PRIVACY-PRESERVING SURVEILLANCE AS AN EDGE SERVICE

BY

ALEM HADDUSH FITWI

DISSERTATION

Submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in Electrical and Computer Engineering
in the Graduate School of
Binghamton University
State University of New York
2021

Accepted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in Electrical and Computer Engineering
in the Graduate School of
Binghamton University
State University of New York
2021

November 17, 2021

Dr. Yu Chen, Chair
Department of Electrical and Computer Engineering, Binghamton University

Dr. Douglas Summerville, Member
Department of Electrical and Computer Engineering, Binghamton University

Dr. Xiaohua (Edward) Li, Member
Department of Electrical and Computer Engineering, Binghamton University

Dr. Kartik Gopalan, Outside Examiner
Department of Computer Science, Binghamton University

Abstract

As enshrined in the constitutional documents of many countries and defined in a number of published literature and textbooks, privacy is the freedom from any form of interference or intrusion to a personal space without one's knowledge and consent. It is the ability of a person or a group of people to have some control over how their personal information is collected and used. However, protecting the privacy of individuals in a highly surveilled world is a very challenging task. Today, video surveillance systems (VSS) using closed-circuit television (CCTV) cameras are widely deployed in many urban and suburban areas to garner a great deal of information about individuals' behavioral patterns and activities. The difficulty to preserve the privacy of individuals in the current practice of surveillance is mainly attributed to the facts that (i) there is no distinctively defined boundary between usability and privacy, (ii) video frames created and collected by the edge CCTV cameras could be intercepted by adversaries while in transit through the public network and spilled into the wider cyber space where there are more than 4.6 billion active users today, (iii) the preexisting VSS garners visual information about individuals indiscriminately due to lack of means to distinguish between criminal, suspicious, and innocuous activities or patterns of individuals, (iv) cameras and stored videos could be abused by personnel in charge of the VSS for personal or institutional gains, and (v) it is difficult to enforce the commonly used compute-intensive standard techniques as-is on the edge cameras owning to the availability of only limited computational resources.

In this work, with the aforementioned challenges being the main impetus, efficient solutions for Privacy-preserving Surveillance as an Edge service (PriSE) are proposed from two aspects based-on a hybrid architecture comprising edge cameras equipped with computational power equivalent to the Raspberry PI 4, fog/cloud servers, permissioned blockchain with connection to off-blockchain cloud storage sites, and surveillance operation centers (SOC). The first solution is based on the detection and scrambling of specific privacy attributes like faces, windows and minors. This scheme, however, is applicable to highly conditioned scenarios. The second facet is based on the practice of selective-surveillance that optimizes the balance between usability and privacy. It includes the design of lightweight deep/machine learning (D/ML)-based models for content-wise frame discrimination and secure end-to-end (E2E) privacy-preserving mechanisms based on computationally-thin chaotic maps. Generally, multiple architectures have been investigated, designed and experimented in this work but the best-performing proposed architecture comprises four major modules: (1) a simplified motion detector designed to differentiate between static or unchanging background objects and actual objects captured at the edge cameras by monitoring the background for significant changes, (2) a lightweight frame classifier designed to label frames as offensive and harmless depending on their contents to ensure the practice of selective surveillance following a frame approximation process, (3) novel chaotic-map-based image scrambling techniques that encipher frames or parts of a frame color-channel wise to ensure E2E privacy of individuals caught on CCTV cameras, and (4) a permissioned blockchain-based solution that enables authentic, authorized, accountable, and controllable access to stored surveillance videos to stymie the rife abuses. In other words, privacy is ensured in this work as the intersection of frame discrimination using ML/DNN-based models, chaotic frame scrambling schemes, and blockchain-based access management. The extensive analysis of the functionality, performance and security of the proposed schemes, and comparisons with pertinent previous works verify that the proposed solutions are valid, more feasible, robust and secure.

Dedication

To

My Exceptionally Caring Mother, **Mihret Gebregziabher**

I am so honored to have the most sagacious, sacrosanctly principled, and exceptionally caring mother in the most unlikely environment. I am through and through the product of your many years' toils, incessant encouragement, and philosophical and principled teachings. Without you, I would have reached virtually no where. In addition, you are the most respectful and caring ever mother-in-law to my wife and a quintessential grandmother that my little princess (Emma Alem Haddush) and my prince (Brook Alem Haddush) will always proudly look up to. Hence, I affectionately and proudly dedicate this work to your name.

Your affectionate son and pupil,

Alem Haddush Fitwi

Acknowledgements

First and foremost, I would like to pour out my profound thanks to my advisor, Dr. Yu Chen, for continuously providing me with most invaluable guidance and timely feedback throughout my Ph.D. research. I feel so honored to have worked with someone like him, who is so friendly, committed, and unflaggingly consistent. Likewise, the very constructive comments and feedbacks from Dr. Douglas Summerville, Dr. Xiaohua Li, and Dr. Kartik Gopalan were so important in helping me navigate my research in the right direction. I, therefore, would like to express my special and sincere gratitude to them, as well.

Moreover, I am most thankful to Dr. Krishnaswami Srihari and Dr. Peter Partell for their indispensable supports, without which I would not have reached at this stage. Generally, I cannot really thank enough all members of the Watson School of Engineering and Applied Science Dean's office for their exceptional daily support and camaraderie since day one, as a result of which my life has been so easy. I would also like to seize the opportunity to thank you my fellow graduate lab-mates in the Intelligent surveillance, Edge-fog-cloud computing, and Security & Privacy laboratory for the most fructuous collaborations, useful discussions and good times we had together. Next, I cannot forget to thank the Department of Electrical and Computer Engineering for all the unreserved supports I had repeatedly received.

Finally yet importantly, I would like to say a special thank you to my family and friends for all the unconditional support during the very intense Ph.D. years. My utmost appreciation goes to my wife for her incessant encouragement and support.

Contents

1 Introduction **1**

1.1 Necessity of Privacy in Video Surveillance 1

1.2 Major Causes of Privacy Breaches 3

1.2.1 Interception Attacks 4

1.2.2 Abuse of Video Surveillance Systems 6

1.2.3 Indiscriminate Surveillance Practice 7

1.2.4 Balance Problem Between Privacy & Usability 8

1.3 Limitations of Edge-Computing Paradigm 8

1.4 Proposed Privacy-preserving Solutions 9

1.5 Requirements and Assumptions 15

1.5.1 Requirements 15

1.5.2 Assumptions 16

2 Survey of Related Works **22**

2.1 Video Surveillance Systems and Privacy 22

2.2 Privacy-sensitive Attributes 25

2.3 Privacy-attributes Detection Methods 26

2.4 Privacy-protection Mechanisms 28

2.5 Blockchain Technology vis-à-vis Privacy 31

2.6 Distance Estimation Using a Single Camera 32

2.7 Survey Summary 34

3 PriSE System Architectures 36

3.1 Overview 36

3.2 Cloud-Based CCTV Architecture 38

3.3 Fog-Based CCTV Service 39

3.4 Edge-Based CCTV Service 40

3.5 Pros and Cons of the various Architectures 42

3.6 Proposed Optimized Hybrid Architectures 44

3.6.1 S1: Loaded Edge & Less-Loaded Fog/Cloud Server 47

3.6.2 S2: Less-Loaded Edge & Loaded Fog/Cloud Server 48

4 Video-frames Discrimination 50

4.1 Overview 50

4.2 A1: Detecting Privacy-sensitive ROIs 51

4.2.1 Morphological Model For Windows Detection 53

4.2.2 Face and Windows Detection Using Haar Method 56

4.2.3 Face and Window Objects Detection Using LDNN 59

4.2.4 Multitasking Face-and-Window-Objects Detectors 61

4.2.5 Minors' Privacy Protection 62

4.2.6 Attributes De-Identification and Identification 66

4.2.7 Performance Analysis 72

4.3 A2: Selective Surveillance 78

4.3.1 Loaded Edge-based Selective Surveillance 79

4.3.2 Loaded Fog/Cloud-based Selective Surveillance 86

4.3.3 Comparative Performance Analysis 96

4.4 Summary 98

5 Ensuring End-To-End Privacy **99**

5.1 Overview 99

5.1.1 Overview of End-to-End Privacy 99

5.1.2 Overview of Chaotic Theory 100

5.2 Chaotic Solutions 103

5.2.1 2D Lightweight Chaotic Maps 103

5.2.2 Improved Logistic Map 109

5.2.3 De Jong Chaos For Attributes Denaturing 113

5.2.4 Novel Sinusoidal Chaotic Map 119

5.2.5 Frame Blocks Shuffling to Enhance Security 122

5.2.6 DAB: DCT-AES-BS Based Scrambling Mechanism 127

5.3 Scrambling Keys Management 131

5.4 Performance and Security Analysis 132

5.4.1 An Attacker Model 133

5.4.2 Performance Analysis 134

5.4.3 Security Analysis 136

5.4.4 Summary of Comparative Performance Analysis 149

6 Video Abuses Control **152**

6.1 Overview 152

6.2 SePriS System Architecture 153

6.3 PBC-based Secure Access To Stored Videos 155

6.4 Performance and Security Analysis 159

6.4.1 Functional Test 159
6.4.2 Security of the DAB-cipher 160
6.4.3 Security and Privacy of The Entire SePriS system 161
6.5 Summary 163

7 Crowd Control 164
7.1 Overview 164
7.2 Unified Model For Crowd Control 166
7.3 Distance Estimation Algorithm 169
7.3.1 Triangle Similarity 169
7.3.2 Pixel Per Metric Method 171
7.4 Area Estimation Algorithm 174
7.5 Performance Analysis 176
7.5.1 Experimental Setup 176
7.5.2 Datasets 179
7.5.3 Distance Estimation 179
7.5.4 Area Estimation 181
7.5.5 Discussion 183
7.6 Summary 185

8 Conclusions 186
8.1 Major Contributions 186
8.2 Future Work 188
8.3 Pictorial Summary 189

Appendices 191

A Appendix: Experimental Environment Setup **192**

B Appendix: Brief Description of NIST Test Suites **193**

C Appendix: Publications **194**

Bibliography **196**

List of Tables

3.1 Pros and Cons of Cloud, Fog, and Edge Architectures 42

4.1 Performance The three Methods 74

5.1 Comparing Serial and Vectorized Bitwise-xor Operator 103

5.2 Control Parameter Specifications 106

5.3 Randomness Test Results 110

5.4 Comparing the vectorized XOR and Multiplication functions 117

5.5 Statistical Descriptions 143

5.6 Comparative Security and Performance Analysis 150

6.1 Security Analysis 161

6.2 Security Analysis 163

7.1 Distance Measurement 181

A.1 Specification of the Raspberry Pi 4 192

B.1 NIST Randomness Test Results 193

List of Figures

1.1 Vulnerabilities at Each Layer of the TCP/IP Architecture. 5

1.2 Computing Paradigms: Use, Pros, & Cons. 8

1.3 Solutions:= ML/DNN-Models $\cap$ Chaotic Scrambling Modules. . . . 11

1.4 DORI Zones of a CCTV Camera. 19

3.1 Hierarchical Computing Paradigm: Cloud, Fog, Edge, or Hybrid. The Video Analytics is performed at a/an (a) distant cloud server , (b) fog server , (c) edge, the point of video creation, with the help of single-board computer (SBC) like Raspberry PI or Thinker Board. 37

3.2 Detailed Hierarchical Architecture comprising Cloud, Network, and Fog Layers on top of the Layer of things connected to the edge network! 37

3.3 Cloud-Based CCTV Service: (a) Edge cameras that create the raw videos; (b) The network infrastructure; (c) Powerful cloud servers in distributed data centers; (d) A surveillance operation center (SOC), aka viewing station. 39

3.4 Fog-Based CCTV Service comprising Edge cameras, Fog-server, & SOC. 40

3.5 Edge-Based CCTV Service: The video analytics is completely performed at the edge and only alerting messages are sent to authorized personnel. 41

3.6 A hybrid Edge-Cloud Video Surveillance Architecture. 44

3.7 S1: Loaded edge-camera and less-loaded fog/cloud server. 47

3.8 S2: Less-loaded edge-camera and loaded fog/cloud server. 49

4.1 Morphological transformation for Window-objects Detection 54

4.2 Window object Detection and Testing processes: (a) The detection process, (b) Testing and Scrambling 55

4.3 Flowchart of the Haar-Cascade based Window and Face Objects Detection Method. 58

4.4 LDNN Designed based on Separable Depthwise and Pointwise Operations. 60

4.5 Multi-tasked Face and Window Object-Detectors 61

4.6 Model structure: This network consists of a batch input and output layer and a deep CNN followed by L2 normalization, which results in the face embedding. This is followed by the triplet loss during training. 65

4.7 Training goal of the FaceNet network. 65

4.8 Dataflow of the proposed model: the MTCNN is used to detect and crop the faces from each input image. Then, the FaceNet is employed to calculate the euclidean distance between the anchor face and positive and negative images to identify minor faces using SVM. (a) Illustrates the scenario where the test input image is not a minor; whereas (b) depicts a case where the input test image is a minor and correctly identified by the SVM model. 67

4.9 Attributes De-Identification and Identification Process: At a server, following the decryption process, an incoming frame ($I = x(t)$) is duplicated into two more frames (I_1 and I_2). I_1 is utilized by a window-detector that outputs bounding boxes if there exists any window. I_2 is used by the face-detector module to check if there is/are face object(s) in it, which also returns their bounding boxes. The FaceNet is used for face embeddings, classified by an SVM model to identify fugitives and alert authorized parties. The Chaotic module performs the scrambling of windows and faces based on the returned bounding boxes. 68

4.10 (a) A sample window image, (b) Lines, and edges that make up the window obtained after filtering and edge detection processes. 68

4.11 Creation a Window-object Detector Using Transfer Learning. 69

4.12 Window Objects: sixteen types of windows, namely Fixed windows, Sliding windows, Pivoted windows, Lantern windows, Skylights, Louvered windows, Metal windows, Sash windows, Ventilators, Double hung windows, Bay windows, Clerestory windows, Corner windows, Casement windows, Gable windows, and Dormer windows are considered in the training and testing processes [26]. 70

4.13 Privacy Attributes Detection: (a) The detection of Window and Face Objects using a multitasked model that comprises Haar-based face detector and a shape-based window detector model (b) The detection of Face and Window objects using LDNN. 73

4.14 CPU and Memory Utilization of the four different models. 75

4.15 ROC Curves for the Window-object Detector, Face-object Detector, and Fugitive Classifier. 78

4.16 Loaded-Edge System Architecture:It comprises edge cameras, and such modules as frame approximation, frame-classifier, and frame-scrambling. 80

4.17 Frame Approximation Using SVD: The input image (a) is compressed by about 25% to produce the output image in (b). 82

4.18 Simplified Frame Classifier Model: It comprises cov2D and dense layers 83

4.19 Frame classifier Model Creation and Optimization Processes. 84

4.20 Dataset: (a) fist-raising dataset, (b) knife-wielding dataset, (c) gun-brandishing dataset, and (d) harmless dataset 85

4.21 Selective Surveillance: Input frames (a), (b), & (c) are input to the classifier model that classified the first three of them as offensive and the 4^{th} one as innocuous with accuracy of 98.5%, 100%, 100%, and 97.35%, respectively. The third row shows that harmless frame is dropped while all other frames are encrypted & forwarded. 86

4.22 Overview of Loaded Fog/Cloud Architecture of Selective Surveillance. 87

4.23 An illustration of foreground object detection and discarding of a frame containing only background object. 88

4.24 Raspberry PI-camera-1 Motion Detection:(a) reference frame, (b) next frame with no change, (c) a null frame which is the difference of (a) and (b), (d) frame with a hand object (e) a black-and-white frame containing only a hand object which is the difference of (a) and (d) 90

4.25 Raspberry PI-camera-2 Motion Detection:(a) reference frame, (b) next frame with no change, (c) a null frame which is the difference of (a) and (b), (d) frame with the shoulder and head part of a fast-moving person (e) a frame containing only the difference of (a) and (d) 91

4.26 Frame Classification and Frame Filtering 93

4.27 An Architecture comprising Edge camera that creates and encrypts frames, intruder that tries to access the video stream in transit, cloud/fog server that deciphers frames, detect windows and faces and mask them, and a viewing station. 94

4.28 Loaded Fog/Cloud-based Window Detection and Denaturing 95

4.29 Sensitivities and Specifities the frame-contents analyzing and classifying models at Edge and Fog/Cloud environments. 97

5.1 Simplified End-to-End Privacy Model. 100

5.2 Chaos: (a) chaotic maps produced by uncontrolled natural processes [114], (b) a chaotic map produced by a controlled system. 101

5.3 Tensors of various ranks. 101

5.4 $\lambda \notin [0.0005, 0.0009]$ in the fourth diagram colored red, at right bottom corner. All the blue ones are random and secure 106

5.5 $\lambda \notin [0.0005, 0.0009]$ in the first diagram colored red, at the left top corner. All the green ones are random and secure. 107

5.6 FPS of the scheme for various frame sizes: $240P = 352 \times 240 \times 3,\ 360P = 480 \times 360 \times 3,\ 480P = 640 \times 480 \times 3,\ 720P = 1080 \times 720 \times 3,\ 1080P = 1920 \times 1080 \times 3,\ 2K = 2048 \times 1080 \times 3,\ 1440P = 2560 \times 1440 \times 3,\ 4K = 3840 \times 2160 \times 3,\ and\ 8K = 7680 \times 4320 \times 3$. 110

5.7 Channel-wise Frame Enciphering: Color channels R, G, and B are scrambled in parallel for fast processing. 113

5.8 Distribution of the De Jong Chaos, Eq. 5.2.13: it is not uniform revealing its weakness against histogram analysis attack. 114

5.9 Histogram: (a) plain ROI (face), (b) cipher of plain ROI (a), (c) histogram of plain ROI (a), and (d) randomized histogram of cipher ROI (b). 117

5.10 Plot of Eq. (5.2.17) for $\alpha \in (1, 255)$, $\beta \in (3.136745612, 3.141592653)$, $\gamma \in (-1, -0.91)$, and $x \in (0, 1)$ 120

5.11 Frame Shuffling and Unshuffling: (a) Original positions of blocks and (b) shuffled positions of blocks at sending end, and (c) restored positions of blocks at receiving end. 125

5.12 Frame shuffling with different block sizes: (a) Clear input frame of Lenna, (b) shuffled with block_size: 32×32, (c) shuffled with block_size:16×16, (b) shuffled with block_size: 8×8, (b) shuffled with block_size: 4×4, (b) shuffled with block_size: 1×1 126

5.13 Frame shuffling with different block sizes: (a) Clear input frame of Mona Lisa, (b) shuffled with block_size: 32×32, (c) shuffled with block_size:16×16, (d) shuffled with block_size: 8×8, (e) shuffled with block_size: 4×4, (f) shuffled with block_size: 2×2, and (g) shuffled with block_size: 1×1 126

5.14 DCT: (a) 2d Gray-scale input frame, (b) Set of 8×8 DCT Blocks of the input frame, and (c) Compressed version of the input frame. 129

5.15 Simplified Diagram of Agents Interaction for key exchange. 131

5.16 Attacker Model Comprising an edge-camera, an intruder, a fog/cloud server, and surveillance operation center (SOC). 133

5.17 Frame scrambling:(a) a clear input image; (b) the cipher of input image (a); (c) clear input frame with semantics; (d) the cipher of input frame (c) 134

5.18 Lyapunov Diagram of the proposed Sinusoidal Chaotic Map (SCM) 138

5.19 Lyapunov Diagram of the proposed Sinusoidal Chaotic Map (SCM), magnified for better illustration and insight. 138

5.20 Lyapunov Diagram of the Logistic Map [77] 139

5.21 Histogram Analysis: Column 1 comprises the clear input image of Lenna and its color channels R, G, and B from top to bottom. Column 2 represents the histograms of the clear input and its R, G, and B channels in the order from top to bottom. Column 3 comprises the ciphers of the input image and its color channels in column 1 in the same order. Column 4 illustrates the histograms of the ciphers of the whole input frame and its channels in column 3. . 142

5.22 Histogram Analysis: (a) is a plain input image of Mona Lisa, (a1) - (a3) are the non-uniform histograms of R, G, and B channels of (a); (b) is cipher of (a) and (b1) - (b3) are the uniform histograms of the R, G, and B channels of (b). The histograms of all channels of the plain and cipher images are assembled together on (a4) and (b4), respectively. On the figure, r, g, and b stand for red, green, and blue colors, respectively. 143

5.23 Input image of Lenna, and corresponding scatter plots of its horizontal, vertical, and diagonal correlations are pictured on (a), (ah), (av), and (ad); whereas (b), (bh), (bv), and (bd) represent the cipher of Lenna input images and the scatter plots of corresponding horizontal, vertical, and diagonal correlations, respectively 147

5.24 Scatter plots: (a) plain input image of Mona Lisa; (c) horizontal correlation of clear frame (a); (e) vertical correlation of clear frame (a); (g) diagonal correlation of clear frame (a); (b) encrypted version of the input frame (a); (d) horizontal correlation of cipher (b); (f) vertical correlation of cipher (b); and (h) diagonal correlation of cipher (b). 148

6.1 PBC-based SePriS Architecture for secure and privacy-aware exchange of surveillance videos stored in distributed off-BC sites. . . . 154

6.2 The work-flow of the PBC-based Solution for secure and privacy-aware exchange of stored surveillance videos. 156

6.3 A sample request made by court, in JSON file format, sent to a BC-node. 157

6.4 A sample request made by court, in JSON file format, an off-BC storage. 158

6.5 Enciphering: (a) Plain frame/Image, (b) cipher created by DAB. . . 159

6.6 Creation of sample blocks with the data part enciphered: data comprises requests made and associated identity information and log of activities. 160

7.1 Unified Model for Human Detection and Estimation of Interpersonal Distance, Number of Social Distance (SD) Violations, Area, and Crowd Density. 167

7.2 Geometrical relationship between actual and virtual human dimensions. 170

7.3 Rectangular estimation of an area occupied by a crowd. 174

7.4 Experimental Setup for obtaining relationship between widths of people and corresponding interpersonal distances. 177

7.5 A camera configured to see areas forward of the point directly below it up to a maximum distance of 10.5m. 178

7.6 A camera setup to see areas forward of a mark at 2m from the point directly below it up to infinity. 178

7.7 Experimental Analysis: a pair of people at least 2m apart from each other at a distance of (a) 15m, (b) 13m, (c), 11m, (d) 9m, (e) 7m, and (f) 5m from the camera perched on a 3m tall pole. 180

7.8 Number of people violating social distancing: 0, Total number of people in a frame: 2, Estimated area: 1.61m^2, Density: 1.24 182

7.9 Number of people violating social distancing: 10, Total number of people in a frame: 13, Estimated area: 92.01m^2, Density: 0.14 . . . 182

7.10 Number of people violating social distancing: 0, Total number of people in a frame: 3, Estimated area: 10.62m^2, Density: 0.28 183

7.11 Number of people violating social distancing: 4, Total number of people in a frame: 7, Estimated area: 47.83m^2, Density: 0.15 184

8.1 Summary of Privacy-Preserving Surveillance As An Edge Service. ECE stands for Edge-Cloud-Edge, EFE stands for Edge-Fog-Edge, and EE stands for Edge-Edge flows. 190

List of Acronyms

2DC	2D Chaos
ACLU	America Civil Liberties Union
AI	Artificial Intelligence
BC	Blockchain
BWC	Body Worn Camera
CAN	Campus Area Network
CCTV	Closed Circuit Television
CDC	Centers for Disease Control
CNN	Convolutional Neural Network
COVID-19	Coronavirus Disease of 2019
DNN	Deep Neural Network
D/ML	Deep/Machine Learning
DORI	Detection, Observation, Monitoring, and Identification
DyCIE	Dynamic Chaotic Image Enciphering
E2E	End-to-End
FPR	False Positive Rate
FPS	Frames Per Second
GIL	Global Interpreter Lock (python)
GPS	Gobal Positioning System
HD	High Definition
HBM	Haar-Based Model
HOG	Histogram Orient Gradient
IDJM	Improved De Jong Map
IIoT	Internet of Industrial of Things
ILM	Improved Logistic Map
IT	Information Technology
LDNN	Lightweight DNN
ML	Machine Learning
MTCNN	Multi-Tasked CNN
OID	Open Image frontal view Dataset
PPF	Pixel Per Foot
PPM	Pixel Per Meter
RCNN	Region Based Convolutional Neural Networks

ROI	Region of Interest
SCM	Sinusoidal Chaotic Map
SCNN	Standard CNN
SD	Social Distancing
SOC	Surveillance Operation Center
SSD	Single-Shot Detection
SVD	Singular value Decomposition
UGV	Unmanned Ground Vehicle
VSS	Video Surveillance Systems
WAN	Wide Area Network
WDM	Window Detection Model
YOLOv3	You Look Only Once version 03

1. Introduction

1.1 Necessity of Privacy in Video Surveillance

The rapid advancement and ramification of electronic technologies over the last two decades have driven urban areas to become a lot smarter. At present, a multitude of cities around the world employ a spectrum of information and communication technologies (ICT), and Internet of Things (IoT) to improve the quality of urban services and to ensure the physical security and safety of their residents. Fixed and mobile mechanical surveillance technologies are among those widely deployed in a number of smart cities and suburban areas to provide public safety and physical security. Surveillance is often practiced through the use of both fixedly deployed closed circuit television (CCTV) cameras and cameras mounted on mobile manned or unmanned aerial and ground vehicles like airplanes, satellites, drones, manned ground patrolling vehicles, and unmanned ground vehicles (UGV). The ubiquitous and versatile deployment of these surveillance cameras in public places, including streets, city corners, stores, and marketplaces, enable first responders, government agencies, or security service providers to garner a great deal of audio-visual information about many individuals indiscriminately without their knowledge and consent.

As a result, there is a mixed public feeling in relation to the practice of mass-surveillance. On one hand, a number of people have a favorable view of the practice of surveillance because they believe it has the potential to deter and reduce crime, and help monitor traffic, in addition to providing footage of crime scenes as an evidence in courts of law. On the other hand, many incidences of privacy breaches and abuse of personal information have been reported, which has caused grave concerns about the invasion of the privacy of individuals through the practice of non-selective surveillance. Hence, the public wants the surveillance system to be equipped with the ability to protect and/or anonymize privacy-sensitive attributes of individuals and the capability to selectively store only those data vital for future use in lieu of mass-storage to cut down on the risk of privacy breaches [33, 46, 60, 121].

With regards to the current practice of mass-surveillance, there are two lines of antithetical arguments. On one hand, the government and pro mass-surveillance people claim that there is no harm to good people that comes from large-scale mass-surveillance. Only bad people have reasons to want to hide things and care about their privacy. Then, they often mention the famous quote *"If you have got nothing to hide, you have got nothing to worry about"* to substantiate their argument. On the other hand, many argue to debunk the previous argument that it is never all about hiding something, it is all about an individual's private things being no one else's business. The quote *"I don't have anything to hide, but I don't have anything I feel like showing to you, either"* is used in an effort to corroborate this position [60]. However, the bottom line here is that there is a real threat of indiscriminate invasion and abuse of privacy and most people are concerned about it. Hence, they want to see privacy-conserving surveillance practices realized that enables the practice of video surveillance without unwarranted interference in the lives of individuals, allowing them to negotiate who they are and how they want to interact with the world around them. The privacy-mechanism should protect individuals from arbitrary and unjustified use of power by states, companies and other actors in relation to the practice of video surveillance. Then, as individuals who champion the practice of privacy-preserving surveillance, we have carried out