IMPROVED HUMAN DETECTION AND CLASSIFICATION IN THERMAL IMAGES

Weihong Wang\textsuperscript{1,2}, Jian Zhang\textsuperscript{1,2}, Chunhua Shen\textsuperscript{1,3}

\textsuperscript{1}NICTA \hspace{1em} \textsuperscript{2}University of New South Wales, Australia \hspace{1em} \textsuperscript{3}Australian National University, Australia

\section*{ABSTRACT}

We present a new method for detecting pedestrians in thermal images. The method is based on the Shape Context Descriptor (SCD) with the Adaboost cascade classifier framework. Compared with standard optical images, thermal imaging cameras offer a clear advantage for night-time video surveillance. It is robust on the light changes in day-time. Experiments show that shape context features with boosting classification provide a significant improvement on human detection in thermal images. In this work, we have also compared our proposed method with rectangle features on the public dataset of thermal images. In this work, we have also compared our proposed method with rectangle features on the public dataset of thermal images. Results show that shape context features are much better than the conventional rectangular features on this task.

\textbf{Index Terms}— Human detection, thermal image, shape context, Adaboost

\section{1. INTRODUCTION}

Human detection is of fundamental importance in computer vision due to its wide applications in video surveillance, smart vehicles and robotics. Feature selection and classification like boosting algorithms have demonstrated promising results for this task. To build a successful human detector under wide ranges of weather and different lighting conditions, thermal imaging can be used for alleviating some problems attached to standard imaging techniques. Still, there are many challenging problems for thermal imaging based detection. The lack of the color and texture information from thermal image will impact on the quality of feature extraction that generates significant features and its descriptor for human detection. Besides, as mentioned in [1], the common ferroelectric thermal sensors also have some technical challenges, including a low Signal-to-Noise Ratio (SNR), white black/hotcold polarity changes, and halos that appear around very hot or cold objects. Therefore, robust feature extraction from thermal images for human detection is critically essential to ensure that a high performance could be achieved.

In this paper, we present a comprehensive experimental study on human detection in thermal images. We apply the Shape Context Descriptor (SCD) for local edge feature extraction and use boosting algorithms with a cascade framework [2] for human classification. The motivation here is to exploit the power and efficiency of edge based SCD as discriminative features on human shapes. To evaluate the proposed method for human detection in thermal images, we adopt a multi-layer boosting algorithm with 1) rectangle feature and 2) edge feature based SCD, which is a multi-dimensional feature descriptor. Experiments show that the SCD based approach can achieve a significantly better performance on the challenging dataset presented in [1].

\section{2. RELATED WORK}

There exist a few approaches for detecting humans in thermal images in the literature. Bertozzi [3] introduced a pedestrian detection method as a part of a driver assistance system. The algorithm is divided into three parts. (1) Candidate generation. The input thermal image is processed to locate warm symmetrical objects with a specific size and aspect ratio. (2) Candidates filtering. The candidates may contain poles, road signs and buildings, which also have symmetry characteristic. These false positive objects can be filtered by analyzing the shape of the vertical histogram in each search window. (3) Validation of candidates. Morphological characteristics of a human are extracted to form a model. Each filtered search window is compared with the model for validation. The weakness of this method is that human should be hotter than its background.

Since human appearance can vary considerably in thermal images due to temperature variations, Davis and Keck presented a two-stage template-based method [4], which takes advantage of the invariance of edge information. In the first stage, human contours are obtained by creating Contour Saliency Map (CSM) [1] of thermal images. CSM represents the belief of each pixel belonging to an edge contour of a person [4]. Then a screening template is produced by averaging the human samples cropped from the CSM images. Last, a multi-resolution screening procedure is applied to obtain candidates. In the second stage, four Sobel filters with different angles are applied to the human samples to get four projected edge images. An Adaboost classifier is trained with the projected images and is applied to new input images. This method proves that edge is a robust feature for object...
Recently Kai and Arens [5] proposed a local-feature based pedestrian detector on thermal data. In the training phase, they used a combination of multiple cues to find interest points in the images and use SURF (Speed Up Robust Features) [6] as features to describe these points. Then a codebook is created by clustering these features and building Implicit Shape Model (ISM) [5] to describe the spatial configuration of features relative to the object center. In the detection stage, SURF features are first located in each image. Then the matching between the features and the codebook is conducted to locate object center. The challenge of this detector is whether a high performance can be achieved when local features are not obvious, for example, in thermal images of poor quality.

3. THE PROPOSED METHOD

In this paper, we propose a shape context descriptor based Adaboost cascade classification for human detection in thermal images.

The boosted cascade classification framework has been successfully applied to object detection applications since introduced by Viola and Jones [7]. Davis and Keck employed it in thermal image area [4]. We use this framework as the classifier in our algorithm. Compared to classification, choosing a promising feature is a more critical step since humans in thermal images could present very different characteristics from those in visual images. We conjecture that appearance-based features, such as Haar features [7], covariance features [8] and histogram of oriented gradient (HOG) features [9, 10], may not be suitable in thermal images based surveillance applications for the following reasons:

1. The size of the target is relatively small,
2. Pixel intensities in human bodies have considerable variations under different weather conditions and,
3. There is lack of texture information.

Motivated by [4], we try to exploit the human shape information. We have evaluated two different image features. For the first one, we follow the second stage of the method in [4]. We apply Sobel filters on images as preprocessing. Then we apply Adaboost cascade on the filtered images with rectangle feature that simply sums up the pixel intensity in a rectangular box in images. For the second one, SC is chosen, for its rich description of shape information. Because of the multi-dimensional nature of the descriptor, to find an optimal linear classifier would require much longer time. As shown in [11], it is possible to use linear support vector machines (SVMs) as weak classifiers. However, to train SVMs could be time-consuming. Here we have adopted a more efficient approach by projecting the multi-dimensional features onto a 1D line using Linear Discriminant Analysis (LDA). The LDA finds a linear projection function which guarantees optimal classification of normally distributed samples of two classes.

Through our experiments, we conclude that the SC descriptor based method can achieve a significantly better performance for human detection in thermal images, compared to the rectangle feature based method.

3.1. Shape Context Descriptor

The SCD [12] was originally invented for shape matching and object recognition. It is composed of two steps to obtain the SC descriptor for an object. The first step is to detect object shapes by using Sobel filter or Canny edge detector and then select points on the shape of an object. In the second step a compact descriptor for each point is obtained by producing a histogram of the relative positions of other points. To make the histogram more distinguishable among nearby pixels, a log-polar coordinate system is used. We call this histogram Shape Context Descriptor. It has been shown that the descriptor is robust and discriminative [12, 13]. The drawback is its high computation cost. Fig. 1 is an example of shape context descriptors on a human image.

3.2. Boosting

Adaboost is an ensemble learning algorithm. The training procedure of Adaboost iteratively and greedily combines a set of weak classifiers to form a strong classifier. The final strong classifier $H(\cdot)$ can be defined as

$$H(x) = \begin{cases} 1 & \text{if } \sum_{i=1}^{T} \alpha_i h_i(x) \geq 0; \\ -1 & \text{otherwise,} \end{cases}$$

where $\alpha_i$ is a weight coefficient; $h_i(\cdot)$ is a weak classifier and $T$ is the number of weak classifiers.

Viola and Jones presented a face detection system based on a cascade of Adaboost classifiers [7]. The cascade has increasing complexity and discriminative power. During the
early detection stages, simple classifiers filter out majority of search windows with no faces inside while in later stages, complex classifiers detect real faces. So efficiency of the whole detection operation is improved greatly.

3.3. Fisher Discriminant Analysis

Fisher’s linear discriminant (FDA) is a method in statistics and machine learning. It can be used for classification or dimensionality reduction. In our algorithm, it is used to reduce the multi-dimensional SC descriptors into one dimension. Assume we have a training set composed of \( x = [x_1, x_2, ..., x_n]^T \). Each training sample belongs to one of two classes, \( C_1 \) and \( C_2 \). We try to find a weight vector \( w = [w_1, w_2, ..., w_n]^T \) and a threshold \( w_0 \) which satisfies

\[
\begin{align*}
& w^T x + w_0 > 0 & (x \in C_1), \\
& w^T x + w_0 > 0 & (x \in C_2).
\end{align*}
\]

The purpose is to find a linear combination of the variables that can separate the two classes as much as possible. The computed linear combination projects the multi-dimensional samples to one dimension. The criterion proposed by Fisher is the ratio of between-class to within-class variances which can be written as

\[
J = \frac{w^T S_b w}{w^T S_w w},
\]

where \( S_b = (m_1 - m_2)(m_1 - m_2)^T \),

\[
S_w = \sum_{x \in C_1} (x - m_1)(x - m_1)^T + \sum_{x \in C_2} (x - m_2)(x - m_2)^T.
\]

Here \( m_1 \) is the mean of class 1, \( m_2 \) is the mean of class 2. \( S_b \) and \( S_w \) are between-class and within-class scatter matrices. We want to find a direction \( w \) such that the between-class variance is maximized and within-class variance is minimized.

3.4. Shape Context based Adaboost cascade

In the training process of our Adaboost cascade, training samples are transformed by Canny edge detector. Then for each pixel in the training samples, the SC descriptor for the pixel is constructed. All of the SC descriptors are candidate features for the Adaboost classifiers. The SC descriptors capture human shape information in a detailed and efficient way. FDA is then applied to project each descriptor to one dimension. This information can be used to represent a distinct part of the human shape. At each Adaboost iteration, a weak classifier is trained from the collection of SC descriptors. Each weak classifier can then be defined as

\[
h(x) = \begin{cases} 
+1 & \text{if } w^T x > \theta; \\
-1 & \text{otherwise},
\end{cases}
\]

where \( h(\cdot) \) is a weak classifier, \( w \) is the weight vector calculated from FDA, \( x \) is a SC descriptor and \( \theta \) is an optimal threshold such that the minimum number of examples is misclassified.

In the test process, test images are transformed by the Canny edge detector and SC descriptors.

4. EXPERIMENTS

To evaluate the proposed method, we conduct two experiments using sliding-windows approach. The first one follows the second stage of the method in [4] which is described in Section 2. However, [4] does not specify how to calculate the feature values based on the projected images. In this paper, we use rectangular feature which calculate the sum of pixel intensity in the rectangle box in images transformed by Sobel filter. In the second experiment, we test our proposed SC based Adaboost cascade. The goal is to compare the performance between the rectangular feature and the SC based methods.

4.1. Dataset

We use the dataset from [4]. For the training process, about 500 positive training samples are cropped from the dataset images. The number of negative samples is about 500 as well. The size of the training samples is scaled to 26 pixels in height and 20 pixels in width. About 170 non-human thermal images are used to bootstrap the negative samples. In the testing process, we use 108 images containing 410 humans as testing dataset. We label bounding boxes on the humans as ground truths. A detection result is considered as correct when the search window and ground truth bounding box is over 50% overlapped.

4.2. Experiment parameters

In the training process of the Adaboost cascade, all of the pixel locations are considered as candidate features to form an Adaboost classifier, the number of which is 520 (= 26 by 20). The scaling factor is set as 1.2 and window shifting step is 1. In each stage of the cascade, the minimum detection rate and the maximum false positive rate are set as 99.5% and 50% respectively. Weak classifiers are combined into an Adaboost classifier until these two conditions are met.

4.3. Experiment results

The ROC curves in Fig. 2 show a comparison of performances of our proposed detector and the rectangle feature based detector in each stage. From the ROC curves, it is obvious that our proposed method significantly outperforms the rectangle feature based classifier overall. In the left most stages, when the number of false positive is about 20 (which corresponds to a 5% false positive rate), our proposed detector can achieve a above 70% detection rate while the rectangle feature based
detector can only obtain a 30% detection rate. When the detection rate reaches 80% for our proposed detector, the false positive is slightly over 40 (which is a 10% false positive rate). In contrast, the rectangle feature based detector can only obtain a 45% detection rate. In summary, by combining the SC and boosting, we can obtain robust features to represent human information in thermal images. It should be noticed that we did not conduct any pre-processing process such as CSM in [4] which, we believe, can significantly filter many false positive samples and bring candidate positive samples to our proposed method for a more promising performance.

Fig. 3 shows some examples of detection results for our proposed detector and the rectangle feature based detector. In the images, a black bounding box indicates a correct detection, a blue box is a missing detection and a yellow box is a false positive. We can see that our proposed method have a much better detection performance.

5. CONCLUSION

We have introduced a novelty classifier for human detection in thermal images. The proposed structure takes advantages of the discriminant power of SC and LDA to achieve a higher human detection performance than the rectangular feature based method. However, the high computation on SC descriptor calculation hampers our algorithm to work in realtime. The reduction of computation cost is our next focus.

6. ACKNOWLEDGEMENT

NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program.

7. REFERENCES