

REAL-TIME HUMAN DETECTION ANALYTICS IN CONSTRAINED IMAGE INPUTS

Heshan Fernando

(188026x)

Thesis/Dissertation submitted in partial fulfillment of the requirements for the degree
Master of Science

Department of Computer Science and Engineering

University of Moratuwa

Sri Lanka

January 2022

DECLARATION

I declare that this is my own work and this thesis/dissertation does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any other university or institute of higher learning and to the best of my knowledge and beliefs, it does not contain any material previously published or written by another person except where the acknowledgment is made in the text.

Moreover, I hereby grant to the University of Moratuwa the non-exclusive right to reproduce and distribute my thesis/dissertation, in whole or in part in print, electronic, or another medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Name	Signature	Date
Mr. Heshan Fernando

The above candidate has carried out research for the Master's dissertation under our supervision.

Name	Signature	Date
Prof. Indika Perera
Dr. Chathura De Silva

ABSTRACT

Real-time video surveillance is a growing trend today. Our surrounding is being monitored daily by an increasing number of surveillance camera systems. Analyzing human movement can be used for the wellbeing of humans. There are a set of analytical tools and algorithms which can be used to detect, track, and analyze humans in images. Human movement analytics has various subdomains including human detection, human recognition, human tracking, human localization, human reidentification, human behavior analysis, and abnormal activity detection. Human detection is the most crucial step among them, and which helps to derive other sub domains.

Human detection analytics in constrained lighting conditions would be a challenging task to apply due to the low contrast of the image context. Currently available systems focused on the daytime. The background light is an essential factor in the camera images, which rigorously affects the quality of the image. We can identify considerable differences if we compare two images at the rich light condition and constrained light condition. Fewer features of the objects can be extracted in constrained light conditions than rich light conditions. Illumination of the background context is an important factor if we focus on such applications. Currently, most researchers have used human detection analytics in visible light. RGB image shows a clear view when there is sufficient light existing, and it is highly sensitive to visible light conditions compared to infrared. In this research, we considered infrared images as constrained image inputs.

Our proposed methodology contains a novel human detection approach based on machine learning and a motion dynamic model. Here we have addressed the problem using a combination of Deep Convolutional Neural Networks (DCNN) for human detection and Kernelized Correlation Filters (KCF) for human tracking. MobileNet pre-trained model is used for frame-wise human detection as the first step. Then the KCF object tracking algorithm is used to increase the human detection accuracy while tracking the human in the context. Furthermore, we applied some preprocessing techniques to reduce the noise effects. Currently, the progress made by this research-based project is sufficient to initiate the development of a complete human detection analysis solution based on live CCTV camera footage. This solution provides the core functionality of human detection analytics and it can be easily adapted to different domain solutions such as customer behavior analytics in a supermarket or worker movement analytics in an industrial premise.

Keywords: Human Detection, Human Tracking, Deep Neural Networks, MobileNet, Kernelized Correlation Filters, Infrared, Realtime Video Feed, Histogram Equalization

ACKNOWLEDGMENT

The research project “Real-time Human Detection Analytics in Constrained Image Inputs” has been a novel experiment done as the Master of Science (Research) project of Eng. Heshan Fernando from the Department of Computer Science and Engineering. First and foremost, my sincere gratitude goes to the Department of Computer Science Engineering for facilitating this opportunity through the curriculum for the MSc. Engineering degree program. Special gratitude to the University of Moratuwa, Sri Lanka for funding this research project under SRC long-term grant SRC/LT/2016/11.

My sincere gratitude is extended to Prof. Indika Perera and Dr. Chathura de Silva, Senior Lecturers, Department of Computer Science and Engineering for supervising the project. The progress of the project made would not have been a possibility unless for their intense supervision and support. The valuable feedback provided to us by Dr. Charith Chitraranjan and Dr. Anjula de Silva during the initial project proposal presentation and progress evaluations were of utmost usefulness for the development of project outcomes. My sincere gratitude is extended to them.

The opportunity provided to us by Dr. Chandana Gamage and Dr. Sulochana Sooriyaarachchi to present the research prototype at Techno 2019 Exhibition Sri Lanka gave us valuable feedback to further polish our research. My sincere gratitude is extended to them. My special thanks go to various support given by all the staff members including academic and non-academic of the University of Moratuwa. My gratitude is also extended to students of the Department of Computer Science and Engineering who took part in the IRANALYTICA dataset and the numerous supports provided to me through the research period.

TABLE OF CONTENTS

Declaration	i
Abstract	ii
Acknowledgment	iii
Table of Contents	iv
List of Figures	ix
List of Tables	xi
List of Abbreviations	xii
1 Introduction	1
1.1 Research Question	2
1.2 Problem Statement	2
1.3 Background and Motivation	2
1.4 Research Objectives	3
1.4.1 Implement a robust human detection analytic system in both daytime and nighttime	3
1.4.2 Develop computer vision techniques to enhance current human detection analytic methods	3
1.4.3 Implement machine learning approaches to enhance human detection analytics	3
1.5 Project Deliverables and Outcomes	3
1.5.1 Novel real-time human detection model	3
1.5.2 IRANALYTICA Infrared Dataset	4
1.5.3 Research paper “Real-time Human Detection and Tracking in Infrared Video Feed”	4
2 Literature Review	5
2.1 Existing Solutions	5
2.2 Related Works	6
2.2.1 Preprocessing	6
2.2.2 Human detection	7
2.2.2.1 AlexNet	10
2.2.2.2 GoogLeNet	11

2.2.2.3	VGGNet	11
2.2.2.4	ResNet	11
2.2.2.5	YOLO	12
2.2.2.6	R-CNN	12
2.2.2.7	Fast R-CNN	13
2.2.2.8	Faster R-CNN	14
2.2.2.9	SSD	14
2.2.2.10	MobileNet	15
2.2.3	Human localization	16
2.2.4	Human recognition	16
2.2.5	Human tracking	16
2.2.6	Human Re-identification	17
2.2.6.1	Short-term and Long-term Re-identification	17
2.2.6.2	Contextual and Non-contextual Re-identification	18
2.2.6.3	Open set Re-identification and Closed set Re-identification	18
2.3	Summary of Literature Reviews	20
3	Dataset	22
3.1	Existing Datasets	22
3.1.1	RGB image datasets	22
3.1.1.1	COCO dataset	22
3.1.1.2	VIPeR dataset	22
3.1.1.3	i-LIDS-static dataset	23
3.1.1.4	PRID2011 dataset	23
3.1.1.5	INRIA person dataset	23
3.1.2	Infrared image datasets	23
3.1.2.1	ETHZ thermal infrared dataset	23
3.1.2.2	KMU-PD dataset	23

3.1.2.3	BU-TIV dataset	24
3.1.3	RGB-Infrared image dataset	24
3.1.3.1	KAIST dataset	24
3.1.3.2	OTCBVS benchmark dataset	24
3.1.3.3	SYSU-MM01 dataset	24
3.2	IRANALYTICA infrared image dataset	25
3.3	Experimental limitations in current datasets and IRANALYTICA	27
4	Methodology	28
4.1	Overall Solution Breakdown	28
4.2	Subproblems and Evaluated Alternative Solutions	29
4.2.1	Obtaining camera inputs	29
4.2.1.1	Native OpenCV implementation available for C++	30
4.2.1.2	Native Python wrapper for OpenCV	30
4.2.1.3	Java wrapper for OpenCV - JavaCV API	30
4.2.1.4	Third-party .net wrapper on OpenCV - EmguCV API	31
4.2.2	Pre-processing on camera input	31
4.2.2.1	Manual removing distorted images	31
4.2.2.2	Resizing of the input frame	32
4.2.2.3	Grayscale conversion	32
4.2.2.4	Denoising using noise filters	33
4.2.2.4.1	Mean filter	33
4.2.2.4.2	Median filter	34
4.2.2.4.3	Adaptive filter	35
4.2.2.4.4	Histogram equalization	36
4.2.2.4.5	Histogram normalization	37
4.2.2.4.6	Morphological transformation operations	38
4.2.3	Human detection	39

4.2.3.1	Histogram of Oriented Gradients for human detection	39
4.2.3.2	Haar cascade detector	41
4.2.3.3	Background subtraction	42
4.2.3.4	OpenPose detector	43
4.2.3.5	Convolution neural network	44
4.2.4	Feature extraction	44
4.2.4.1	HOG feature	44
4.2.4.2	SIFT feature	45
4.2.4.3	SURF feature	45
4.2.4.4	Haar features	46
4.2.4.5	Color-based features	47
4.2.4.6	Local Binary Patterns	47
4.2.5	People tracking	47
4.2.5.1	TLD tracker	47
4.2.5.2	Kalman filter	48
4.2.5.3	MIL tracker	48
4.2.5.4	KCF tracker	48
4.3	Proposed Solution	50
4.3.1	The camera feed agent	50
4.3.2	The image processing agent	51
4.3.3	The human detection agent	52
4.3.4	The human tracking agent	54
4.3.5	Deploy and set up the solution in the real-time environment	55
4.3.6	Research assumptions and limitations	55
5	Experimental Evaluation And Discussion	56
5.1	Dataset	56
5.2	Experimental Setup	56
5.2.1	Transfer learning the DCNN	57

5.2.2	Evaluation of the model	57
5.3	Intersection Over Union	57
5.4	Results and Evaluation	59
5.4.1	Evaluation of camera feed methods	59
5.4.2	Evaluation of preprocessing methods	60
5.4.3	Evaluation of human detection methods	60
5.4.4	Evaluation of human tracking methods	60
5.4.5	Comparison of the Results with State art Methods	61
6	Conclusion And Recommendation	62
6.1	Conclusion	62
6.2	Recommendation	63
6.2.1	Multiple person detection and tracking	63
6.2.2	Extend the methodology with multiple camera systems	63
6.2.3	Try out different human analytics subdomains	63
6.2.4	Human movement analytics in RGB – Infrared cross-modality	63
7	References	64

LIST OF FIGURES

		Page
Figure 1	Three different types of images (a) RGB Image (b) Infrared Image (c) Depth Image	6
Figure 2	Images of IRANALYTICA Dataset	20
Figure 3	The overall architecture of the proposed system	21
Figure 4	Distorted images of the collected dataset	25
Figure 5	Mean filter example	27
Figure 6	Median filter example	28
Figure 7	The preprocess using histogram equalization (a) raw image (b) enhanced image after histogram equalization	30
Figure 8	Comparison of the histogram equalization and histogram normalization	32
Figure 9	Morphological operation steps (a) original image (b) after global threshold (c) after applying erode (d) after applying dilate	32
Figure 10	Human detection using HOG detector in (a) RGB image (b) infrared image (c) infrared image (d) infrared image	34
Figure 11	Human detection using (a) haarcascade_mcs_upperbody.xml (b) haar cascade_fullbody.xml classifiers	35
Figure 12	Background subtraction (a) frame difference image (b) detected RoIs of humans	36
Figure 13	Examples of human detection using OpenPose	37
Figure 14	Neural network architecture of AlexNet	38
Figure 15	Neural network architecture of ResNet	39
Figure 16	The architecture of RCNN	40
Figure 17	The architecture of F-RCNN	41
Figure 18	The network architecture of SSD	42
Figure 19	Concurrent video stream processing	50
Figure 20	The preprocessing using histogram equalization (a) raw image (b) enhanced image after histogram equalization	50
Figure 21	The neural architecture of the proposed MobileNet CNN	51
Figure 22	Motion based adaptive detection model	53
Figure 23	Proof of concept displayed at Techno 2019	53
Figure 24	The IP camera used for the dataset	54

Figure 25	Intersect over Union calculation	55
Figure 26	Results obtained in proposed system	57

LIST OF TABLES

		Page
Table I	Comparison of available image datasets	19
Table II	Comparison of currently available tracking models	47
Table III	The layer structure of the modified MobileNet	52
Table IV	Detection accuracy, precision, and F1 score as related to the effect of preprocessing & motion mode	60
Table V	Comparison of accuracy, precision, recall, and processing time of HOG detector, YOLO, and our proposed method	60

LIST OF ABBREVIATIONS

Abbreviation	Description
DCNN	Deep Convolution Neural Network
IP	Internet Protocol
KCF	Kernelized Correlation Filters
CCD	Charge Coupled Device
HOG	Histograms of Oriented Gradients
SVM	Support Vector Machine
FLIR	Forward Looking Infrared
YOLO	You Only Look Once
ABMS	Adaptive Boolean Map based Saliency
BMS	Boolean Map based Saliency
SCA	Stel Component Analysis
CWBTFs	Cumulative Weighted Brightness Transfer Functions
IR	Infrared
SAD	Sum of Absolute Differences
SIFT	Scale Invariant Feature Transform
SURF	Speeded Up Robust Feature
LBP	Local Binary Pattern
RoI	Region of Interest
MIL	Multiple Instance Learning
IoU	Intersection over Union
FLIR	Forward Looking Infrared