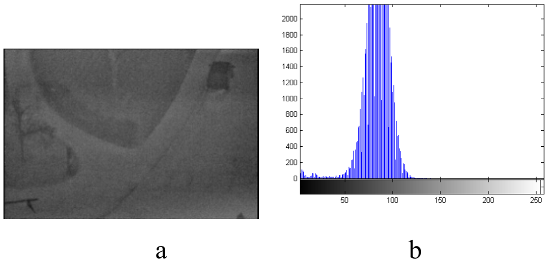

Figure 4. (a) Imagery (with no pedestrian present), (b) The histogram of the imagery

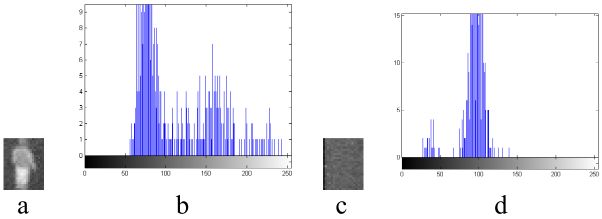

Figure 5. (a) Positive example (b) The histogram of the positive example (c) Negative example as the same conditions as positive example (d) The histogram of the negative example

We use the following characteristics in our experiments, where similar characteristics have been used in the previous literature [4, 6]:

1) *Mean Contrast*

The average gray value of the example minus that of the imagery contains the sample, which is divided by the average gray value of the imagery, then takes the absolute value of its. The calculation formula is as follows:

$$\left| \frac{meangray\_subimg - meangray}{meangray} \right| \tag{6}$$

Where

$meangray\_subimg$ the average gray value of the sample;

$meangray$ the average gray value of the imagery containing the sample.

2) *Standard Deviation*

The standard deviation of the sample reflects the statistical features of the example pixels. The calculation formula is as follows:

$$S = \sqrt{\frac{\Sigma\,(f(x,y) - \overline{f(x,y)})}{N-1}} \tag{7}$$

Where

$f(x,y)$ the gray value of the point (x, y);

$\overline{f(x,y)}$ the average gray value of the sample;

N the number of pixels.

3) Ratio Bright Pixels/Total Pixels

The number of bright pixels is the number of pixels whose gray value is 10% smaller than the largest gray value or 10% larger than the smallest gray value of the sample.

The combination of HOG features and Geometric characteristics is our implement feature vector.

IV. DATASET DESCRIPTION AND RESULTS

A. Dataset Description

Our training and testing dataset is OSU thermal pedestrian database, containing 10 categories of images at different times of different weather. The total number of images is 284 and the typical size of each image is 360*240. Those images are obtained from the same place, pedestrian intersection on the Ohio State University campus, at the same infrared sensor.

For a typical size image, most of the pedestrians in the image are around 15 to 25 pixels in width and around 20 to 30 pixels in height. An image region with size 20*25 pixels can cover most pedestrians in the imagery. Hence, sub-images of size 20*25 pixels centered at pedestrian centroids are built into the positive example base. Since the pedestrians are represented by so few pixels, their detection is highly sensitive to the surrounding context. In the detection process, we detect the imagery at different scales to accommodate different sizes of the pedestrian. We select 744 of the images as positive examples, together with their left-right reflections (1488 images in all). Meanwhile, in order to adapt to all weather conditions, 1000 positive training examples are selected out by the same proportion from the 10 categories of images, and the remaining 488 examples as the test examples. Fig. 6 shows some samples.

The negative examples are generated through scanning the imagery without pedestrian by the window-stride 4*5 with the same size as the positive examples. We get 3784*10 of the images as the negative examples. Similarly, 3000*10 of the negative training examples is selected by the same proportion from the 10 categories of images, and the remaining 784*10 examples as the test examples.

B. Training and Detection

Support vector machine (SVM) is a machine learning algorithm introduced by Vladimir Vapnik, which has been shown to be highly effective at two-category target categorization. SVM shows good performance on high dimensional patter recognition problem. The basic idea of SVM is to map the data x into a high-dimensional feature space F via a nonlinear mapping and find close neighbors to a query sample to do linear regression in this space. The classifier has the largest distance to the nearest training data points of any class.

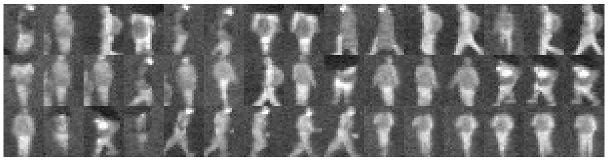

Figure 6. Some sample images from our pedestrian detection database

In experiment we use a linear SVM trained with LIBSVM package which is called by MATLAB and the cost (C of C-SVC) is 0.01. We get the SVM weight after training, which is used to search the prepared test samples exhaustively for false positives, and then we re-train using this augmented set to produce the final detector.

As the typical size of the examples are 20*25 pixels, but the scale of the pedestrians is different in the actual imagery. So it is useful to scale the image to different sizes while detecting. The detecting process is shown as Table I.

TABLE I. DETECTING PROCESS, THE SCALES ARE SAVED IN SCALE[LEVELS]

```
●  Given the final detector and the images that need to be detected.
   The width of the image is diw and the height of the image is dih.
●  Initialize the detection window size dw = 20, dh = 25, where dw
   means the width of detection window and dh means the height of
   detection window, the maximum levels ml = 24, the value of scale
   s = 1, the step of scale s_step = 1.05, level l = 0.
●  While l < ml:
       1.  diw <- diw / s
           dih <- dih / s
       2.  if (dw > diw or dh > dih)
                 break;
           else
                 s <- s*s_step
                 scan across the imagery at multiple positions
                 calculate the features and detect whether it is a
                 pedestrian or not
                 calibrate out the pedestrian
        l ++
```

We get the final pedestrians detected in the imagery after combining overlapping detections into a single detection. Example images showing detections from the combining process are shown in Fig. 7.

C. Detection Results

The detection results of our experimental with OTCBVS benchmark – OSU thermal pedestrian database is shown in Table II.

We give the true positive (TP) and false positive (FP) of the detection, likewise, we calculate the Sensitivity and Positive Predictive Value (PPV) [5] to quantify our detector performance. The Sensitivity reports the fraction of people that are correctly identified by our experiment, where a high Sensitivity corresponds to a higher detection rate of people. The PPV reports the fraction of detections that actually are people, where a high PPV corresponds to a low number of false positives. Some detection results are shown in Fig. 8.

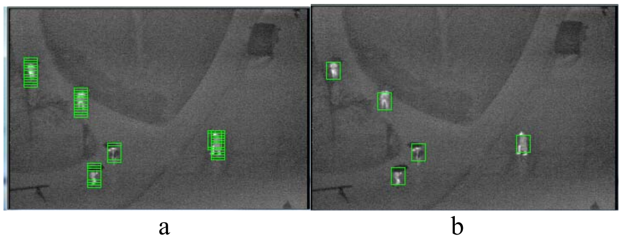

Figure 7. (a) Detection results before combining (b) Final detection results

For the analysis of experimental results we found high PPV measurements in such a complex background and high Sensitivity measurements in most cases. But the Sensitivity is less than 0.7 in two categories of images. This may be due to the excessive noise which has a great impact on the gradient after analysis of those images. Overall, our experiment has achieved the target of detecting pedestrian in thermal imagery excellently.

TABLE II. DETECTION RESULTS

No.	#Fra.	#People	#TP	#FP	Sen.	PPV
1	31	91	78	2	0.86	0.98
2	28	100	95	3	0.95	0.97
3	23	101	70	13	0.69	0.84
4	18	109	109	10	1	0.92
5	23	101	91	6	0.83	0.94
6	18	97	88	2	0.91	0.98
7	22	94	64	2	0.68	0.94
8	24	99	82	0	0.83	1
9	73	95	91	9	0.96	0.91
10	24	97	77	0	0.79	1
1-10	284	984	845	41	0.86	0.95

I. SUMMARY AND CONCLUSIONS

In this paper, we present an effective approach to detect pedestrian in thermal imagery combining the HOG features with geometric characteristics. Experimental results with OSU Thermal Pedestrian Database have demonstrated that the detection performance is very promising. Simultaneously, with minor modifications this approach can be adapted to training using representative datasets, such as tanks and other small targets in thermal imagery. This is based on the efficient and robust nature of the HOG features and the general nature of Geometric characteristics.

Future work: Although our experiment has achieved a satisfactory result, there is also some experiment results are not that excellent. For example, the Sensitivity is too low in #3 and #6, and the PPV is not that high in some categories of images. We will study and analyze the reason for these and design related algorithms to improve the detection rate and reduce the false positive in future work.

ACKNOWLEDGMENT

This work is supported by the national natural science foundation of China (No. 61073130) and the project of National High Technology Research and Development Program of China (863 Program) (No. 2011AA01A207).

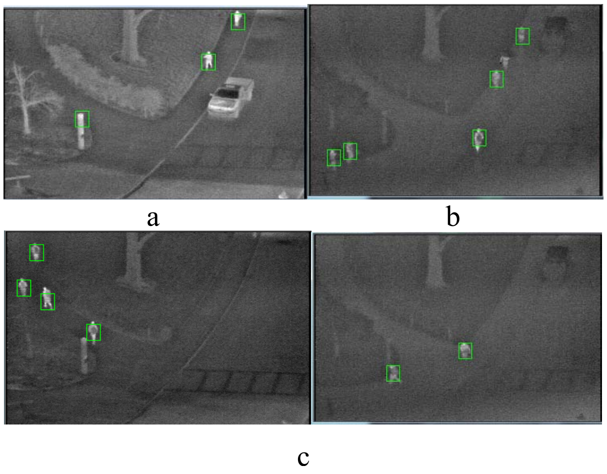

Figure 8. (a) example of false positive (b) example of false negative (c) examples of true positive

REFERENCES

[1] P. Viola, M. J. Jones, and D. Snow “Detecting Pedestrians Using Patterns of Motion and Appearance”, In Proc. Int. Conf. Comp. Vis., 2003, pp.734-741

[2] N. Dalal, and B. Triggs, “Histograms of oriented gradients for human detection”, In CVPR, vol. 1, 2005, pp. 886-893

[3] Q. Zhu, M. C. Yeh, K. T. Cheng and S. Avidan, “Fast human detection using a cascade of histograms of oriented gradients” , in CVPR, vol. 2, 2006, pp. 1491-1498

[4] X. Y. Jin, and C. H. Davis, “Vector-Guided Vehicle Detection from High-Resolution Satellite Imagery”, IEEE, 2004, pp. 1095-1098

[5] J. W. Davis, and M. A. Keck, “A Two-Stage Template Approach to Person Detection in Thermal Imagery”, Applications of Computer Vision and the IEEE Workshop on Motion and Video Computing, IEEE Workshop on, vol. 1, pp. 364-369

[6] B. N. Nelson, “Automatic Vehicle Detection in Infrared Imagery Using a Fuzzy Inference-Based Classification System”, IEEE TRANSACTIONS ON FUZZY SYSTEMS, VOL. 9, NO. 1, FEBRUARY, 2001, pp. 53-61

[7] Q. J. Wang, and R. B. Zhang, “LPP-HOG: A New Local Image Descriptor for Fast Human Detection”, IEEE, 2008, pp. 640-643

[8] H. Fujiyoshi, A. J. Lipt, and R. S. Patil, “Moving Target Classification and Tracking from Real time Video”, Processing of IEEE Workshop on Applications of Computer Vision, 1998, pp. 8-14

[9] M. Oren, C. Papageorgiou, and P. Sinha, “Pedstrian Detection Using Wavelet Templates”, In IEEE Conference on Computer Vision and Pattern Recognition, 1997, pp. 193-199

[10] A. Torralba, and A. A. Efros, “Unbiased Look at Dataset Bias”, In Conference on Computer Vision and Pattern Recognition, 2011, pp. 1521-1528

[11] P. Dollar, C. Wojek, B. Schieele, and P. Perona, “Pedestrian detection: A benchmark”, In Conference on Computer Vision and Pattern Recognition, 2009, pp. 304-311

[12] P. Viola, and M. Jones “Rapid Object Detection using a Boosted Cascade of Simple Features”, In IEEE Computer Vision and Pattern Recognition, Vol. 1, 2001, pp. 511-518

[13] V. Paul, and M. Jones, “Robust Real-time Face Detection”, International Journal of Computer Vision, 2004, pp. 137-154

[14] K. Mikolajczyk, C. Schmd, and A. Zisserman, “Human Detection Based on a Probabilistic Assembly of Robust Part Detector”, European Conference on Computer Vision, 2004, pp.69-82

[15] Q. Tang, “Research of Threshold Segmentation Algorithms and Pedestrians Detection in Infrared Image”, South China University of Technology, 2010