Human Detection Through Infrared Vision with Python and OpenCV with trained Linear SVM Model on HOG Descriptors

Anurag Kulkarni
Department of Computer Science and Engineering
at University of Nevada, Reno

Marcus Casey
Department of Computer Science and Engineering
at University of Nevada, Reno

Abstract-- **Human detection through Infrared Vision can be utilized for various realistic real-world applications ranging from search and rescue, image/video processing to identify humans, and many more. Without infrared, current deep learning algorithms allow for human detection with a lower accuracy if the trained dataset varies from the actual dataset through visible camera frames. Using thermal vision and infrared vision, humans can be detected based on their thermal heat emittance. This way, the presence of humans can be detected regardless of their pre-trained posture datasets, hence reducing the room for errors. Main objective of this research project is to develop an efficient computer vision algorithm that can enable human detection from lightweight low-resolution IR cameras.**

Keywords-- HOG, SVM, Linear SVM, Cascades, classifiers, infrared vision, computer vision, thermal vision, robotics, algorithm, normalization;

I. INTRODUCTION

Human detection through Infrared vision or thermal vision can be utilized for various realistic real-world applications. These applications range from (but are not limited to) search and rescue, image and video processing to search for certain features like identification of humans in low-light conditions, as well as many more.

The perception of environments is incredibly critical in any of those applications in regards to mobile unmanned autonomous robots and systems. This research project will utilize the HOG-SVM linear trained model approach to image (and video) real-time human detection through thermal and infrared vision images.

The proposed approach in the beginning of the project was to utilize frame-by-frame dissection of images and video feeds using IR and thermal vision; however, after several trial-and-error methods, we were unable to achieve the frame-by-frame due to bugs and code errors. Instead, this research paper proposes an approach that utilizes linear SVM (support vector machine) models trained on HOG (histogram of oriented gradient) descriptor features in order to detect humans in low-resolution IR vision and thermal vision images and video streams. The project is built on Python programming language using the open-source computer vision library OpenCV on Rodeo Python IDE for data science on the Anaconda3 distribution.

II. HUMAN DETECTION

2.1 Previous Work

Detection and tracking of people in visible light images has been subject to extensive research in the past decades. There are various research papers that have assisted in the algorithm development for this project. The works in [1] describe a detailed approach to dissecting a video stream or an image frame-by-frame. The approach seemed as the proper method of breaking down a frame one at a time in order to analyze it. However, a combined approach to training a support vector machine model on histogram of oriented gradients features and descriptors provided much more accurate and precise results[1], [2].

Furthermore, there are also other classifiers available in OpenCV that specialize in face recognition. For example: the HAAR cascade could be utilized in order to recognize pedestrians and their faces [4]. For the scope of this project, the factor with utmost importance was recognizing full human bodies regardless of their posture through computer vision algorithm. Additionally, there were other, much more sophisticated approaches to human detection in thermal and IR vision through aerial views. In research paper [2], a two-stage template approach was introduced for human detection in thermal vision. The first stage consisted of the screening process in order to “hypothesize only the locations of the people in the image” [2]. This was followed by the creation of a CSM--Contour Saliency Map which “represents the confidence/belief of each pixel belonging to an edge contour of a foreground object” [2]. This is followed by the second stage of the process: AdaBoost classification with adaptive filters in order to seperate the best person matches from other candidates for increased accuracy and precision. For the simplicity of this research paper and project, a simple trained SVM model on HOG descriptors will be utilized.

III. METHODOLOGY

3.1 OpenCV Human Detection

At the fundamental level of human beings, we all have similar features that make us a “human”. These features--or “descriptors” define us all at a structural level. That is: all of the human beings (or most) have “a head, two arms, a torso, and two legs” [4]. Computer vision algorithms can be used in order to exploit this semi rigid physical structure to extract these features and consequently identify or “quantify” the human body [4]. These features can be “passed on” to machine learning and deep learning models that can be trained to “track” and “detect” these features in images and video streams to detect human presence [4]. When coupled with low-resolution IR and thermal vision cameras, humans can be detected in low-light areas and areas where there is extremely low visibility--for example: a room filled with smoke, or a cave with no natural lighting, and so on.

OpenCV is widely-used open-source computer vision library with interfaces for Python, C++, Java, etc. with built-in methods to perform human detection. OpenCV “ships with a pre-trained HOG + Linear SVM model that can be used to perform human detection in images and video streams. In this project, we will be working with OpenCV on Python programming language on HomeBrew environment.

3.2 Histogram of Oriented Gradients (HOG)

The Histogram of Oriented Gradients method suggested by Navneet Dalal and Bill Triggs in their seminal 2005 paper demonstrated with clarity that trained linear SVM model on HOG descriptors could be used to “train highly accurate object classifiers” -- or human detectors. The Histogram of Oriented Gradients (HOG) approach utilizes feature descriptors “--image” and evaluates well-normalized local histograms of image gradient orientations in a dense grid [6].

The HOG descriptor method algorithm to train an a human detector using Histogram of Oriented Gradients is as follows:

Fig. 1: The overall process of Histogram of Oriented Gradients [7].

1. **Step 1:** Sample P positive samples from your training data of the humans/objected to be detected. Extract HOG descriptors from these samples [5]. This is the processing stage of the overall process. These features or “descriptors” will be critical in detecting human presence.
2. **Step 2:** Sample N negative samples from a negative training set that does not contain any humans that we need to detect. Extract HOG descriptors from these samples as well in order to find the difference between those samples that contain humans and those that do not. Calculate the gradient images in order to find the descriptors. Convolve the image with discrete derivative mask.

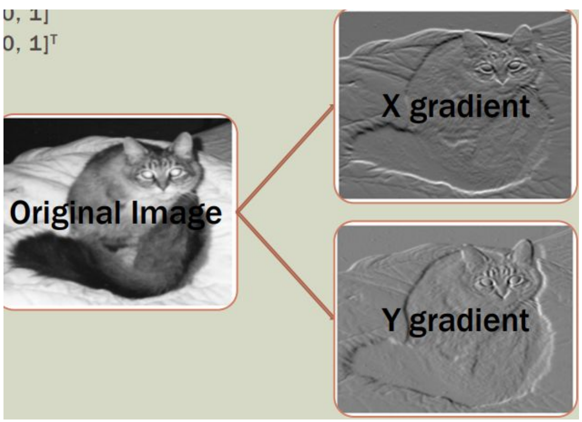


Fig. 2: Convolving the image with discrete X gradient and Y gradient mask. In our case, humans were used [7].

3. **Step 3:** split the image in 8x8 cell bins in order to dissect and analyze each cell individually in order to match and detect the descriptors.

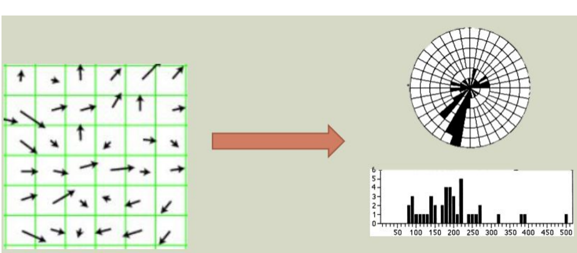


Fig. 3: Dividing into 8x8 cells and count the gradient angles [7].

4. **Step 4:** normalize data set using 16x16 normalisation blocks in order to work with linear SVM models. Overlapping blocks yields better results and higher percentage of object matching and detection. Local contrast normalisation is needed to find “true” weight of an edge.

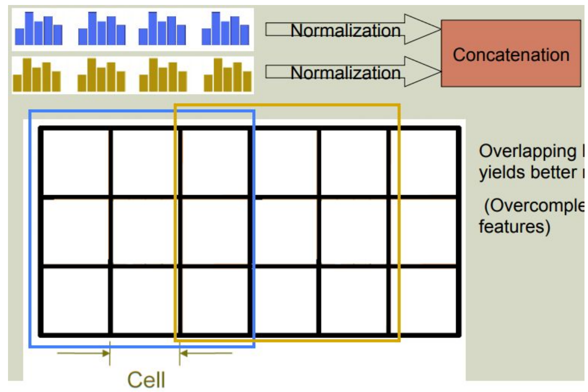


Fig. 4: Normalization of cells followed by concatenation of ‘blocks’ [7].

5. **Step 5:** calculate the HOG feature descriptor vector. At each window extract HOG descriptors and apply classifier.

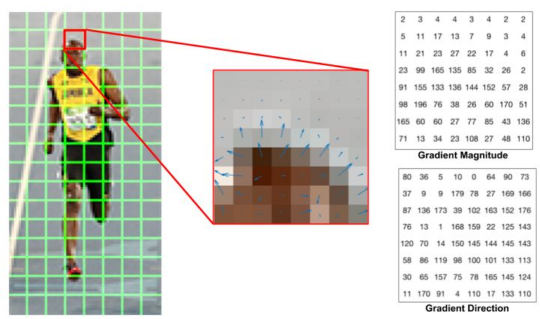


Fig. 5: The general process of HOG feature extraction.

3.3 Linear Support Vector Machine (SVM)

For simplicity, a linear Support Vector Machine was trained on HOG descriptors and features in order to detect humans in low-res IR and thermal vision cameras.

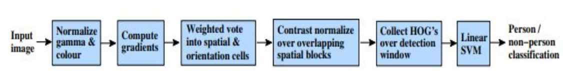


Fig. 6: An overview of our feature extraction and object detection chain. The detector window is tiled with a grid of overlapping blocks in which Histogram of Oriented Gradient feature vectors are extracted. The combined vectors are fed to a linear SVM for object/non-object classification. The detection window is scanned across the image at all positions and scales, and conventional non-maximum suppression is run on the output pyramid to detect object instances.

The primary process of training a Linear SVM on HOG descriptor is as follows:

1. **Step 1:** Detection window from an input image.
2. **Step 2:** Normalize gamma and color.
3. **Step 3:** Compute Gradients.
4. **Step 4:** Accumulate weighted votes for gradient orientation over spatial cells.
5. **Step 5:** Normalize contrast within overlapping blocks of cells. Collect HOG's for all blocks over detection window.

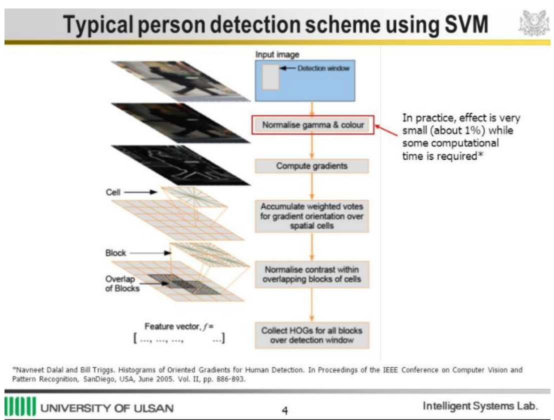


Fig. 7: An overall person detection scheme using trained linear SVM on HOG descriptors.

The overall algorithm for SVM focuses on finding the hyperplane that gives the largest minimum distance to the training examples. The distance receives the important name of margin twice. Optimal separating hyperplane maximizes the margin of the training data.

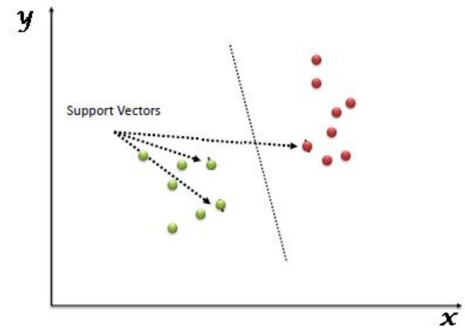


Fig. 8: Finding the hyperplane in SVM that yields the largest min. distance to training examples.

IV. IMPLEMENTATION AND EXPERIMENTS

Implementation of the project began with exploring different libraries of code and resources available to the implement the IR detection. Initially, we experimented with haag cascades to detect facial features so we could start to begin progress on human recognition. After discovering many different SVM's with human recognition we finally implemented DefaultPeopleDetector() which was trained on the INRIA database and implemented in OpenCV. Using the SVM applied with our HOG we were able to have the program load in large amount of images and export them applying the detection to the entire collection and saving a new copy. The python script to grab all executables is listed as Fig. 8

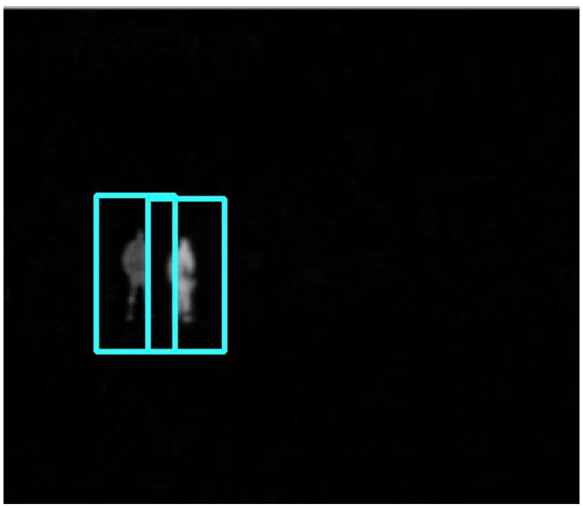
```
#grab images from image directory
def locateFile(path, filter=''):
    #file_list is for storing all files in a directory into an []
    file_list = []
    #grab all files in directory, name doesn't matter, filter applies relevant directory listing semantics
    for (root, dirs, files) in os.walk(path):
        if path == root:
            for file in files:
                if file.find(filter) > -1:
                    file_list.append(os.path.join(root,file).replace("\\", "/"))
    return file_list
```

Fig 8: Python script to grab all images in a folder to apply hog_svm.

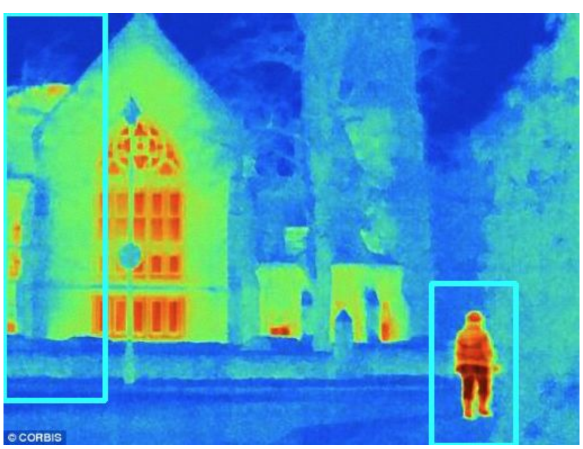
Other SVM's that were trained in human detection as well as infrared application were acknowledged and tested with future iterations of the program most likely utilizing more libraries and open source platforms to help bring results even closer to optimal results.

V. RESULTS AND PERFORMANCE

Results of the program were satisfactory to conventional imagery. Simple infrared photos without too much static or conflicting overlays were easy to detect.



Some more benchmark images suffered detection accuracy but still hit the mark most of the time.



When an image was small and had incredibly dense and low light conditions, the image suffered major detection failure.

This is most likely due to the fact that the SVM trained in human detection with IR modifications to help better detect light patterns was not noticing features due to high light. These photos taken from the infrared photo database benchmark have especially high light signatures.

VI. CONCLUSION

Overall our project was a success in detection of humans in infrared images. However, the project requires **extensive** reworks to ensure accurate detection in even worse lighting and disaster situations. As it stands now, the program could not simply be used in real life situations so that it could be useful, as it is not accurate enough for rescue situations. With modifications and future implementations, the creators will be making improvements to the software to ensure more accurate detection. Some of these improvements will be implementing a Countour Saliency Map and a Multi Resolution Screening with Adaptive Filters. Additionally, if we wanted to keep the scope of the project similar, we would implement LatentSVM on top of our current SVM which would allow for more accurate image detection. These might require haass cascade paired with a dpm module, which has been a popular method of implementing IR detection of humans in some recent research studies.

ACKNOWLEDGEMENT

The authors are incredibly grateful to Prof. (Dr.) Kostas Alexis and his vital guidance and insight that made this overall research, simulation and project a success.

REFERENCES

[1] Fernández-Caballero, Antonio, et al. “Optical Flow or Image Subtraction in Human Detection from Infrared Camera on Mobile Robot.” Elsevier Robotics and Autonomous Systems Journal, 26 June 2010, ruidera.uclm.es/xmlui/bitstream/handle/10578/2101/fi_1332170738-sdarticle.pdf?sequence=1.

[2] Davis, James, Keck, Mark. “A two-stage template approach to person Detection in Thermal Imagery.” Ohio State University Computer Science department. Applications of Computer Vision, 2005.

[3] Portmann, J, Lynen, S, Chli, M & Siegwart, R 2014, People Detection and Tracking from Aerial Thermal Views. in Proceedings of the IEEE International Conference on Robotics and Automation (ICRA). IEEE, pp. 1794 - 1800. DOI: 10.1109/ICRA.2014.6907094

[4] Rosebrock, Adrian. “Pedestrian Detection OpenCV.” PyImageSearch, 9 Nov. 2015, www.pyimagesearch.com/2015/11/09/pedestrian-detection-opencv/.

[5] Rosebrock, Adrian. “Pedestrian Detection OpenCV.” PyImageSearch, 10 Nov. 2015, www.pyimagesearch.com/2014/11/10/histogram-oriented-gradients-object-detection/

[6] Triggs, Bill, Dalal, Navneet. “Histograms of Oriented Gradients for Human Detection.” INRIA. http://lear.inrialpes.fr/people/triggs/pubs/Dalal-cvpr05.pdf

[7] Wang, Tao, Cosman, Rafael. “Histograms of Oriented Gradients for Human Detection.” Stanford Vision. http://vision.stanford.edu/teaching/cs231b_spring1213/slides/hog_rafael_tao.pdf