Using Cross-examples in Viola-jones Algorithm for Thermal Face Detection

Hong Tran, Chunhua Dong, Masoud Naghedolfeizi, Xiangyan Zeng
Fort Valley State University
Fort Valley, Georgia, USA
htran1@wildcat.fvsu.edu,dongc@fvsu.edu,feizim@fvsu.edu,zengx@fvsu.edu

ABSTRACT

Detection of the face region is a key step in a face recognition system. Thermal images are widely used in many applications where normal visibility is reduced, impaired or ineffective, such as night surveillance and fugitive searches. However, low spatial resolution brings challenges to face detection in thermal images. Viola-Jones is an object detection method widely used for face detection. The algorithm suffers from missed faces and wrongly detected non-face objects due to low resolution of thermal images. To improve the face detection performance for thermal images, we propose to incorporate cross-examples into our framework. In addition to using negative samples of non-face thermal images, we utilize non-face visible images as part of the negative samples (cross-examples). Cross-examples effectively increase the discriminability between the positive samples and negative samples. Experimental results show that the proposed scheme can effectively reduce the non-face objects and thus improve accuracy of face detection.

CCS CONCEPTS

• **Computing methodologies** → **Object detection**; **Object identification**; **Object recognition**.

KEYWORDS

Face Detection, Infrared Image, Thermal Image, Object Detection, Cross-examples, Viola-Jones

ACM Reference Format:

Hong Tran, Chunhua Dong, Masoud Naghedolfeizi, Xiangyan Zeng. 2021. Using Cross-examples in Viola-jones Algorithm for Thermal Face Detection. In *2021 ACM Southeast Conference (ACMSE 2021), April 15–17, 2021, Virtual Event, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3409334.3452083

1 INTRODUCTION

Face detection is a critical step in a face recognition system [5–7, 9, 14], with the purpose of localizing and extracting the region of interest (ROI) that contains faces from the background. In the past decade, most of face detection and recognition research was conducted in the visible spectrum. Most recently, deep leaning has been applied to face detection in visible images and achieved high performance [8, 17]. However, the quality of visible images is highly dependent on the lighting conditions of the environment. The performance of face detection and recognition is less than desirable when the lighting conditions are not proper, e.g. low light in evening or irregular illumination. On the other hand, thermal imaging systems produce pictures from heat, not visible light. Sensors in thermal cameras convert temperature information into images called thermogram, which uses a color palette to indicate the range of temperatures present. Compared to visible cameras, thermal cameras can create images in dark areas.

The human face is a dynamic object and has a high degree of variability in its appearance under different illumination, which makes face detection a difficult problem in computer vision. Recently, numerous techniques have been proposed for face detection [1, 2, 10, 11, 16] using thermal images and visible images, ranging from simple edge-based algorithms, knowledge-based methods, feature invariant approaches, and template matching methods to composite high-level approaches utilizing advanced pattern recognition methods. Knowledge-based methods use pre-defined rules based on human knowledge; feature invariant approaches aim to find face structure features robust to pose and lighting variations; template matching methods use pre-stored face templates to determine where a human face is depicted in an image. Y. Zheng proposed a method for face detection and eyeglass detection based on region growing and morphological operations [19]. J. Mekyska et al. proposed a detection method that binarizes an image and calculates the vertical and horizontal projections, from which the coordinates of the rectangle that contains a face are determined [12]. Based on J. Mekyska's method, Y.K. Cheong used the horizontal projection of an image to determine the global minimum that helps identify the height and width of the head region [3]. These methods were designed and tested on images that contain one person, e.g. frontal portrait pictures. Viola and Jones proposed a learning-based object detection algorithm that has been successfully used for real-time face detection [18]. The experiments conducted by Reese et al. [13] show that the Viola-Jones algorithm is the best for face detection in the thermal spectrum. However, the performance of the Viola-Jones algorithm may suffer from missed faces and wrongly detected non-face objects.

In this paper, we propose to incorporate cross-examples into the Viola-Jones face detection framework to improve the detection accuracy. Our training database contains positive samples that are thermal face images and negative samples (non-face objects) that include thermal non-face images and visible non-face images called "cross-examples", which means the negative samples is a set of thermal and visible images. Cross-examples aim to increase the

ACMSE 2021, April 15–17, 2021, Virtual Event, USA

ACM ISBN 978-1-4503-8068-3/21/04…$15.00
https://doi.org/10.1145/3409334.3452083

discrimination between the positive samples and negative samples, and improve the performance of the Viola-Jones algorithm that is based on classification.

The remainder of this paper is organized as follows. Section 2 presents a brief overview of the Viola-Jones method and elaborates our proposed strategy of adding cross examples in this framework. Section 3 contains experimental results and discusses the implementation of our method, followed by the conclusion (Section 4).

2 METHODOLOGY

Viola-Jones algorithm developed by Paul Viola and Michael Jones back in 2001 [18] is a robust face detection method with high speed to practically apply in real time. Viola-Jones method is adopted to detect faces in thermal images as a baseline method. The algorithm uses AdaBoost (Adaptive Boosting) for selecting the best subset of features, and a cascade classification scheme. We replace the original haar-like features using the histogram of oriented gradient (HOG) features. In addition, we propose to add cross-examples in the training of AdaBoost classifiers.

2.1 Feature Descriptor of Histogram of Oriented Gradient (HOG)

HOG is largely invariant to global intensity changes and is capable of capturing geometric properties of faces that are difficult to capture with linear feature descriptors such as haar-like features. The algorithm is comprised of the following steps [4]:

- Step 1: Calculate the gradient values in the horizontal and vertical directions, and then the magnitude and direction of gradient.
- Step 2: Create the cell (8×8) histogram where each pixel within a cell casts a weighted vote for an orientation-based histogram channel.
- Step 3: Construct the HOG descriptor by concatenating normalized cell histograms from the cells in a block.

2.2 AdaBoost Algorithm for Feature Selection and Classification

The AdaBoost algorithm constructs a strong classifier from a set of weak classifiers and selects the most discriminative features. A weak binary classifier is built for a single feature f through a thresholding operation:

$$h(x) = \begin{cases} 1 & pf(x) < p\theta \\ 0 & otherwise \end{cases} \tag{1}$$

where $f(x)$ is the defined feature of image x, θ is the threshold and p is the polarity indicating the direction of the inequality.

Given a set of l positive and m negative training samples (x_i, y_i), where $i = 1, \ldots, m + l$, $y_i = 0, 1$ for negative and positive samples respectively. The AdaBoost algorithm can be described as follows:

Algorithm 1 AdaBoost Algorithm

Initialization:

$t = 0$; the sample weights $w_i = \frac{1}{2m}, \frac{1}{2l}$ for negative and positive samples;

Iteration:

for $t = 1, \ldots, T$ **do**

(1). $t = t + 1$;

(2). Normalize the weights w_i so that they sum to one;

(3). Find the best weak classifier h_t that yields the minimal weighted classification error $\varepsilon_t = \sum_i^T w_i \left| h_t(x_i) - y_i \right|$;

(4). Update the weights $w_i = w_i \beta_t^{1-e_t}$, where $\beta_t = \frac{\varepsilon_t}{1-\varepsilon_t}$, and $e_t = 0$ if x_i is correctly classified, otherwise $e_t = 1$;

end for

The final strong classifier is a linear combination of the T weak classifiers:

$$H(x) = \begin{cases} 1 & \sum_{t=1}^{T} a_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} a_t \\ 0 & otherwise \end{cases};$$

2.3 Cascade Classifier

Face detection is performed by a cascade classifier as shown in Figure 1. The Cascade classifier is a multi-stage classifier, each stage containing a strong classifier trained using the AdaBoost algorithm on the samples that pass through all the previous stages. Cascade of classifier can perform detection very quickly and accurately by rejecting negative input and spending more computation on positive (face) input.

2.4 Addition of Cross-examples in Training

Face detection in thermal images is more challenging due to the low spatial resolution in the images. The difference between positive and negative samples are relatively smaller than in visible images. Therefore, the classification margin defined as the difference between the weight assigned to the correct label and the maximal weight assigned to any single incorrect label is also smaller. The idea that larger margins can improve the generalization error of a classifier was previously suggested and studied [15]. In order to increase the margin in training data, we propose to add visible images of non-face objects as part of the negative image set, as shown in Figure 2. As shown in Figure 2, a visible negative image is very different from the thermal images and thus will increase the margin in the training data.

3 EXPERIMENTAL RESULTS

3.1 Dataset

The pictures used in the experiment were acquired at Fort Valley State University (FVSU) using a FLIR T650sc infrared camera that produces thermal images of 640×480 pixels. The dataset includes 72 thermal infrared frontal images of 21 subjects (8 males and 13 females). The subjects were required to pose with three to five different expressions, wearing glasses and without glasses.

3.2 Classifier Training

We manually extracted 71 positive face samples from the thermal images using the Image Labeler app in MATLAB. The bounding box was placed around each face just underneath the chin, about half way of the forehead lengthwise, and excluded the ears. We generated 95 negative samples by extracting the non-face regions

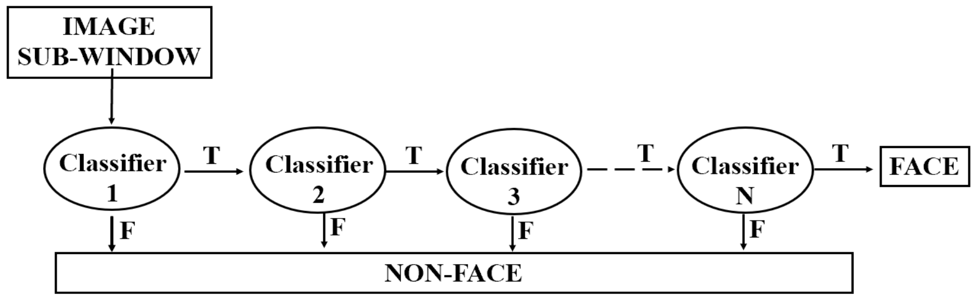

Figure 1: Cascade Scheme

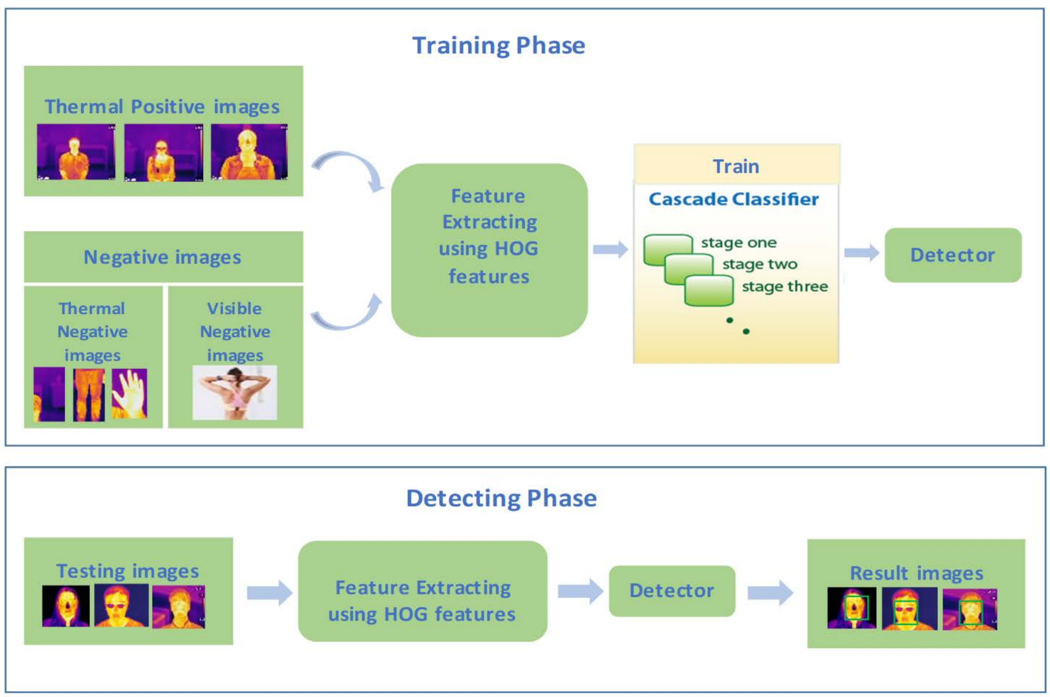

Figure 2: Thermal Face Detecting Process with Cross-examples

of the thermal images. In addition, five non-face visible images were used as negative cross-examples.

Our experiment was performed in MATLAB. The classifier training process was implemented using MATLAB function $train - CascadeObjectDetector$. Parameters for the function were set as follows: false alarm rate acceptable at each stage ($FalseAlarmRate$) was 0.2; $TruePositiveRate$ was the default value 0.995; the numbers of negative samples and positive samples used at each stage were even, in other words, the scalar ratio of negative and positive samples were 1; the number of cascade stages to train was 9; the type of feature to use was HOG features.

3.3 Detection Results and Discussion

In order to evaluate the detection capacity of the algorithm, we downloaded thermal images from the website https://github.com/albinjenith/Thermal-Imaging and use them as test images. These images were acquired using a different FLIR thermal camera. The trained cascade classifier was applied to detecting faces in 25 test images, which include one image from the FVSU dataset and 24 downloaded images. Three examples of detected faces in thermal images are shown in Figure 3. Although there are still some false positives detected by the classifier trained with cross-examples, the false positive rate is much lower than the classifier without cross-examples.

Furthermore, we use the detection rate (DR) to quantitatively evaluate the results. The DR is defined as $100 \times (N_f/N)$, where N_f is the number of correct faces and N is the number of detected faces. A correct face must contain at least eyebrows, both eyes and the lips. An example of DR calculation is shown in Figure 4. The DR for the 25 test images is calculated using the total number of correct faces and detected faces in these images. The face detection in thermal images by the Viola-Jones algorithm without cross examples yielded

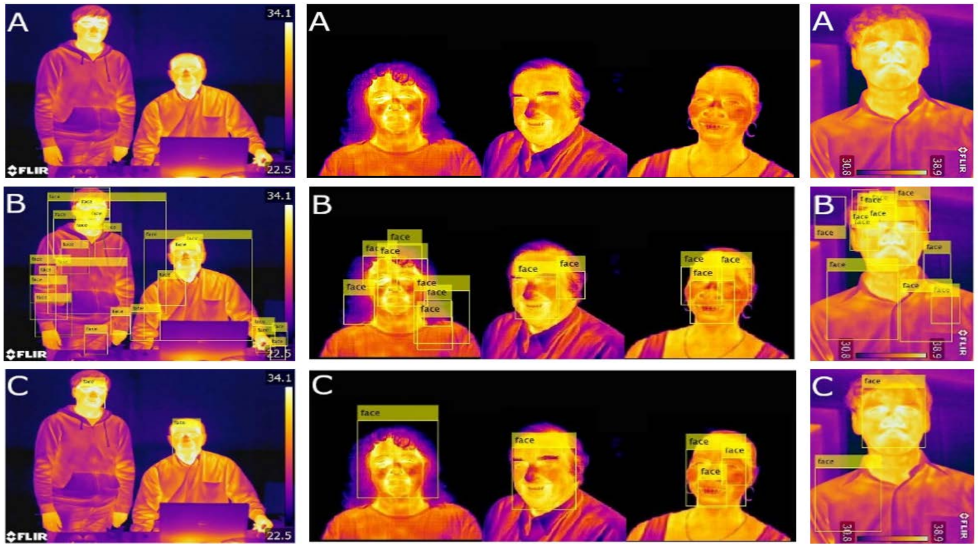

Figure 3: Thermal Face Detection. A: Original Thermal Images; B: Face Detection without Cross-examples; C: Face Detection with Cross-examples

a low DR of 15% due to the large number of false positives. When the five cross examples were added, the DR increased to 67%. We had a relatively small dataset in our experiment. Our future work is to expand the dataset to more images with more imaging conditions. In addition, we will also investigate the impact of positive cross examples that are visible face images.

4 CONCLUSION

This paper explores a strategy to improve the performance of the well know Viola-Jones algorithm for face detection in thermal images. The capability of the Viola-Jones framework in detecting faces is hindered by low resolution of thermal images. Thermal images have less textural details that constitute the features in Viola-Jones framework. Aiming to increase the classification margin, we include cross-examples in the training stage. The experimental results demonstrate consistent performance improvements of the cascade classifier trained with cross examples over the classifiers trained without cross-examples.

ACKNOWLEDGMENTS

This work was supported by the Army Research office under award No. W911NF1810457.

REFERENCES

[1] A. M. Basbrain, J. Q. Gan, and A. Clark. 2017. Accuracy Enhancement of the Viola-jones Algorithm for Thermal Face Detection. In *Proceedings of Intelligent Computing Methodologies. Lecture Notes in Computer Science*, Vol. 10363.

[2] Y. K. Cheong, V. V. Yap, and H. Nisar. 2013. A Robust Face Recognition Algorithm Under Varying Illumination Using Adaptive Retina Modeling. In *Proceedings of AIP Conference Proceedings 1559*. 155.

[3] Y. K. Cheong, V. V. Yap, and H. Nisar. 2014. A Novel Face Detection Algorithm Using Thermal Imaging. In *Proceedings of IEEE Symposium on Computer Applications and Industrial Electronics*. 208–213.

[4] N. Dalai and B. Triggs. 2005. Histograms of Oriented Gradients for Human Detection. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 208–213.

[5] X. H. Di, B. S. Riggan, S. Hu, N. J. Short, and V. M. Patel. 2019. Polarimetric Thermal to Visible Face Verification via Self-attention Guided Synthesis. In *Proceedings of 2019 International Conference on Biometrics*. 1–8.

[6] S. Hu, N. J. Short, B. S. Riggan, C. Gordon, K. P. Gurton, M. Thielke, P. Gurram, and A. L. Chan. 2016. A Polarimetric Thermal Database for Face Recognition Research. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 119–126.

[7] S. M. Iranmanesh, A. Dabouei, H. Kazemi, and N. M. Nasrabadi. 2018. Deep Cross Polarimetric Thermal-to-visible Face Recognition. In *Proceedings of 2018 International Conference on Biometrics*. 166–173.

[8] H. Jiang and E. Learned-Miller. 2017. Face Detection with the Faster R-CNN. In *Proceedings of 2017 12th IEEE International Conference on Automatic Face and Gesture Recognition*. 650–657.

[9] Y. Jin, J. Lu, and Q. Ruan. 2015. Coupled Discriminative Feature Learning for Heterogeneous Face Recognition. *IEEE Transactions on Information Forensics and Security* 10, 3 (2015), 640–652.

[10] M. Kopaczka, L. Breuer, Justus Schock, and D. Merhof. 2019. A Modular System for Detection, Tracking and Analysis of Human Faces in Thermal Infrared Recordings. *Sensors* 19, 19 (2019), 4135.

[11] C. Ma, T. Thanh, H. Uchiyama, H. Nagahara, A. Shimada, and R. Taniguchi. 2017. Adapting Local Features for Face Detection in Thermal Image. *Sensors* 7, 12 (2017), 2741.

[12] J. Mekyska, V. Espinosa-Duro, and M. Faundez-Zanuy. 2010. Face Segmentation: A Comparison Between Visible and Thermal Images. In *Proceedings of IEEE International Carnahan Conference on Security Technology*. 185–189.

[13] K. Reese, Y. Zheng, and A. S. Elmaghraby. 2012. A Comparison of Face Detection Algorithms in Visible and Thermal Spectrums. In *Proceedings of International Conference on Advances in Computer Science and Application*. 49–53.

[14] M. S. Sarfraz and R. Stiefelhagen. 2017. Deep Perceptual Mapping for Cross-modal Face Recognition. *International Journal of Computer Vision* 122, 3 (2017), 426–43.

[15] R. E. Schapire, Y. Freund, P. Bartlett, and W. S. Lee. 1998. Boosting the Margin: A New Explanation for the Effectiveness of Voting Methods. *The Annals of Statistics* 26, 5 (1998), 1651–1686.

[16] C. P. Sumathi, T. Santhanam, and M. Mahadevi. 2012. Automatic Facial Expression Analysis: A Survey. *International Journal of Computer Science and Engineering Survey* 3, 6 (2012), 47–59.

[17] X. D. Sun, P. C. Wu, and S. C. H. Hoi. 2018. Face Detection Using Deep Learning: An Improved Faster RCNN Approach. *Neurocomputing, Research Collection School Of Information Systems* 299 (2018), 42–50.

[18] P. Viola and M. Jones. 2004. Robust Real-time Face Detection. *International Journal of Computer Vision* 57, 2 (2004), 137–154.

[19] Y. Zheng. 2012. Face Detection and Eyeglasses Detection for Thermal Face Recognition. In *Proceedings of SPIE 8300, Image Processing: Machine Vision Applications V*. 83000C.

Figure 4: An Example of DR Calculation: $N_f = 4$, $N = 7$, $DR = 57\%$